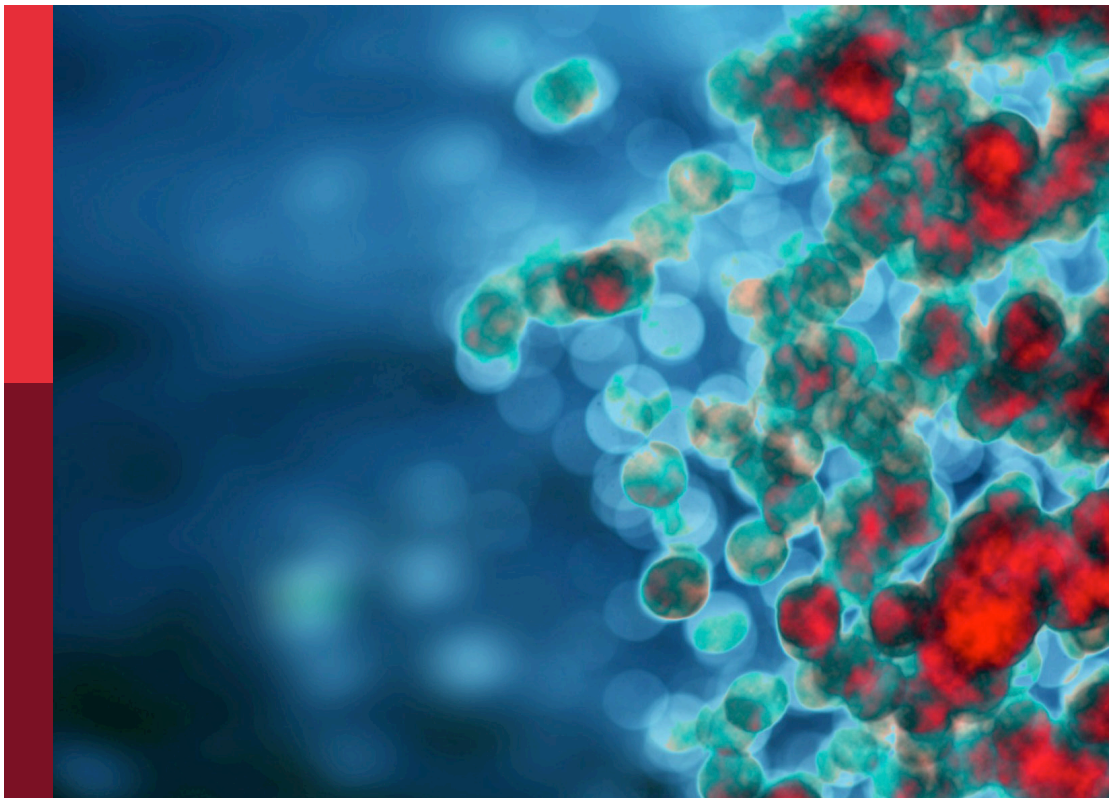
frontiers | Research Topics

Systems immunology to advance vaccine development

Edited by
Joe Hou and Helder Nakaya

Published in
Frontiers in Immunology



ISSN 1664-8714
ISBN 978-2-8325-5810-2
DOI 10.3389/978-2-8325-5810-2

About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

Systems immunology to advance vaccine development

Topic editors

Joe Hou — Fred Hutchinson Cancer Center, United States

Helder Nakaya — University of São Paulo, Brazil

Citation

Hou, J., Nakaya, H., eds. (2024). *Systems immunology to advance vaccine development*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-5810-2

Table of contents

frontiers | Frontiers in Immunology

TYPE Editorial
PUBLISHED 09 December 2024
DOI 10.3389/fimmu.2024.1527238

OPEN ACCESS

EDITED AND REVIEWED BY
Simon Mitchell,
Brighton and Sussex Medical School,
United Kingdom

*CORRESPONDENCE
Jue Hou
joseph.houjue@gmail.com

RECEIVED 13 November 2024
ACCEPTED 25 November 2024
PUBLISHED 09 December 2024

CITATION
Hou J and Nakaya HI (2024) Editorial: Systems immunology to advance vaccine development.
Front. Immunol. 15:1527238.
doi: 10.3389/fimmu.2024.1527238

Editorial: Systems immunology to advance vaccine development

Jue Hou[1]* and Helder I. Nakaya[2,3]

[1]Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Research Center, Seattle, WA, United States, [2]Department of Clinical and Toxicological Analyses, School of Pharmaceutical Sciences, University of São Paulo, São Paulo, Brazil, [3]Hospital Israelita Albert Einstein, São Paulo, Brazil

KEYWORDS

vaccine, systems immunology and AI, vaccine development, computational biology, bioinformatics, modeling

Editorial on the Research Topic
Systems immunology to advance vaccine development

The field of systems immunology has emerged as an essential interdisciplinary approach for understanding immune responses on a comprehensive, system-wide level. By integrating high-throughput technologies such as transcriptomics, proteomics, and computational modeling, systems immunology extends beyond traditional research methods. Systems immunology enables researchers to explore the intricate interactions within the immune components and make prediction about vaccine outcomes, accelerating the development of immunizations against pathogens, including rapidly evolving viruses like SARS-CoV-2. The studies in this Research Topic reflects these advancements, offering fresh insights into the molecular and computational aspects of immune system dynamics, which are crucial for improving vaccine efficacy.

Our Research Topic has brought together 133 authors worldwide, culminating in 16 articles showcasing cutting-edge research. These contributions cover a range of themes, from immune receptor dynamics and biomarker discovery to computational modeling and predictive analytics in vaccine responses.

Immune receptor dynamics and antigen-specific responses

Richardson et al. explored the B cell receptor (BCR) repertoires of 40 participants from the EBL2001 clinical trial, focusing on responses to the Ad26.ZEBOV/MVA-BN-Filo Ebola vaccine. Through bulk sequencing and bioinformatic mining, the authors mapped BCR clonotypes and identified antigen-specific responses, including IGHV3-15 antibodies targeting Ebola glycoprotein, — underscoring the role of systems immunology in decoding antibody-mediated immunity. Similarly, Akhmatova et al. investigated a synthetic disaccharide conjugated to BSA (bovine serum albumin), designed to mimic Streptococcus pneumoniae serotype 3 polysaccharides. Their study demonstrated enhanced IL-17A production and γδ T cell expansion in mice, highlighting how synthetic carbohydrate-based vaccines stimulate both innate and adaptive immunity.

Haralambieva et al. explored the transcriptional profiles of B cells after a third MMR (Measles, Mumps, and Rubella) vaccine dose, identifying genes like IL20RB and BEX2 as

correlates with measles-specific neutralizing antibody responses. These findings point to early biomarkers that could predict vaccine efficacy, advancing personalized vaccinology by supporting tailored vaccine schedules.

In another study, Costa-Gouvea et al. compared immune responses elicited by a *Plasmodium vivax* circumsporozoite protein malaria vaccine formulated with two different adjuvants: Poly I:C and Alhydrogel. They demonstrated that Poly I:C induced broader and stronger humoral and cellular responses, including higher levels of IgG antibodies and a more diverse IgG isotype profile, compared to Alhydrogel. The study also revealed enhanced memory B cell formation, highlighting Poly I:C's potential to improve vaccine efficacy for malaria.

Computational models and machine learning approaches

Several papers in this Research Topic explore computational models for designing and predicting vaccine efficacy. Khan et al. applied molecular simulations and structure-guided engineering to enhance the binding affinity of a monoclonal antibody targeting the aP2 antigen, which is linked to type 2 diabetes. Their engineered T94M mutant demonstrated superior binding strength, illustrating how computational models can advance therapeutic antibody design. Høie et al. developed DiscoTope-3.0, a computational tool that uses inverse folding latent representations to predict B cell epitopes. Benchmarked against multiple datasets, DiscoTope-3.0 excelled, particularly in predicting conformational epitopes critical for vaccine design.

In another study, Parizi et al. introduced PANDORA v2.0, a software designed to model peptide-MHC class II complexes. Their study demonstrated that PANDORA's computational efficiency and accuracy make it a valuable tool for vaccine design and immunotherapy, especially for predicting antigenic peptides that drive immune responses. Together, these studies illustrate the transformative role of computational models in refining vaccine candidates, leveraging systems biology to predict immune outcomes and optimize antigen design.

Biomarker discovery and vaccine reactogenicity

Carvalho et al. applied machine learning algorithms to identify baseline gene signatures associated with reactogenicity to the rVSVDG-ZEBOV-GP Ebola vaccine. By analyzing gene expression data from cohorts across four countries, the authors identified 22 critical genes associated to adverse events, offering valuable insights into how molecular profiles might predict vaccine side effects. Building on this, Martinez-Murillo et al. refined an innate plasma signature associated with the same Ebola vaccine. They identified 11 additional biomarkers, including CXCL10 and IL-15, that correlated with reactogenicity and long-term immune responses, enhancing adverse event prediction across diverse populations.

Naidu and Lulu S. investigated the immune responses to enteric infections in endemic versus non-endemic settings, finding that GRB2, a key adaptor molecule in T cell receptor (TCR) signaling, as a major immunomodulatory response in endemic regions, highlighting the importance of regional immune variations in vaccine design. This study demonstrates how systems immunology can inform the development of region-specific vaccines by identifying immune modulation mechanisms.

T cell dynamics and antigen-specific responses

Mark et al. investigated the phenomenon of "hidden public" TCRs, which emerge following acute viral infections like lymphocytic choriomeningitis virus (LCMV) and SARS-CoV-2. Their analysis revealed that viral infections drive the expansion of shared TCRs, particularly in effector T cells, adding a new layer of understanding to how TCR repertoires function during infections. Mosmann et al. applied the SWIFT (Scalable Weighted Iterative Flow-clustering Technique) clustering algorithm to analyze intracellular cytokine staining data from the HVTN 105 HIV vaccine trial, identifying novel antigen-specific T cell populations and correlating them with antibody responses. This work provides a deeper understanding of the T cell dynamics driving vaccine-induced protection.

Vaccine design and epitope prediction

Ali et al. used reverse vaccinology to design a multi-epitope vaccine targeting the newly identified Songling virus. By screening the viral proteome and validating epitopes through molecular docking and dynamics simulations, they identified a promising vaccine candidate with broad coverage potential. Farriol-Duran et al. introduced Brewpitopes, a bioinformatics pipeline that refines B cell epitope predictions in public health emergencies. Validated with the SARS-CoV-2 proteome, Brewpitopes achieved a fivefold enrichment in predicted neutralizing epitopes, demonstrating its potential for real-time vaccine development.

Díaz-Dinamarca et al. investigated two protein-based adjuvants, rSIP from Streptococcus agalactiae and FLH from Fissurella latimarginata, as Toll-like receptor 4 (TLR4) ligands. Their study showed that these adjuvants activate both MyD88- and TRIF-dependent signaling pathways, enhancing antigen cross-presentation and suggesting their potential as vaccine adjuvants. In another vaccine design study, Imon et al. used immunoinformatic tools to create a multi-epitope vaccine against Merkel cell polyomavirus, the causative agent of Merkel cell carcinoma. Computational simulations demonstrated strong interactions with TLR4, indicating a robust immune response, though the vaccine requires further experimental validation.

Conclusion

The collective research presented in this Research Topic highlights the transformative potential of systems immunology in

advancing vaccine development. By integrating cutting-edge techniques such as immune receptor profiling, biomarker discovery, computational modeling, and machine learning, these studies illustrate how systems immunology can unravel the complexities of immune responses. The insights gained are paving the way for more effective, personalized vaccines, improved strategies for predicting and managing adverse reactions, and the identification of novel antigenic targets. As systems immunology continues to evolve, it will remain a cornerstone in addressing global health challenges, enabling the development of next-generation vaccines that are more precise, adaptable, and capable of protecting diverse populations against both existing and emerging infectious diseases.

Author contributions

JH: Writing – original draft, Writing – review & editing. HN: Writing – original draft, Writing – review & editing.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

frontiers | Frontiers in Immunology

TYPE Original Research
PUBLISHED 30 May 2023
DOI 10.3389/fimmu.2023.1199064

OPEN ACCESS

EDITED BY
Joe Hou,
Fred Hutchinson Cancer Research Center, United States

REVIEWED BY
Keyue Ma,
Zai Lab, China
Adriana Tomic,
Boston University, United States

*CORRESPONDENCE
Michal Mark
michal.mark@weizmann.ac.il
Asaf Madi
asafmadi@tauex.tau.ac.il

†These authors have contributed equally to this work

‡Deceased

RECEIVED 02 April 2023
ACCEPTED 16 May 2023
PUBLISHED 30 May 2023

CITATION
Mark M, Reich-Zeliger S, Greenstein E, Biram A, Chain B, Friedman N and Madi A (2023) Viral infection reveals hidden sharing of TCR CDR3 sequences between individuals. *Front. Immunol.* 14:1199064. doi: 10.3389/fimmu.2023.1199064

Viral infection reveals hidden sharing of TCR CDR3 sequences between individuals

Michal Mark[1*], Shlomit Reich-Zeliger[1], Erez Greenstein[1], Adi Biram[1], Benny Chain[2], Nir Friedman[1†‡] and Asaf Madi[3*†]

[1]Department of Immunology, Weizmann Institute of Science, Rehovot, Israel, [2]Division of Infection and Immunity, Department of Computer Science, University College London, London, United Kingdom, [3]Department of Pathology, Tel-Aviv University, Tel-Aviv, Israel

The T cell receptor is generated by a process of random and imprecise somatic recombination. The number of possible T cell receptors which this process can produce is enormous, greatly exceeding the number of T cells in an individual. Thus, the likelihood of identical TCRs being observed in multiple individuals (public TCRs) might be expected to be very low. Nevertheless such public TCRs have often been reported. In this study we explore the extent of TCR publicity in the context of acute resolving Lymphocytic choriomeningitis virus (LCMV) infection in mice. We show that the repertoire of effector T cells following LCMV infection contains a population of highly shared TCR sequences. This subset of TCRs has a distribution of naive precursor frequencies, generation probabilities, and physico-chemical CDR3 properties which lie between those of classic public TCRs, which are observed in uninfected repertoires, and the dominant private TCR repertoire. We have named this set of sequences "hidden public" TCRs, since they are only revealed following infection. A similar repertoire of hidden public TCRs can be observed in humans after a first exposure to SARS-COV-2. The presence of hidden public TCRs which rapidly expand following viral infection may therefore be a general feature of adaptive immunity, identifying an additional level of inter-individual sharing in the TCR repertoire which may form an important component of the effector and memory response.

KEYWORDS

TCR - T cell receptor, LCMV (lymphocytic choriomeningitis virus), epitope-specific T cell, effector T cells, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)

1 Introduction

T cell receptor (TCR) antigen recognition is a key step in cellular immunity. The ability to recognize a wide range of different pathogens depends on the huge αβ TCR repertoire diversity generated by the stochastic and imprecise recombination of variable, diversity and joining (VDJ) genes (1). The estimated number of possible TCRs which could be generated has been estimated as greater than 10^{14} (2), exceeding by many orders of magnitude the

number of T cells in the human body. Nevertheless, TCR sequences shared between many individuals, often referred to as public TCRs, have been reported in both human (3, 4); and mouse (5, 6). Although some public sequences have been annotated as specific to viral or bacterial antigens (7, 8), most studies have focused on repertoires from healthy individuals, and less is known about the balance between public and private TCRs in the context of acute infection.

Lymphocytic choriomeningitis virus (LCMV) offers an excellent and well-described model in which to study the TCR repertoire associated with acute infection. The Armstrong strain of LCMV is cleared by eight days post-infection, which corresponds to a strong expansion of CD4+ and CD8+ virus-specific T cells (9, 10). This is followed by a contraction phase, giving rise to a subset of long-lived memory T cells maintained by antigen-independent homeostatic proliferation (11). CD4+ memory T cells subsequently decline slowly, while the CD8+ memory population remains relatively stable (12). The magnitude of the CD8+ response is greater than the CD4+ response throughout the response (13). However, CD4+ T cells are essential for an optimum CD8+ memory response. For example, the TCR signal strength of anti-viral CD4+ LCMV specific T cells has been shown to be critical to memory differentiation during the primary response (14, 15).

In C57BL/6 mice infected with LCMV, both CD4+ and CD8+ T cell epitopes have been identified. These epitopes are derived from the viral glycoprotein (GP) or nucleoprotein (NP). Some regions of the viral antigens can stimulate both CD4+ and CD8+ T cells. For example, the GP 66-77 region is dually restricted by both MHC class I and II molecules (16). The immunodominance hierarchy of the epitopes has been characterized in some detail. At the peak of infection, the CD8+ T cell response is dominated by cells that recognize NP396-404, a peptide that binds with high affinity with both H-2Db and H-2Kb (17, 18), followed by the intermediate epitopes NP205-212 and GP92-101 (19).

In this study, we combine antigen-specific tetramer sorting with bulk TCR sequencing of different phenotypic populations of T cells to characterize the T cell receptor (TCR) repertoire at different phases of the LCMV response. We demonstrate that LCMV infection drives a convergent CD8+ effector response across mice, resulting in the detection of emerging shared (public) TCR CDR3 sequences whose publicity cannot be observed in the unimmunized repertoire. A similar phenomenon of emerging public CDR3s was observed in humans infected with SARS-COV-2. These "hidden" public TCRs reveal an under-appreciated level of constraint on the naive TCR repertoire, with important consequences for our understanding of the interaction between the T cell repertoire and viral infection.

2 Materials and methods

2.1 Animals

Female C57BL/6 mice at five weeks old (Envigo) were injected intravenously with $2X10^5$ PFU of the Armstrong LCMV strain (20). Mice were collected after 8 or 40 days of infection. Healthy control mice were injected with PBS and collected eight days post-treatment. All animals were handled according to regulations formulated by The Weizmann Institute's Animal Care and Use Committee and maintained in a pathogen-free environment.

2.2 SARS-COV-2 consortium and study design

We undertook a case control study nested within our COVID consortium healthcare worker cohort. Participant screening, study design, sample collection, and sample processing have been described in detail previously (21). Briefly, healthcare workers were recruited (between 23rd and 31st March 2020) and underwent weekly evaluation using a questionnaire and biological sample collection for up to 16 weeks when fit to attend work at each visit, with further follow up samples collected at 6 months.

Participants with available blood RNA samples who had PCR-confirmed SARS-COV-2 infection (Roche cobas® diagnostic test platform) at any time point were included. A subset of consecutively recruited participants without evidence of SARS-COV-2 infection on nasopharyngeal swabs and who remained seronegative by both Euroimmun anti S1 spike protein and Roche anti-nucleocapsid protein throughout follow-up were included as uninfected controls.

2.3 Sample preparation and T cell isolation

Spleens were dissociated with a syringe plunger, and single-cell suspensions were treated with ammonium-chloride potassium lysis buffer to remove erythrocytes.

Bone marrow cells were extracted from mice femur and tibia bones and were purified with CD3+ T isolated kit (CD3ϵ MicroBead Kit, mouse, 130-094-973, Miltenyi Biotec). Splenic CD4+ and CD8+ cells were purified in two steps: (1) Selection of CD4+ cells (CD4+ T Cell Isolation Kit, mouse, 130-104-454, Miltenyi) (2) Unbound cells were purified for CD8+ cells (CD8a+ T Cell Isolation Kit, mouse, 130-104-07, Miltenyi Biotec). For the tetramers binding reaction, we pooled splenocytes from previously vaccinated mice (5 mice after 8 days post infection) and purified their T cells using the untouched isolation kit (Pan T Cell Isolation Kit II, mouse, 130-095-130, Miltenyi Biotec).

2.4 Flow cytometry analysis and cell sorting

The following fluorochrome-labeled mouse antibodies were used according to the manufacturers' protocols: PB or Percp/cy5.5 anti -CD4, PB or PreCP/cy5.5 anti- CD8, PE or PE/cy7 anti- CD3, APC anti-CD62L, Fitc or PE/cy7 anti- CD44 (Biolegend). Cells were sorted on a SORP-FACS-AriaII and analyzed using FACSDiva (BD Biosciences) and FlowJo (Tree Star) software. Sorted cells were centrifuged (450g for 10 minutes) before RNA extraction.

2.5 LCMV -tetramers staining and Flow cytometry sorting

Four monomers (NIH Tetramer Core Facility) with different LCMV epitopes were used: MHCII -GP66–77(H-2Bb), MHCI-NP396-404(H-2Db), MHCI- NP205-212(H-2Kb), MHCI- GP92-101 (H-2Db). Tetramers were constructed via binding Biotinylated monomers to PE/APC – conjugated- streptavidin (according to the NIH protocol). Purified T cells were stained with FITC anti-CD4+ and PB anti-CD8+ and followed by tetramers staining (two tetramers together), for 30 min at room temperature (0.6ug/ml). CD4+ and CD8+ epitope-specific cells were sorted from single-positive gates for one type of tetramer. Using two tetramers together for staining provided a control for nonspecific binding, in addition to using cells collected from the unbinding population (SI Figure 1B).

2.6 Library preparation for TCR-sequencing

All libraries in this work were prepared according to the published method (22), with minor adaptations for mice and an in-house pipeline for pre-processing of the data. The pipeline introduces unique molecular identifiers attached to individual cDNA molecules, which allows correction for sequencing error PCR bias, and provides a quantitative and reproducible method of library preparation. Full details pre-processing pipeline are published (23).

We used sequences that were fully annotated (both V and J segments assigned), in-frame (i.e., they encode for a functional peptide without stop codons), and with copy number greater than one.

2.7 Analysis

All statistical analysis was performed using R Statistical Software (version 4.0.0). The Cosine similarity was computed with the package "coop" (version 0.6-3) (24). With the Olga tool (25) we computed the generation probability for each CDR3βAA sequence.

T cell repertoires were sub-sampled for equal size (n=1000 CDR3AAβ clones in spleen). CDR nucleotide sequences were replicated according to the UMI count number, and then randomly sampled. The average Renyi scores for each k (k = 0, 0.25, 0.5, 1, 2, 4) were calculated from 30 repeats of this random sampling.

The package "vegan" (version 2.5-7) (26) was used to project the Nonmetric Multidimensional Scaling (27) Epitope-specific TCRs were filtered based on: 1) top 1,000 sequences, and 2) absence in the unbinding-tetramer populations and across multiple epitope-specific types. Only the filtered TCRs were annotated to the to the bulk samples (SI Table 2).

The five amino acid motifs were computed for each CDR3AA by locating the center base and driving from it two additional amino acids from each direction. The amino acids motif sequences logo and charge were calculated with the packages "ggseqlogo" (28) and "Peptides" (29), respectively.

The probability of generation (pGen) for each CDR3AA β chain was commuted using the Olga package (25). The convergent recombination was inferred by counting the number of CDR nucleotide sequences matched V and J segments for each CDR3AA sequence.

SARS-COV-2 expanded TCRs were defined as any TCR which changed significantly between any two time points. The significance boundaries were defined as the maximum TCR abundance which might be observed at time 2, given its abundance at time 1, given Poisson distribution of counts with $p < 0.0001$, to give a false discovery rate of <1 in 1000. TCR abundances are normalized for the total number of TCRs sequenced in each sample and expressed as counts/million. From these maximal values at any time point, we calculated the expanded TCRβ frequency.

2.8 Data availability

All DNA sequences from young and adult mice have been submitted to the Sequence Read Archive under the identifier PRJNA954849. https://www.ncbi.nlm.nih.gov/sra/PRJNA954849.

3 Results

3.1 LCMV infection promotes clonal expansion within the CD8+ and CD4 + effector and CD8+ central memory repertoire

We sequenced the TCR repertoire of naive, central memory and effector memory CD4+ and CD8+ T cells from the spleen and bone marrow of three to four C57BL/6 mice at 8 - and 40-days post LCMV infection (summarized in Figure 1A). The library preparation incorporates molecular identifiers (UMI) for each cDNA molecule, which allows subsequent correction for PCR bias and sequencing error, allowing a robust and quantitative annotation of each sequence in terms of CDR3 sequence and frequency (22, 23, 30); About ~1.89 x10^6 annotated CDR3 nucleotide beta chains were obtained, including a varied number of sequences between compartments, tissues, and infection status (SI Table 1), which positively correlates with the number of sorted cells (SI Figure 1C). Our analysis focuses mainly on the amino acid sequence of the TCR beta complementarity determining region 3 (CDR3βAA), which is the most diverse region of the TCR molecule and is associated with antigen epitope recognition (1).

The abundance distribution profile of the repertoires showed the presence of highly expanded TCRs in the spleen of both CD8+ and CD4+ effector, and in CD8+ central memory T cells 8 days following infection (9, 10). After 40 days of infection, clonal expansion could still be observed in the CD4+ effector, but not the CD8 central memory populations (SI Figure 1D). Clonal expansion following infection can also be captured more quantitatively by the set of Renyi diversities, which are shown in Supplementary Figure 1E.

Overall, the changes in TCR repertoire in memory and effector populations reflect the known rapid proliferative expansion of memory

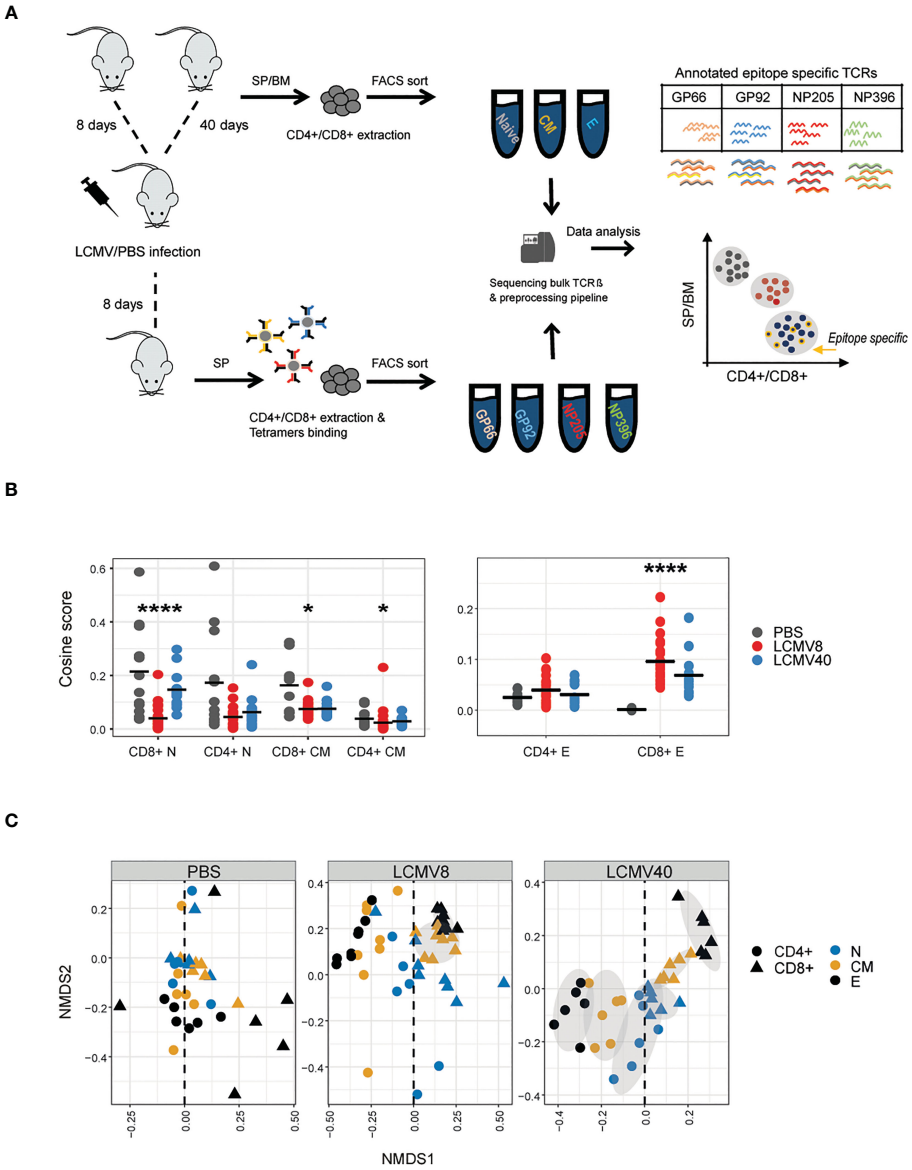


FIGURE 1
LCMV infection promotes organized TCRβ clonal structure of T cell states, mainly in the expanded CD4+ and CD8+ compartments. **(A)** The experimental design: immunizations, T cell isolation, and TCR repertoire sequencing and analysis pipeline. **(B)** T cell effector repertoires increased clonal similarity during LCMV Infection. Cosine similarity between CDR3AAβ across tissues and mice in each T cell state (effector, central memory and naive) and condition (healthy vs. mice after 8- or 40-days post infection). Horizontal black lines show the mean. Significant differences between mice and tissues are denoted in asterisks (p-values: * < 0.01, ****<0.0001 Kruskal-Wallis test, fdr corrections). **(C)** Non-metric multidimensional scaling (NMDS) representation of similarity between repertoires of different compartments. Each dot represents a T cell state (effector, central memory, and naive in black, orange, and blue, respectively), class (CD4+ in circle, CD8+ in triangle) from a single healthy or LCMV-infected mouse. CDR3AAβs distances between mice, tissues, and compartments were calculated using the cosine similarity index and projected on a plane using NDMS. The grey ellipses on the NDMS panel were computed using the normal confidence ellipses.

and effector T cells following infection, providing confidence in the quantitative output of the TCR sequencing pipeline.

3.2 Increased CDR3βAA sharing following LCMV infection

We were interested in the impact of infection on driving convergence (increased sharing) versus divergence (decreased TCR sharing) between repertoires. In order to quantify repertoire overlap, while incorporating TCR abundance, we used the pairwise cosine distance between the abundance vectors for each repertoire (see M&M in (23)) to create a matrix of similarities between all pairs of repertoires. We have previously shown that this measure is highly correlated to the Morisita overlap index. LCMV infection drives increased similarity (i.e. increased overlap) within CD4+ and CD8+ effector repertoires 8 days post-infection (peak response), which decreases towards baseline by day 40 (Figure 1B -right). No such effect was observed in naive or memory populations (Figure 1B -left). An alternative way to visualize the overall pairwise similarity

matrix between all the repertoires is to display the matrix in two-dimensional space using multi-dimensional scaling (Figure 1C). While the PBS immunized mice show a disordered pattern, dominated by highly divergent effector distributions (perhaps reflecting the heterogeneous previous immunological history of each mouse), infection drove a strong pattern of repertoire convergence, with tight segregation between CD4+ and CD8+ repertoires, tightly clustered effector populations furthest away from naive populations and memory populations in between naive and effectors. This overall pattern was maintained at 40 days post-infection, reflecting long-term stable changes to the repertoire organization following infection.

To further validate whether these long-term repertoire organizational changes are driven by common TCRs, we used the same measurements described in Figures 1B, C to evaluate the clonal overlap in mice at different immune states (healthy vs. infected mice at day 8 vs. infected mice at day 40). Indeed, the clonal overlap was increased only in CD4+ and CD8+ effector T cells between day 8 and 40 post-infection and not in the other T cell states and between PBS and infected mice (Figures 2A, B). Thus, the repertoire organizational changes are driven at least in part by shared effectors TCRs detected upon infection.

3.3 Expansion and increased sharing in LCMV– specific TCRs

The increased sharing following infection observed in the data (Figures 1, 2) did not distinguish between antigen-specific or potential bystander T cells activated by the infection. We, therefore, identified a set of antigen-specific TCRs, using tetramer purification and subsequent stringent bioinformatic filtering (see **methods** section), resulting in good reproducibility and high sequence overlap between biological replicates (SI Figure 1F). We used this pipeline to sort and sequence TCRs specific for 3 CD8+ and 1 CD4+ LCMV epitopes from mice at day 8 post-infection (Figure 1A).

A summary of the selected annotated epitope-specific TCRs is presented in Supplemental Table 2. A set of Herpes simplex virus CD8 specific TCRs (31) served as a control for these analyses.

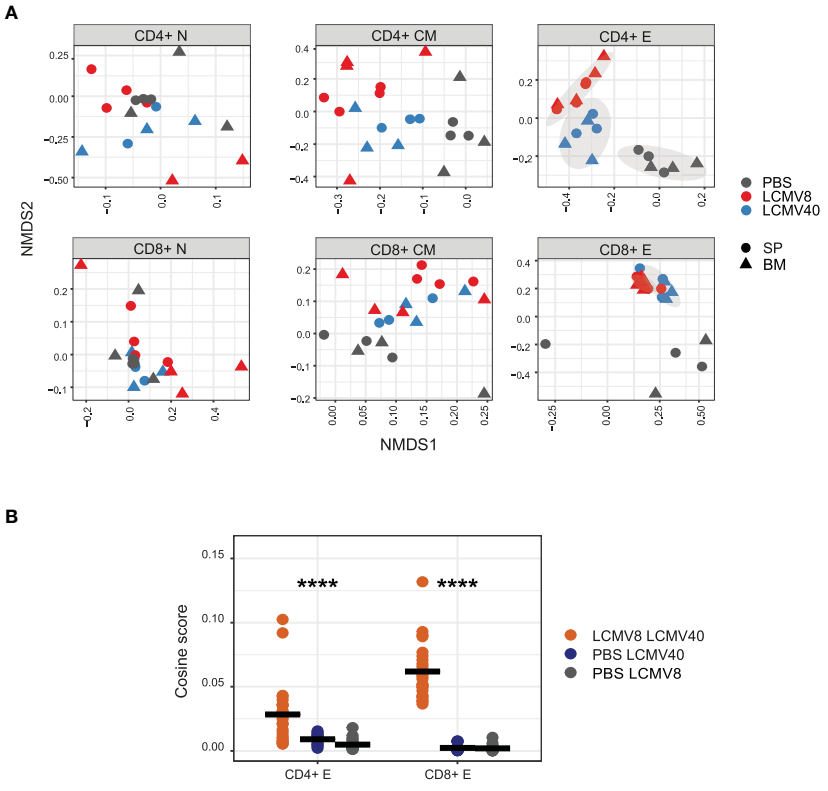


FIGURE 2
LCMV infection induces common long-lasting T cell effector clones. **(A)** Clonal similarity evaluation between infected and uninfected mice in each T cell compartment. Pairwise cosine similarity scores were projected on the NMDS plane for each T cells compartment (sub-plots), tissue (shape), and a single mouse in different conditions (colored dots). Healthy-PBS injected mice are marked in grey dots (PBS), mice 8 days post-infection in red dots (LCMV8), and 40 days post-infection in blue dots (LCMV40). The grey ellipses on the NDMS panel of the CD4+ and CD8+ effector subplots are computed using the normal confidence ellipses. **(B)** CD4+ and CD8+ effector CDR3AAβs are highly shared between mice at day 8 and day 40 post LCMV infection. Cosine similarity was computed between effector CDR3AAβ across tissues and mice in different conditions; day 8- and 40-days post-infection (red dots), 40 days post-infection, and PBS control (blue dots), 8 days post-infection and PBS control (grey dots). All the effector sequences are in the left panel, and the effector epitope-specific clones are in the right panel. The mean is shown in black lines (n=number of paired mice cross treatments and tissues). Significant differences between mice and tissues are denoted in asterisks (p-value ****<0.0001 Kruskal-Wallis test).

We then looked for this set of antigen specific TCRs in the bulk repertories from the different subpopulations of T cells (Figure 3A). Out of the set of epitope-specific CDR3AAβs, a high fraction was found in at least one repertoire from LCMV-infected mice (SI Table 2, out of filtered TCRs: GP66- 43%, GP92- 65%, NP205- 92%, NP396- 64%). As expected, the maximum enrichment of the antigen-specific TCR sequences was seen in the day 8 effector and memory population. At the peak of the infection, day 8, splenic CD8+ effector and memory repertoires contained a higher fraction of NP396 and GP92 specific clones (1-2%) than NP205 clones (~0.5-0.6%), reflecting the known immunodominance hierarchy (19). We did not observe significant enrichment of CD4+ GP66 epitope-specific T cells in the CD4+ effector population. Similarly, we did not observe any significant enrichment of Herpes simplex virus type 1 (HSV1)-specific TCRs in either the effector or memory compartments. We focused on the splenic effector cells, which contained the highest fraction of epitope-specific clones and plotted the abundance profile of the annotated TCRs (Figure 3B). Clonal expansion, as evidenced by the presence of TCRs present at high abundance compared to unimmunized mice was observed for all epitopes and was especially pronounced at the peak of infection. No expansion of HSV1 annotated TCR sequences was observed.

We next examined sharing between the epitope specific TCR repertoires, as described in Figure 1B for the bulk repertoires. We observed a similar increase in repertoire similarity at day 8 post infection (Figure 3C) within the effector T cells for all four epitopes, although the CD4+ changes in the peak of infection were smaller and did not reach statistical significance (Figure 3C). Infection did not alter sharing in the control HSV1-annotated TCR set. Overall, we confirmed that infection induced a concurrent expansion and convergence of TCR sequences in effector cells, including the epitope specific repertoire.

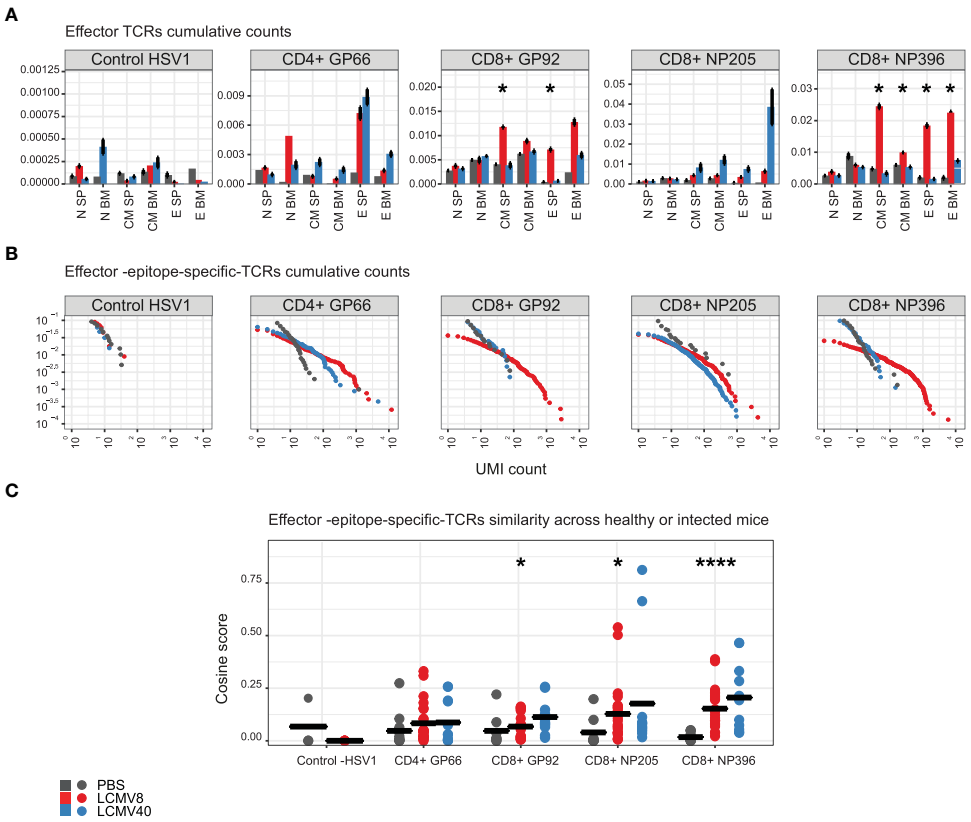


FIGURE 3
Epitope-specific CDR3AAs are mainly found in the effector state of mice after eight days of LCMV infection. The epitope-specific CDR3AAβs are annotated to healthy-PBS injected mice (grey dots and bars), mice at 8 (red dots and bars), or 40 days post LCMV infection (blue dots and bars). Each epitope-specific group is labeled above or on the X-axis. The control epitope-specific sequences are labeled "Control -HSV1". **(A)** The mean fraction of epitope specific CDR3AAβs in each compartment, tissue, and mice condition. Error bars are SEM (n=mice number). Significant differences between mice after 8 days of infection and healthy control mice are denoted by asterisks (p-values: * <0.05, Kruskal-Wallis test). **(B)** The cumulative frequency of the effector - epitope-specific sequence. The plots show the cumulative proportion of the repertoire (y-axis) made up of TCR sequences observed once, twice, etc. (x-axis). Significant differences were obtained between 8 days post infection and PBS treated mice, in effector epitope-specific CD8+ NP396, CD8+ GP92,CD8+ NP205, and CD4+ GP66 cells (p-value=5.4e-9, p-value=6.3e-5, p-value= 1.3e-3, p-value=4.9e-6, respectively, Kolmogorov-Smirnov test). Significant differences were obtained between 40 days post infection and PBS treated mice, in effector - epitope-specific CD8+ NP205 and CD4+ GP66 cells (p-value= 5.0e-4,p-value=4.9e-6, respectively, Kolmogorov-Smirnov test). **(C)** Effector – epitope-specific sequences are highly shared across LCMV infected mice. Cosine similarity scores were calculated for each type of epitope-specific repertoire between mice and tissues. The mean is shown in black lines (n= number of paired mice and tissues). Significant differences between mice and tissues are denoted in asterisks (p-values: * <0.05, **** <0.0001 Kruskal-Wallis test).

3.4 Acute LCMV infection reveals patterns of CDR3 sequence sharing, mainly among the effector T cells of infected mice

We identified 1149 "public" CDR3 sequences which were shared between most T effector repertoires from LCMV infected mice (4-9 mice, Figure 4A). We hypothesized that if these TCR sequences were classical public sequences (6, 32) they would be frequently observed in repertoires of unimmunized mice. We therefore searched for these common TCRs in a repertoire database of 28 uninfected "control" mice investigated previously (6). 1093 CDR3 sequences were detected in the reference cohort and showed very variable degrees of sharing. 481 TCR sequences were shared between 22-28 mice in the reference cohort and defined as classical public TCRs. As expected, these CDR3s were enriched in the uninfected control mice in our experiment (grey bars in Figure 4B).

Out of the shared TCRs (1149) that were not classical public, 668 were defined as "hidden public TCRs". Interestingly, CDR3s detected in less than 14 reference repertoires were significantly enriched in both 8- and 40-days post-infection mice (Figure 4B). A similar pattern was observed in the subset of the 1149 public CDR3s which were also identified as LCMV-specific by tetramer staining, although the number of such CDRs was much smaller (Figure 4B, lower panel). The proportion of the shared LCMV CDR3s which bound HLA-tetramer is shown in Figure 4C. Thus, we conclude that there is a substantial proportion of CDR3s which is highly public when comparing the effector repertoires of LCMV-infected mice but have intermediate levels of sharing in unimmunized repertoires. We refer to these as hidden public CDR3s.

The degree of sharing between repertoires in different individuals is determined in part by the probability of generating a particular TCR during somatic recombination (pGen), which can be inferred from the CDR3 sequence (25). This repertoire bias results in highly frequent naive populations encoded by many different CDR nucleotide sequences (convergent recombination degree - CR). Public CDR3s have been shown to have a much higher pGen, CR and frequencies distribution and shorter lengths than private CDR3s, explaining in part how they can be observed in many independent repertoires. We calculated these measurements for all the CDR3s shared between all LCMV-infected repertoires and stratified them according to their publicity within the control uninfected repertoires (Figures 4D, E). The hidden public CDR3s had pGen and length distributions which lay between that of private and public CDR3s (Figure 4D; SI Figure 2B). Hidden public CDR3s were also detected with intermediate levels of naive frequencies and CR degrees (Figure 4E), suggesting they hold unique repertoire bias properties, which can be fully revealed upon viral expansion. To better understand these dynamic changes, we focused on overlapped clones from effector cells of infected mice and naive cells from healthy mice. This allowed us to follow a clone- trajectory based on the average clonal frequency change from the healthy to day 8 and 40 post-infection (Figure 4F). While public TCRs were reduced, hidden public TCRs increased at 8 days post-infection. After 40 days of infection, the hidden public TCR changed their dynamics, CD4+ reduced, and CD8+ maintained high frequency. We note that private TCRs cannot be linked to this trajectory as they contained unique CDR3AA sequences in each mouse (Figure 4F, marked in blue dashed lines). However, private TCRs can represent a reference, which showed, on average, lower clonal frequency compared to the shared TCRs. Similar patterns were observed in both spleen and bone-marrow tissues, and by computing the repertoire fraction in each public, hidden public, and private populations (SI Figure 2A). The differences between the private, hidden public and public CDR3s were further explored via the physicochemical properties of the CDR3 amino acids.

We visualized the relative contribution of each of the central five amino acids of the CDR3, the region most likely to contact the peptide epitope (33). As shown in Figure 4G, serine is over-represented at the beginning of the sequence in the fully public CDR3s, while both private and hidden public sequences were more diverse (Figure 4G). A lower average basic amino acid was observed in the public and hidden public motifs than in the private motifs (Figure 4H).

3.5 Hidden public TCRs in the context of SARS-COV-2 infection

We hypothesized that hidden public TCRs may emerge more generally as a response to acute infection. We therefore examined the TCR repertoires of 39 individuals who tested PCR positive for SARS-COV-2 during the first wave of the pandemic in the UK (Manisty), as well as 6 individuals who remained PCR negative and seronegative throughout. As described in detail previously in (21), we identified a wave of TCRs which expanded within the first few weeks of infection in most infected individuals.

To compare the level of publicity of these expanding CDR3s between the COVID-infected individuals, and uninfected individuals we utilized a reference cohort of 786 healthy individuals (34), referred to here as the Emerson data set, Figure 5A) collected several years prior to the SARS-COV-2 pandemic. Most of the expanded SARS-COV-2 CDR3 sequences were found in the Emerson data set (59.4%, 2794). Within the expanded set of TCRs we identified a set of classical public TCRβ sequences, which are highly shared across many healthy and SARS-COV-2 infected individuals (92 TCRs shared in more than 65% of individuals in both data sets). However, we also identified a set of CDR3 sequences that are highly shared only among the SARS-COV-2 infected individuals (21 TCRs found shared in more than 65% of SARS-COV-2 infected individuals and below 5.3% of healthy individuals) (Figure 5B). This set of TCRs is analogous to the hidden public TCRs from mice, which were highly shared only among LCMV infected individuals and not in the 28 reference mice. The hidden public TCRs were present at a significantly higher abundance in the repertoires of the SARS-COV-2 infected individuals (13 per million TCR) than the classical public TCRs (4 per million TCR, p-value < 2.2e-16, Wilcoxon test).

We further examined the few hidden public-TCRs which were also detected in the PCR negative individuals and found them to be

10.3389/fimmu.2023.1199064

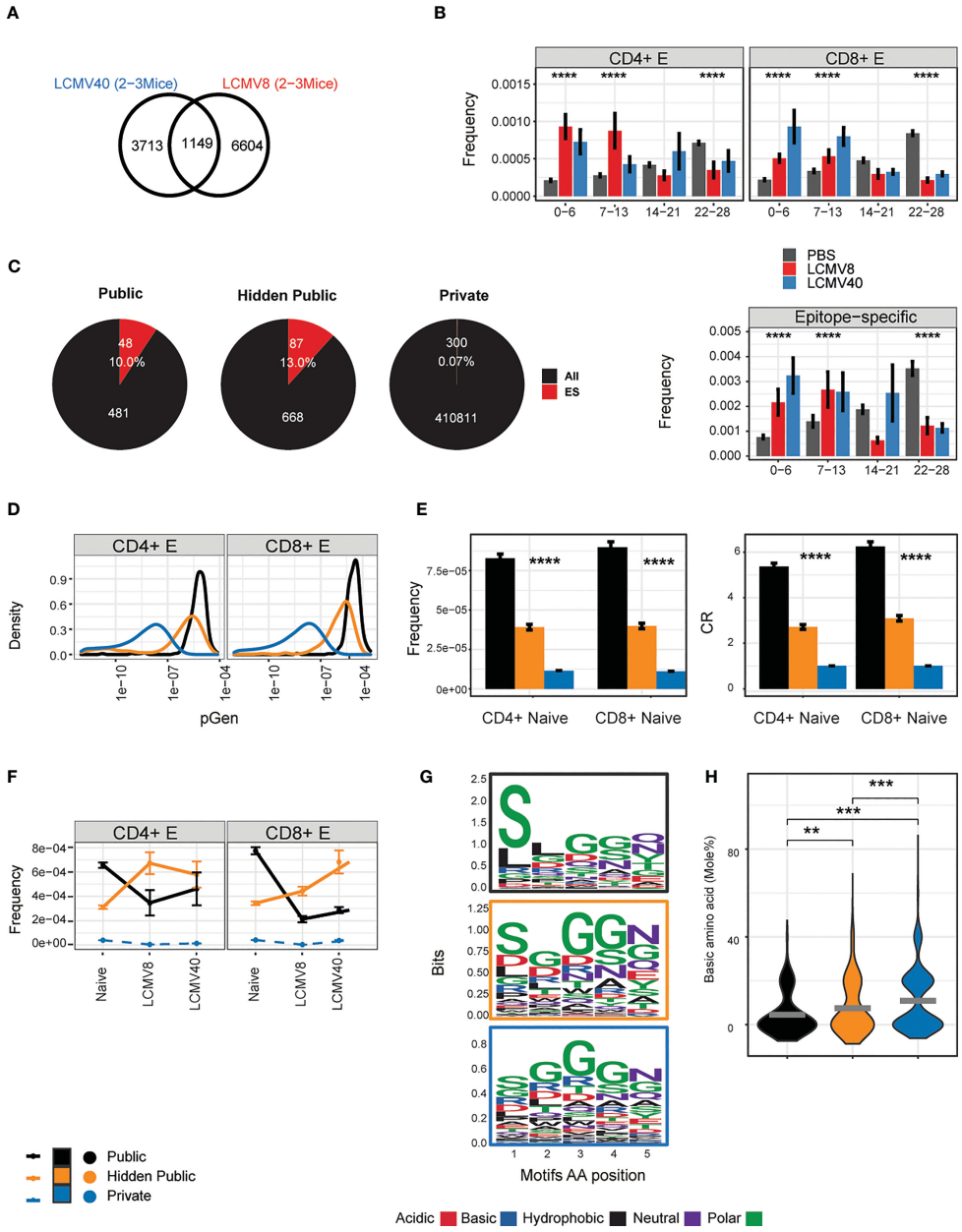


FIGURE 4
Defining properties of LCMV driven- hidden- public clones. **(A)** The number of CD4+ and CD8+ effector CDR3AAβ sequences that overlapped with most mice after 8 (LCMV8) and 40 (LCMV40) days of infection (4-9 mice, 1149," LCMV- long-lasing TCRs"). **(B)** The sharing distribution of LCMV-long-TCRs, across the 28 mice reference cohort. The frequencies of the CD4+ or CD8+ LCMV- long-lasting CDR3βAA (1149) and the epitope-specific sequences among them (lower panel) that were undetected (0) or found in a 1-28 mice reference data set (6). Healthy-PBS injected mice are marked in grey bars (PBS), and mice 8 or 40 days post-infection are marked in red and blue bars, respectively. The frequency was calculated by normalizing the CDR3AA UMI count from each class (CD4+/CD8+) and immune state (PBS/LCMV8/LCMV40) by the total counts in all mice and tissues. Presented in the mean frequency in each sharing group (0-6,7-13,14-21,22-25). Error bars are SEM (n=sequences number). **(C)**. LCMV- long-lasing TCRs (n=1149, A) are divided into two groups according to the sharing hierarchy found in the reference data set: 1) public TCRs shared by 22-28 mice, 2) hidden public TCRs undetected (0) or found shared by 1-21 mice. CD4+ and CD8+ effector TCRs that are termed private are sequences that appeared in one mouse from the current dataset and not in the reference cohort. The total ("All") and the epitope specific TCRs ("ES") number and fraction are marked in white text and red color. **(D)** The probability generation (pGen) scores and CDR3AAβ for each CD4+ and CD8+ TCRs population. **(E)** CD4+ and CD8+ naïve precursor frequency and convergent recombination (CR) mean number across public, hidden, and private TCRs population. Error bars are SEM (n=sequences number). **(F)** Clonal evolution from the naïve state to 8 up to 40 days post-infection. For each TCRs population (out of 1093 TCRs) in the different immune states, points represent the mean frequency. The connected lines describe the clone time-based trajectory. Private CDR3AA in each immune state were subsampled (500) to avoid the size variation between the TCRs populations. The dashed lines represent the private population's unique CDR3AA sequences in each immune state trajectory. **(G)** Chemical properties of the five amino acid motifs from the public, hidden public, and private (indicated by the frame colors). Significant differences were obtained between pGen distribution of hidden public TCRs and public TCRs and between hidden public TCRs and private TCRs (p-value < 2.2e-16, Kolmogorov-Smirnov test). **(H)** Each point represents a basic (H + K + R) amino acid mole percentage in each 5AA motif of the public, hidden public or private TCRs populations. The mean is shown in (n=number CDR3AAs in each group). Significant differences between public, hidden public and private TCRs are denoted in asterisks (p-values: ** < 0.01, *** <0.001, **** <0.0001 Kruskal-Wallis test).

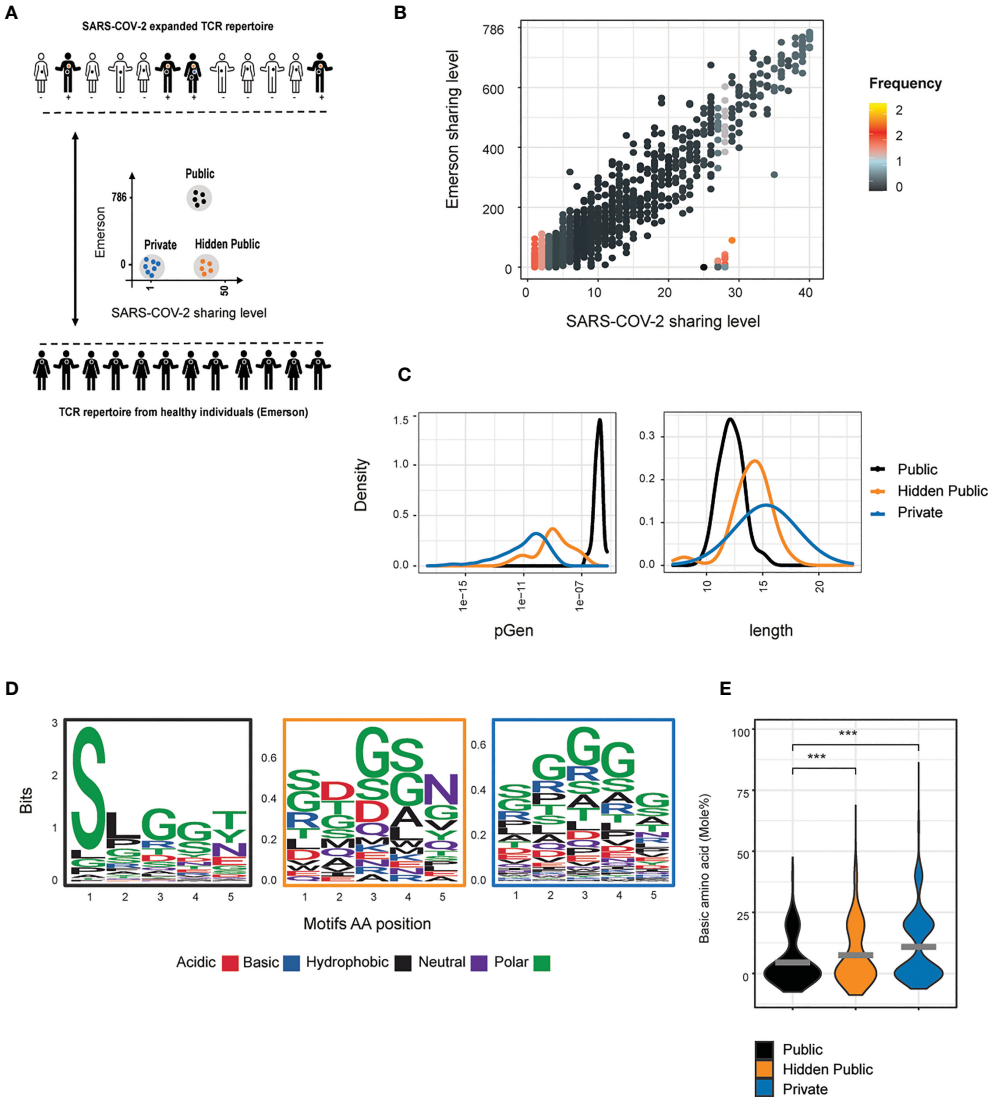


FIGURE 5
Hidden public TCRs revealed in SARS-COV-2 patients. **(A)** An overview of the collected data and analysis design **(B)** Comparison between the sharing levels of CDR3 sequences found across individuals from the Covid and the Emerson data sets. The color represents the log 10 median frequency of all CDR3AAs in each Covid sharing level (high= orange, low = black). Three TCRs populations were defined: 1) Public TCRs highly shared in both data sets (above 524 and 26 individuals in the Emerson and Covid patients, respectively, 92 CD3AAs). 2) "Hidden public" TCRs which were highly shared only among the SARS-COV-2 cohort (above 26 and below 50 individuals from the Covid and Emerson cohort, respectively, 21 CD3AAs). 3) Private TCRs exclusively detected in one patient from the Covid data set. **(C)** The probability generation scores, CDR3AAβ length distributions in reach of the defined population. **(D, E)** The chemical property of the 5 middle amino acid motifs in each of the defined populations. **(D)** Amino acid sequences logo. **(E)** Each point represents the mole percentage of basic (H + K + R) amino acid in each public, hidden public or private motifs. The mean is shown in (n=number CDR3AAs in each group). Significant differences between public, hidden public and private TCRs are denoted in asterisks (p-value *** <0.001 Kruskal-Wallis test).

present at significantly lower abundance than in the PCR positive individuals (7 CDR3AA with frequency means of 17.1 vs. 1.17 PCR positive vs. negative individuals, p-value < 2.2e-16, Wilcoxon test)(SI Figure 2C). The increased abundances in the PCR positive individuals support their association with antigen-driven expansion.

SARS-COV-2 driven hidden public TCR were also found in an additional higher resolution independent dataset, generated from 39 individuals prior to the SARS-COV-2 pandemic (35). This dataset has an average 2.2-fold higher number of TCRβ per individual (409519), in comparison to the Emerson data set (183211). Here as well, the SARS-COV-2 associated hidden public sequences showed intermediate abundancy, levels between public and private TCRs (SI Figure 2D).

The SARS-COV-2- associated hidden public CDR3s were found to have pGen and length distributions intermediate between the public and the private CDR3s (Figure 5C), as we observed for the LCMV hidden public sequences. Lastly, we calculated the central five amino acids usage and their average percentage of basic amino acids (Figure 5D). Public CDR3s showed a more constrained amino acid usage pattern than the private and hidden public CDR3s, and the public and hidden public motifs showed lower average scores of basic amino acids than the private motifs (Figure 5E).

Taken together, these results suggest that hidden public CDR3 sequences, with distinct properties from classical public CDR3s can be observed in different acute viral infections and host species. Thus, this phenomenon may be a generalized feature of the adaptive immune system, revealing some unexpected constraints on the diversity generated by somatic recombination in T cells.

4 Discussion

The well-characterized model of acute LCMV infection allowed us to probe the reactive T cell repertoire during the peak and memory phases of the viral infection. We demonstrated that viral infection drove convergent evolution in the TCR repertoire, which could be detected in both the total and the antigen-specific effector compartment. Convergence was driven by the expansion of a set of shared CDR3 sequences, which could only be detected after the antigen-specific response. These antigen-dependent shared CDR3s were seen less often than classical "public" CDR3s in unimmunized repertoires, consistent with their lower probability of generation. These observations suggest that the degree of sharing between individuals is greater than was previously thought, but that many of the shared sequences are "hidden" by being present at low abundance in the naive repertoire, and are therefore not observed in typical sampling of unimmunized mice, which sequence only a tiny proportion of the total repertoire. Strikingly, hidden public TCRs were also identified in SARS-COV-2 infected individuals, supporting the notion that these findings represent a broader and conserved phenomenon.

We examined in greater detail a subset of shared LCMV-dependent effector T cells which persisted in the repertoire until at least day 40 post-immunization. We searched for these TCRs in an independent cohort of 28 antigen-naïve mice (6). These persistent shared CDR3s were found in zero to six of these control repertoires, defining a new intermediate level of publicity. We hypothesize that the "hidden public" TCRs originate from naive cells which are generated at a sufficiently high frequency to be present in many naive repertoires, but are present at low abundance in the naive repertoire, resulting in them not being detected in routine TCR sampling. However, following infection, T cells expressing these shared CDR3 consistently expand and differentiate into effector cells as a result of exposure to LCMV peptides. As a consequence, their abundance reaches a critical level at which they are consistently detected in the repertoire samples we analyze. Consistent with this hypothesis, we find that the "hidden public" CDR3s have higher naive precursor frequencies, more convergent recombination, and higher generation probabilities than random sets of CDR3s (which are mostly private to a single mouse and compartment). However, they have lower levels of these metrics than classical "public" CDR sequences.

The differences between the public and "hidden public" CDR3s may reflect different functional properties. Indeed, while public TCRs were shown to be more self-immunity-associated (6), the hidden public TCRs react to viral infections. Although the mechanisms remain incompletely understood, increasing levels of naive precursor T cell frequencies have been shown to drive more significant peptide MHC responding capacities (19, 31). The range of naive precursor frequencies and the phenotype heterogeneity (36) has yet to be fully determined but might explain the hidden public pre-exposure antigenic preferences.

The hidden public TCRs appear to be a broader phenomenon found also in other viral infections and species. The first SARS-COV-2 pandemic wave offered a good model for a primary viral infection in humans. We searched the expanded SARS-COV-2 TCRs (37) in a large cohort of healthy humans, and detected a set of TCRs that were highly shared across SARS-COV-2 infected individuals but showed less publicity in a cohort of pre-pandemic TCR repertoires. The detection of sharing even in the genetically heterogeneous HLA-diverse human setting is interesting, and will merit further study. Similar to the LCMV hidden public population, these TCRs had intermediate generation probabilities.

We investigated whether the "hidden public" CDR3s also showed distinct amino acid composition, which might explain their more frequent selection in the thymus (38) or their higher abundance in the naive repertoires. Since these hidden public TCRs originated from a diverse set of HLA genotype, we focused on the five amino acid middle of the CDR3AA, a region associated with binding the peptide within the MHC complex (33). The Covid and LCMV hidden public motifs showed higher amino acid diversity than the public motifs. In addition, we found that public and hidden public motifs tend to include less positively charged amino acids compared to private motifs, suggesting they hold conserved binding properties. We can speculate that the hidden public amino acid constraints might provide an evolutionary cross-reactive advantage, allowing them to react to foreign and self-antigens (39). However, further study is required to better understand the developmental process, driving the generation preference of the hidden public TCRs.

The study we present here has several limitations. The number of individuals analyzed and epitope-specific sequences were relatively small, limiting the amount of robust statistical analysis that could be carried out. Another limitation is that the analysis of the post-infection repertoires was limited to two time points. We also recognize that the effector functional state we defined was based on a rather simplistic and limited panel of cell surface markers, which could result in heterogeneous effector memory phenotypic states, especially at late post-infection time. In addition, the bulk TCRβ chain analysis cannot capture the absolute clonal identity which comprises paired α and β chains. The TCR α chain is less diverse and can be expressed twice (Dual TCRα) in virus-specific CD4+ and CD8+ T cells during acute responses (up to 60%) (40), highlighting the complexity of using the TCRα chain as a clone identifier.

This study describes a naive precursor population carrying a shared set of CDR3s capable of providing a rapid response to viral infections. We coin the term "hidden public" to describe this population. Our results suggest that the TCR repertoire may be more constrained, and hence more similar between individuals, than current dogma supposes. Deeper understanding of the processes which shape this repertoire, and determine the level of inter-individual sharing is important for understanding the antiviral response and in rational design of next-generation vaccines.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.ncbi.nlm.nih.gov/sra/PRJNA954849.

Ethics statement

The animal study was reviewed and approved by Council for experiments on animals (IACUC) Weizmann institute. Written informed consent was not obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

Author contributions

MM designed the study, prepared and analyzed the data, and wrote the manuscript. SR-Z: 1. AB: 2. BC: 3. NF: 1,3. AM: 1,3. Contributed with: 1. Design and conception of the study 2. Experimental preparation of the data 3. Data analysis 4. Writing the manuscript. All authors contributed to the article and approved the submitted version.

Funding

AM was supported by The Alon fellowship for outstanding young scientists, Israel Council for Higher Education, from the Israel Science Foundation (1700/21). NF was supported by the Applebaum Family Foundation.

Acknowledgments

This study was initiated by our friend, mentor, and colleague NF. The phenomena of public T cell receptors captured NF's imagination, and much effort was devoted to further uncovering its mysteries.

In memoriam

Sadly, NF died after a long battle with illness before the study could be completed. We have attempted to complete this study in the spirit in which it was undertaken but acknowledge that we raised more questions than answered. We dedicate this study to his memory. We greatly appreciate that the Armstrong LCMV viral strain was a kind gift from Prof. Matteo Iannacone (San Raffaele Institute).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2023.1199064/full#supplementary-material

SUPPLEMENTARY FIGURE 1
(A) Representative sorting gates of CD4+ cells from one mouse in each condition (PBS/LCMV8/LCMV40). **(B)** Representative sorting gates from CD8+ T cells specific for NP205 peptide. From the CD8+ population, 2.86% are positive for the MHC class I NP205 tetramer (lower right panel) and almost all CD4+ cells are negative (0.018%, lower left panel). The CD8+ cells that are negative for the tetramer were also sorted and analyzed. **(C)** The number of UMIs correlates with the sorted cell number. Dots correspond to the sum of UMI count versus the sorted cell number in mice 8 days post LCMV infection. **(D)** Cumulative frequency distributions in CD8+ and CD4+ naive central memory and effector repertoires from spleen and bone marrow. Healthy control, and mice after 8- or 40-days of infection are marked in colored dots (gray, red and blue dots, respectively). Significant differences were obtained between day 8 post-infection and PBS treated mice, in the bone-marrow CD8+ and CD4+ effectors (p-value < 2.2e-16, p-value=3.1e-6, respectively, Kolmogorov-Smirnov test) and in the following splenic compartments: CD8+ central memory CD8+ effector, CD4+ effector (p-value < 2.2e-16, Kolmogorov-Smirnov test). Significant differences were obtained between day 40 post-infection and PBS treated mice, in splenic and bone-marrow CD4+ effector (p-value =1.3e-9 and 2.2e-11, respectively, Kolmogorov-Smirnov test) and bone-marrow CD8+ central memory and CD8 + effector (p-value =3.7e-11 and 8.9e-4, Kolmogorov-Smirnov test). **(E)** The Renyi diversities of order 0, 0.25, 0.5, 1, 2, 4 Renyi values were computed from sequence frequencies at equal sizes (1000 in the spleen and 100 in the bone marrow), averaging values over 100 repeated samplings. Each color represents one CD4 + or CD8+ compartment from one mouse in a single condition. See legend for symbols and color code. **(F)** NP396 epitope-specific clones from two biological repetitions are positively correlated in the obtained UMI counts. Each point is the UMI count of a single CDR3AAβ found in the two repetitions (Rep1/Rep2).

SUPPLEMENTARY FIGURE 2
(A) Similar frequencies in spleen and bone marrow tissues of a single mouse, immune state (PBS/LCMV8/LCMV40), CD4+ or CD8+ class, and TCRs population. Frequencies are calculated by the sum of UMI counts per TCRs

population (public/hidden public/private) divided by the total UMI count sum in each mouse, immune state, tissue, and T cell class. R2 coefficient scores are marked in each subplot. **(B)** CDR3AAβ length distributions in reach of the defined population. **(C)** The frequency of hidden public TCRs found in healthy individuals (7 CDR3AAs, PCR negative) and SARS-COV-2 infected individuals (PCR positive). Mean values are marked in black lines. Significant differences are marked in p value (Wilcoxon test). (D)SARS-COV-2 –associated hidden public TCRs detected in high resolution pre- SARS-COV-2 pandemic dataset. The three-populations identified SARS-COV-2 individuals were searched in the Britanova et al. data set (35). Each bar represents the mean frequency of SARS-COV-2 associated- public TCRs (all 92 detected, black bar), or hidden public TCRs (7 out of 21 detected, orange bar) and private TCRs (blue bars). Error bars are SEM (n=sequences number). Significant differences between public, hidden public and private TCRs are denoted in asterisks (p-values: *< 0.05, *** <0.001, Kruskal-Wallis test).

SUPPLEMENTARY TABLE 1

Number of CDR3NT and CDR3AA sequences, and sum of UMI counts for each T cell state ("ClassState"), tissue ("Tissue"), and mouse number ("Mice") in healthy ("PBS") or infected conditions (8 or 40 days post- LCMV infection, LCMV8 and LCMV40, respectively).

SUPPLEMENTARY TABLE 2

Number of CDR3AA sequences and sum of UMIs for each epitope-specific type during the filtering process. The sequences were obtained from the tetramer isolation experiment (see Methods section, "All" column). Epitope-specific TCRs were filtered based on: 1) top 1,000 sequences, and 2) absence in the unbinding-tetramer populations and across multiple epitope-specific types ("Filtered"). The number of filtered epitope-specific TCRs found in the bulk samples (LCMV40, LCMV8, PBS, and multiple T cell states) is presented in the "Found in bulk" column.

References

1. Davis MM, Bjorkman PJ. T-Cell antigen receptor genes and T-cell recognition. *Nature* (1988) 334:395–402. doi: 10.1038/334395a0

2. Robins HS, Campregher PV, Srivastava SK, Wacher A, Turtle CJ, Kahsai O, et al. Comprehensive assessment of T-cell receptor beta-chain diversity in alphabeta T cells. *Blood* (2009) 114:4099–107. doi: 10.1182/blood-2009-04-217604

3. Iglesias MC, Almeida JR, Fastenackels S, van Bockel DJ, Venturi V, Gostick E, et al. Escape from highly effective public CD8 + T-cell clonotypes by HIV escape from highly effective public CD8 ƨT-cell clonotypes by HIV. *Blood* (2011) 118:2138–49. doi: 10.1182/blood-2011-01-328781

4. Becher LRE, Nevala WK, Sutor SL, Abergel M, Hoffmann MM, Parks CA, et al. Public and private human T-cell clones respond differentially to HCMV antigen when boosted by CD3 copotentiation. *Blood Adv* (2020) 4:3–5. doi: 10.1182/bloodadvances.2020002255

5. Gordin M, Philip H, Zilberberg A, Gidoni M, Margalit R, Id CC, et al. Breast cancer is marked by specific, public T-cell receptor CDR3 regions shared by mice and humans. (2021). *PLoS Comput Biol* 17(1):e1008486 doi: 10.1371/journal.pcbi.1008486

6. Madi A, Shifrut E, Reich-Zeliger S, Gal H, Best K, Ndifon W, et al. T-Cell receptor repertoires share a restricted set of public and abundant CDR3 sequences that are associated with self-related immunity. *Genome Res* (2014) 24:1603–12. doi: 10.1101/gr.170753.113

7. Huisman W, Hageman L, Leboux DAT, Khmelevskaya A, Efimov GA, Roex MCJ, et al. Public T-cell receptors (TCRs) revisited by analysis of the magnitude of identical and highly-similar TCRs in virus-specific T-cell repertoires of healthy individuals. *Front Immunol* (2022) 13:851868. doi: 10.3389/fimmu.2022.851868

8. Uchida AM, Boden EK, James EA, Shows DM, Konecny AJ, Lord JD. Escherichia coli-specific CD4+ T cells have public T-cell receptors and low interleukin 10 production in crohn's disease. *Cell Mol Gastroenterol Hepatol* (2020) 10:507–26. doi: 10.1016/j.jcmgh.2020.04.013

9. Masopust D, Murali-Krishna K, Ahmed R. Quantitating the magnitude of the lymphocytic choriomeningitis virus-specific CD8 T-cell response: it is even bigger than we thought. *J Virol* (2007) 81:2002–11. doi: 10.1128/JVI.01459-06

10. McDermott DS, Varga SM. Quantifying antigen-specific CD4 T cells during a viral infection: CD4 T cell responses are larger than we think. *J Immunol* (2011) 187:5568–76. doi: 10.4049/jimmunol.1102104

11. Youngblood B, Hale JS, Kissick HT, Ahn E, Xu X, Wieland A, et al. Effector CD8 T cells dedifferentiate into long-lived memory cells. *Nature* (2017) 552:404–9. doi: 10.1038/nature25144

12. Homann D, Teyton L, Oldstone MBA. Differential regulation of antiviral T-cell immunity results in stable CD8+ but declining CD4+ T-cell memory. *Nat Med* (2001) 7:913–9. doi: 10.1038/90950

13. Whitmire JK, Murali-Krishna K, Altman J, Ahmed R. Antiviral CD4 and CD8 T-cell memory: differences in the size of the response and activation requirements. *Philos Trans R Soc B Biol Sci* (2000) 355:373–9. doi: 10.1098/rstb.2000.0577

14. Oxenius A, Bachmann MF, Zinkernagel RM, Hengartner H. Virus-specific MHC-class II-restricted TCR-transgenic mice: effects on humoral and cellular immune responses after viral infection. *Eur J Immunol* (1998) 28:390–400. doi: 10.1002/(SICI)1521-4141(199801)28:01<390::AID-IMMU390>3.0.CO;2-O

15. Brooks DG, Teyton L, Oldstone MBA, McGavern DB. Intrinsic functional dysregulation of CD4 T cells occurs rapidly following persistent viral infection. *J Virol* (2005) 79:10514–27. doi: 10.1128/JVI.79.16.10514-10527.2005

16. Dow C, Oseroff C, Peters B, Nance-Sotelo C, Sidney J, Buchmeier M, et al. Lymphocytic choriomeningitis virus infection yields overlapping CD4+ and CD8+ T-cell responses. *J Virol* (2008) 82:11734–41. doi: 10.1128/JVI.00435-08

17. van der Most R. Changing immunodominance patterns in antiviral CD8 T-cell responses after loss of epitope presentation or chronic antigenic stimulation. *Virology* (2003) 315:93–102. doi: 10.1016/S0042-6822(03)00594-4

18. Van Der Most RG, Murali-Krishna K, Whitton JL, Oseroff C, Alexander J, Southwood S, et al. Identification of db- and kb-restricted subdominant cytotoxic T-cell responses in lymphocytic choriomeningitis virus-infected mice. *Virology* (1998) 240:158–67. doi: 10.1006/viro.1997.8934

19. Kotturi MF, Scott I, Wolfe T, Peters B, Sidney J, Cheroutre H, et al. Naive precursor frequencies and MHC binding rather than the degree of epitope diversity shape CD8+ T cell immunodominance. *J Immunol* (2008) 181:2124–33. doi: 10.4049/jimmunol.181.3.2124

20. Biram A, Winter E, Denton AE, Zaretsky I, Dassa B, Bemark M, et al. B cell diversification is uncoupled from SAP-mediated selection forces in chronic germinal centers within peyer's patches. *Cell Rep* (2020) 30:1910–1922.e5. doi: 10.1016/j.celrep.2020.01.032

21. Chandran A. Rapid synchronous type 1 IFN and virus-specific T cell responses characterize first wave non-severe SARS-COV-2 infections ll ll rapid synchronous type 1 IFN and virus-specific T cell responses characterize first wave non-severe SARS-COV-2 infectio. *Cell Rep Medicine* (2022) 3(3), 100557 doi: 10.1016/j.xcrm.2022.100557

22. Oakes T, Heather JM, Best K, Byng-Maddick R, Husovsky C, Ismail M, et al. Quantitative characterization of the T cell receptor repertoire of naive and memory subsets using an integrated experimental and computational pipeline which is robust, economical, and versatile. *Front Immunol* (2017) 8:1267. doi: 10.3389/fimmu.2017.01267

23. Mark M, Reich-Zeliger S, Greenstein E, Reshef D, Madi A, Chain B, et al. A hierarchy of selection pressures determines the organization of the T cell receptor repertoire. *Front Immunol* (2022) 13:939394. doi: 10.3389/fimmu.2022.939394

24. Schmidt D. (2019), Package T, Operations S. Co-Operation: fast correlation, covariance, and cosine similarity. Available at: https://cran.r-project.org/package=coop.

25. Sethna Z, Elhanati Y, Callan CG, Walczak AM, Mora T. OLGA: fast computation of generation probabilities of B- and T-cell receptor amino acid sequences and motifs. (2019) *Bioinformatics* 35(17):2974–81. doi: 10.1093/bioinformatics/btz035

26. Dixon P. VEGAN, a package of r functions for community ecology. *J Veg Sci* (2003) 14:927–30. doi: 10.1111/j.1654-1103.2003.tb02228.x

27. Faith DP, Minchin PR, Belbin L. Compositional dissimilarity as a robust measure of ecological distance. *Vegetatio* (1987) 69:57–68. doi: 10.1021/ja00731a055

28. Wagih O. ggseqlogo: a versatile R package for drawing sequence logos. *Bioinformatics* (2017). Available at: https://doi.org/10.1093/bioinformatics/btx469.

29. Osorio D, Rondón-Villarreal P, Torres R. Peptides: a package for data mining of antimicrobial peptides. *R J* (2015) 7:4. doi: 10.32614/RJ-2015-001

30. Uddin I, Woolston A, Peacock T, Joshi K, Ismail M, Ronel T, et al. Quantitative analysis of the T cell receptor repertoire. *Methods Enzymol* (2019) 629:465–92. doi: 10.1016/bs.mie.2019.05.054

31. Wallace ME, Bryden M, Cose SC, Coles RM, Schumacher TN, Brooks A, et al. Junctional biases in the naive TCR repertoire control the CTL response to an immunodominant determinant of HSV-1. *Immunity* (2000) 12:547–56. doi: 10.1016/s1074-7613(00)80206-x

32. Madi A, Poran A, Shifrut E, Reich-Zeliger S, Greenstein E, Zaretsky I, et al. T Cell receptor repertoires of mice and humans are clustered in similarity networks around conserved public CDR3 sequences. *Elife* (2017) 6:1–17. doi: 10.7554/eLife.22057

33. Egorov ES, Kasatskaya SA, Zubov VN, Izraelson M, Nakonechnaya TO, Staroverov DB, et al. The changing landscape of naive T cell receptor repertoire with human aging. *Front Immunol* (2018) 9:1618. doi: 10.3389/fimmu.2018.01618

34. Emerson RO, DeWitt WS, Vignali M, Gravley J, Hu JK, Osborne EJ, et al. Immunosequencing identifies signatures of cytomegalovirus exposure history and

HLA-mediated effects on the T cell repertoire. *Nat Genet* (2017) 49:659–65. doi: 10.1038/ng.3822

35. Britanova OV, Shugay M, Merzlyak EM, Staroverov DB, Putintseva EV, Turchaninova MA, et al. Dynamics of individual T cell repertoires: from cord blood to centenarians. *J Immunol* (2016) 196:5005–13. doi: 10.4049/jimmunol.1600005

36. van den Broek T, Borghans JAM, van Wijk F. The full spectrum of human naive T cells. *Nat Rev Immunol* (2018) 18:363–73. doi: 10.1038/s41577-018-0001-y

37. Milighetti M, Peng Y, Tan C, Mark M, Nageswaran G, Byrne S, et al. Large Clones of pre-existing T cells drive early immunity against SARS-COV-2 and LCMV infection. *bioRxiv* (2022) 2022. doi: 10.1101/2022.11.08.515436

38. Lu J, Van Laethem F, Bhattacharya A, Craveiro M, Saba I, Chu J, et al. Molecular constraints on CDR3 for thymic selection of MHC-restricted TCRs from a random pre-selection repertoire. *Nat Commun* (2019) 10:1019. doi: 10.1038/s41467-019-08906-7

39. Nelson RW, Beisang D, Tubo NJ, Dileepan T, Wiesner DL, Nielsen K, et al. T Cell receptor cross-reactivity between similar foreign and self peptides influences naive cell population size and autoimmunity. *Immunity* (2015) 42:95–107. doi: 10.1016/j.immuni.2014.12.022

40. Yang L, Jama B, Wang H, Labarta-Bajo L, Zúñiga EI, Morris GP. TCRα reporter mice reveal contribution of dual TCRα expression to T cell repertoire and function. *Proc Natl Acad Sci* (2020) 117:32574–83. doi: 10.1073/pnas.2013188117

frontiers | Frontiers in Immunology

TYPE Original Research
PUBLISHED 27 June 2023
DOI 10.3389/fimmu.2023.1160260

OPEN ACCESS

EDITED BY
Joe Hou,
Fred Hutchinson Cancer Research Center,
United States

REVIEWED BY
Sunil Gairola,
Serum Institute of India, India
Xiongye Xiao,
University of Southern California,
United States
Sahil Jain,
Tel Aviv University, Israel

*CORRESPONDENCE
Foysal Ahammad
foah48505@hbku.edu.qa
Farhan Mohammad
mohammadfarhan@hbku.edu.qa

RECEIVED 07 February 2023
ACCEPTED 30 May 2023
PUBLISHED 27 June 2023

CITATION
Imon RR, Samad A, Alam R, Alsaiari AA, Talukder MEK, Almehmadi M, Ahammad F and Mohammad F (2023) Computational formulation of a multiepitope vaccine unveils an exceptional prophylactic candidate against Merkel cell polyomavirus. *Front. Immunol.* 14:1160260. doi: 10.3389/fimmu.2023.1160260

Computational formulation of a multiepitope vaccine unveils an exceptional prophylactic candidate against Merkel cell polyomavirus

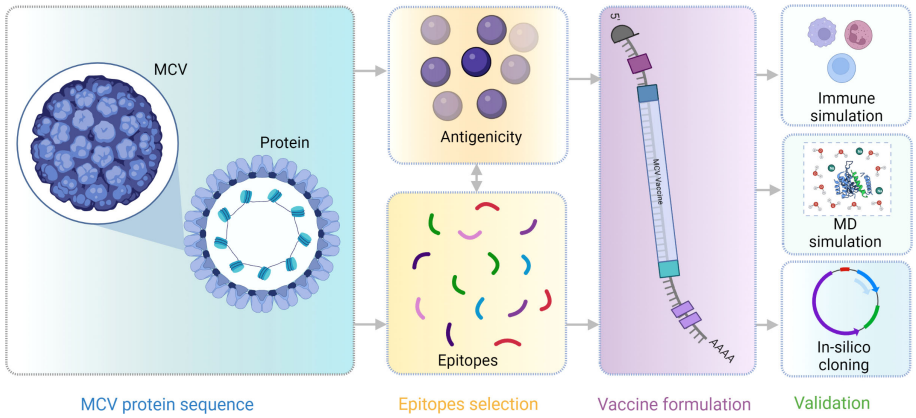
Raihan Rahman Imon[1,2], Abdus Samad[1,2], Rahat Alam[1,2], Ahad Amer Alsaiari[3], Md. Enamul Kabir Talukder[1,2], Mazen Almehmadi[3], Foysal Ahammad[1,4]* and Farhan Mohammad[4]*

[1]Laboratory of Computational Biology, Biological Solution Centre (BioSol Centre), Jashore, Bangladesh, [2]Department of Genetic Engineering and Biotechnology, Jashore University of Science and Technology, Jashore, Bangladesh, [3]Clinical Laboratories Science Department, College of Applied Medical Science, Taif University, Taif, Saudi Arabia, [4]Division of Biological and Biomedical Sciences (BBS), College of Health and Life Sciences (CHLS), Hamad Bin Khalifa University (HBKU), Doha, Qatar

Merkel cell carcinoma (MCC) is a rare neuroendocrine skin malignancy caused by human Merkel cell polyomavirus (MCV), leading to the most aggressive skin cancer in humans. MCV has been identified in approximately 43%–100% of MCC cases, contributing to the highly aggressive nature of primary cutaneous carcinoma and leading to a notable mortality rate. Currently, no existing vaccines or drug candidates have shown efficacy in addressing the ailment caused by this specific pathogen. Therefore, this study aimed to design a novel multiepitope vaccine candidate against the virus using integrated immunoinformatics and vaccinomics approaches. Initially, the highest antigenic, immunogenic, and non-allergenic epitopes of cytotoxic T lymphocytes, helper T lymphocytes, and linear B lymphocytes corresponding to the virus whole protein sequences were identified and retrieved for vaccine construction. Subsequently, the selected epitopes were linked with appropriate linkers and added an adjuvant in front of the construct to enhance the immunogenicity of the vaccine candidates. Additionally, molecular docking and dynamics simulations identified strong and stable binding interactions between vaccine candidates and human Toll-like receptor 4. Furthermore, computer-aided immune simulation found the real-life-like immune response of vaccine candidates upon administration to the human body. Finally, codon optimization was conducted on the vaccine candidates to facilitate the *in silico* cloning of the vaccine into the pET28+(a) cloning vector. In conclusion, the vaccine candidate developed in this study is anticipated to augment the immune response in humans and effectively combat the virus. Nevertheless, it is imperative to conduct *in vitro* and *in vivo* assays to evaluate the efficacy of these vaccine candidates thoroughly. These evaluations will provide critical insights into the vaccine's effectiveness and potential for further development.

KEYWORDS

Merkel cell polyomavirus (MCV), Merkel cell carcinomas (MCC), immunoinformatics, vaccine design, multiepitope vaccine, molecular dynamics simulation (MD), molecular docking



GRAPHICAL ABSTRACT

1 Introduction

Merkel cell polyomavirus (MCV) is one of the seven currently known human oncoviruses in the human polyomaviruses (HPV) family. It has drawn massive attention due to its link to rare human cancer. The virus induces cancer in its natural host and is a primary agent known to cause Merkel cell carcinoma (MCC) (1, 2). MCV is a causative agent in approximately 43%–100% of MCCs, leading to more incidences in aged and immunocompromised patients (3). MCC, which is an aggressive type of skin cancer, was first described by Cyril Toker in 1972 (1, 4). He discovered that the development of neuroendocrine carcinoma of the skin, also referred to as a "trabecular tumor of the skin," is associated with MCV infection. The viral infection triggers an abnormal increase in Merkel cells (MCs) and skin mechanoreceptor cells, leading to uncontrolled proliferation (5). The MCs are found deep in the epidermis of the top layer of the skin as innervated clusters of cells close to the nerve endings receiving touch and pressure sensations (6). The MCC is considered the second deadliest form of skin cancer after malignant melanoma, with a mortality rate of 35% (7). Skin cancer ranks as the 17th most prevalent cancer globally and one of the most diagnosed cancers worldwide. In the United States alone, an estimated 9,500 new cases of skin cancer are diagnosed daily (8). Particularly, MCC contributes to approximately 700 annual fatalities (9). However, the etiology and pathogenesis of MCC remain elusive (10, 11).

MCV is a small, circular, non-enveloped, double-stranded DNA virus highly prevalent in humans and causes skin malignancy (12). The virus is classified within the ortho-polyomaviruses family, which encompasses various mammalian polyomaviruses, including simian virus (SV40), murine polyomavirus, and the human BK polyomaviruses and John Cunningham virus (13–16). The prototype genomic sequence of MCV encodes characteristic polyomavirus proteins from opposite strands, including early genes encoding large T antigen and small T antigen and late genes encoding viral capsid proteins (VP1, VP2, and VP3 genes) (11, 17). MCV viral T antigens are oncoproteins expressed in human MCC tumors (18). The oncoproteins, namely, large and small T, play a pivotal role in the transformation of normal cells into cancer cells. They exert their influence by activating tumor suppressor proteins, contributing to the development and progression of cancer (19). The viral proteins 1, 2, and 3 (VP1, VP2, and VP3) are expressed by the virus through three open reading frames, functioning as capsid proteins. The VP1 protein constitutes 70% of the total virus protein particles, and the protein is a major immunogenic component found in the host immune system required for producing pseudo virions (20). MCV causes abnormalities in the skin's MCs and transforms normal cells into cancer cells. In MCC tumors, VP1 is the major viral protein required to form viral particles and to bind to the site for infection. Anti-VP1 antibodies in the blood indicate chronic disease with MCV (21). Vaccines have successfully been developed against HPV and HBV, targeting the different structural proteins of the viruses (22, 23). The limited understanding of MCC etiology has prevented us from achieving similar successes for MCV, necessitating exploring innovative approaches and treatments for MCC (24). Developing a therapeutic vaccine can be considered a success for the disease that may provide support to enhance the activity of cancer-specific T cells and promote antitumor immunity. The therapeutic vaccines will enhance cellular response by activating antigen-specific CD8 + T cells of patients with MCC-positive tumors.

This study aims to design an efficient multiepitope vaccine against MCV using computational immunoinformatics approaches to provide novel treatment options for MCC. The multiepitope vaccine will generate a more robust immune response to viral particles and peptides (25, 26). It will produce fewer fatal consequences than vaccines developed using complete viral proteins and peptides (27, 28). As MCV T antigens are tumor suppressors in the human body and form cancer cells by altering typical MCs, this newly developed vaccine may help prevent the transformation of MCs to cancer cells in human skin (29). We have designed a vaccine candidate against MCV that binds to MC's receptor site and can potentially fight against MCV in the human body. Previously, DNA vaccines were developed targeting large and

small T antigens or VP1. They produced antitumor effects by inducing cytotoxic and helper T lymphocyte (CTL and HTL) responses in mice (30–32). In this study, we have used selected epitopes of capsid proteins (VP1-VP3) and large and small T antigens and designed multiple epitope vaccine candidates, showing computationally more robust immune responses. However, further *in vitro* and *in vivo* experiments must be conducted to confirm the efficacy of the designed multiepitope vaccines produced using predicted epitopes.

2 Materials and methods

A flow chart of the overall procedure applied in these studies, initiated from antigenic protein selection to vaccine construction and evaluation, is illustrated in Figure 1.

Also, we have provided detailed information about the servers used in the design of MCV vaccine candidates in Table 1. It includes their functions, parameters, and thresholds, which are crucial for predicting antigenicity, epitopes, protein structures, and optimized vaccine design. Table 1 indicates the specific parameters and thresholds each server uses during its prediction processes. This valuable information will serve as a comprehensive reference guide, highlighting essential servers, their functionality, and the parameters and thresholds used during the computational-based design process of MCV vaccine candidates.

2.1 Proteome retrieval and antigenicity prediction

We obtained the protein sequences of MCV from the UniProt website, a widely accessible database of experimentally characterized protein sequences. UniProt offers comprehensive information regarding these protein sequences, facilitating our research and analysis (51). Five protein sequences, including large T antigen, small T antigen, VP1, VP2, and VP3 of the MCV, were retrieved from the UniProt (Proteome ID: UP000154903). All protein sequences were downloaded in the FASTA file format and submitted to the VaxiJen v2.0 server for antigenicity prediction (34). We utilized a web-based tool to align independent protein sequences, enabling the identification of antigens that exhibited performance based on auto cross-covariance transformation and aiding in the determination of uniform vectors of equal lengths. We selected the proteins with the highest antigenicity for subsequent analysis. The threshold value was 0.5 to predict 12 MHC supertypes, including supertypes A26 and B39 of MHC. Additionally, ANTIGENpro was also used to indicate the antigenicity of the selected proteins (35).

2.2 Epitope identification

2.2.1 Cytotoxic T lymphocyte epitope evaluation and selection

We submitted the selected antigenic proteins with the highest antigenicity scores to the NetCTL 1.2 server for CTL epitope prediction, which has a higher predictive capability and sensitivity than other available methods (36, 52). We analyzed CTL epitopes within the 12 HLA-I supertypes (A1, A2, A3, A24, A26, B7, B8, B27, B39, B44, B58, and B62) to select specific antigenic proteins. The proteins with the highest NetCTL scores were selected. A default NetCTL value of 0.75 was used as a cut-off to predict and select a CTL epitope (53). This method combines the prediction of peptide major histocompatibility (MHC) class I binding, proteasomal C_terminal cleavage, and transporter associated with antigen

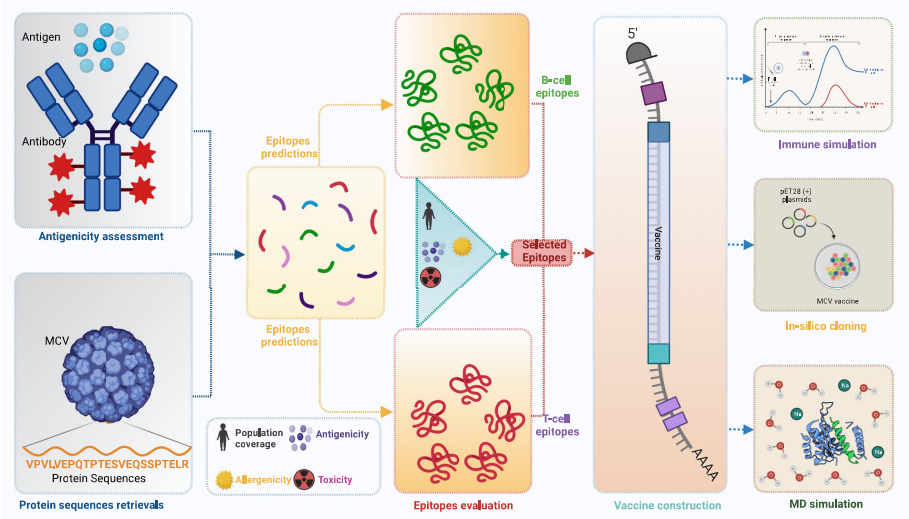


FIGURE 1
This schematic diagram illustrates the comprehensive workflow employed in the current study for computational multiepitope vaccine design against Merkel cell polyomavirus (MCV). The workflow utilized vital steps, including target antigen identification, epitope prediction, epitope selection, design of multiepitope constructs, structural modeling and validation, *in silico* cloning, and *in silico* evaluation of immunogenicity and efficacy. These steps collectively contribute to developing an optimized multiepitope vaccine candidate for MCV.

TABLE 1 This table provides a comprehensive compilation of servers employed in the design of Merkel cell polyomavirus (MCV) vaccine candidates, including their functions, parameters, and thresholds.

Server Name	Function	Parameters	Threshold	Reference
UniProt	Protein sequence retrieval	–	–	(33)
VaxiJen v2.0	Antigenicity prediction	Default	0.5	(34)
ANTIGENpro	Antigenicity prediction	Default	–	(35)
NetCTL 1.2	CTL epitope predictions	Default	0.75	(36)
AllergenFP v.1.0	Allergenicity prediction	Default	–	(37)
ToxinPred	Toxicity prediction	Default	0.5	(38)
MHC-I immunogenicity	Immunogenicity prediction	Default	–	(39)
Immune epitope database and analysis resource (IEDB)	T-cell epitope prediction	Default	–	(40)
IFNepitope	IFN-γ inducing epitope prediction	Default	–	(41)
BepiPred 2.0	B-cell epitope prediction	Default	0.5	(42)
ProtParam	Physicochemical property prediction of a protein	Default	–	(43)
SOLpro	Solubility	Default	–	(44)
I-TASSER	Protein structure prediction	–	–	(45)
GalaxyRefine	Refinement and optimization of protein structures	Default	–	(46)
ProSA	Validation of protein structures	Default	–	(47)
Protein Data Bank (PDB)	Experimentally determined 3D protein structures	–	–	(48)
ClusPro 2.0	Protein–protein docking	Default	–	(49)
JCat	Optimizing codon	Default	–	(50)

The servers in the table perform various tasks related to antigenicity prediction, epitope prediction, protein structure prediction, refinement, and optimization.

processing (TAP_) transport (54). We evaluated selected CTL epitopes for immunogenic, antigenic, allergenic, and toxicity properties. The immunogenic response of CTLs is the main requirement for vaccine construction. First, the selected epitopes were submitted to the MHC-I immunogenicity tool of the IEDB website to evaluate immunogenic properties (40). Second, selected epitopes were analyzed using the VaxiJen 2.0 server for antigenic evaluation (34). Third, the allergenicity of the selected epitopes was predicted using the AllergenFP v.1.0 server for CTL epitope evaluation (37). Finally, the toxicity of CTL epitopes was evaluated using the ToxinPred server (38). In most cases, we utilized the default parameters of the server for epitope evaluations. In this study, we chose CTL epitopes with immunogenic, antigenic, non-allergenic, and non-toxic properties for the final vaccine constructs.

2.2.3 Helper T lymphocyte epitope evaluation and selection

Helper T lymphocyte (HTL) cells play a crucial role in adaptive immunity, stimulating both humoral and cellular immune responses against foreign antigens (53). To identify HTL epitopes of the MCV protein, we utilized the MHC-II binding allele-IEDB Analysis Resource website as a resource in this study (55). We used the consensus method of 5% percentile for HTL epitope prediction and selection, and 15-mer peptide epitopes were selected. Subsequently, the chosen HTL epitopes were evaluated based on interferon-gamma, interleukin-4, interleukin-10, and antigenicity properties. The interferon-gamma is a type of cytokine critical to both innate and adaptive immunity that plays an essential role in vaccine construction. First, selected epitopes were submitted to the IFNepitope server for interferon-gamma secretion property analysis, which utilized a hybrid method [support vector machine (SVM) and motif method] to analyze the properties (41). Second, the interleukin-4- and interleukin-10-producing ability of the HTL epitopes were predicted by using the IL4Pred and IL10Pred servers (56, 57). Based on the induction and non-induction properties, interleukin-4 and interleukin-10 were selected. Finally, we analyzed the antigenic properties of the HTL epitopes using the VaxiJen 2.0 server (34). The HTL epitopes selected based on the induction ability of interferon-gamma, interleukin-4, and interleukin-10 and antigenicity properties were used for the vaccine constructs.

2.2.3 B-Cell lymphocyte epitope evaluation and selection

Linear B-cell epitopes are crucial in antibody production and the construction of peptide-based vaccines (58). To identify linear B-cell epitopes, we submitted the selected antigenic proteins to the BepiPred 2.0 web tool (42, 59). This tool successfully identified 12-mer peptide epitopes corresponding to the MCV protein. The threshold parameter that has been set was 0.5. The selected B-Cell

lymphocyte (BCL) epitopes were further evaluated based on antigenicity, allergenicity, and toxicity properties. Finally, the best BCL epitopes with the highest antigenicity, non-allergenicity, and non-toxicity properties were selected for vaccine construction.

2.3 Estimation of population coverage

The distribution of human leukocyte antigen (HLA) alleles and their expression patterns vary country by country and worldwide according to the differences in genomic regions and ethnicities (60). In computational vaccine design, population coverage directly indicates the worldwide effectiveness of the vaccine by evaluating the prevalence of HLA alleles related to the epitope of interest. Therefore, the population coverage was calculated using the T-cell epitopes with their respective HLA-binding alleles. To achieve this, we submitted the selected epitopes along with their allelic information to the IEDB population coverage tool. This tool allows for assessing the coverage provided by the selected epitopes across different populations (61). Population coverage scores were calculated using the HLA hit score derived from the relative allele frequency at a specific locus within a particular population.

2.4 Formulation of multiepitope vaccine

The multiepitope vaccine candidate was formulated by properly utilizing previously selected CTL, HTL, and LBL epitopes initiated with a suitable adjuvant linked by a different linker, including EAAAK, AAY, GPGPG, and KK. In this study, the adjuvant, linker, and epitope of the protein were ordered in a way that can elicit maximum immune cell–specific responses and confer protection against the virus (53, 62). However, the epitopes of the vaccines were shuffled and appointed in a different order. Based on the antigenicity and physiochemical properties, the best confirmation was selected for further evaluation. Initially, an ideal adjuvant receptor was identified through an advanced literature search to enhance the immunogenicity of MCV protein fuse with Fc of human IgG. It has been found that TLR agonists TLR 2, 4, 5, 7, and 9 play an essential role in the pattern recognition of MCV protein (63). However, in this study, the TLR4 agonist was used as an adjuvant due to the maximal rate of synthesis ability and activating the highest immune responses against the MCV (64). The TLR4 agonist known as 50S ribosomal protein L7/L12 of *Mycobacterium tuberculosis* was retrieved from the UniProtKB (ID: P9WHE3) and used as the adjuvant to enhance the immunogenicity of the vaccine candidate (65). Specific linker molecules were employed to fuse the peptide sequences in the study. The front of the adjuvant was attached with a bifunctional linker EAAAK. Subsequently, CTL, HTL, and LBL epitopes were linked together through AAY, GPGPG, and KK linkers, respectively (66). Initially, the vaccine adjuvant was attached to the front of the vaccine using the EAAAK linker, which consists of helix-forming peptides of various lengths. This linker serves to separate the two weakly interacting β-domains (67). On the other hand, selected CTL was linked using Ala-Ala-Tyr (AAY) linkers, while HTL was linked with Gly-Pro-Gly-Pro-Gly (GPGPG) linkers. In addition, LBL has linked to Lys-Lys (KK) linkers (53). The AAY linker, which is a cleavage site for the proteasome, was used to affect protein stability, reduce immunoreactivity, and enhance epitope presentation. The GPGPG linker, known as the glycine–proline linker, prevents the formation of "junctional epitopes" and facilitates the immunological process (68). In addition, the bi-lysine KK linker helps preserve independent immunological activities during the vaccine formulation (69). The peptides of the construct were fused with each other using the selected linker due to their ability to provide support for structure flexibility, improve protein stability, and play an important role in increasing the biological activity of the vaccine construct (53, 70).

2.5 Physicochemical and immunological properties analysis

The efficacy of the vaccine candidate was assessed by evaluating its physiological, antigenic, immunogenic, allergenic, and soluble properties. The physicochemical properties of the vaccine construct were analyzed using the ProtParam tool, enabling a comprehensive examination of its characteristics (43).. The tool calculated the physiological properties, including molecular weight, theoretical pi, and the number of positively charged residues (Arg + Lys). The chemical formula of the vaccine, the whole number of atoms, coefficient extinction, *in vitro* and *in vivo*, half-life, instability, aliphatic index, and grand average of hydropathicity (GRAVY) value were also determined by using the tool. We evaluated the antigenic, immunogenic, and allergenic properties of the vaccine constructs by utilizing specific web tools. The VaxiJen 2.0 tool was employed to assess antigenicity, the MHC-I immunogenicity tool from the IEDB was used to determine immunogenicity, and the AllergenFP v.1.0 web tool was utilized to evaluate allergenicity. These analyses provided valuable insights into the properties of the vaccine constructs (34, 37, 55). The solubility of the vaccine construct was also evaluated with the help of the SOLpro web-based tool (44).

2.6 Vaccine structure prediction, refinement, and validation

2.6.1 Secondary structure prediction

For analyzing the extended strand, alpha-helix, and random coils for the secondary structure of the constructed vaccine, the PSIPRD web-based tool was used in this study (71). The PSIPRED web-based tool offers a user-friendly interface and employs a machine-learning approach to analyze protein sequences and predict their secondary structures. This tool also utilizes a cross-validation approach to validate its performance (72).

2.6.2 Tertiary structure prediction and refinement

The three-dimensional (3D) structure of the final multiepitope vaccine construct was predicted by using the Iterative Threading

ASSEmbly Refinement (I-TASSER) homology modeling server (45). The initial model of the MCV vaccine identified from the I-TASSER server was further validated and refined using the GalaxyRefine web server developed based on a refinement method that has been successfully implicated and investigated in CASP10 (46). Schrödinger Maestro (Schrödinger Release 2022-3: Maestro, Schrödinger, LLC) tools were used to visualize the obtained initial and refined 3D structure of the vaccine candidate.

2.6.3 Structure validation

Validation of the protein structure that has been predicted through homology modeling is the core of structural determination methods. Validation of the protein 3D structure provides a more extraordinary idea about the compatibility of a structural model with its amino acid (AA) residues. It helps to determine the missing AA residues of the protein (73). Therefore, to validate the structural confirmation of the proposed MCV vaccine, the 3D structure of the protein was submitted to the ProSA-web server (47). The overall quality of the protein structure was accessed based on the z-score value provided by the server. If the z-scores of the anticipated model fall outside compared to the construction of the native protein, it indicates an erroneous protein. Additionally, the Ramachandran plot evaluation of the proposed vaccine candidate was performed by utilizing the Ramachandran Plot Server developed by ZLab to check the main-chain conformational tendencies of AA residues (74).

2.7 Molecular docking

Molecular docking is a very commonly used computational method that simulates the interaction of a ligand with its receptor and consequently forecasts the energy score generated during the interaction (75). The technique can determine the binding affinity of two molecules based on certain scoring functions. For molecular docking, the desired TLR4 receptor was retrieved from the RCSB Protein Data Bank (PDB) having a PDB ID: 4G8A. The TLR4 receptor was docked with the vaccine candidates that were defined as a ligand during the docking simulation. The TLR4 receptor was prepared by removing water and heteroatom and adding hydrogen through Schrödinger's protein preparation wizard (76). To evaluate the binding affinity, molecular docking was performed by using ClusPro 2.0 web server (49). The performance of the server was assessed based on the ability to cluster the lowest energy structure, rigid body docking, and structural refinement process depending on energy minimization. The best-docked complex was selected and retrieved based on the binding affinity between the ligand–receptor complex. The interaction between the receptor TLR4 and vaccine construct was visualized by using the PyMOL visualization tool (77).

2.8 Complex structural stability evaluation through molecular dynamics simulation

The stability of the protein–protein complex refers to stable protein dynamics (more association and less dissociation of a protein–protein complex). The binding strength of the receptor and ligand (vaccine candidate) complex system and their dynamic behavior can be evaluated using different computational tools and animal model systems. To ascertain the constancy of the predicted vaccine and vaccine–receptor (VR) complex, a computational molecular dynamics (MD) simulation approach of the refined vaccine and VR complex was performed using 'Desmond v6.3 Program' in Schrödinger (Academic version) under the Linux operating system. The thermodynamic stability of the vaccine and VR complex was calculated using this computational approach, where a predefined TIP3P water model was used to emulate water molecules using the OPLS3e force field (78). Orthorhombic periodic boundary conditions were set up to specify the shape and size of the repetition unit safeguarded at 20 Å distances. To achieve electrical neutralization, the system was balanced by adding suitable sodium and chlorine ions, ensuring a minimized charge within the Desmond module. This process was carried out utilizing the OPLS3e force field. Molecular dynamic simulations were carried out with periodic boundary conditions in the constant number of particles, pressure, and temperature (NPT) ensemble (79). The temperature and pressure were kept at 300 K and 1 atm using Nose–Hoover temperature coupling and isotropic scaling (80). The operation was followed by running the 200 ns simulation and saving the configurations thus obtained at 200 ps intervals. The vaccine and vaccine complex stability was further evaluated using statistical parameters like root mean square deviation (RMSD), root mean square fluctuation (RMSF), the radius of gyration (rGyr), and hydrogen bond (HB) values. The superimposition of the vaccine and VR complexes was also evaluated in this study. The entire molecular dynamics (MD) simulation was executed in the Linux (Ubuntu-20.04.1 LTS) operating system and Intel Core i7-10700K processor CPU, 3200 MHz DDR4 RAM, and RTX 3080 DDR6 8704 CUDA core GPU.

2.9 Immune response simulation

In silico immune simulations were used to estimate the possible immunogenic profile of multiepitope vaccine candidates in real-life conditions by using the C-IMMSIM server (81). The output of the immune responses was salvaged for comprehensive observation. For ideal vaccine candidates, the minimum recommended interval between doses 1 and 2 is 3–4 weeks (22). Therefore, a minimum gap of 30 days between two dosages was taken into consideration in this study. Three injections of the vaccine candidates were administered computationally with time steps of 1, 84, and 168, where the one-time step was considered eight h in real life. The immune simulation was carried out for a total of 300 steps, and the rest of the simulation parameters were kept defaults.

2.10 Codon optimization and *in silico* cloning

Codon optimization is a gene engineering technique that employs synonymous codon modifications to enhance protein

expression (82). Optimization of codon should be performed based on the specific host organism or expression system because the expression pattern of a foreign gene depends on the type of host organisms or expression system (53). To optimize the codon of the desired vaccine candidate, the JCat tool was used in this study (50). The tool uses an algorithm to maximize codons based on the codon adaptation index (CAI) (83). In this study, the widely used *E. coli* K12 was considered the host, and based on the expression system, codon optimization was performed. The following criteria were skipped during the optimization steps: (i) restriction enzyme (RE) cleavage sites, (ii) rho-independent termination of transcription, and (iii) binding sites of the prokaryotic ribosome. The final and optimized sequence was evaluated based on the CAI value and guanine–cytosine (GC) content. Finally, the optimized nucleotide sequence of the vaccine construct was inserted into the pET28a (+) vector using SnapGene 3.2.1 software.

3 Results of the study

3.1 Proteome retrieval and antigenicity prediction

The target sequence of MCV was retrieved in FASTA format from the UniProt database. Five proteins were recovered from the database: large T-antigen, small T-antigen, VP1, VP2, and VP3. The VaxiJen 2.0 and ANTIGENpro tools predicted the antigenic potency of the selected proteins listed in Table 2. All the primary sequences of the chosen protein have good antigenic properties that were used for further analysis.

3.2 Epitope evaluation and selection

The selected five antigenic proteins with better antigenicity scores were submitted to a different server that predicted the different number of CTL, HTL, and linear BCL epitopes. Subsequently, the antigenic, immunogenic, toxic, and non-allergenic properties of the epitope's candidates were evaluated, which found a high number of potential epitopes. However, we selected 30 (10 CTL, 10 HTL, and 10 linear BCL) epitopes for further evaluation. After considering the antigenic, immunogenic, and non-toxic properties, the selection process determined the best 10 epitopes for constructing a multiepitope vaccine against MCV. In the case of each antigenic protein found in MCV, two CTL epitopes, two HTL epitopes, and two linear BCL epitopes were explicitly chosen, listed in Table 3.

3.2.1 Potential cytotoxic T lymphocyte epitopes

Using the NetCTL v1.2 server, unique CTL epitopes (9-mer) were predicted from the MCV-selected five antigenic proteins. A total of 90 (29, 8, 16, 21, and 16 CTL epitopes from the large T-antigen, small T-antigen, VP1, VP2, and VP3, respectively) unique epitopes were identified that were antigenic, immunogenic, non-toxic, and non-allergenic (Table S1). The best two CTL epitopes for each protein (total 10) were selected and considered for further evaluation (Table 3).

3.2.2 Potential helper T lymphocyte epitopes

A total of 47 unique HTL epitopes (15-mer) were predicted using the IEDB MHC-II prediction tool. Among the 47 unique epitopes, 6, 7, 13, 11, and 10 HTL epitopes were identified from the large T-antigen, small T-antigen, VP1, VP2, and VP3, respectively (Table S2). The epitopes were evaluated based on cytokine (IFN-γ, IL-4, and IL-10)-inducing ability and antigenic properties. Based on the aforementioned properties, a careful analysis was conducted, leading to the selection of the top two HTL epitopes for each protein. These epitopes were chosen for further evaluation and are presented in Table 3.

3.2.3 Potential BCL epitopes

Specific antigenic regions of a protein that ultimately trigger antibody formation are known as BCL epitopes. The BepiPred 2.0 tool was used to predict linear B-cell (12-mer) epitopes from the selected proteins. A total of 70 (22, 8, 18, 12, and 10 epitopes from the large T-antigen, small T-antigen, VP1, VP2, and VP3, respectively) linear B-cell unique epitopes were identified, which were antigenic, non-allergenic, and non-toxic (Table S3). Here, we also selected the top two B-cell epitopes from each protein (total 10) for further evaluation (Table 3).

3.3 Worldwide population coverage

The worldwide population coverage ability of the vaccine candidates has been evaluated based on the selected CTL and

TABLE 2 The selected proteins of MCV along with their corresponding antigenicity scores, which were identified using the VaxiJen 2.0 and ANTIGENpro tools.

NCBI ID	Protein Name	Antigenicity Score		Remark
		VaxiJen server	AnitgenPro server	
B6DVW7	Large T antigen	0.4762	0.889	Selected
B0G0V7	Small T antigen	0.5042	0.761	Selected
B0G0 W3	VP1	0.4374	0.942	Selected
B0G0 W4	VP 2	0.6649	0.697	Selected
A0A0N9DRI5	VP 3	0.5721	0.5	Selected

TABLE 3 The top two selected cytotoxic T lymphocyte (CTL), helper T lymphocyte (HTL), and linear BCL epitopes of MCV, as predicted by the NetCTL 1.2, IEDB MHC-II, and BepiPred 2.0 servers, respectively.

Protein name	CD8 epitope	CD4 epitope	Linear B-cell epitope
Large T antigen	LPFELGCAL	FKVDFKSRHACELGC	PEEPPSSRSSPR
	FELEFALDK	VIMMELNTLWSKFQQ	NKPLLNYEFQEK
Small T antigen	TLEETDYCL	CFCYQCFILWFGFPP	GCMLKQLRDSKC
	LNRKEREAL	VIMNELNTVFSKFQQ	CKLSRQHCSLKT
VP 1	PRYFNVTLR	CDTLQMWEAISVKTE	GLVLDYQTEYPK
	SVAPAAVTF	FNVTLRKRWVKNPYP	FAIGGEPLDLQG
VP 2	LVNYPASWV	AQLGFTAEQFSNFSL	GQDIFNSLSPTS
	QLGFTAEQF	ATTGVTLEAILTGKA	LAQLGFTAEQFS
VP 3	LVNRDVSWV	RHALMAFSLDPLQWE	NSRWVFQTTASQ
	QLGCLGEQF	VNLILNSRWVFQTTA	SLVNRDVSWVGS

HTL epitopes depicted in Figure 2. CTL and HTL epitopes showed a considerably high percentage (%) of population coverage. The combined world population coverage found for the CTL and HTL epitopes was 99.33%, where CTL individually shows a world coverage of 97.77% and HTL shows a world coverage of 70.14%. The identified epitopes are also prone to a high number of HLA alleles originating from different countries, such as Germany, Europe, the United States, South Asia, and India, with a combined (CTL and HTL) population coverage of 99.96%, 99.86%, 99.74%, 96.30%, and 95.75%, respectively (Figure 2). Therefore, the vaccine candidates that have been designed by utilizing the selected epitopes will cover most of the population around the world.

3.4 Formulation of multiepitope vaccine

To design multiple epitope vaccine candidates, initially, the 10 best highly antigenic CTL epitopes that were immunogenic, non-allergenic, and non-toxic were selected from each of the five (large T-antigen, small T-antigen, VP1, VP2, and VP3 of MCV) proteins (Table 3). Based on the cytokine-inducing properties, the best 10 HTL epitopes were selected from five proteins, which were highly antigenic and had the potential to generate cytokines. At last, the 10 best linear B-cell unique epitopes were identified from the structural protein of MCV, which were antigenic, non-allergenic, and non-toxic. The vaccine construct was formulated by using the selected 30 epitopes belonging to three different classes (10 CTL, 10 HTL, and 10 LBL). The vaccine constructs were initially accompanied by the TLR4 agonist 50S ribosomal protein L7/L12 as an adjuvant, positioned before the constructs connected to the first CTL epitope using EAAAK linkers. The selection of 30 epitopes, comprising 10 CTL, 10 HTL, and 10 BCL epitopes, was joined by the utilization of AAY, GPGPG, and KK linkers, respectively, to establish the desired connections between the epitopes. The total AA residue count in the final vaccine construct was 592. The sequential arrangement of the different epitopes and their corresponding linkers is shown in Figure 3.

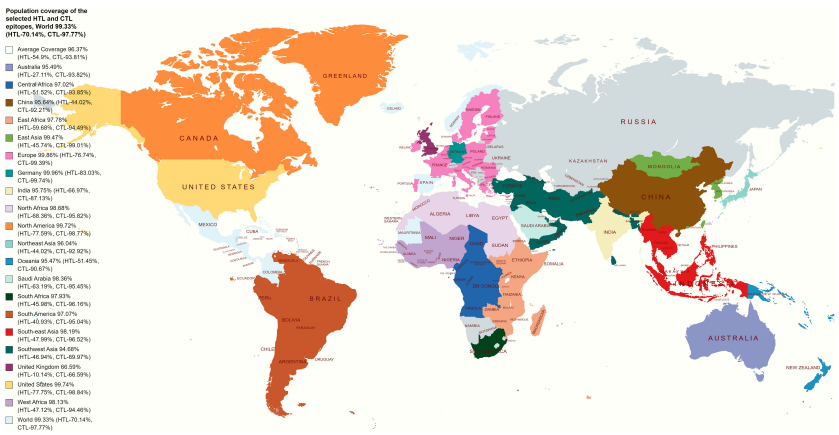


FIGURE 2
Illustrates a global population coverage map predicted using specific CTL and HTL epitopes. The map showcases the estimated coverage in various regions worldwide. These predictions rely on the chosen CTL and HTL epitopes, essential for stimulating cellular immune responses. The map offers valuable information about the potential effectiveness and coverage of the epitopes in diverse populations, assisting in evaluating and optimizing vaccine design strategies.

3.5 Physicochemical and immunological properties of the vaccine

The ProtParam server was used to analyze the physicochemical properties of the multiepitope vaccine construct (Table 4). It exhibited an antigenic score of 0.6730, indicating a significant ability to elicit an immune response and effectively initiate interactions between antigens and antibodies. Based on the analysis, the vaccine candidate showed a molecular weight of 64,118.85 Da, which suggests a moderately sized construct. This size has implications for several important aspects of the vaccine's development, including manufacturing, formulation, and stability (43). The theoretical isoelectric point that represents the pH of the vaccine was calculated to be 8.72, suggesting alkaline or basic nature of the construct. The alkaline nature of the construct has significant implications for various aspects such as its stability, solubility, and interactions with other molecules or components present in the formulation. The vaccine shows an instability index (II) of 30.77, which indicates a good post-expression stability of the construct. The thermostability of the construct was determined by assessing the aliphatic index, which yielded a value of 77.55. This range falls between 70 and 100, indicating that the proteins within the construct possess a notable degree of thermal stability. The server calculated the GRAVY as -0.210, which indicates a strong correlation with the highly hydrophilic nature of the construct. This hydrophilicity is expected to facilitate significant protein–protein interactions. The analysis also revealed that the vaccine had an estimated half-life of 30 h in mammalian reticulocytes in an *in vitro* setting. In yeast cells, the vaccine showed a half-life of over 20 h in an *in vivo* environment. Similarly, in *Escherichia coli*, the estimated half-life exceeded 10 h in an *in vivo* setting. These results suggest that the vaccine exhibits a comparatively long-lasting presence and stability across various biological systems, emphasizing its potential effectiveness and durability. Evaluation of immunogenicity provided a value of 1.24781. Moreover, the analysis of allergenicity properties shows the absence of allergenic features in the vaccine candidate. Additionally, the candidate showed a high solubility rate of 0.98246 as determined by the SOLpro server indicates that the candidate is expected to have good solubility in aqueous solutions (44). It implies that the vaccine construct has a high likelihood of dissolving well and remaining in solution, which is advantageous for its formulation and administration.

3.6 Vaccine structure prediction, refinement, and validation

3.6.1 Secondary structure prediction

The secondary structures of the vaccine candidate were composed of extended strands, alpha helices, and random coils. The secondary structure of the vaccine construct was estimated *via* the PSIPRED 3.2 server. The analysis yielded an average Q3 score of

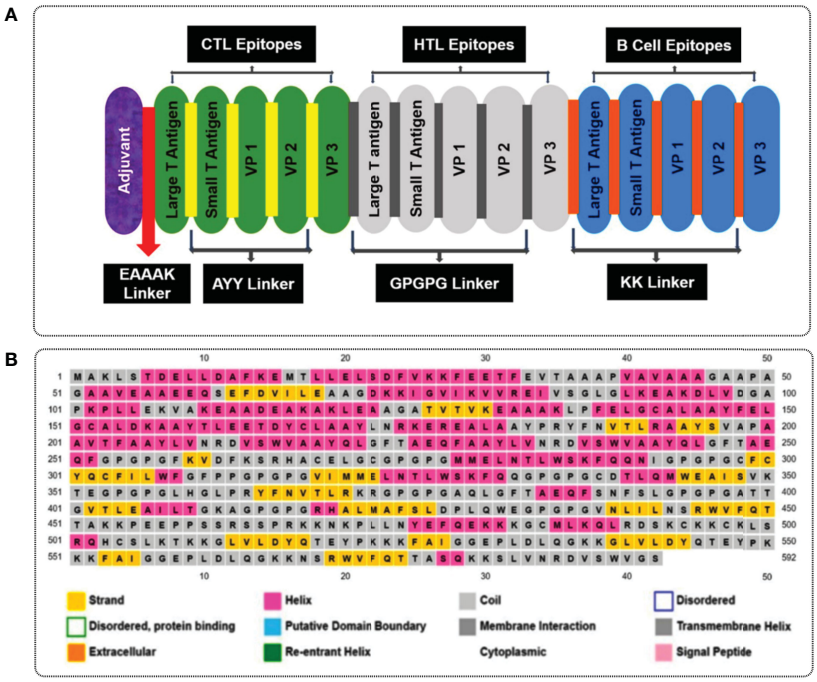


FIGURE 3
(A) A visual representation of the MCV vaccine constructs. Different colors are used to denote the adjuvant (purple), cytotoxic T lymphocyte (CTL, green), helper T lymphocyte (HTL, white), and linear B-cell epitope (linear BCL, blue) epitopes. The adjuvant and CTL epitopes are connected by the EAAAK linker (indicated in red), while AYY (gold color), GPGPG (gray boxes), and KK (orange boxes) linkers are employed to join the CTL, HTL, and linear BCL epitopes, respectively. **(B)** represents the secondary elements, including α-helices (pink), β-strands (yellow), and random coils (blue) of the MCV vaccine candidate.

TABLE 4 List of the physiochemical parameters, antigenicity, immunogenicity, allergenicity, and solubility of the final vaccine candidates.

Parameters	Evaluation of properties
Number of amino acids	592
Molecular weight	64,118.85 Da
Theoretical pi	8.72
Total number of positively charged residues (Arg + Lys)	72
Total number of atoms	9,050
Extinction coefficient (at 280 nm in H2O)	82,570
Estimated half-life (mammalian reticulocytes, *in vitro*)	30 h
Estimated half-life (yeast cells, *in vivo*)	>20 h
Estimated half-life (*Escherichia coli*, *in vivo*)	>10 h
Instability index	30.77
Aliphatic index	77.55
Grand average of hydropathicity (GRAVY)	-0.210
Antigenicity	0.6730
Immunogenicity	1.24781
Allergenicity	Non-allergen
Solubility	0.982460

81.6% for the helix, sheet, and loop). The Q3 score serves as a valuable metric to assess the accuracy of secondary structure prediction methods like PSIPRED (71). It quantifies the proportion of correctly predicted secondary structure elements (helix, sheet, or loop) in relation to the known experimental structure of a protein. The study obtained a Q3 score of 81.6%, reflecting a high level of accuracy in predicting the secondary structure. This score signifies that approximately 81.6% of the amino acids (AA) in the construct were correctly assigned to their respective secondary structure elements (helix, sheet, or loop) by the prediction algorithm. Notably, our observations revealed a notable prevalence of alpha-helices in the construct, visualized by the pink color in Figure 3B. Alpha-helices are widely acknowledged for their remarkable structural stability and often play a critical role in protein folding and stability. Additionally, the presence of loops, depicted by the gray color, indicates flexible regions that contribute to conformational variability and can actively participate in protein–protein interactions and antigenic determinants. The construct consisted of 592 AAs in total, and the α-helix, β-strands, and random coils found in the structure indicated by pink, yellow, and gray colors, respectively, are represented in Figure 3B.

3.6.2 Tertiary structure prediction

The vaccine construct's tertiary structure was generated using the I-TASSER server. The server provided the top five 3D models of the vaccine construct with different C-score values (Table S4). The C-score is a confidence score for estimating the quality of predicted models generated by I-TASSER. The study considered the model with the lowest C-score (−1.37), as recommended by the server and visualized by Schrodinger Maestro (Figure 4A).

3.6.3 Tertiary structure refinement

The Galaxy Refine server was used to refine the projected tertiary structure of the final vaccine construct. The initial protein model retrieved from the I-TASSER was submitted for refinement. The protein-refining server provided five refined models with the presence of an increased number of AA residues in the favorable region listed in Table S5. The study selected the best refined model based on the Ramachandran favored score. In this study, model-5 (Table S5) shows a highly Ramachandran-favored score of 88.5% with a GDT-HA score of 0.9396, an RMSD value of 0.451, a MolProbity score of 2.363, and a clash score of 19 selected for further evaluation (Table S5). The refined vaccine model was visualized *via* Schrodinger Maestro represented in Figure 4A.

3.6.4 Tertiary structure validation

The tertiary structure of the initial vaccine construct (before refinement) and final vaccine construct (after refinement) were validated by analyzing the output found from the Ramachandran Plot Server and ProSA-Web server. Ramachandran plot analysis of the initial vaccine model found that a total of 86.992% amino acid residues was in the favorable region of the plot (Figure 4B). However, after the refinement rampage server generated a Ramachandran plot, where a total of 94.512% of residues were in the favorable region of the plot (Figure 4B).

The prose-web server was used to assess the validation quality and potential errors in a crude tertiary structure model (Table S6). To validate the final vaccine model, its agreement with experimental data was assessed using the Z-score. The Z-score is a quantitative measure that evaluates the alignment between a model and experimental information. Its range varies depending on factors like the protein and its size. Generally, Z-scores fall within a range -4 to +4. When the Z-score approaches zero, it indicates a close resemblance of the model's energy to experimentally determined structures (84). For the initial model, the Z-score was calculated as -2.75, indicating a moderate deviation from the experimental data. However, through refinement, the model achieved a slightly improved Z-score of -2.59 (Figure 4C). This suggests that the refined model exhibits better alignment with the experimental data, although the improvement is relatively minor.

3.7 Molecular docking

The binding affinity of the receptor (TLR-4) and ligand (refined vaccine) was calculated by using the ClusPro 2.0 server. The server provided a total of nine complex confirmational structures along with different binding energy scores. The lowest and central energy

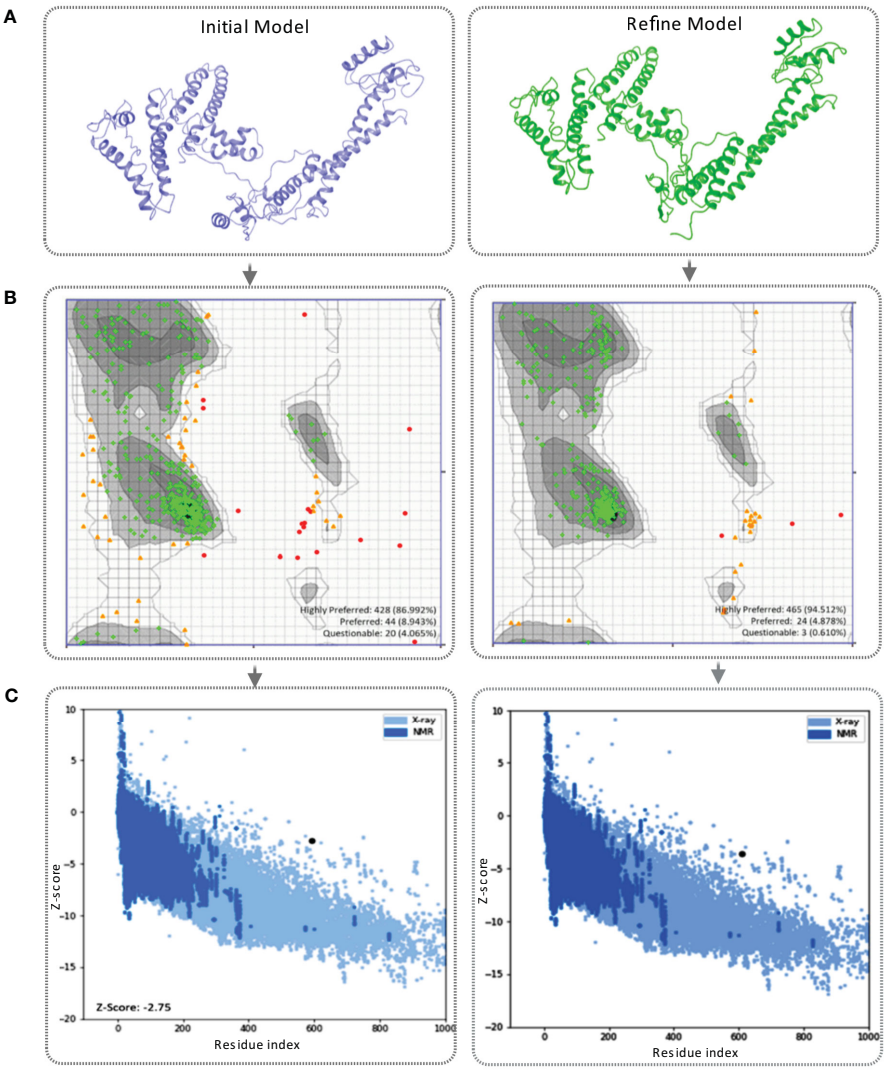


FIGURE 4
(A) This figure showcases the tertiary structure of the MCV vaccine model. The left side displays the initial model of the vaccine, while the right side represents the refined vaccine construct. **(B)** The Ramachandran plot of the final vaccine model, initial vaccine model (right), and refined vaccine model (left). Highly preferred conformations are represented by black, dark gray, and gray, while preferred conformations are depicted by white with a black grid. Questionable conformations are shown as white with a gray grid. **(C)** The validation of the final vaccine model was performed based on the Z-score. The Z-score for the initial model was -2.75, whereas the refined model attained a Z-score of -2.59.

of the cluster found for each complex structure is listed in Table S7. The best complex confirmation structure has been chosen based on the lowest energy value. In this study, Cluster-7 shows lowest binding energy value -1122.9 kcal/mol, which was retrieved for further analysis (Figure 5A). The interaction between the TLR-4 receptor and vaccine construct was analyzed from the VR docking complex and shown in Figures 5B, C. The interaction residue participant in the complex formation is also listed in Table S8.

3.8 Molecular dynamics simulation analysis

MD simulation is a convenient way that was used to analysis the structural stability of the vaccine and VR complex structure. The strength of the complex interface was evaluated based on RMSD, RMSF, Rg, intramolecular HBs (Intra HB), and ligand-protein contacts.

3.8.1 Root mean square deviation of vaccine construct

RMSD of the vaccine construct was measured to evaluate the average change happened due to the displacement of a selected atoms from the vaccine frame comparing to a reference frame. During the simulation of the vaccine construct, the highest fluctuation was 16.162 Å, the lowest was 3.109 Å, and the average was 11.94 Å (Figure 6A). A minor notch of fluctuation was observed for the vaccine structure after 160 ns dynamic simulation indicating structural stability of the vaccine construct.

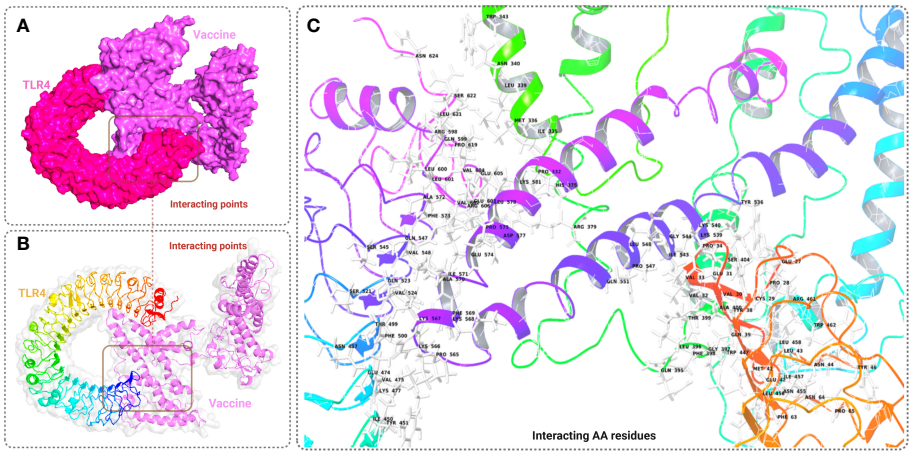


FIGURE 5
This figure depicts a graphical representation of the molecular interaction between the MCV vaccine candidates and the TLR-4 receptor. The molecular interaction is presented in three different views: **(A)** surface view, **(B)** cartoon view, and **(C)** specific amino acid interactions. The surface view in A provides an overall visual representation, while the cartoon view in B offers a simplified depiction. In C, specific amino acid interactions between the vaccine candidates and the TLR-4 receptor are highlighted.

3.8.2 Root mean square fluctuation of vaccine construct

RMSF of the vaccine construct was calculated to examine the change of structural flexibility occurred due to the displacement of a specific AA residue in the protein. The RMSF plot of the vaccine construct showed a fluctuation peak between 95 and 570 AA residual positions. The highest fluctuation was 19.636 Å observed at GLU122 AA residual position, the second-highest fluctuation found at VAL124, and the third-highest fluctuation found at LYS123 AA residual position with an RMSF score of 19.413 and 18.95 Å, respectively, shown in Figure 6B. These fluctuations indicate regions of the protein that may have increased mobility or flexibility, potentially influencing its conformational changes, protein–protein interactions, and overall stability. These fluctuations are important for assessing the functional implications and optimizing the design and performance of the vaccine construct.

3.8.3 Radius of gyration of vaccine construct

The distribution of atoms in the vaccine construct around its axis was measured based on the radius of gyration (Rg) value throughout the 200 ns simulation run. Analysis of the Rg profile found a higher deviation between 15 and 200 ns, where the average Rg score of the construct was 39.25 Å. The Rg score of the study provided information concerning the compactness of the vaccine. Herein, we found an average lower score of the Rg value vaccine construct indicating the tightest packing characteristic of α/β-proteins (Figure 6C).

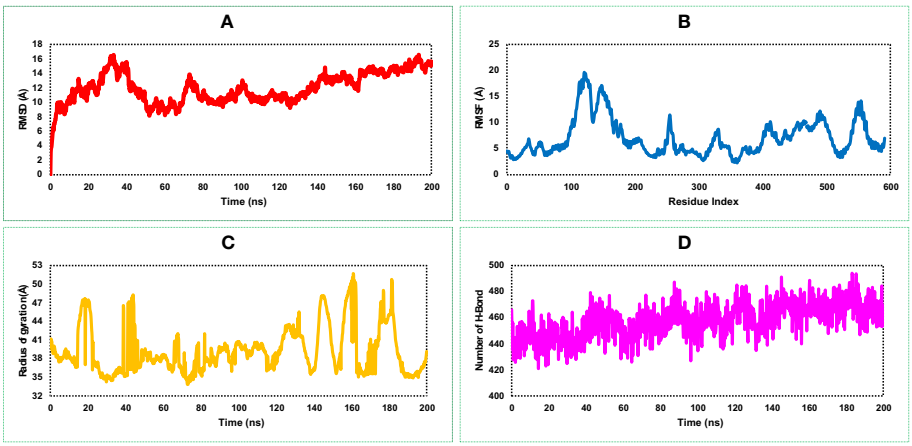


FIGURE 6
Representing four dynamic properties of the MCV vaccine construct obtained from a 200 ns molecular dynamics (MD) simulation. **(A)**, the root mean square deviation (RMSD) plot demonstrates the deviation of the vaccine construct from its initial conformation, indicating any structural changes or fluctuations during the simulation. **(B)** The root mean square fluctuation (RMSF) plot, revealing the residue-wise flexibility or fluctuation of the vaccine construct throughout the simulation. **(C)** The radius of gyration (Rg) quantifies the compactness or size of the vaccine construct during the simulation. **(D)** presents the number of HBs formed within the vaccine construct, highlighting the interactions and stability of the structure.

3.8.4 Number of hydrogen bonds of vaccine construct

Most of the direct contacts require protein folding, protein structure, and molecular recognition depends on the HBs of the structure. The number of HBs in a protein structure can be used to understand the protein structure and motions. Therefore, to understand the structure and motion of the vaccine candidate, the number of HBs found during the 200 ns simulations was analyzed and represented in Figure 6D. Analysis of the simulation data found the highest number of HBs between 90 and 200 ns simulation time. The average number of HBs found in this study was 457 for 200 ns simulation run indicating the vaccine construct will maintain active configurations by connecting protein structure in a fluxional equilibrium.

3.8.5 Root mean square deviation of the vaccine–receptor complex structure

The highest RMSD value of the VR complex found in this study was 16.918 Å. The VR complex structure shows the lowest and average RMSD value of 1.323 Å, and 6.34 Å, respectively during the 200 ns simulation run. The complex structure of the protein shows a stable and optimum fluctuation after 55 ns represented in Figure 7A. The RMSD of the vaccine complex structure (Figure 7A) was lower than the RMSD of the vaccine (Figure 6A) construct indicating stability of the complex structure.

3.8.6 Root mean square fluctuation of the vaccine–receptor complex structure

Analysis of the RMSF plot of the VR complex found the most fluctuation peak between 650 and 800 AA residues represented in Figure 7B. The vaccine candidate that was in complex with the receptor showed the highest fluctuation between 700 and 710 AA residue with an average fluctuation of 20.1 Å. The second highest RMSF value was 8.57 Å found between 900 and 1,200 AA residual position, and the rest of the time, the VR complex shows an optimum fluctuation rate of a complex structure (Figure 7B). A comprehensive understanding and analysis of these RMSF fluctuations have played a critical role in the evaluation of the structural dynamics, ultimately facilitating the optimization of the MCV vaccine candidate's design. This knowledge helped us to enhance the vaccine's effectiveness and immunogenicity, leading to improved protective immune responses and potentially increasing its potential as a prophylactic measure against the MCV.

3.8.7 Rg of vaccine–receptor complex structure

The VR-complex structure shows the average Rg value of 43.78 Å, and a high deviation of the score observed between the range of 5–55 ns simulation run. The lowest and stable Rg value of the VR complex was observed after 70–200 ns simulation run indicating tight packaging of the system (Figure 7C).

3.8.4 Number of hydrogen bonds of the vaccine–receptor complex

The highest number of HBs found for the vaccine and receptor complex structure was between 0 and 20 ns and 120 and 200 ns simulation run (Figure 7D). The number of the HBs reduced in this study during 80–100 ns simulation run. However, throughout the 200 ns simulation, the complex structure of the protein consistently maintains an optimal number of hydrogen bonds (HB). This observation indicates the protein's significant contribution to the free energies of VR complexes.

3.9 Superimposition of vaccine and vaccine–receptor complex

Different structural and conformational changes of the vaccine and VR complex were analyzed from the 200 ns simulation trajectory as shown in Figure 8. Conformational changes were observed for each 50 ns time interval during the simulation of the vaccine (Figure 8A) and the VR complex (Figure 8B). Very low confirmation change was found from the very beginning to 200 ns simulation time of the vaccine and VR complex. Therefore, the vaccine and vaccine complex remained stable in a 200 ns dynamic simulation trajectory (Figure 8).

3.10 Immune response analysis

The computational immune simulation of the vaccine candidates found a response similar to the actual immune responses of a human, as shown in Figure 9. Secondary and tertiary responses generated during the immune simulation process were higher than the primary immune response. Analysis of the immune simulation initially identified higher concentrations of IgM in the case of the primary immune response, where the secondary and tertiary responses show higher levels of immunoglobulin activities (*i.e.*, IgG1 + IgG2, IgM, and IgG + IgM antibodies) with concomitant antigen reduction represented in Figure 9A. The results found in this study indicate the ability of the vaccine candidates to form memory T cells. The immune simulation also found some long-lasting B-cell isotypes that can help with potential isotype switching, resulting in the formation of memory cells (Figures 9B, C). In the case of TH (helper) and T.C. (cytotoxic) cells, a similar elevated response along with the respective memory development was also observed (Figures 9D, E). This indicates that the emergence of immune memory results in a high level of antigen clearance upon subsequent exposure (Figure 9F). During the exposure time, the immune system showed increased macrophage activity, with simultaneous proliferating dendritic cells (Figures 9G, H). High levels of IFN-γ and IL-2 were also observed during exposure, suggesting that it will help to promote the development of T regulatory cells. This profile suggests immune memory development and, therefore, natural immune protection against the virus (Figure 9).

3.11 Codon optimization and *in silico* cloning

This study utilized an *in silico* molecular cloning approach to analyze and modify the target vaccine sequence for compatibility with the selected vector. The process involved identifying suitable

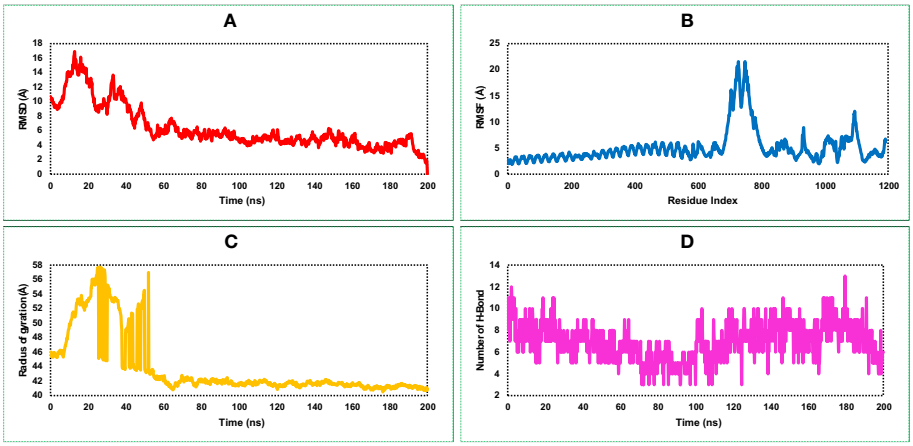


FIGURE 7
This figure illustrates the dynamic properties of the MCV vaccine–receptor (VR) (TLR-4) complex obtained from a 200 ns MD simulation. It examines the **(A)** RMSD plot, indicating structural changes and fluctuations in the MCV VR complex during the simulation. **(B)** displays the RMSF plot, revealing residue-wise flexibility and fluctuations within the complex. **(C)** quantifies the complex's compactness and size using the radius of gyration. **(D)** highlights the interactions and stability of the structure through the number of hydrogen bonds (HBs).

RE recognition sites, optimizing codon usage, and considering factors that could influence gene expression and protein production. Initially, the sequences of the vaccine candidate were optimized by using the codon optimization process to maximize the expression of the vaccine candidate in the *E. coli* K12 expression system. Codon optimization was performed by utilizing the 1,776 (bp) nucleotide sequences retrieved by converting the protein sequences of the construct. The task was completed by using the JCat tool and accessed based on the GC content and CAI value. The GC content found for the vaccine construct was 52. 36% lies between the normal range of 30%–70%. The CAI value found for the construct was 0.98, which also lies in the ideal range between 0.8 and 1.0. Based on the content of GC and CAI value, the MCV vaccine will be expressed highly whenever the *E. coli* expression system is utilized as a host. Two restriction digestion endonucleases, the EcoRI and BamHI, were used to cut the vaccine and vector pET28a (+) vector sequence (Figure 10). Herein, the cloned vaccine sequence's absolute length was 7,143 bp after RE digestion and ligation, shown in Figure 10A. The steps and outcomes of this *in silico* molecular cloning process are illustrated in Figure 10B.

4 Discussion

Given the elevated mortality rate associated with MCC, exploring and advancing preventive measures have become an urgent and imperative matter (85). In such circumstances, vaccination is the most effective and suitable strategy for developing immunity against

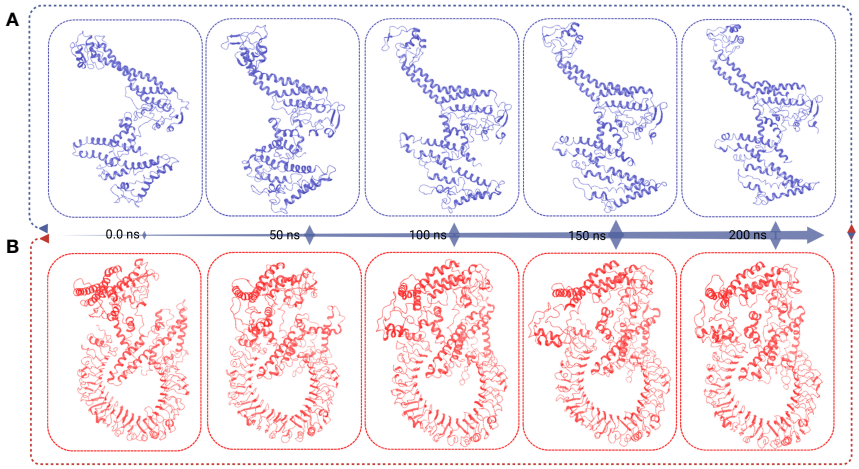


FIGURE 8
This figure presents the superimposition frames at different simulation times (0, 50, 100, 150, and 200 ns) for both the MCV vaccine and the vaccine receptor's complex structure. **(A)** The superimposition frames illustrate the alignment and comparison of the MCV vaccine structure at different time points during the simulation. **(B)** showcases the superimposition frames of the complex structure formed between the MCV vaccine and its receptor, providing insights into their conformational changes and interactions over the simulation duration.

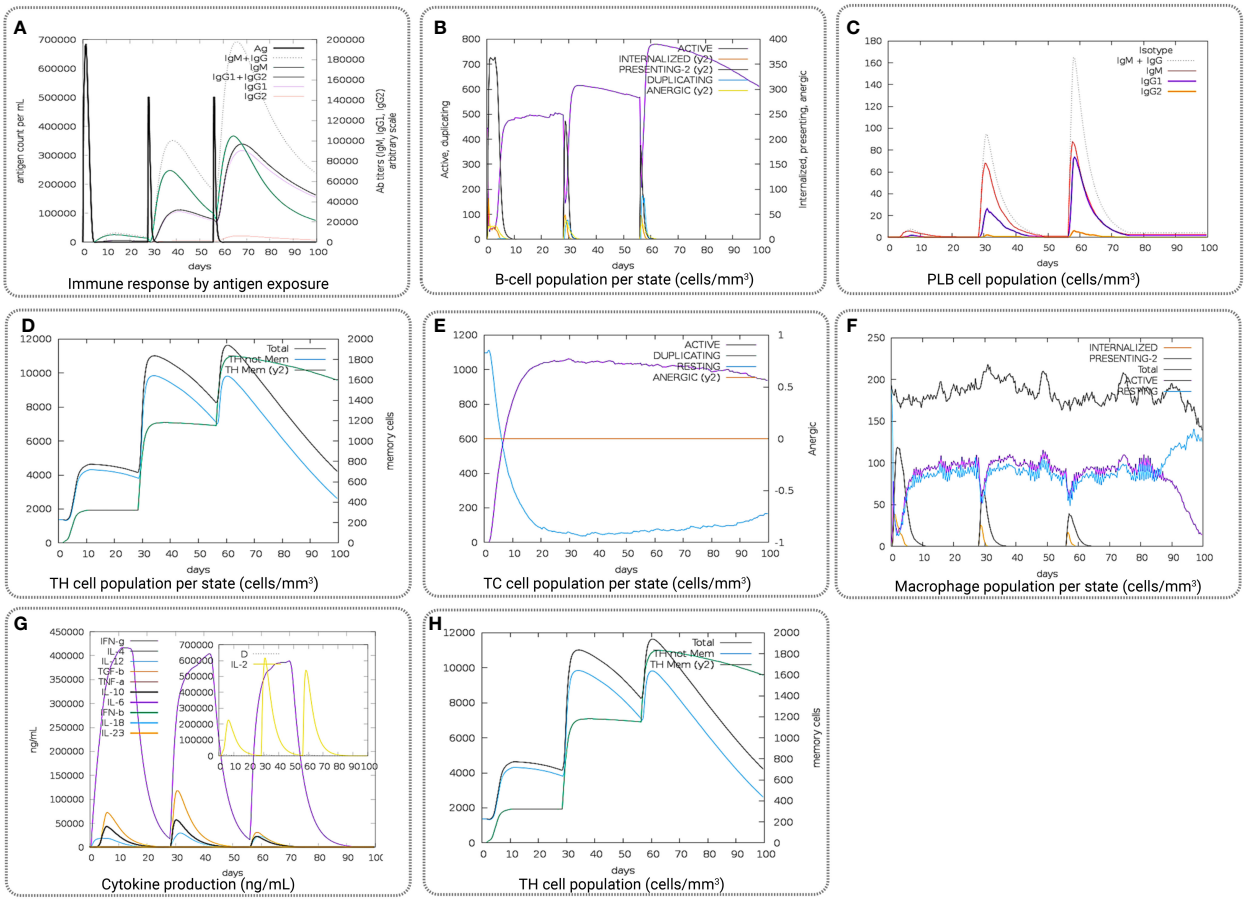


FIGURE 9
Representing the overall immune response of the vaccine candidate act as an antigen: **(A)** generation of immunoglobulins and B-cell isotypes upon exposure to an antigen; **(B)** amount of active B-cell populations per state; **(C)** amount of plasma B-lymphocytes and their isotypes per state; **(D)** state of helper T-cell population; **(E)** cytotoxic T-cell population per state of antigen exposure; **(F)** activity of macrophage population; **(G)** production of cytokine and interleukins in different states with the Simpson index, and **(H)** T.H. cell population (cells/mm^3).

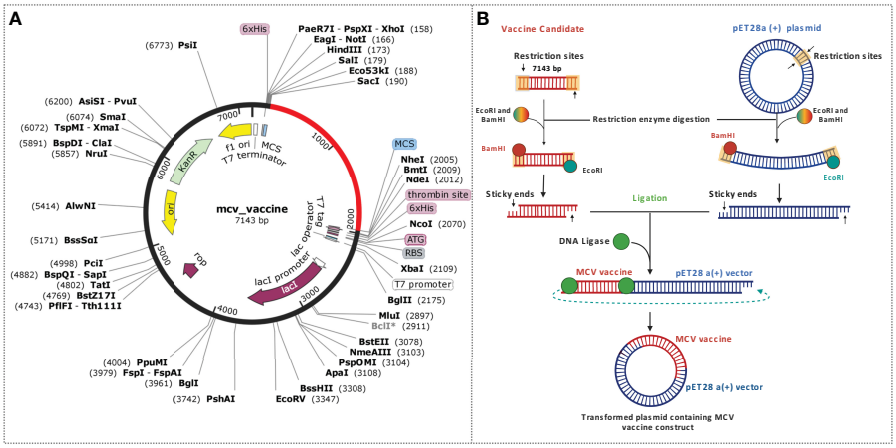


FIGURE 10
Schematic representation of the *in silico* cloning procedure for the MCV vaccine candidate into the pET28a (+) vector. **(A)** The coding gene sequence of the designed vaccine is depicted in red, while the vector backbone sequence of the designed vaccine is represented in black. **(B)** Illustration of the complete cloning process, encompassing restriction enzyme (RE) digestion and ligation steps.

viral pathogens (32, 86). However, the conventional approach to designing and developing vaccines against viruses is financially demanding and time-consuming (87). It necessitates a complex and intricate selection process for identifying suitable immunodominant epitopes, antigens, and efficient delivery systems, presenting significant challenges and difficulties. With the advent of immunoinformatics and computational approach formulation of prophylactic vaccines against a specific disease or pathogens has become the fastest, easiest, and most cost-effective (70, 88). The immune system appears to play a critical role in MCC biology, with increasing evidence of virus-specific cellular and humoral immune responses that influence the prognosis of MCC patients (89). In recent decades, many treatment strategies have been applied to treat cancer, but the specific treatment option for the disease remains elusive (90). Previously, different peptide vaccines have served as promising anticancer candidates due to their ability to target tumor cells and induce specific T-cell responses (91, 92). Therefore, the study aimed to design a multiepitope peptide vaccine candidate to fight against MCC, a widely viral-causing skin cancer.

We predicted effective epitopes as antigens and their correspondence alleles for both B and T cells to generate a sufficient immune response against MCC-positive tumors (Tables S9 and S10). The study initially identified and retrieved the sequences of five MCV proteins: large T antigen, small T antigen, VP1, VP2, and VP3. These proteins play crucial roles in the MCV infection process and contribute to the understanding of the virus' mechanisms (11). All five proteins (large T-antigen, small T-antigen, VP1, VP2, and VP3) exhibited notably high antigenicity scores. Consequently, we utilized all of these proteins to identify the most potent CD8, CD4, and linear B-cell epitopes, leading to the subsequent construction of a vaccine candidate targeting the virus. Several linear orders were applied to construct the multiepitope vaccine, and most potential vaccine structures were prepared by joining the adjuvant through the linker with CTL-HTL-BCL epitopes according to their higher to lower antigenic scores. The constructed vaccine candidate has a molecular weight of 64118.85 D with a theoretical PI of 8.72, indicating the basic properties of the protein. The aliphatic index provides insight into the relative presence of aliphatic side chains, such as alanine, valine, isoleucine, and leucine, within the protein structure. With a value of 77.55, the aliphatic index indicates a high level of thermal stability for the protein (87, 93). Additionally, the GRAVY value was determined to be -0.210, suggesting a hydrophilic nature of the construct and strong interactions with water molecules (88, 94). Finally, the study identified linear contiguous AA sequence fragments and confirmed AA fragments as potential epitopes for BCL that are immunogenic and antigen in nature and utilized for multiepitope vaccine design.

The 3D tertiary structure of the vaccine candidates was predicted and validated through different approaches. Subsequently, the vaccine candidates were refined, and structural validity was checked. The crude model of the vaccine candidates shows a Ramachandran score of 86.992% of the AA residues in the favorable region. After the refinement of the vaccine construct, the Ramachandran plot generated a better result of 94.512%, which means that most of the AA residues of the refined vaccine candidates were in the favorable regions. In addition, the Ramachandran plot shows that 94.512% of residues clustered tightly in the most favored region with very few residues in outliers. A good-quality model would probably exceed 90% in the most favored regions (95). The Z-score of the refined model was -2.59, indicating a satisfactory quality of the overall model. The Z-scores of the anticipated model were outside the scale of the property for local proteins, which shows the incorrect structure; thus, the MCV vaccine model is inside the scale property for local proteins (96, 97). Therefore, the structure of the vaccine candidate was deemed acceptable based on our evaluation. The length of the vaccine construct was determined to be 592 AA, acknowledging that the ideal length of a vaccine can vary based on factors including the target pathogen, desired immune response, and antigen or epitope characteristics (98). In the case of MCV, which exhibits genetic diversity and antigenic variability, it is often necessary to include multiple epitopes or larger antigenic regions to ensure comprehensive protection against various MCV strains (89). Incorporating multiple epitopes has the advantage of enhancing immune recognition, preventing immune evasion, and improving cross-reactivity and cross-protection (99). Therefore, the use of a 592 AA construct containing multiple epitopes supports an effective immune response and provides protection against MCV strains.

We also employed molecular docking simulation to determine the binding affinity between the vaccine candidate and the TLR-4 receptor (75, 100). We found that the vaccine can properly bind with the receptor TLR-4 and has the lowest binding energy score. A comprehensive structural analysis of the vaccine candidate and its receptor complex was also performed through MD simulation approaches to determine the binding stability of the complex system (76, 101). The vaccine conformation showed an average RMSD change of 11.94 Å, with fluctuations ranging from 3.109 Å to 16.162 Å. RMSF analysis identified peak fluctuations between amino acid residues 95–570, mainly at GLU122, VAL124, and LYS123. The compactness of the structure was confirmed by radius of gyration (Rg) analysis, which yielded an average Rg score of 39.25 Å. Analysis of HBs revealed a significant increase, peaking between 90 and 200 ns simulation time. For the vaccine–receptor complex, stable fluctuations were observed after 55 ns with an RMSD value of 16.918 Å. RMSF analysis highlights fluctuations between amino acid residues 650–800, particularly at position 700–710. The Rg analysis shows tight packing with deviation observed from the 5–55 ns simulation run (102). The number of hydrogen bonds was observed to fluctuate, with the highest number observed between 0 and 20 ns and 120 and 200 ns simulation time. Simulated microscale changes in the protein backbone and mild fluctuations of the side chain residues were observed in this study, which altogether confirmed the stability of the vaccine–TLR4 complex. These findings deepen our understanding of the conformation of the vaccine and the structural dynamics and interactions within its receptor complex. Stable and fluctuating regions provide valuable insights into conformational changes and intermolecular associations, facilitating further optimization and development of vaccines.

In addition, the immunological response of the vaccine candidate was also evaluated, which showed higher B- and T-cell activity, indicating the typical immune response. The evaluation of the immune response was performed by using the vaccine as an antigen, where a high level of immunoglobulin and B-cell isotype formation was observed upon exposure. Upon exposure, the number of active B-cell populations, plasma B-lymphocytes and

their isotype, helper T-cell and cytotoxic T-cell population per state; the number of plasma B-lymphocytes and their isotypes per state; state of helper T-cell population; and cytotoxic T-cell population, macrophage population, production of cytokine and interleukins per state were improved substantially, indicating a better memory formation ability of the vaccine candidates. In the final stage, the vaccine candidates underwent computational cloning into the pET28a (+) plasmid vector. Subsequently, the recombinant vaccine constructs were subjected to *in silico* expression within the *E. coli* K12 expression systems for subsequent analysis and evaluation. Before being computationally expressed into the host system, the vaccine candidate was optimized through the codon adaptation method (83). The GC content of the sequence was 52.36% in the optimized DNA sequence, indicating the optimal range (30%–70%) for expression (103). Additionally, the CAI value of the sequence was 0.98, close to 1.0, which indicates the higher expression probability of the vaccine candidate in the expression vector (104). Consequently, adequate adaptation was accomplished for the large-scale production of vaccine candidates.

Previously, programmed death-1 (PD-1) cell surface receptor inhibitors have found as a valuable treatment option for MCC, particularly in cases where cancer has spread or is not responsive to other therapies. Although Programmed death ligand 1 (PD-L1) blockade is highly effective, ~50% of infected patients with skin carcinoma either do not respond to PD-L1 therapy or develop PD-L1 refractory disease and, thus, do not experience long-term benefit (105, 106). Few other studies performed *in silico* analysis and constructed multiepitope peptide vaccines, although it is known that Merkel cell polyoma–mediated skin cancer is caused by the pathogenic proteins including the large T antigen, small T antigen, and viral capsid proteins (V.P. 1, V.P. 2, and Vp3) and all are involved in viral pathogenicity (107). However, the previous study has only either targeted only VP1 or T-antigenic proteins, for the epitope selection (31, 32). The current study identified and selected epitopes from all major pathogenic and antigenic proteins that will increase the efficacy of the vaccine candidates. Eventually, the study formulated a multiepitope vaccine candidate that will help to fight against MCV and boost the immune system of humans.

5 Conclusion

Recent groundbreaking developments in immunoinformatics have introduced novel techniques for disease prevention. Considering past outbreaks of viral infections in humans, computational approaches have been embraced to identify swift treatment strategies for diverse viral diseases. Peptide vaccines currently expressed as the most successful treatment option for viral infections can be designed by using either free peptides or peptides coated on dendritic cells. At this instant, peptide-based vaccines that are being designed by computational methods can play a critical role in the treatment of different infectious viral diseases. MCC is an aggressive infectious disease for which effective vaccine candidates are not available, and hence, it is necessary to develop an effective vaccine candidate. Therefore, in this study, we designed and identify a potential peptide vaccine candidate against MCV by using computational approaches that can be further utilized for subsequent vaccine construction. The study successfully identified peptide candidates against the virus and designed a valid multiepitope vaccine construct to fight against the aggressive MCC caused by MCV. However, further *in vitro* and *in vivo* investigations are suggested to finally determine to ensure the candidate vaccine's true potential in combating against MCV.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding authors.

Author contributions

FM, AA, and FA designed the project; RI, AS, MA, and RA performed the analysis of data; RI, AS, FM, RA, MA, and FA evaluated and interpreted the results; RI, RA, MA, and AS prepared the draft manuscript; RI, AS, RA, and FA performed data curation and visualization; FM, RI, AA, MA, and FA. critically reviewed and finalized the manuscript; FM, FA, and AA performed investigation and supervision of the manuscript. All authors contributed to the article and approved the submitted version.

Acknowledgments

The research received partial support from intramural funding allocated to FM by the College of Health and Life Sciences (CHLS) at Hamad Bin Khalifa University, Qatar Foundation. FA also received support from the same college. The researchers would also like to acknowledge the Deanship of Scientific Research (DSR) at Taif University for funding this work. Furthermore, they deeply appreciate the valuable technical support offered by the Biological Solution Centre (BioSol Centre).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2023.1160260/full#supplementary-material

References

1. Feng H, Shuda M, Chang Y, Moore PS. Clonal integration of a polyomavirus in human merkel cell carcinoma. *Sci (80-)* (2008) 319:1096–100. doi: 10.1126/science.1152586

2. Moens U, Macdonald A. Effect of the large and small t-antigens of human polyomaviruses on signaling pathways. *Int J Mol Sci* (2019) 20:3914. doi: 10.3390/ijms20163914

3. Liu W, MacDonald M, You J. Merkel cell polyomavirus infection and merkel cell carcinoma. *Curr Opin Virol* (2016) 20:20–7. doi: 10.1016/j.coviro.2016.07.011

4. Bayrou O, Avril MF, Charpentier P, Caillou B, Guillaume JC, Prade M. Primary neuroendocrine carcinoma of the skin: clinicopathologic study of 18 cases. *J Am Acad Dermatol* (1991) 24:198–207. doi: 10.1016/0190-9622(91)70027-Y

5. Walsh NM, Cerroni L. Merkel cell carcinoma: a review. *J Cutan Pathol* (2021) 48:411–21. doi: 10.1111/cup.13910

6. Woo SH, Lumpkin EA, Patapoutian A. Merkel cells and neurons keep in touch. *Trends Cell Biol* (2015) 25:74–81. doi: 10.1016/j.tcb.2014.10.003

7. Moshiri AS, Nghiem P. Milestones in the staging, classification, and biology of merkel cell carcinoma. *JNCCN J Natl Compr Cancer Netw* (2014) 12:1255–62. doi: 10.6004/jnccn.2014.0123

8. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer statistics, 2022. *CA Cancer J Clin* (2022) 72:7–33. doi: 10.3322/CAAC.21708

9. Pietropaolo V, Prezioso C, Moens U. Merkel cell polyomavirus and merkel cell carcinoma. *Cancers (Basel)* (2020) 12:1774. doi: 10.3390/cancers12071774

10. Stakaityte G, Wood JJ, Knight LM, Abdul-Sada H, Adzahar NS, Nwogu N, et al. Merkel cell polyomavirus: molecular insights into the most recently discovered human tumour virus. *Cancers (Basel)* (2014) 6:1267–97. doi: 10.3390/cancers6031267

11. Spurgeon ME, Lambert PF. Merkel cell polyomavirus: a newly discovered human virus with oncogenic potential. *Virology* (2013) 435:118–30. doi: 10.1016/j.virol.2012.09.029

12. MacDonald M, You J. Merkel cell polyomavirus: a new DNA virus associated with human cancer. In: *Advances in experimental medicine and biology* (2017). (Singapore: Springer Nature) p. 35–56. doi: 10.1007/978-981-10-5765-6_4

13. Vilchez RA, Butel JS. Emergent human pathogen simian virus 40 and its role in cancer. *Clin Microbiol Rev* (2004) 17:495–508. doi: 10.1128/CMR.17.3.495-508.2004

14. Burke JM, Bass CR, Kincaid RP, Ulug ET, Sullivan CS. The murine polyomavirus MicroRNA locus is required to promote viruria during the acute phase of infection. *J Virol* (2018) 92:e02131–17. doi: 10.1128/jvi.02131-17

15. Rinaldo CH, Tylden GD, Sharma BN. The human polyomavirus BK (BKPyV): virological background and clinical implications. *APMIS* (2013) 121:728–45. doi: 10.1111/apm.12134

16. Lauver MD, Lukacher AE. JCPyV VP1 mutations in progressive multifocal leukoencephalopathy: altering tropism or mediating immune evasion? *Viruses* (2020) 12:1156. doi: 10.3390/v12101156

17. Asseri AH, Alam MJ, Alzahrani F, Khames A, Pathan MT, Abourehab MAS, et al. Toward the identification of natural antiviral drug candidates against merkel cell polyomavirus: computational drug design approaches. *Pharm* (2022) 15:501. doi: 10.3390/PH15050501

18. Shuda M, Kwun HJ, Feng H, Chang Y, Moore PS. Human merkel cell polyomavirus small T antigen is an oncoprotein targeting the 4E-BP1 translation regulator. *J Clin Invest* (2011) 121:3623–34. doi: 10.1172/JCI46323

19. Houben R, Shuda M, Weinkam R, Schrama D, Feng H, Chang Y, et al. Merkel cell polyomavirus-infected merkel cell carcinoma cells require expression of viral T antigens. *J Virol* (2010) 84:7064–72. doi: 10.1128/jvi.02400-09

20. Schowalter RM, Buck CB. The merkel cell polyomavirus minor capsid protein. *PloS Pathog* (2013) 9:e1003558. doi: 10.1371/journal.ppat.1003558

21. Pastrana DV, Tolstov YL, Becker JC, Moore PS, Chang Y, Buck CB. Quantitation of human seroresponsiveness to merkel cell polyomavirus. *PloS Pathog* (2009) 5:e1000578. doi: 10.1371/journal.ppat.1000578

22. Meites E, Kempe A, Markowitz LE. Use of a 2-dose schedule for human papillomavirus vaccination [[/amp]]mdash; updated recommendations of the advisory committee on immunization practices. *MMWR Morb Mortal Wkly Rep* (2016) 65:1405–8. doi: 10.15585/mmwr.mm6549a5

23. Udomkarnjananun S, Takkavatakarn K, Praditpornsilpa K, Nader C, Eiam-Ong S, Jaber BL, et al. Hepatitis b virus vaccine immune response and mortality in dialysis patients: a meta-analysis. *J Nephrol* (2020) 33:343–54. doi: 10.1007/s40620-019-00668-1

24. Zhang X, Xin L, Li S, Fang M, Zhang J, Xia N, et al. Lessons learned from successful human vaccines: delineating key epitopes by dissecting the capsid proteins. *Hum Vaccines Immunother* (2015) 11:1277–92. doi: 10.1080/21645515.2015.1016675

25. Kumari S, Kessel A, Singhal D, Kaur G, Bern D, Lemay-St-Denis C, et al. Computational identification of a multi-peptide vaccine candidate in E2 glycoprotein against diverse hepatitis c virus genotypes. *J Biomol Struct Dyn* (2023), 1–12. doi: 10.1080/07391102.2023.2212777

26. Jain S, Baranwal M. Conserved peptide vaccine candidates containing multiple Ebola nucleoprotein epitopes display interactions with diverse HLA molecules. *Med Microbiol Immunol* (2019) 208:227–38. doi: 10.1007/s00430-019-00584-y

27. Ayub G, Waheed Y, Najmi MH. Prediction and conservancy analysis of promiscuous T-cell binding epitopes of Ebola virus l protein: an in silico approach. *Asian Pacific J Trop Dis* (2016) 6:169–73. doi: 10.1016/S2222-1808(15)61007-6

28. Terry FE, Moise L, Martin RF, Torres M, Pilotte N, Williams SA, et al. Time for T? immunoinformatics addresses vaccine design for neglected tropical and emerging infectious diseases. *Expert Rev Vaccines* (2014) 14:21–35. doi: 10.1586/14760584.2015.955478

29. Kwun HJ, Shuda M, Feng H, Camacho CJ, Moore PS, Chang Y. Merkel cell polyomavirus small T antigen controls viral replication and oncoprotein expression by targeting the cellular ubiquitin ligase SCF Fbw7. *Cell Host Microbe* (2013) 14:125–35. doi: 10.1016/j.chom.2013.06.008

30. Zeng Q, Gomez BP, Viscidi RP, Peng S, He L, Ma B, et al. Development of a DNA vaccine targeting merkel cell polyomavirus. *Vaccine* (2012) 30:1322–9. doi: 10.1016/j.vaccine.2011.12.072

31. Gomez B, He L, Tsai YC, Wu TC, Viscidi RP, Hung CF. Creation of a merkel cell polyomavirus small T antigen-expressing murine tumor model and a DNA vaccine targeting small T antigen. *Cell Biosci* (2013) 3:1–8. doi: 10.1186/2045-3701-3-29

32. Xu D, Jiang S, He Y, Jin X, Zhao G, Wang B. Development of a therapeutic vaccine targeting merkel cell polyomavirus capsid protein VP1 against merkel cell carcinoma. *NPJ Vaccines* (2021) 6:119. doi: 10.1038/s41541-021-00382-9

33. Boutet E, Lieberherr D, Tognolli M, Schneider M, Bairoch A. "UniProtKB/Swiss-prot.,". *Methods Mol Biol* (2007) 30:846–51. doi: 10.1007/978-1-59745-535-0_4

34. Doytchinova IA, Flower DR. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinf* (2007) 8:1–7. doi: 10.1186/1471-2105-8-4

35. Magnan CN, Zeller M, Kayala MA, Vigil A, Randall A, Felgner PL, et al. High-throughput prediction of protein antigenicity using protein microarray data. *Bioinformatics* (2010) 26:2936–43. doi: 10.1093/bioinformatics/btq551

36. Larsen MV, Lundegaard C, Lamberth K, Buus S, Lund O, Nielsen M. Large-Scale validation of methods for cytotoxic T-lymphocyte epitope prediction. *BMC Bioinf* (2007) 8:424. doi: 10.1186/1471-2105-8-424

37. Dimitrov I, Naneva L, Doytchinova I, Bangov I. AllergenFP: allergenicity prediction by descriptor fingerprints. *Bioinformatics* (2014) 30:846–51. doi: 10.1093/bioinformatics/btt619

38. Gupta S, Kapoor P, Chaudhary K, Gautam A, Kumar R, Raghava GPS. In silico approach for predicting toxicity of peptides and proteins. *PloS One* (2013) 8:e73957. doi: 10.1371/journal.pone.0073957

39. Calis JJA, Maybeno M, Greenbaum JA, Weiskopf D, De Silva AD, Sette A, et al. Properties of MHC class I presented peptides that enhance immunogenicity. *PloS Comput Biol* (2013) 9:e1003266. doi: 10.1371/journal.pcbi.1003266

40. Fleri W, Paul S, Dhanda SK, Mahajan S, Xu X, Peters B, et al. The immune epitope database and analysis resource in epitope discovery and synthetic vaccine design. *Front Immunol* (2017) 8:278. doi: 10.3389/fimmu.2017.00278

41. Dhanda SK, Vir P, Raghava GPS. Designing of interferon-gamma inducing MHC class-II binders. *Biol Direct* (2013) 8:1–15. doi: 10.1186/1745-6150-8-30

42. Jespersen MC, Peters B, Nielsen M, Marcatili P. BepiPred-2.0: improving sequence-based b-cell epitope prediction using conformational epitopes. *Nucleic Acids Res* (2017) 45:W24–9. doi: 10.1093/nar/gkx346

43. Garg VK, Avashthi H, Tiwari A, Jain PA, Ramkete PWR, Kayastha AM, et al. MFPPI – multi FASTA ProtParam interface. *Bioinformation* (2016) 12:74–7. doi: 10.6026/97320630012074

44. Magnan CN, Randall A, Baldi P. SOLpro: accurate sequence-based prediction of protein solubility. *Bioinformatics* (2009) 25:2200–7. doi: 10.1093/bioinformatics/btp386

45. Zhang Y. I-TASSER server for protein 3D structure prediction. *BMC Bioinf* (2008) 9:1–8. doi: 10.1186/1471-2105-9-40

46. Heo L, Park H, Seok C. GalaxyRefine: protein structure refinement driven by side-chain repacking. *Nucleic Acids Res* (2013) 41:W384. doi: 10.1093/nar/gkt458

47. Wiederstein M, Sippl MJ. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res* (2007) 35:W407. doi: 10.1093/NAR/GKM290

48. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The protein data bank. *Nucleic Acids Res* (2000) 28:235–42. doi: 10.1093/nar/28.1.235

49. Kozakov D, Hall DR, Xia B, Porter KA, Padhorny D, Yueh C, et al. The ClusPro web server for protein-protein docking. *Nat Protoc* (2017) 12:255–78. doi: 10.1038/nprot.2016.169

50. Grote A, Hiller K, Scheer M, Münch R, Nörtemann B, Hempel DC, et al. JCat: a novel tool to adapt codon usage of a target gene to its potential expression host. *Nucleic Acids Res* (2005) 33:526–31. doi: 10.1093/nar/gki376

51. Apweiler R. The universal protein resource (UniProt). *Nucleic Acids Res* (2008) 36:D190. doi: 10.1093/nar/gkm895

52. Kar PP, Araveti PB, Kuriakose A, Srivastava A. Design of a multi-epitope protein as a subunit vaccine against lumpy skin disease using an immunoinformatics approach. *Sci Rep* (2022) 12:1–11. doi: 10.1038/s41598-022-23272-z

53. Samad A, Ahammad F, Nain Z, Alam R, Imon RR, Hasan M, et al. Designing a multi-epitope vaccine against SARS-CoV-2: an immunoinformatics approach. *J Biomol Struct Dyn* (2020) 40:14–30. doi: 10.1080/07391102.2020.1792347

54. Larsen MV, Lundegaard C, Lamberth K, Buus S, Brunak S, Lund O, et al. An integrative approach to CTL epitope prediction: a combined algorithm integrating MHC class I binding, TAP transport efficiency, and proteasomal cleavage predictions. *Eur J Immunol* (2005) 35:2295–303. doi: 10.1002/eji.200425811

55. Dhanda SK, Mahajan S, Paul S, Yan Z, Kim H, Jespersen MC, et al. IEDB-AR: immune epitope database - analysis resource in 2019. *Nucleic Acids Res* (2019) 47: W502–6. doi: 10.1093/nar/gkz452

56. Dhanda SK, Gupta S, Vir P, Raghava GP. Prediction of IL4 inducing peptides. *Clin Dev Immunol* (2013) 2013:263952. doi: 10.1155/2013/263952

57. Nagpal G, Usmani SS, Dhanda SK, Kaur H, Singh S, Sharma M, et al. Computer-aided designing of immunosuppressive peptides based on IL-10 inducing potential. *Sci Rep* (2017) 7:1–10. doi: 10.1038/srep42851

58. El-Manzalawy Y, Dobbs D, Honavar VG. In silico prediction of linear b-cell epitopes on proteins. In: *Prediction of protein secondary Structure. Methods in molecular biology*. Springer Nature (2017). p. 255–64. doi: 10.1007/978-1-4939-6406-2_17

59. Weyant KB, Oloyede A, Pal S, Liao J, De Jesus MR, Jaroentomeechai T, et al. A modular vaccine platform enabled by decoration of bacterial outer membrane vesicles with biotinylated antigens. *Nat Commun* (2023) 14:1–15. doi: 10.1038/s41467-023-36101-2

60. Shiina T, Hosomichi K, Inoko H, Kulski JK. The HLA genomic loci map: expression, interaction, diversity and disease. *J Hum Genet* (2009) 54:15–39. doi: 10.1038/jhg.2008.5

61. Misra N, Panda PK, Shah K, Sukla LB, Chaubey P. Population coverage analysis of T-cell epitopes of neisseria meningitidis serogroup b from iron acquisition proteins for vaccine design. *Bioinformation* (2011) 6:255–61. doi: 10.6026/97320630006255

62. Maleki A, Russo G, Parasiliti Palumbo GA, Pappalardo F. In silico design of recombinant multi-epitope vaccine against influenza a virus. *BMC Bioinf* (2021) 22:1–18. doi: 10.1186/s12859-022-04581-6

63. Jouhi L, Koljonen V, Böhling T, Caj H. The expression of toll-like receptors 2, 4, 5, 7 and 9 in merkel cell carcinoma. *Anticancer Res* (2015) 35:1843–9.

64. Bhatia S, Miller NJ, Lu H, Longino NV, Ibrani D, Shinohara MM, et al. Intratumoral G100, a TLR4 agonist, induces antitumor immune responses and tumor regression in patients with merkel cell carcinoma. *Clin Cancer Res* (2019) 25:1185–95. doi: 10.1158/1078-0432.CCR-18-0469

65. Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, et al. Deciphering the biology of mycobacterium tuberculosis from the complete genome sequence. *Nature* (1998) 393:537–44. doi: 10.1038/31159

66. Alam R, Samad A, Ahammad F, Nur SM, Alsaiari AA, Imon RR, et al. In silico formulation of a next-generation multiepitope vaccine for use as a prophylactic candidate against Crimean-Congo hemorrhagic fever. *BMC Med* (2023) 21:1–19. doi: 10.1186/s12916-023-02750-9

67. Chen X, Zaro JL, Shen WC. Fusion protein linkers: property, design and functionality. *Adv Drug Del Rev* (2013) 65:1357–69. doi: 10.1016/j.addr.2012.09.039

68. Dey J, Mahapatra SR, Singh PK, Prabhuswamimath SC, Misra N, Suar M. Designing of multi-epitope peptide vaccine against acinetobacter baumannii through combined immunoinformatics and protein interaction–based approaches. *Immunol Res* (2023) 1:1–24. doi: 10.1007/s12026-023-09374-4

69. Gu Y, Sun X, Li B, Huang J, Zhan B, Zhu X. Vaccination with a paramyosin-based multi-epitope vaccine elicits significant protective immunity against trichinella spiralis infection in mice. *Front Microbiol* (2017) 8:1475. doi: 10.3389/fmicb.2017.01475

70. Bhuiyan MA, Quayum ST, Ahammad F, Alam R, Samad A, Nain Z. Discovery of potential immune epitopes and peptide vaccine design - a prophylactic strategy against rift valley fever virus. *F1000Research* (2020) 9:999. doi: 10.12688/f1000research.24975.1

71. McGuffin LJ, Bryson K, Jones DT. The PSIPRED protein structure prediction server. *Bioinformatics* (2000) 16:404–5. doi: 10.1093/bioinformatics/16.4.404

72. Atapour A, Ghalamfarsa F, Naderi S, Hatam G. Designing of a novel fusion protein vaccine candidate against human visceral leishmaniasis (VL) using immunoinformatics and structural approaches. *Int J Pept Res Ther* (2021) 27:1885–98. doi: 10.1007/s10989-021-10218-8

73. Pražnikar J, Tomić M, Turk D. Validation and quality assessment of macromolecular structures using complex network analysis. *Sci Rep* (2019) 9:1–11. doi: 10.1038/s41598-019-38658-9

74. Anderson RJ, Weng Z, Campbell RK, Jiang X. Main-chain conformational tendencies of amino acids. *Proteins Struct Funct Genet* (2005) 60:679–89. doi: 10.1002/prot.20530

75. Opo FADM, Rahman MM, Ahammad F, Ahmed I, Bhuiyan MA, Asiri AM. Structure based pharmacophore modeling, virtual screening, molecular docking and ADMET approaches for identification of natural anti-cancer agents targeting XIAP protein. *Sci Rep* (2021) 11:4049. doi: 10.1038/s41598-021-83626-x

76. Pokhrel S, Bouback TA, Samad A, Nur SM, Alam R, Abdullah-Al-Mamun M, et al. Spike protein recognizer receptor ACE2 targeted identification of potential natural antiviral drug candidates against SARS-CoV-2. *Int J Biol Macromol* (2021) 191:1114–25. doi: 10.1016/J.IJBIOMAC.2021.09.146

77. Yuan S, Chan HCS, Hu Z. Using PyMOL as a platform for computational drug design. *Wiley Interdiscip Rev Comput Mol Sci* (2017) 7:e1298. doi: 10.1002/wcms.1298

78. Roos K, Wu C, Damm W, Reboul M, Stevenson JM, Lu C, et al. OPLS3e: extending force field coverage for drug-like small molecules. *J Chem Theory Comput* (2019) 15:1863–74. doi: 10.1021/acs.jctc.8b01026

79. Kim M, Kim E, Lee S, Kim JS, Lee S. New method for constant- NPT molecular dynamics. *J Phys Chem A* (2019) 123:1689–99. doi: 10.1021/acs.jpca.8b09082

80. Basconi JE, Shirts MR. Effects of temperature control algorithms on transport properties and kinetics in molecular dynamics simulations. *J Chem Theory Comput* (2013) 9:2887–99. doi: 10.1021/ct400109a

81. Rapin N, Lund O, Castiglione F. Immune system simulation online. *Bioinformatics* (2011) 27:2013–4. doi: 10.1093/bioinformatics/btr335

82. Mauro VP, Chappell SA. A critical analysis of codon optimization in human therapeutics. *Trends Mol Med* (2014) 20:604–13. doi: 10.1016/j.molmed.2014.09.003

83. Sharp PM, Li WH. The codon adaptation index-a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res* (1987) 15:1281–95. doi: 10.1093/nar/15.3.1281

84. Roy AA, Dhawanjewar AS, Sharma P, Singh G, Madhusudhan MS. Protein interaction z score assessment (PIZSA): an empirical scoring scheme for evaluation of protein-protein interactions. *Nucleic Acids Res* (2019) 47:W331–7. doi: 10.1093/nar/gkz368

85. Liang E, Brower JV, Rice SR, Buehler DG, Saha S, Kimple RJ. Merkel cell carcinoma analysis of outcomes: a 30-year experience. *PloS One* (2015) 10:e0129476. doi: 10.1371/journal.pone.0129476

86. Hasan MR, Alsaiari AA, Fakhurji BZ, Molla MHR, Asseri AH, Sumon MAA, et al. Application of mathematical modeling and computational tools in the modern drug design and development process. *Molecules* (2022) 27:4169. doi: 10.3390/molecules27134169

87. Plotkin S, Robinson JM, Cunningham G, Iqbal R, Larsen S. The complexity and cost of vaccine manufacturing – an overview. *Vaccine* (2017) 35:4064–71. doi: 10.1016/j.vaccine.2017.06.003

88. Sunita, Sajid A, Singh Y, Shukla P. Computational tools for modern vaccine development. *Hum Vaccines Immunother* (2020) 16:723–35. doi: 10.1080/21645515.2019.1670035

89. Bhatia S, Afanasiev O, Nghiem P. Immunobiology of merkel cell carcinoma: implications for immunotherapy of a polyomavirus-associated cancer. *Curr Oncol Rep* (2011) 13:488–97. doi: 10.1007/s11912-011-0197-5

90. Tang T, Huang X, Zhang G, Hong Z, Bai X, Liang T. Advantages of targeting the tumor immune microenvironment over blocking immune checkpoint in cancer immunotherapy. *Signal Transduct Target Ther* (2021) 6:1–13. doi: 10.1038/s41392-020-00449-4

91. Lin MJ, Svensson-Arvelund J, Lubitz GS, Marabelle A, Melero I, Brown BD, et al. Cancer vaccines: the next immunotherapy frontier. *Nat Cancer* (2022) 3:911–26. doi: 10.1038/s43018-022-00418-6

92. Hu Z, Ott PA, Wu CJ. Towards personalized, tumour-specific, therapeutic vaccines for cancer. *Nat Rev Immunol* (2018) 18:168–82. doi: 10.1038/nri.2017.131

93. Ikai A. Thermostability and aliphatic index of globular proteins. *J Biochem* (1980) 88:1895–8. doi: 10.1093/oxfordjournals.jbchem.a133168

94. Chang KY, Yang JR. Analysis and prediction of highly effective antiviral peptides based on random forests. *PloS One* (2013) 8:70166. doi: 10.1371/journal.pone.0070166

95. Sobolev OV, Afonine PV, Moriarty NW, Hekkelman ML, Joosten RP, Perrakis A, et al. A global ramachandran score identifies protein structures with unlikely stereochemistry. *Structure* (2020) 28:1249–1258.e2. doi: 10.1016/j.str.2020.08.005

96. Benkert P, Biasini M, Schwede T. Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics* (2011) 27:343–50. doi: 10.1093/bioinformatics/btq662

97. Aljahdali MO, Molla MHR, Ahammad F. Compounds identified from marine mangrove plant (Avicennia alba) as potential antiviral drug candidates against WDSV, an in-silico approach. *Mar Drugs* (2021) 19:253. doi: 10.3390/md19050253

98. Graham BS, Gilman MSA, McLellan JS. Structure-based vaccine antigen design. *Annu Rev Med* (2019) 70:91–104. doi: 10.1146/annurev-med-121217-094234

99. Pollard AJ, Bijker EM. A guide to vaccinology: from basic principles to new developments. *Nat Rev Immunol* (2021) 21:83–100. doi: 10.1038/s41577-020-00479-7

100. Ahammad F, Alam R, Mahmud R, Akhter S, Talukder EK, Tonmoy AM, et al. Pharmacoinformatics and molecular dynamics simulation-based phytochemical screening of neem plant (Azadiractha indica) against human cancer by targeting MCM7 protein. *Brief Bioinform* (2021) 2021:1–15. doi: 10.1093/bib/bbab098

101. Bouback TA, Pokhrel S, Albeshri A, Aljohani AM, Samad A, Alam R, et al. Pharmacophore-based virtual screening, quantum mechanics calculations, and molecular dynamics simulation approaches identified potential natural antiviral drug candidates against MERS-CoV S1-NTD. *Mol* (2021) 26:4961. doi: 10.3390/MOLECULES26164961

102. Alam R, Rahman Imon R, Enamul M, Talukder K, Akhter S, Hossain MA, et al. GC-MS analysis of phytoconstituents from ruellia prostrata and senna tora and identification of potential anti-viral activity against SARS-CoV-2. *RSC Adv* (2021) 11:40120–35. doi: 10.1039/D1RA06842C

103. Ranaghan MJ, Li JJ, Laprise DM, Garvie CW. Assessing optimal: inequalities in codon optimization algorithms. *BMC Biol* (2021) 19:1–13. doi: 10.1186/s12915-021-00968-8

104. Sen A, Kargar K, Akgün E, Plnar MC. Codon optimization: a mathematical programing approach. *Bioinformatics* (2020) 36:4012–20. doi: 10.1093/bioinformatics/btaa248

105. Sabbatino F, Marra A, Liguori L, Scognamiglio G, Fusciello C, Botti G, et al. Resistance to anti-PD-1-based immunotherapy in basal cell carcinoma: a case report and review of the literature. *J Immunother Cancer* (2018) 6:1–8. doi: 10.1186/s40425-018-0439-2

106. Patrinely JR, Dewan AK, Johnson DB. The role of anti-PD-1/PD-L1 in the treatment of skin cancer. *BioDrugs* (2020) 34:495–503. doi: 10.1007/s40259-020-00428-9

107. Liu W, You J. Molecular mechanisms of merkel cell polyomavirus transformation and replication. *Annu Rev Virol* (2020) 7:289–307. doi: 10.1146/annurev-virology-011720-121757

frontiers | Frontiers in Immunology

TYPE Original Research
PUBLISHED 18 September 2023
DOI 10.3389/fimmu.2023.1186188

OPEN ACCESS

EDITED BY
Joe Hou,
Fred Hutchinson Cancer Research Center, United States

REVIEWED BY
Mohanned Alhussien,
Technical University of Munich, Germany
Alexis Labrada,
National Center of Bioproducts (BIOCEN), Cuba

*CORRESPONDENCE
Abel E. Vásquez
avasquez@ispch.cl
María Inés Becker
mariaines.becker@fucited.cl
Diego A. Díaz-Dinamarca
dadiaz@ispch.cl

RECEIVED 14 March 2023
ACCEPTED 29 August 2023
PUBLISHED 18 September 2023

CITATION
Díaz-Dinamarca DA, Salazar ML, Escobar DF, Castillo BN, Valdebenito B, Díaz P, Manubens A, Salazar F, Troncoso MF, Lavandero S, Díaz J, Becker MI and Vásquez AE (2023) Surface immunogenic protein from Streptococcus agalactiae and *Fissurella latimarginata* hemocyanin are TLR4 ligands and activate MyD88- and TRIF dependent signaling pathways. *Front. Immunol.* 14:1186188. doi: 10.3389/fimmu.2023.1186188

Surface immunogenic protein from Streptococcus agalactiae and *Fissurella latimarginata* hemocyanin are TLR4 ligands and activate MyD88- and TRIF dependent signaling pathways

Diego A. Díaz-Dinamarca[1,2,3]*, Michelle L. Salazar[2], Daniel F. Escobar[1], Byron N. Castillo[2], Bastián Valdebenito[1], Pablo Díaz[1], Augusto Manubens[4], Fabián Salazar[2,4,5], Mayarling F. Troncoso[6], Sergio Lavandero[6,7], Janepsy Díaz[8], María Inés Becker[2,4]* and Abel E. Vásquez[1,9]*

[1]Sección de Biotecnología, Subdepartamento, Innovación, Desarrollo, Transferencia Tecnológica (I+D+T) y Evaluación de Tecnologías Sanitarias (ETESA), Instituto de Salud Pública, Santiago, Chile, [2]Laboratorio de Inmunología, Fundación Ciencia y Tecnología para el Desarrollo (FUCITED), Santiago, Chile, [3]Facultad de Ciencias Químicas y Farmacéuticas, Universidad de Chile, Santiago, Chile, [4]Investigación y Desarrollo, BIOSONDA S.A., Santiago, Chile, [5]Medical Research Council Centre for Medical Mycology, University of Exeter, Exeter, United Kingdom, [6]Advanced Center for Chronic Diseases (ACCDiS), Facultad Ciencias Químicas y Farmacéuticas and Facultad de Medicina, Universidad de Chile, Santiago, Chile, [7]Department of Internal Medicine (Cardiology Division), University of Texas Southwestern Medical Center, Dallas, TX, United States, [8]Departamento Agencia Nacional de Dispositivos Médicos, Innovación y Desarrollo, Instituto de Salud Pública de Chile, Santiago, Chile, [9]Facultad de Ciencias de la Salud, Escuela de Medicina, Universidad del Alba, Santiago, Chile

The development of vaccine adjuvants is of interest for the management of chronic diseases, cancer, and future pandemics. Therefore, the role of Toll-like receptors (TLRs) in the effects of vaccine adjuvants has been investigated. TLR4 ligand-based adjuvants are the most frequently used adjuvants for human vaccines. Among TLR family members, TLR4 has unique dual signaling capabilities due to the recruitment of two adapter proteins, myeloid differentiation marker 88 (MyD88) and interferon-β adapter inducer containing the toll-interleukin-1 receptor (TIR) domain (TRIF). MyD88-mediated signaling triggers a proinflammatory innate immune response, while TRIF-mediated signaling leads to an adaptive immune response. Most studies have used lipopolysaccharide-based ligands as TLR4 ligand-based adjuvants; however, although protein-based ligands have been proven advantageous as adjuvants, their mechanisms of action, including their ability to undergo structural modifications to achieve optimal immunogenicity, have been explored less thoroughly. In this work, we characterized the effects of two protein-based adjuvants (PBAs) on TLR4 signaling via the recruitment of MyD88 and TRIF. As models of TLR4-PBAs, we used hemocyanin from *Fissurella latimarginata* (FLH) and a recombinant surface immunogenic protein (rSIP) from *Streptococcus*

10.3389/fimmu.2023.1186188

agalactiae. We determined that rSIP and FLH are partial TLR4 agonists, and depending on the protein agonist used, TLR4 has a unique bias toward the TRIF or MyD88 pathway. Furthermore, when characterizing gene products with MyD88 and TRIF pathway-dependent expression, differences in TLR4-associated signaling were observed. rSIP and FLH require MyD88 and TRIF to activate nuclear factor kappa beta (NF-κB) and interferon regulatory factor (IRF). However, rSIP and FLH have a specific pattern of interleukin 6 (IL-6) and interferon gamma-induced protein 10 (IP-10) secretion associated with MyD88 and TRIF recruitment. Functionally, rSIP and FLH promote antigen cross-presentation in a manner dependent on TLR4, MyD88 and TRIF signaling. However, FLH activates a specific TRIF-dependent signaling pathway associated with cytokine expression and a pathway dependent on MyD88 and TRIF recruitment for antigen cross-presentation. Finally, this work supports the use of these TLR4-PBAs as clinically useful vaccine adjuvants that selectively activate TRIF- and MyD88-dependent signaling to drive safe innate immune responses and vigorous Th1 adaptive immune responses.

KEYWORDS

protein-based adjuvants (PBAs), TLR4 agonist, MyD88, TRIF, antigen-presenting cells, vaccines, recombinant surface immunological protein from *Streptococcus agalactiae* (rSIP), hemocyanin from *Fissurella latimarginata* (FLH)

1 Introduction

In recent decades, TLR agonists have been investigated as possible vaccine adjuvants. Most compounds with adjuvant effects, such as lipopolysaccharide and oligonucleotides, are nonprotein microbial components. However, many studies have reported that TLR-dependent immunomodulation can be activated by numerous xenogeneic proteins in a thymus-dependent manner (1, 2). However, the potential role of these proteins as adjuvants requires a deeper understanding of their mechanisms of action to enable the creation of adjuvants with more powerful and more specific immunological effects. Among the many known TLR agonists, TLR4 ligand-based adjuvants are the most commonly used for developing commercial vaccines (3, 4). However, an improved understanding of TLR4 receptor–ligand interactions, signaling pathways, and biological/immunological mechanisms is needed to develop safe and potent vaccine formulations (5–7).

TLR4 is a transmembrane protein in leukocytes that belongs to the leucine-rich repeat family of proteins. It is activated by lipopolysaccharide (LPS), which triggers innate responses against gram-negative pathogens (8). Interaction of a TLR with its corresponding ligand agonist results in TLR dimerization, which triggers the recruitment of adapter proteins to the Toll IL-1 receptor (TIR) in the cytoplasmic domain of TLR4. This dimerization-based signaling process is an essential step in TLR4 signaling in which cytosolic TIR domains are activated to recruit adapter molecules, such as myeloid differentiation primary response factor 88 (MyD88), adapter-like MyD88 (MAL), the TIR domain-containing interferon-β inducer adapter (TRIF), and TRIF-related adapter molecule (TRAM), which then facilitate downstream signaling (9–12).

MyD88 signaling is associated with the rapid production of proinflammatory cytokines and innate immune responses to infectious threats (13, 14). In contrast, TRIF signaling is associated with processes that can promote adaptive immune responses essential for effective vaccination (10, 11, 15). Considering the roles of TLR4 agonists, most studies have used LPS-based ligands. However, a growing number of TLR4 protein agonists that could be used as vaccine adjuvants have been described (2, 16). Indeed, protein TLR4 agonists have several unique properties, including the ability to undergo structural modulation, optimal immunogenicity, and minimal toxicity (2, 16, 17). In this context, we analyzed two protein-based adjuvant agonists of TLR4 in terms of activation of the MyD88 and TRIF signaling pathways: one of bacterial origin, namely, the surface immunogenic protein (SIP) of Group *B Streptococcus* (GBS), and the other of molluskan origin, namely, hemocyanin from *Fissurella latimarginata* (FLH).

Gastropod hemocyanins are large metallo-glycoproteins of high molecular weight (approximately 8 to 13 MDa) that possess a complex quaternary structure and induce humoral and cell-mediated responses of the Th1 type in mammals, including humans. Due to this property, hemocyanins are widely used in biomedicine (17–22). In addition, different mollusk hemocyanins with immunological effects have also been characterized, such as FLH from *Fissurella latimarginata* (23–25). Studies performed in our laboratory showed that FLH has antitumor effects in murine melanoma and oral cancer models (20). In addition, FLH binds to TLR4 and induces the expression and secretion of Th1-type proinflammatory cytokines (17, 23, 25). A remarkable characteristic of hemocyanins is their carbohydrate content,

which is fundamental to their structure and immunological efficacy (23, 26). Recent work showed that the enzymatic N-deglycosylation of FLH influences its immunogenic effects on macrophages (23), leading to a decrease in its binding to C-type lectin receptors, such as mannose receptor (MR), macrophage galactose lectin receptor (MGL), DC-specific intracellular adhesion molecule (ICAM)-grabbing nonintegrin (DC-SIGN), and TLR4 (17, 23, 25).

In contrast to hemocyanins, including the FLH used in this study, the recombinant surface immunogenic protein (rSIP) from Group B *Streptococcus* is a small protein. It was previously expressed in *Escherichia coli* and *Pichia pastoris* and had a molecular weight of 53 kDa with a β-folded structure and the ability to form dimers (27). This recombinant protein was analyzed as a vaccine against GBS in a preclinical trial. It was shown that this protein has excellent immunogenic capacity, as it binds TLR4 and induces a Th1-type response against GBS (28–31). Furthermore, given that SIP less complex structure than hemocyanins, it can undergo genetic fusion with other protein antigens to ensure joint antigen–adjuvant delivery (2). Previously, immunization with rSIP without adjuvant was shown to decrease GBS vaginal colonization and induce secretion of opsonizing antibodies evaluated by *in vitro* opsonophagocytosis (OPA) assays (30). Furthermore, rSIP was found to promote humoral immunity in a murine model using ovalbumin (OVA) as an antigen (27). Thus, considering that rSIP immunogenicity studies are in the advanced preclinical stage, this protein is a new candidate vaccine adjuvant.

In this work, we focused on characterizing two TLR4 protein agonists, their associations with MyD88 and TRIF recruitment and their contributions to antigen cross-presentation to CD8+ T lymphocytes. Since rSIP and FLH differ in origin, structure, and size, we hypothesized that MyD88 and TRIF recruitment is important for generating the TLR4-dependent Th1 effects of these protein-based adjuvants (PBAs). For this purpose, we studied the role of rSIP and FLH in the recruitment of MyD88 and TRIF in antigen-presenting cells (APCs) by characterizing molecular targets involved in TLR4 activation, as well as their adjuvant effects on antigen cross-presentation in bone marrow-derived dendritic cells (BM-DCs).

2 Materials and methods

2.1 Hemocyanin, rSIP, and ovalbumin antigen

F. latimarginata hemocyanin (FLH) was provided by Biosonda SA (Santiago, Chile). This protein was isolated and purified under sterile, pyrogen-free conditions in phosphate-buffered saline ([PBS] containing sodium phosphate 0.1 M NaCl), pH 7.2 (24), and Tris buffer for FLH containing 50 mM Tris, pH 7.4, 5 mM CaCl 2.5 mM $MgCl_2$, and 0.15 mM NaCl (25). All chemicals were analytical reagent grade, and solutions were prepared with human irrigation water (Baxter Healthcare, Charlotte, NC, USA) and filtered through a 0.2 μm membrane filter (Millipore).

rSIP was obtained according to a procedure previously published by our group (27, 31). Briefly, rSIP was expressed in *E. coli* BL21 (DE3) and transformed into the plasmid pET21a::sip. Then, rSIP was expressed as a soluble protein and purified using nickel-nitrilotriacetic acid (NI-NTA) resin by low-pressure chromatography and high-performance liquid chromatography (HPLC) using a molecular exclusion column. The system consisted of a BioSep-SEC-s2000 300 x 21.2 mm Preparative Column 00H-2145-P0 (PHENOMENEX) and Smartline UV detector 2520 (Knauer, WissenschaftlicheGeräte GmbH, Germany). rSIP had a purity > 98%.

rSIP and FLH has endotoxin levels less than 0.5 EU/mL, which was determined using the ToxinSensor™ Chromogenic LAL Endotoxin Assay Kit. Additionally, protein concentrations were determined using the Pierce 660 nm Protein Assay Reagent (Thermo Scientific, Waltham, MA) according to the manufacturer's instructions with a Pierce™ Bovine Serum Albumin Standard (Thermo Scientific).

Endotoxin-free OVA protein (Invivogen, cat. vac-stova) was used as the model antigen.

The stimuli and inhibitors used in this work did not induce cell toxicity in any cells used. Viability was determined with Trypan Blue and Annexin-V/propidium iodide (data not shown). Additionally, rSIP and FLH concentrations are reported as molar concentrations due to their significant differences in size and molar mass (rSIP ≈ 53 kDa; FLH ≈ 8,000 kDa).

2.2 Experimental animals

Mice of the wild-type C57BL/6 strain, C57BL/6-Tg (TcraTcrb) 1100Mjb/J mice (OT-I), and B6.Cg-Tg(TcraTcrb)425Cbn/J (OT-II) mice were purchased from Jackson Laboratory. In addition, OT-I and OT-II were supplied by Fundación Ciencia & Vida (Chile). All mouse experiments followed international ethical standards and Chilean Animal Protection Law 20380 (2009). The Institutional Committee reviewed the experimental protocol in accordance with the Care and Use of Laboratory Animals of the Institute of Public Health of Chile, codes C110322-01 and C120421-01. The mice were housed in the Facility of the Laboratory Animal Maintenance and Experimentation Room (MEAL) of the Biotechnology Section of ISPCh. The mice were maintained following the regulations established by the Institutional Committee for the Use and Care of Animals of the laboratory.

2.3 Acquisition and culture of BM-DCs

BM-DCs were prepared using a modified procedure based on Lutz et al. (32). Briefly, bone marrow was extracted from the femurs and tibias of mice, washed with Hanks saline solution (HBSS), and cultured in BM-DC-specific medium containing Roswell Park Memorial Institute (RPMI-1640, Cytiva, cat. SH30027.02) supplemented with 10% inactivated fetal bovine serum (GIBCO, cat. 26140079), 2 mM L-glutamine, 1 mM sodium pyruvate, penicillin (50 U/ml), streptomycin (50 mg/ml), 50 mM β-mercaptoethanol, and 20 ng/ml granulocyte-macrophage colony-stimulating factor (GM-CSF; Peprotech, Cat. 315-03) to generate BM-DCs. Cells were seeded in a Petri dish at 2 x 10^6/mL and incubated at 37°C. On days 3, 6, and 8, 10 ml of BM-DC medium

was added to the cultures. On days 6 and 8, 10 ml of BM-DC medium was removed and replaced with fresh medium. On day 10, nonadherent BM-DCS were harvested. The BM-DCs were phenotypically characterized using flow cytometry (FACSVerse) after reaching > 85% expression of the phenotype CD11c+ CD11b+ MHCII+ CD86low CD80low CD4− CD8- B220− GR1−.

2.4 Cytokine secretion assay

The measurement of cytokines in the supernatant of BM-DCs cultured with FLH and rSIP was carried out according to Kolb et al. (12). BM-DCs (1 x 10^5 per well) were incubated in flat-bottom 96-well plates for 2 h at 37°C before TLR4 protein agonists (FLH 120 nM and rSIP 40 nM) or PBS (vehicle control) was added. TRIF and MyD88 inhibition experiments were performed as described by Chen et al. (33) in which 75 µM Pepinh-TRIF (Humimmu LLC), 75 µM Pepinh-Control (Negative Control, Humimmu LLC), 75 µM MyD88 peptide control (Novus Biological), and 75 µM antennapedia control peptide (Novus Biological, negative control) were added 18 h before the addition of FLH and rSIP.

For the TLR4 signaling inhibitor, TAK242 was used according to the supplier's recommendations (34). TAK 242 (10 µg/mL, Invivogen) was added 2 h before the addition of FLH and rSIP. After 18 hours of stimulation with FLH and rSIP at 37°C, the supernatants were collected. The concentrations of IL-6 and IP-10 were measured using a commercial enzyme-linked immunosorbent assay (ELISA) kit (Mouse IL-6 ELISA Kit: BMS603-2; Mouse IP-10 ELISA Kit: BMS6018, Invitrogen) according to the manufacturer's instructions and with all necessary controls. The sensitivity of both ELISA kits was 6.5 pg/mL. The assay range for the IL-6 ELISA kit was 31.3-20,000 pg/mL. The assay range for the IP-10 ELISA kit was 7.8-500 pg/mL.

The inhibitors did not induce cell toxicity in the cells used in these assays (data not shown).

2.5 Determination of gene expression by real-time reverse-transcriptase polymerase chain reaction (RT–qPCR)

The measurement of mRNA from BM-DCs was carried out according to Kolb et al. (12). BM-DCs (1x10^5 per well) were incubated in flat-bottom 96-well plates for 2 hours at 37°C before TLR4 protein agonists or PBS (vehicle control) were added. MyD88 and TRIF pathway inhibitors were used in the manner described above. Once stimulated, the cells were washed with cold HBSS. Cell lysis and total RNA isolation were performed with NucliSENS® easyMAG equipment; Biomérieux and complementary DNA (cDNA) were synthesized with SuperScript™ III Reverse Transcriptase (Invitrogen). Assays were performed using 1 µl of gDNA template and a Stratagene Mx3000P thermocycler (Agilent Technologies). Increases in mRNA abundance in treated cells relative to control cells were calculated using the 2−ΔΔCt method and normalized to β-actin mRNA. The sequences of primers and probes used for detecting mRNAs can be found in Supplemental Table 1.

2.6 THP-1 Dual cell culture

THP-1 Dual cells (InvivoGen, thpd-nfis), TRIF KO Dual Reporter THP1 Cells (InvivoGen, thpd-kotrif), THP1-Dual™ KO-MyD cells (InvivoGen, thpd-komyd), and TLR4 KO Dual Reporter THP- 1 Cells (InvivoGen thpd-koTLR4) were cultured in RPMI-1640 medium supplemented with 10% fetal bovine serum (Gibco), L-glutamine, 100 U/ml penicillin, 100 mg/ml streptomycin (Gibco), 100 mg/mL normocin (InvivoGen, ant-nr-1), 100 mg/mL zeocin (InvivoGen, ant-zn-1), and 10 mg/mL blasticidin (InvivoGen, ant-bl-1). Dual THP1 cells were incubated at 37°C and 5% CO_2.

THP1-Dual™ KO-TLR4, THP1-Dual™ KO-TRIF, and THP1-Dual™ KO-MyD88 cells were generated from THP1 Dual cells™ via knockout (KO) of TLR4, TRIF and MyD88 (InvivoGen). Previously, these cells were validated to characterize the functionality of SEAP and LUCIA expression. The activation of NF-kB and IRF was previously analyzed using a NOD1 agonist and TLR3 agonist. The NOD1 ligand generated an increase in SEAP in all cell lines. For LUCIA, all lines showed activation of IRF in the presence of the TLR3 agonist (data not shown).

2.7 SEAP and LUCIA assays in THP-1 and Hek-blue cells

THP-1 cells were seeded in 96-well plates (Corning Costar) at a density of 100,000 cells per well. First, cells were stimulated for 18 h with rSIP, FLH, LPS, a NOD 1 ligand (positive control for NF-kB; C12-iE-DAP, InvivoGen), and a PRR agonist (positive control for IRF; Poly (dA:dT)/LyoVec™, InvivoGen). The supernatant was then subjected to a colorimetric enzyme assay to measure alkaline phosphatase (AP) activity using the commercial QUANTI-Blue™ solution (InvivoGen). The supernatant was then incubated at 37°C for 3 h, and the optical density was read at 650 nm in an Epoch 2 reader (BioTek). On the other hand, luciferase activity (LUCIA) was measured using the commercial solution QUANTI-Luc ™ (InvivoGen), which has a coelenterazine substrate and stabilizing agents for the luciferase reaction. The light signal produced was then quantified using a Berthold luminometer (Model LB9515), and the signal was expressed as relative light units (RLUs).

Hek-Blue cells (hkb-mtlr4 and hkb-htlr4, InvivoGen) express SEAP under the control of promoters containing binding elements for the NF-κB transcription factor (35). Hek-Blue cells were seeded in 96-well plates (Corning Costar) at a density of 25,000 cells per well in HEK-Blue™ Detection medium (InvivoGen). Then, the cells were stimulated for 48 h with rSIP, FLH, and LPS, and SEAP was quantified using an Epoch 2 reader (Biotek).

2.8 Exogenous antigen presentation assays in an *in vitro* model

The presentation of exogenous antigens in an *in vitro* model was evaluated as described by Alloatti et al. (36). BM-DCs (1 x 10^5 per well) were incubated in flat-bottom 96-well plates for 2 h at 37°C before OVA and TLR4 agonists were added. The BM-DCs

were incubated with OVA, OVA + FLH, OVA +SIP, or PBS. After 24, 48 and 72 h, the BM-DCs were washed three times with a 0.1% (vol/vol) PBS/BSA solution and then labeled with the 25-D1.16 antibody that detects peptide–major histocompatibility class (MHC)-I (SIINFEKL: MHC-I) complexes (37).

2.9 Antigen cross-presentation assay

The antigen cross-presentation assay was adopted and modified from Alloati et al. (36). Following TLR4-induced maturation of DCs, antigen cross-presentation is first enhanced and then modulated downstream of antigen internalization and cytosolic delivery (36, 38). It was previously reported that antigen cross-presentation capacity increased in the initial hours after TLR4 activation (39). To evaluate antigen cross-presentation, BM-DCs were seeded at 40 x 10^3 cells per well in a 96-well plate and then pulsed with the antigens at different concentrations for 3 h. At the end of the pulsing period, the cells were washed three times to remove excess antigen and cocultured with 1 x 10^5 CD8 T cells purified from the spleens of OT-I mice using the MojoSort™ Mouse CD8 T-Cell Isolation Kit according to the manufacturer's protocol. CD8 T cells from OT-I mice were labeled with CellTrace Violet™ (Molecular Probes™, ThermoFisher Scientific). Three days later, the proliferation of CD8 T cells was measured using flow cytometry. As a proliferation control, OT-I lymphocytes were cocultured with BM-DCs pulsed with the ovalbumin peptide SIINFEKL (Invivogen).

For MyD88 and TRIF inhibition, the peptides were added to BM-DCs 18 hours before pulsing with antigen and adjuvant. TAK-242 (10 μg/mL, Invivogen) was added 2 hours before washing and pulsing with more antigen adjuvant. The BM-DCs were cocultured with 30x10^4 CD8 T cells, and cell proliferation was characterized as described above.

A similar approach was used to assess the effect of the proteasomal, vacuolar, and endoplasmic reticulum trafficking processes using several inhibitors: (A) Brefeldin A at 1 μM, (B) Epoxomicin at 5 nM, (C) Leupeptin at 10 μM, (D) Pepstatin A at 40 nM, (E) MG132 at 4 nM, (F) Bafilomycin at 10 nM, and (G) Simvastatin at 10 nM. All inhibitors were obtained from Enzo. The inhibitors were added to BM-DCs 1 h before washing, pulsing with antigen and adjuvant, and coculture with 30 x 10^4 CD8 T cells (OT-I) as described above.

2.10 Classical antigen presentation assay

The classical presentation of antigens was evaluated similarly to antigen cross-presentation, with minor modifications. BM-DCs were seeded at 40 x 10^3 cells per well in a 96-well plate and then pulsed with the antigens at different concentrations for 18 h. At the end of the pulsing period, the cells were washed three times to remove excess antigen and cocultured with 1 x 10^5 CD4 T cells purified from the spleens of OT-II mice using the MojoSort™ Mouse CD4 T-Cell Isolation Kit according to the manufacturer's protocol. CD4 T cells from OT-II mice were labeled with CellTrace Violet™ (Molecular Probes™, ThermoFisher Scientific). Four days later, the proliferation of CD4 T cells was measured using flow cytometry. As a proliferation control, OT-II lymphocytes were cocultured with BM-DCs pulsed with the OVA peptide ISQAVHAAHAEINEAGR (InvivoGen).

2.11 Immunoblotting

Immunoblotting was performed according to Jiménez et al., with modifications (17). BM-DCs (4×10^6) were incubated for 2 h at 37°C in polystyrene tubes and were then exposed to rSIP or FLH. The cells were lysed at the indicated time points using RIPA lysis buffer supplemented with protease inhibitor cocktail (5 mg/mL; Roche). Proteins were separated in polyacrylamide gels (SDS–PAGE, 10-15%), electrotransferred to nitrocellulose membranes, and blocked with 5% bovine serum albumin (BSA) in 0.1% (v/v) TBS-Tween 20 (TBST). Primary antibodies were dissolved in 5% BSA and incubated with the blocked membranes overnight at 4°C. After exposure of the membranes to horseradish peroxidase-conjugated anti-rabbit secondary antibodies for 1 h in BSA, bands were visualized using the Odyssey system (Li-Cor Bioscience) detection system, and band intensities were analyzed with LI-COR Image Studio Software. SuperSignal West Femto Maximum Sensitivity Substrate (Thermo Fisher) and EZ-ECL (Biological Industries) were the chemiluminescent substrates used for developing the Western blots. The antibodies used for immunoblotting were a recombinant anti-IRF3 antibody [EPR2418Y] (ab68481), goat anti-rabbit IgG H&L (HRP) (Abcam, ab205718) and phospho-IRF-3 (Ser396) (4D4G) rabbit mAb #4947 (Cell Signaling Technology, Danvers, USA).

2.12 Statistical analysis and determination of log (EC_{50}) values

The log half maximal effective concentration (EC_{50}) values for each agonist-induced response were calculated according to Ehlert et al. (40), generating nonlinear fits for four parameters defined as the baseline response line (Bottom), maximum response (Top), slope of the curve (HillSlope), and concentration of protein agonist that elicited a median response between the baseline and upper response (EC_{50}). Dose–response data were analyzed using GraphPad Prism software with the following equation: Y=Bottom + (X^Hillslope)*(Top-Bottom)/(X^HillSlope + EC50^HillSlope).

Differences between log (EC_{50}) values were analyzed using GraphPad Prism 9 software by applying a two-tailed t test (for comparisons between two sets of proteins). In addition, the statistical significance of differences in inhibition by TAK-242, MyD88, and TRIF inhibitors was evaluated with the one-tailed Mann–Whitney U test or the Kruskal–Wallis test followed by *post hoc* tests for multiple comparisons.

10.3389/fimmu.2023.1186188

3 Results

3.1 FLH and rSIP differ in their potency as TLR4 agonists

Understanding the modulation of TLR4 agonist-mediated signaling is pivotal for deciphering the immune mechanisms to develop vaccine adjuvants (41). To gain insight into the mechanism underlying TLR4 protein agonists, we performed dose–response analysis of HEK-blue reporter cells expressing either mouse TLR4 (mTLR4) or human TLR4 (hTLR4). Analyses were performed to evaluate LPS stimulation, including a full TLR4 agonist and PBS as a negative control. In the raw data, rSIP induced stronger stimulation than FLH according to the dose–response curve (Figure 1A). rSIP and FLH were partial agonists of mTLR4 and reached a maximum activation value of approximately 80% compared to the full agonist LPS (4), as shown in Figure 1B. Therefore, the half-maximal effective concentrations of the two proteins were compared to determine whether a difference in the immunological potency of these partial agonists could be observed. The findings showed that rSIP had a lower EC_{50} than FLH and a mean log (EC_{50}) value of –1, compared to FLH, which had a mean log (EC_{50}) value of 0.1; therefore, these agonists stimulated mTLR4 at lower protein concentrations (Figure 1C).

Different mTLR4 agonists used to characterize preclinical animal models differ significantly from those observed in human cell systems (42). In this context, the agonist effects of rSIP and FLH on Hek-Blue hTLR4 cells was characterized (Supplemental Figure 1A). Compared to LPS (100% activation), rSIP was found to be a partial agonist of hTLR4, while FLH acted as a lower efficacy agonist of hTLR4. This conclusion was confirmed by the observation that LPS activated HEK-Blue-TLR4 cells at 18 h post-stimulation, while rSIP and FLH stimulated the cells at 48 h post-stimulation, with rSIP reaching approximately 90% activation and FLH reaching 25% activation. The above result suggests that TLR4-PBAs are ligands of hTLR4, with partial and weak partial agonist effects for rSIP and FLH, respectively. Although a difference in the potency of hTLR4 activation was found, these agonists presented similar log(EC_{50}) values (Supplemental Figure 1B). HEK-Blue Null1-v cells, which do not express mTLR4 or hTLR4, were used as a negative control for the cell line and were not activated in the presence of FLH or rSIP (data not shown). Notably, both rSIP and FLH were able to activate NF-κB in the HEK-Blue-TLR4 cell line. Furthermore, both proteins were able to activate IRF3 in BM-DCs (Supplemental Figure 1C). Our results show the specificity of mTLR4 for the protein ligands described above, and we propose that the structural differences between these ligands, i.e., FLH is very large and glycosylated, and rSIP is small and not glycosylated, will allow them to serve as new models to study their contributions to MyD88 and TRIF signaling.

3.2 TLR4-PBAs show no bias toward MyD88 or TRIF signaling at minimal activation doses

To determine whether rSIP and FLH are agonists biased toward the TRIF or MyD88 pathway, BM-DCs were activated with an extensive dilution series of rSIP and FLH. The potencies of these agonists in activating a panel of TRIF-dependent and MyD88- and TRIF-codependent proteins were measured by calculating the log (EC_{50}) values. We evaluated IP-10 as a representative TRIF-dependent protein and IL-6 as a MyD88- and TRIF-codependent cytokine (12). As expected, based on the log (EC_{50}) value, rSIP was more active than FLH in inducing the expression of TRIF-dependent and TRIF-codependent proteins (MyD88 and TRIF) (Figures 2A–C). rSIP produced mean log (EC_{50}) values of 1.8 and 2.1 for IL-6 and IP-10, respectively, while FLH produced mean log (EC_{50}) values of 6.8 and 7.1 for IL-6 and IP-10, respectively. However, no log (EC_{50}) difference for the comparison of IL-6 and IP-10 expression between rSIP and FLH was found (Figure 2C). Taken together, these results suggest that at minimal activation doses, rSIP has a lower Log (EC_{50}) than FLH, which is consistent with the results presented in Figure 1. No differences in the preferential activation pathway were found when the log (EC_{50}) values for IL-6 and IP-10 were compared. FLH generated the same effect. Therefore, FLH and rSIP induce MyD88- and TRIF-codependent pathways, as reflected by IL-6 and IP-10 expression.

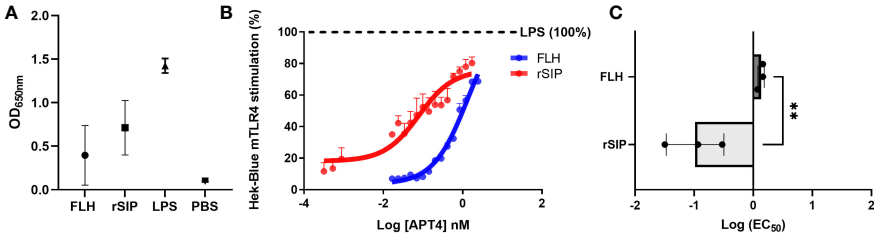


FIGURE 1
Partial agonism of mTLR4 by protein-based adjuvants. **(A, B)** Raw data and determination of the concentrations of protein agonists needed to activate mTLR4. The HEK-Blue-mTLR4 reporter cell line was exposed to different concentrations of rSIP and FLH. Dose–response curves were generated for cells exposed to a maximum concentration of 2540 nM rSIP or FLH for 48 h. Data show normalized HEK-Blue mTLR4 cell responses considering treatment with lipopolysaccharide (LPS) as 100% stimulation; 100% = maximum dose plateau of the LPS agonist. **(C)** Comparison of the log EC_{50} values of protein agonists in the activation of mTLR4. Log (EC_{50}) values for rSIP and FLH were determined according to the relative abundance of soluble alkaline phosphatase (AP) secreted by Hek-Blue-mTLR4 cells. Individual log (EC_{50}) values and mean values from three independent experiments are shown. The statistical significance of differences was analyzed using an unpaired t test (**p < 0.01).

10.3389/fimmu.2023.1186188

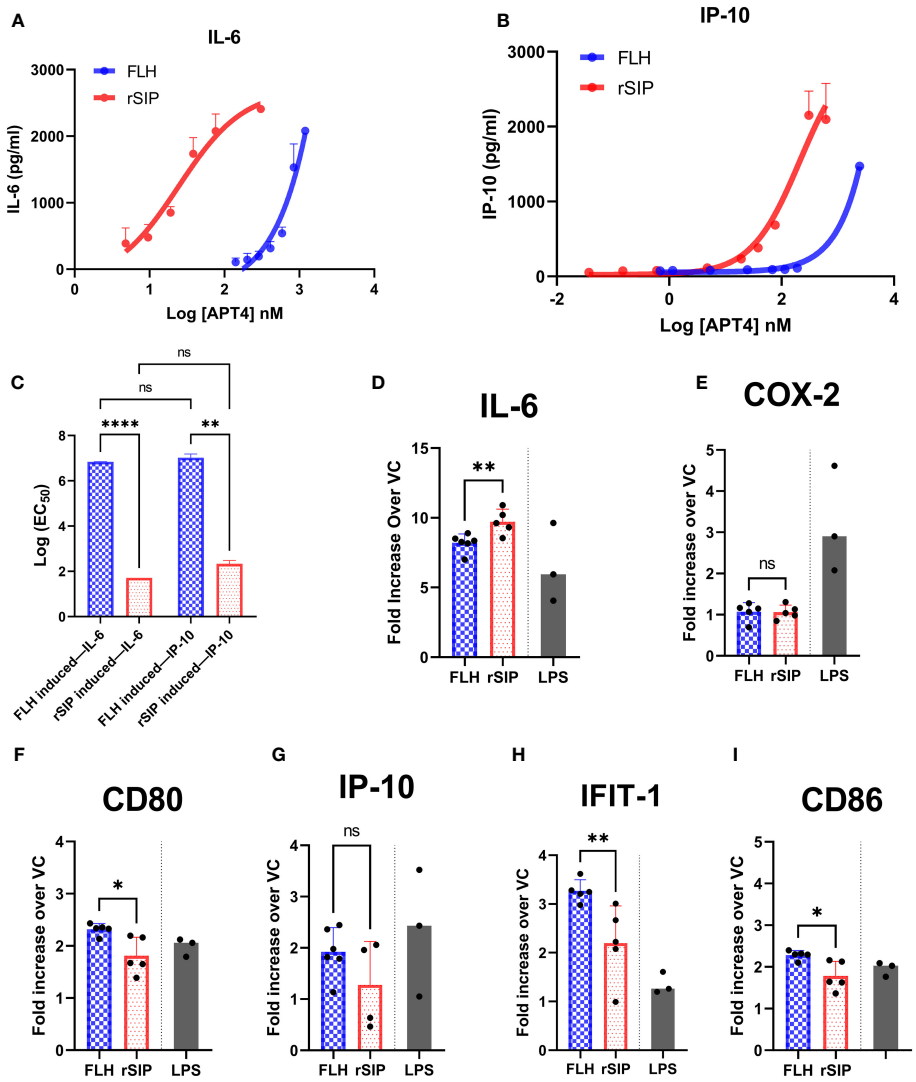


FIGURE 2
FLH and rSIP induce differential expression patterns for molecules associated with the MyD88 and TRIF pathways. Bone marrow dendritic cells (BM-DCs) from C57BL/6 mice were treated with the indicated concentrations of rSIP and FLH. **(A, B)** Analysis of IL-6 and IP-10. After 18 h, the concentrations of IL-6 and IP-10 were determined using enzyme-linked immunosorbent assay (ELISA). Data are expressed as the means ± standard deviations (SD) of three independent experiments. **(C)** Log (EC_{50}) comparison. The log (EC_{50}) values of the indicated rSIP- and FLH-stimulated changes in interleukin 6/(IL-6/IP-10) expression from experiments A and B were compared. Individual log (EC_{50}) values are shown. Values are the means of three independent experiments. Statistical differences were analyzed for the data in **(C)** using the Mann–Whitney U test (**p<0.01; ****p<0.0001; ns, statistically not significant). **(D–I)** Analysis of MyD88– and TRIF–codependent expression of target genes. Wild-type BM-DCs were stimulated with rSIP (40 nM) and FLH (120 nM) for 4 h. The mRNA abundances of **(D)** IL-6, **(E)** cyclooxygenase 2 (COX-2), **(F)** cluster of differentiation 80 (CD80), **(G)** IP-10, **(H)** IFIT-1, and **(I)** CD86 were analyzed by RT–qPCR. LPS-stimulated BM-DCs were used as a positive control for each of the mRNAs analyzed. Data are expressed as the mean fold increase in mRNA abundance in cells stimulated with protein adjuvants compared to cells treated with PBS. Each dot represents an independent experiment. Data are the means ± SDs of 4 or 5 independent experiments. Statistical significance was determined using the Mann–Whitney U test (*p<0.05; **p<0.01; ns, statistically not significant).

Furthermore, we aimed to characterize other molecules associated with the recruitment of MyD88 and TRIF. The MyD88- and TRIF-codependent pathway is characterized by the gene expression of IL-6, Cox-2, and cluster of differentiation 80 (CD80). Expression of the IP-10, IFIT1, and CD86 mRNAs is associated with the TRIF-dependent pathway. BM-DCs were stimulated with EC_{50} doses of rSIP and FLH, and the mRNAs of IL-6, cyclooxygenase (COX-2), CD80, IP-10, interferon-induce protein with tetratricopeptide repeats (IFIT-1), and CD86 were evaluated (Figures 2D–I). RT–qPCR revealed that FLH induced higher expression of CD80, IFIT-1, and CD86 than rSIP. On the other hand, rSIP induced higher expression of IL-6 than did FLH. In contrast, COX-2 and IP-10 expression showed no variations associated with either rSIP or FLH. This result suggests that different profiles associated with the recruitment of MyD88 and TRIF could occur for rSIP and FLH at their respective EC_{50} values. This fine signaling regulation mediated by TLR4 implies some cross-regulation of these pathways.

10.3389/fimmu.2023.1186188

3.3 The regulation of FLH– and rSIP–TLR4 activation is associated with MyD88- and TRIF-dependent genes

Since a preference for the MyD88 or TRIF pathway was not observed, we decided to further examine the effects of these pathways on the immune response induced by both adjuvant proteins. To characterize the contribution of MyD88 to TLR4 activation by rSIP and FLH, BM-DCs were pretreated for 18 h with the MyD88 inhibitor, an inhibitory peptide that blocks MyD88 signaling, or a control peptide prior to stimulation of BM-DCs for 4 h, and RT–qPCR was performed. Inhibition of MyD88 decreased IL-6, CD80, CD86, and IFIT1 transcript levels but not COX-2 transcript levels when BM-DCs were stimulated with FLH (Figure 3). For rSIP, inhibition of MyD88 induced a decrease in IL-6, CD80, CD86, IFIT1, and COX-2 transcript levels (Figure 3). On the other hand, no variation in the number of IP-10 transcript levels upon stimulation with rSIP and FLH was found, suggesting that the TLR4 agonistic effects of rSIP and FLH are MyD88 dependent.

Similar to the approach used for MyD88 inhibition, we decided to further examine the effects of TRIF on the immune response induced by the rSIP and FLH proteins. The BM-DCs were pretreated for 18 h with Pepinh-TRIF, an inhibitory peptide that blocks TRIF signaling, or Pepinh-Control, a control peptide, and then treated with rSIP and FLH. After the cells were stimulated for 4 h, RT–qPCR was performed. TRIF inhibition decreased the transcription of IL-6, COX-2, and IP-10 after stimulation with rSIP (Figure 4). Regarding FLH, TRIF inhibition decreased IL-6, CD80, IFIT1, and IP-10 expression; however, no effects on CD86 transcript levels in response rSIP and FLH were found. Therefore, TRIF is required for the activation of IL-6 and IP-10 expression by FLH and rSIP (Figures 4A, D). These results indicate that TLR4, MyD88, and TRIF are important for the signaling patterns associated with TLR4-PBAs in BM-DCs.

3.4 rSIP activates MyD88– and TRIF–dependent proteins, while FLH activates TRIF– dependent proteins

After establishing an association between TLR4-PBAs and genes with MyD88- and TRIF-dependent expression at minimal activation concentrations, no preference for the TRIF and MyD88 pathways was observed. Therefore, we next sought to address whether inhibition of MyD88 and TRIF influences IL-6 and IP-10 secretion by BM-DCs. First, we characterized the effects of TLR4 on cytokine secretion. For this purpose, the cells were treated for 2 h with TAK-242 or dimethyl sulfoxide (DMSO) as the negative control. TAK-242 induced complete and partial inhibition of the IL-6 and IP-10 secretion induced by both rSIP and FLH, respectively (Figures 5A, B). These results suggest that the effect of rSIP and FLH on cytokine secretion by BM-DCS is dependent on TLR4 in BM-DCs. This effect was also reflected by the expression of CD86 by flow cytometry, which was dependent on TLR4 (data not shown).

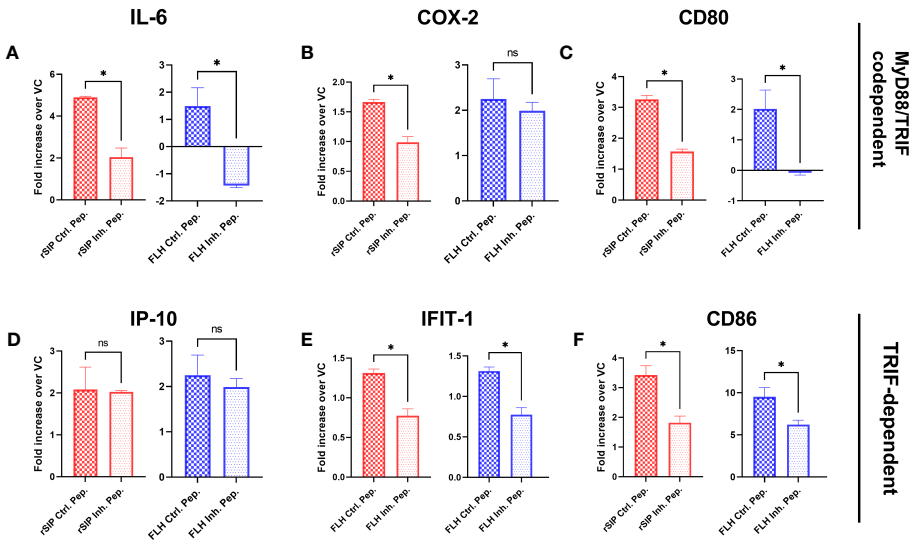


FIGURE 3
MyD88 is required for the activation of signaling pathways by protein-based adjuvants. Wild-type BM-DCs were pretreated with peptide inhibitors of MyD88 or a peptide control and stimulated with rSIP (40 nM) and FLH (120 nM) for 4 h. **(A–F)** Quantification of RNA. The mRNA abundances of **(A)** IL-6, **(B)** COX-2, **(C)** CD80, **(D)** IP-10, **(E)** IFIT-1, and **(F)** CD86 were analyzed by reverse-transcriptase polymerase chain reaction (RT–qPCR). Data are the mean fold increase in mRNA abundance in cells stimulated with protein adjuvants compared to cells treated with PBS (vehicle control, VC) and averaged from three independent experiments. The activation of TLR4 by rSIP is dependent on MyD88 recruitment for the activation of **(A)** IL-6, **(B)** COX-2, **(C)** CD80, **(E)** IFIT-1, and **(F)** CD86 expression. On the other hand, the activation of TLR4 by FLH is dependent on the recruitment of MyD88 for the activation of **(A)** IL-6, **(C)** CD80, **(E)** IFIT-1, and **(F)** CD86 expression, but the activation of **(D)** IP-10 expression was not dependent on MyD88 after activation by FLH and rSIP. Data are represented as the means ± SDs of three independent experiments. Statistical significance was determined using the Mann–Whitney U test (*p<0.05; ns: not statistically significant).

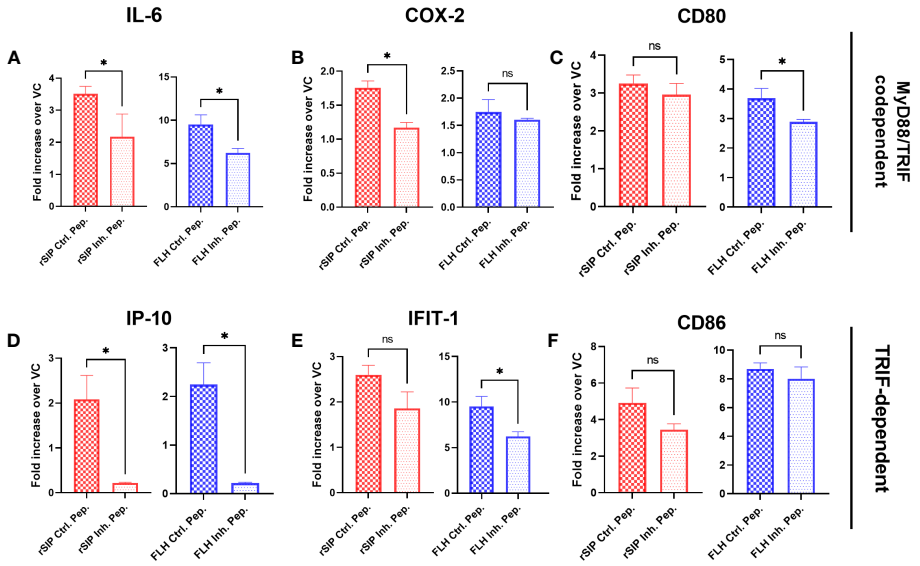


FIGURE 4
Activation of IP-10 and IL-6 by TLR4-PBAs is dependent on TRIF. Wild-type BM-DCs were pretreated with the TRIF peptide inhibitor or a peptide control and stimulated with rSIP (40 nM) and FLH (120 nM) for 4 h. **(A–F)** Quantification of RNA. The mRNA abundances of **(A)** IL-6, **(B)** cyclooxygenase 2 (COX-2), **(C)** cluster of differentiation 80 (CD80), **(D)** IP-10, **(E)** interferon-induced protein with tetratricopeptide repeat 1 (IFIT-1), and **(F)** CD86 were analyzed by RT–qPCR. Data are expressed as the mean fold increase in mRNA abundance in cells stimulated with the protein adjuvants compared to cells treated with PBS (vehicle control, VC) and averaged from three independent experiments. The activation of TLR4 by rSIP is dependent on the recruitment of TRIF for the activation of **(A)** IL-6, **(B)** COX-2 and **(D)** IP-10 expression. In contrast, the activation of FLH depends on the recruitment of TRIF to activate **(A)** IL-6, **(C)** CD80, **(D)** IP-10, and **(E)** IFIT-1 expression. Data are the means ± SDs of three independent experiments. Statistical significance was determined using the Mann–Whitney U test (*p<0.05; ns, not statistically significant).

To characterize the effect of MyD88, we used the methods described above, and we treated BM-DCs for 18 h with a MyD88 inhibitor peptide and then pulsed them with rSIP and FLH. Inhibition of MyD88 caused dramatic inhibition of IL-6 and IP-10 expression in rSIP-stimulated BM-DCs (Figures 5C, D). Conversely, IL-6 and IP-10 production induced by FLH was not affected by inhibition of MyD88 recruitment. Next, to characterize the effect of TRIF, we used peptides that inhibit the recruitment of TRIF (Pepinh-TRIF) in BM-DCs. As with the MyD88 inhibition methods, we pretreated cells with the TRIF-inhibiting peptide for 18 h and pulsed them with rSIP and FLH. TRIF inhibition, how, had a less dramatic but significant effect on the expression of IL-6 and IP10 in BM-DCs stimulated with rSIP and FLH (Figures 5E, F). These data suggest differential regulation of TLR4 signaling in terms of the recruitment MyD88 and TRIF by rSIP and FLH. They also highlight a slight difference in that MyD88 is important for rSIP-induced IL-6 and IP-10 secretion, whereas TRIF is essential for FLH-induced IL-6 and IP-10 secretion.

3.5 MyD88 and TRIF are required for NF-κB- and IRF-associated signaling during TLR4 activation by FLH and rSIP

To obtain more information in a human model, we used THP1-Dual™ cells derived from the human THP-1 monocyte cell line to characterize the NF-kB and IRF pathways. These cells show stable integration of two inducible reporter constructs that allow the concurrent study of the NF-κB pathway by monitoring the activity of SEAP and the IRF pathway by assessing the activity of a secreted luciferase (LUCIA). Consistent with our previous data, the results showed that the two model proteins significantly induced the secretion of SEAP and LUCIA in the PBS control group (Figures 6A, B). Additionally, LPS was analyzed in the assay and stimulated both NF-kB and IRF at levels similar to those seen for our proteins. In addition, nucleotide-binding oligomerization domain-containing protein 1 (NOD1/C12-iE-DAP) and TLR3 [poly (I:C)] agonists were used as positive controls for the secretion of SEAP and LUCIA, respectively. The data suggest that protein agonists activate the NF-kB and IRF pathways.

To characterize the effects of TLR4 protein agonists on TRIF and MyD88 recruitment, THP1-Dual™(WT), KO-TRIF, KO-MyD88, and KO-TLR4 cells were pulsed for 18 h with rSIP and FLH at 0.02 μM and 2.45 μM, respectively. These concentrations are the minimum concentrations needed to achieve stimulation of the THP1 line via the NF-kB and IRF pathways. Then, the supernatants were used to evaluate the levels of SEAP associated with activation of the NF-κB pathway (Figures 6C, D) and the levels of LUCIA (Figures 6E, F) associated with activation of the IRF pathway. SEAP induction in response to rSIP and FLH were abolished in the THP1 Dual KO-TRIF, KO-MyD88, and KO-TLR4 cells, with levels approximately 9- and 3-fold lower than those in the WT control, respectively. Moreover, LUCIA signals in response to rSIP and FLH were abolished in the THP1 Dual KO-TRIF, KO-MyD88, and KO-TLR4 cells, with levels approximately 40-fold and 10-fold lower than those of the WT control, respectively.

Notably, LPS can activate the TRIF-independent NF-κB pathway and MyD88-independent IRF pathway (38, 43, 44). In

10.3389/fimmu.2023.1186188

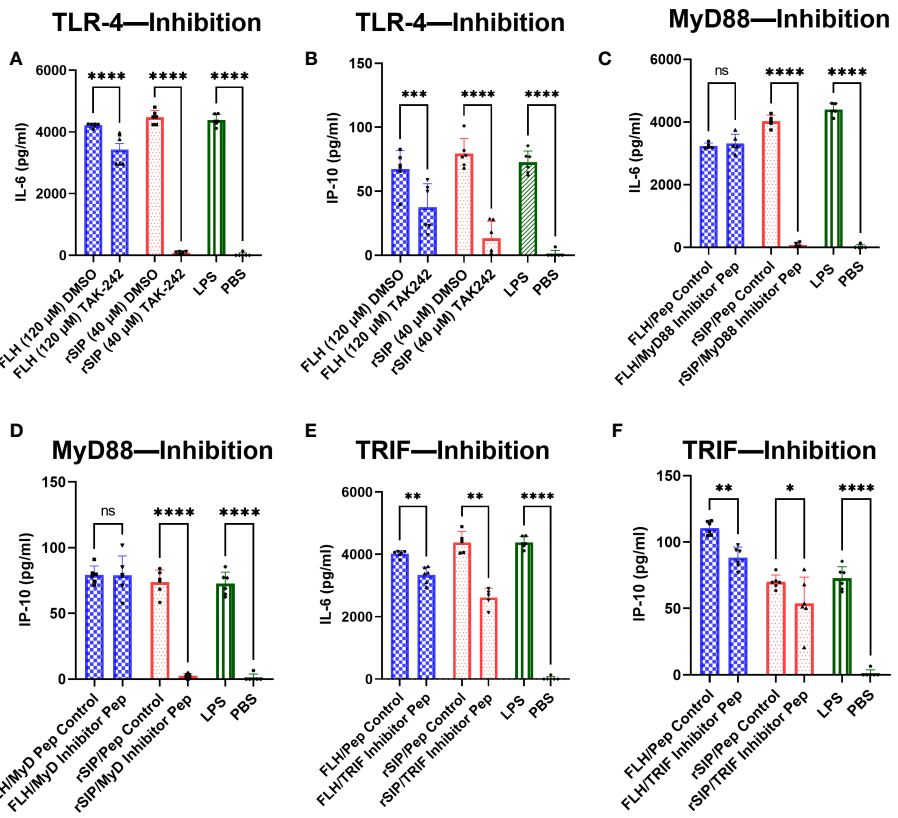


FIGURE 5
rSIP and FLH differ in signaling due to the recruitment of MyD88 and TRIF. Wild-type BM-DCs were pretreated with **(A, B)** TLR4 inhibitors and peptide inhibitors of **(E, F)** TRIF and **(C, D)** MyD88 and stimulated with rSIP (40 nM) and FLH (120 nM) for 18 h. Next, IL-6 and IP-10 were analyzed using ELISA. Assays were validated using LPS as a positive control and PBS as a negative control. Each dot represents an independent experiment. Data are the means ± SDs of six independent experiments. Statistical significance was determined using repeated measures one-way analysis of variance (ANOVA) and the *post hoc* Sidak test (*p<0.05; **p<0.01; ***p<0.001; ****p<0.0001; ns, not statistically significant).

this context, the activation of NF-κB and IRF was compared in the THP1-Dual KO-TRIF and THP1-Dual KO-MyD88 cell lines, and it was observed that the rSIP and FLH agonist proteins activate the NF-κB and IRF pathways (Figures 6G, H). LPS led to a 2.5-fold increase in SEAP levels compared to the control in the TRIF-KO THP1 cell line, while rSIP and FLH induced only a 1-fold increase compared to the control. The similar result obtained for IRF indicated that LPS activated the MyD88-KO THP1 cell line, with a 1.9-fold increase compared to the control, while FLH and rSIP caused only 1- and 0.8-fold activation, respectively, compared to the control. These results suggest that TLR4-PBAs are equally affected by the MyD88 and TRIF pathways and that NF-κB and IRF are essential for rSIP and FLH signaling.

3.6 rSIP and FLH promote antigen cross-presentation by recruiting MyD88- and TRIF-dependent proteins

After establishing a pattern associated with the recruitment of MyD88 and TRIF by rSIP and FLH for signaling, we decided to characterize the adjuvant effects of these proteins on antigen cross-presentation. Following TLR4-induced maturation of DCs, antigen cross-presentation is first enhanced and then modulated downstream of antigen internalization and cytosolic delivery (36). We wanted to investigate whether these two TLR4 ligands exerted an adjuvant effect on antigen cross-presentation; to this end, we pulsed BM-DCs for 3 h with OVA, OVA + LPS, OVA + FLH, and OVA + rSIP formulations. The BM-DCs were washed with PBS and cocultured for three days with CellTrace Violet (CTV)-labeled naïve CD8+ T cells (OT-I). Dye dilution in proliferative cells was used to characterize the activation of naïve CD8 T lymphocytes based on flow cytometry. FLH and rSIP promoted CD8+ T-cell proliferation compared to control OVA (Figure 7A). However, FLH generated an effect at 1 and 0.5 mg/mL, whereas rSIP only did so at 1 mg/mL. Additionally, enhancement of the antigen-specific response induced by both PBAs was revealed by the 25D1.16 mAb antibody that recognizes MHC-I loaded OVA peptide (H-2Kb-SIINFEKL), and at 72 h post-stimulation with rSIP and FLH, there was promoted an increase in the population of CD11c+ 25D1.16+ cells (data not shown). Furthermore, rSIP and FLH induced classical MHC-II presentation to CD4+ T cells from OT-II mice. Their effects were similar to those observed for antigen cross-presentation, with FLH and rSIP enhancing T-cell activation compared to that observed with OVA alone (Supplemental Figure 2). Antigen cross-presentation is relevant because it confirms that the protein

10.3389/fimmu.2023.1186188

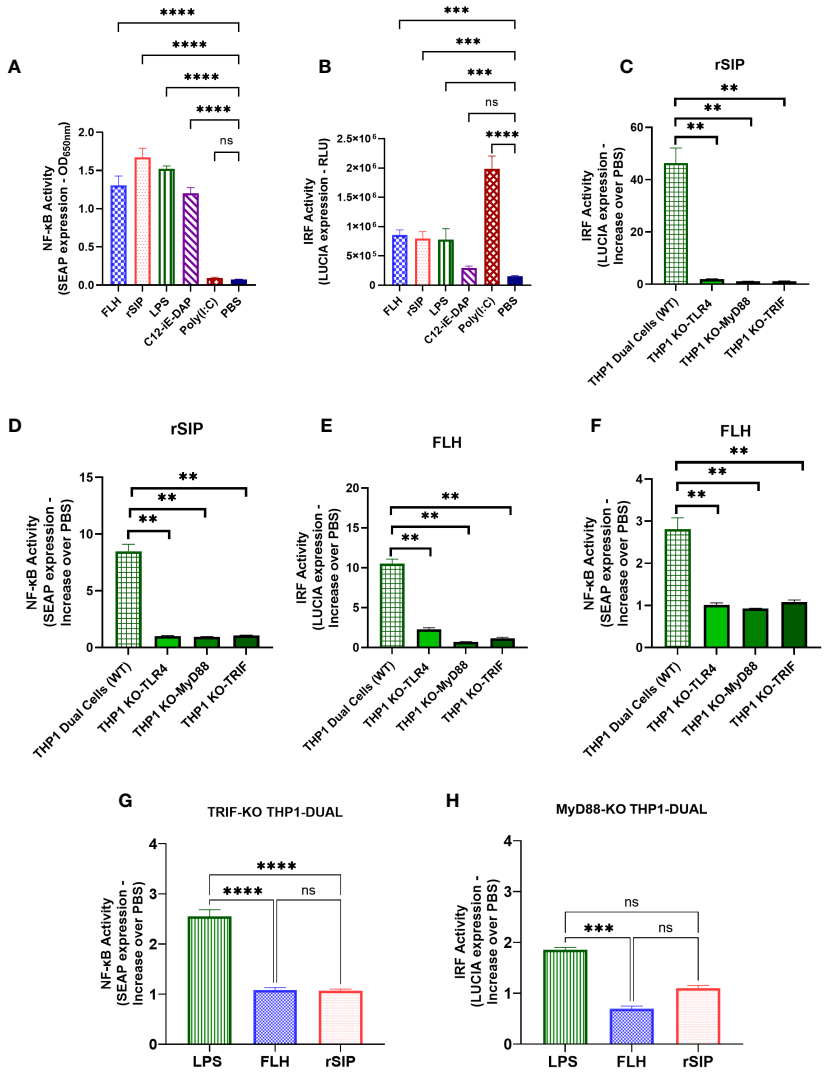


FIGURE 6
MyD88 and TRIF are required for NF-κB- and IRF-associated signaling after TLR4-PBA activation. THP1 Dual (wild-type) cell lines were treated with rSIP (0.02 μM) and FLH (2.45 μM) for 18 h. **(A, B).** The activity of **(A)** secreted alkaline phosphatase (SEAP) and **(B)** luciferase (LUCIA) was characterized after stimulation with rSIP, FLH, LPS (10 μg/ml), C12-iE-DAP (1 μg/ml; a nucleotide-binding oligomerization domain-containing protein 1 [NOD1] ligand), and poly(I:C) (1 μg/ml; a TLR3 ligand). The data are the average of the OD at 600 nm, and the relative light units (RLUs) are the average of three independent experiments. Statistical significance was determined by one-way analysis of variance (ANOVA) compared to the PBS control (***p<0.001; ****p<0.0001; ns, statistically not significant). **(C–G).** The THP1 Dual (wild-type), MyD88-KO, TLR4-KO, and TRIF-KO cell lines were treated with rSIP and FLH for 18 h. SEAP activity was characterized after stimulation with **(C)** rSIP and **(E)** FLH in the THP1 Dual (wild-type), MyD88-KO, TLR4-KO, and TRIF-KO cell lines. Data are the average increase in SEAP induction compared to the negative control (PBS) and are averaged from three independent experiments. LUCIA activity in response to **(D)** rSIP and **(F)** FLH was then characterized in THP1 Dual (wild-type), MyD88-KO, TLR4-KO, and TRIF-KO cells. Data are the average increase in LUCIA induction compared to the negative control (PBS) and are averages of three independent experiments. Statistical significance was determined using one-way ANOVA and the Mann–Whitney U test (**p < 0.01). **(G, H).** The THP1-Dual TRIF-KO cell line and the THP1-Dual MyD-KO cell line were pulsed with rSIP (0.02 μM), FLH (2.45 μM), LPS (10 μg/ml), and PBS for 18 h. Then, SEAP activity and IRF activity were characterized. Data are expressed as the average of the increase compared to the negative control (PBS) for **(G)** SEAP induction in the TRIF-KO THP1 cell line and **(H)** LUCIA luciferase induction in the MyD88-KO THP1 cell line. The results are from three independent experiments. Statistical significance was determined by one-way analysis of variance (ANOVA) with Tukey's multiple comparisons tests (***p<0.001; ****p<0.0001; ns, statistically not significant).

ligands had characteristics that were different from each other, which could be associated with their molecular structures and their affinities for TLR4.

To determine whether the adjuvant effect of rSIP and FLH was due to TLR4, we pretreated BM-DCs with DMSO or TAK242 to inhibit TLR4 signaling. The results showed that the activation of CD8 T-lymphocytes was inhibited, from approximately 80% to 20%, for both proteins (Figures 7B, G). With the recruitment inhibition approach for MyD88 and TRIF, both rSIP and FLH decreased CD8 T lymphocyte proliferation (Figures 7C, D, G). Remarkably, FLH stimulated approximately 80% activity in the control, and when MyD88 was inhibited, proliferation fell to approximately 40%. A similar effect was observed with rSIP, with proliferation decreasing from approximately 80% to 45% after inhibition of MyD88 recruitment (Figures 7C, G). Similarly, FLH stimulated approximately 40% activity after inhibition of TRIF recruitment, while the control proliferation was up to

10.3389/fimmu.2023.1186188

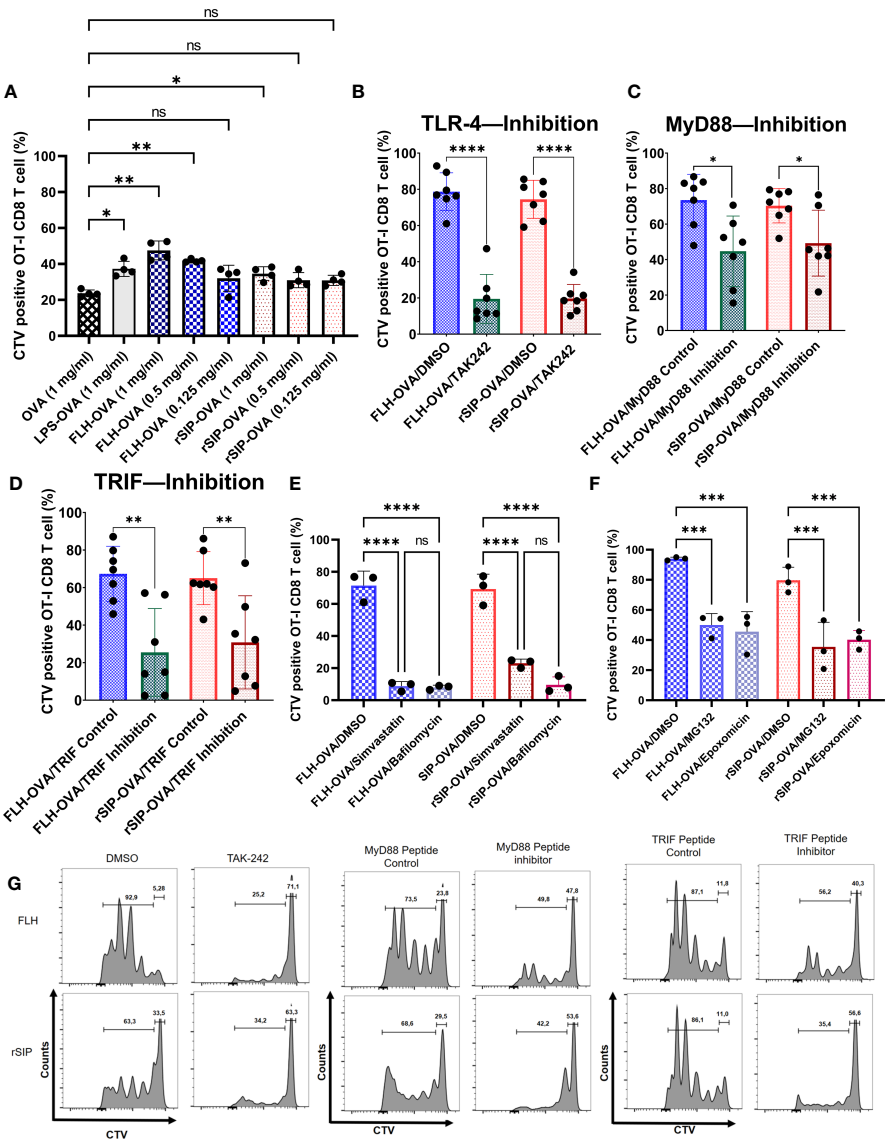


FIGURE 7
rSIP and FLH induce antigen cross-presentation dependent on MyD88 and TRIF recruitment. **(A)** Proliferation induced by OVA. BM-DCs were stimulated for 3 h with rSIP, FLH, and coadministered increasing concentrations of OVA (1 mg/mL, 0.5 mg/mL, and 0.125 mg/mL). Naïve OT-I CD8 T-cell (1 x 10^5 cells) proliferation was measured via CellTrace Violet staining after three days of coculture with treated BM-DCs. Data are the means ± SDs of four independent experiments. Statistical significance was determined using repeated measures one-way analysis of variance (ANOVA) and the *post hoc* Sidak test (*p<0.05; **p<0.01; ns, not statistically significant). **(B, E, F)** Effect of pharmacological inhibitors on FLH and rSIP processing. BM-DCs were pretreated for 2 h with TAK-242 (10 µg/mL) or for 1 h with bafilomycin (10 nM), simvastatin (10 nM), epoxomicin (5 nM), MG132 (4 nM) or DMSO and stimulated for three hours with rSIP + OVA (1 mg/mL) and FLH + OVA (1 mg/mL). Naïve OT-I CD8+ T-cell (2.5 x 10^5 cells) proliferation was measured via CellTrace Violet (CTV) staining after three days of coculture with treated BM-DCs. For bafilomycin (10 nM), simvastatin (10 nM), epoxomicin (5 nM), and MG132 (4 nM), the data are the means ± SDs of three independent experiments. For TAK-242, the data are the means ± SDs of seven independent experiments. Statistical significance was determined using **(B)** the Mann–Whitney U test and **(E, F)** repeated measures one-way analysis of variance (ANOVA) and the *post hoc* Sidak test (****p<0.0001; ns, not statistically significant). **(C, D)** Dendritic cells were pretreated for 18 h with Pepinh-TRIF or Pepinh-MyD and stimulated for three days with rSIP + OVA (1 mg/mL) and FLH + OVA (1 mg/mL). Naïve OT-I CD8+ T-cell proliferation (2.5 x 10^5 cells) was measured via CTV staining after three days of coculture with treated BM-DCs. Data are the means ± SDs of seven independent experiments. Statistical significance was determined using the Mann–Whitney U test (*p<0.05; **p<0.01). **(G)** Flow cytometry analyses show representative CTV dilution profiles for the experiments shown in figures **(B–D)**. Representative histograms of CTV dilution in gated CD8+ OT-I cells represent the inhibition of TAK-242, MyD88 and TRIF. ***p<0.001.

approximately 80%. An effect similar to that seen for rSIP was observed, with FLH stimulating up to approximately 80% of control cells, and inhibition of TRIF recruitment inducing proliferation in 45% of cells (Figures 7D, G).

Since the adjuvant effect of TLR4 on antigen cross-presentation depends on vacuolar processes (38), we decided to use two vacuolar inhibitors, simvastatin and bafilomycin. Pretreatment with these two inhibitors decreased CD8 T lymphocyte proliferation in response to FLH and rSIP (Figures 7E, G). In the case of FLH, the DMSO control stimulated proliferation in approximately 80% of CD8 T lymphocytes, while pretreatment with simvastatin and bafilomycin generated values of 10%. In the case of rSIP, DMSO

induced proliferation in approximately 80% of cells, while pretreatment with simvastatin and bafilomycin generated values of approximately 16% and 10%, respectively.

Since TRIF is involved in the proteasome pathway associated with antigen cross-presentation, we decided to characterize the effects of proteasomal inhibitors (43). As described for bafilomycin and simvastatin, we pretreated cells with epoximicin (a proteasome inhibitor) and MG132 (a proteasome inhibitor) for one hour and pulsed them with rSIP plus OVA and FLH plus OVA. We characterized antigen cross-presentation through CD8 T lymphocyte proliferation. The proteasomal inhibitor influenced antigen cross-presentation stimulated by rSIP and FLH (Figure 7F). In the case of FLH, the DMSO control stimulated proliferation in approximately 95% of CD8 T lymphocytes, while pretreatment with epoxomicin and MG132 generated values of approximately 45%. In the case of rSIP, DMSO induced proliferation in ~80% of cells, while pretreatment with epoxomicin and MG132 generated values of approximately 38% and 40%, respectively.

Additionally, given the relevance of vacuolar inhibitors in reducing crossover, we decided to characterize the influence of lysosomal proteases and intermediates on endoplasmic reticulum (ER) to Golgi vesicular transport (44). In this context, similar to Bafilomycin and Simvastatin, we pretreated cells with Brefeldin A (an ER-Golgi traffic inhibitor), Leupeptin (a Cathepsin B inhibitor), and Pepstatin A (a Cathepsin D and E inhibitor) for one hour and pulsed them with rSIP plus OVA and FLH plus OVA and characterized antigen cross-presentation through CD8 T lymphocyte proliferation. Cathepsin D and E inhibitors affected SIP- and FLH-stimulated antigen cross-presentation (Supplemental Figure 3). In the case of FLH, the DMSO control stimulated proliferation in approximately 90% of CD8 T-lymphocytes, while pretreatment with pepstatin A generated a value of approximately 50%. In the case of rSIP, DMSO induced proliferation in approximately 80% of cells, while pretreatment with Pepstatin generated values of approximately 38%. Conversely, in the case of Cathepsin D, Leupeptin was only significant inhibitor of FLH and generated OT-I lymphocyte proliferation values of 40%. In the case of inhibition of traffic from the endoplasmic reticulum (ER) to the Golgi, Brefeldin A generated CD8 T-lymphocyte proliferation values of 40% and 50% after stimulation with rSIP and FLH, respectively. Together, these results suggest that rSIP and FLH generate an adjuvant effect on antigen cross-presentation and depend on MyD88 and TRIF recruitment. Moreover, vacuolar and cytosolic pathways are essential for these effects on antigen cross-presentation.

4 Discussion

Few adjuvants currently used in licensed vaccines are known to elicit potent cytotoxic T-lymphocyte (CTL) responses. Thus, the development of new vaccine adjuvants is considered one of the slowest processes in the history of medicine (1). Nevertheless, the results of several studies are consistent with the idea that modulation of the TLR4 signaling pathway using Lipid A or monophosphoryl lipid A (MPL) can be used to dissociate beneficial immune responses from harmful LPS side effects, which are attributed to the stronger activation of NF-κB than MPL in APCs (45–48), leaving a gap in our understanding of how downstream signaling is affected by different protein agonists of this receptor. Therefore, elucidating the contributions of TLR4 agonist protein adjuvants that modulate proinflammatory activity and immunomodulation will help researchers understand the adjuvant effects of these molecules on the immunological synapse between APCs and T cells.

The rationale for using two model adjuvant proteins, FLH and rSIP, whose use as potential adjuvants has been previously documented through *in vivo* studies in murine models, is based on their similarities and differences. Several similarities have been described: (i) both are TLR4 agonist proteins, (ii) they promote the maturation of DCs, (iii) a limited understanding of the TLR4-associated cell signaling pathway exists, (iv) their contributions to the presentation of exogenous antigens needs to be better understood, and (v) both induce the development of adaptive responses of the Th1 type. Among their differences, two are worth mentioning: (i) species of origin: FLH comes from a mollusk, while rSIP is bacterial, and (ii) the structure of hemocyanin is a very large glycosylated oligomeric protein, unlike rSIP, which is small and lacks oligosaccharides. One of the advantages of TLR4-PBAs is that they can ensure a shared antigen–adjuvant load. rSIP can be expressed in a heterologous system in conjunction with the antigen, while FLH must be conjugated to the antigen. However, it is unknown how rSIP and FLH affect the immune system by binding to TLR4, a receptor that activates multiple signal transduction pathways via MyD88, and TRIF. In this study, we compared these PBAs of TLR4, revealing that their immunomodulatory effects are codependent on MyD88 and TRIF in.

Subunit vaccines containing highly purified recombinant pathogen components are safe; however, they are poorly immunogenic and thus require the use of adjuvants to increase their immunogenicity (42). The protection provided by the most effective vaccines depends on the induction of neutralizing antibodies. Unfortunately, most currently used adjuvants are poorly effective in inducing strong cellular immunity (1, 2, 7). For diseases requiring neutralizing antibodies and T-cell immunity, such as acquired immunodeficiency syndrome (AIDS), severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), tuberculosis, and malaria, it is essential to incorporate immune adjuvants that elicit strong T-cell immunity (1, 2, 7). To trigger the induction of robust CD8 T-cell immunity by vaccines, it is necessary to engage the antigen processing pathway for cross-presentation by APCs, as previously described (1, 2, 7). Although rSIP and FLH activate TLR4 signaling pathways that depend on MyD88 and TRIF recruitment, both proteins undergo finely tuned regulation of their adjuvant effects, which is associated with the intrinsic molecular properties of each protein. Indeed, although there are no crystallographic data for these proteins, considering the available published data, it is possible to confirm that they are very different, as one is a very large, glycosylated protein with a complex quaternary structure (20), and the other is a small nonglycosylated

protein without multiple subunits (49). These differences strongly suggest that the interactions of these proteins with TLR4 could be different. Indeed, the interaction of FLH with TLR4 occurs due to the oligosaccharide residues of FLH (as a viral protein) because when FLH glycosylations are removed, the interaction decreases significantly (23). In contrast, the binding of rSIP could be facilitated by CD14 and the contribution of the MD-2 protein stably associated with the extracellular fragment of the receptor.

The function of DCs stimulated with TLR4 is linked to the greater abundance of costimulatory molecules on their surface, cytokines, and receptors in addition to chemokines and promotes adaptive immunity by activating specific T-lymphocytes. MyD88 signaling is associated with proinflammatory and innate immune responses (50). In contrast, TRIF signaling is associated with the development of an adaptive immune response, which is essential for effective vaccination (10). Although preliminary studies characterizing MyD88 and TRIF interactions with TLR4-LPS have been published (12), in this work, we characterized two protein agonists from different species for the first time. Furthermore, this study establishes that MyD88 and TRIF are essential for the adjuvant effects of these proteins. Specifically, one of the most notable effects is that rSIP and FLH generate IL-6 and IP-10 transcripts in a manner dependent on MyD88 and TRIF. However, in terms of IL-6 and IP-10 secretion, only rSIP depends on MyD88 and TRIF, while FLH is TRIF dependent. These differences can be explained by the fact that the genome-wide correlation between mRNA expression levels has an explanatory power of approximately 40% and can be attributed to other levels of regulation between the transcript and the protein product (51–53).

TLR4 can interact with other pattern recognition receptors (PRRs) to mediate intracellular signaling and interactions with C-type lectin receptors, such as MR and DC-SIGN, to promote, in some cases, antigen cross-presentation (54–56). Following TLR4 agonist-induced DC maturation, processes associated with antigen cross-presentation, such as scavenging receptor-mediated phagocytosis and phagolysosomal fusion, are enhanced during the initial hours of TLR4 activation, after which a loss of antigen internalization and the molecular components necessary for cytosolic delivery of antigen occurs (57). Gupta et al. found that MHC-I molecules are not derived from the endoplasmic reticulum–Golgi intermediate compound (ERGIC) upon TLR stimulation because ERGIC components are recruited to phagosomes independent of TLR signaling (38). However, stimulation of TLR4 results in the accumulation of MHC class I molecules derived from the endocytic recycling compartment (ERC; marked by Rab11a and vesicle-associated membrane proteins 3 and 8 [VAMP3 and 8, respectively]) in phagosomes (44, 58). In addition, TLR-mediated MyD88-dependent IKK phosphorylation of synaptosome-associated protein 23 (SNAP23) mediates endosomal recycling compartment (ERC)–phagosome fusion (38). Alloatti et al. also showed that TLR4 activation delays phagosome maturation and antigen degradation, which induces Rab34-mediated intracellular perinuclear pool formation (36). On the other hand, concerning the endosome-to-cytosol pathway, it is known that the activity of the translocon protein Sec61 in the ER is mediated by TRIF because this step is essential for translocation from the endosome to the cytosol (43). Our results suggest that TLR-based adjuvants likely engage vacuolar pathways to potentiate effective CD8 T-cell responses. However, rSIP and FLH may also be involved in the endosome-to-cytosol pathway via TRIF and the Sec61 protein. This assumption was supported, given that different proteasome inhibitors decreased the proliferation of CD8 OT-I lymphocytes.

This work supports the conclusion that rSIP and FLH mediate TLR4 activation and that this modulation depends on the recruitment of MyD88 and TRIF because each protein can induce finely tuned signaling patterns. This characteristic seemed dependent on the structure of the TLR4 agonist and its potency because FLH can mediate cytokine secretion independent of MyD88 recruitment, and FLH can mediate antigen cross-presentation in a manner dependent on MyD88 recruitment. In contrast, rSIP is totally dependent on MyD88 and TRIF for cytokine secretion and antigen cross-presentation. Notably, the TRIF pathway is essential for rSIP- and FLH-induced secretion of IP-10 and IL-6, and the MyD88 pathway is only essential for rSIP-induced secretion of IP-10 and IL-6 (Figure 5). However, IL-6 secretion by FLH was dependent on MyD88 recruitment, which could be attributed to the stimulation time of FLH in BM-DCs, since FLH is influenced by the recruitment of MyD88 during the first 4 hours of stimulation (Figure 3A). However, after prolonged stimulation times, the inhibition of MyD88 recruitment did not exert a significant effect on the expression of IL-6 (Figures 5C). Another explanation is that the MyD88-adapter-like (MAL) protein could be involved in signaling, as previously described for FLH (17). MAL could be recruited by TRAF6, suggesting that after longer stimulation with FLH, the TRAF6 protein would be activated differently by rSIP, enabling the secretion of IP-10 and IL-6.

Regarding the regulation of TLR4, it was previously shown using iterative mathematical models that the pathways mediated by MyD88 and TRIF provide are dependent on the concentrations of ligands that transmit information about the threat of the pathogen (59). These changes in signaling are supported by the fact that the start of TLR4 signaling involves oligomerization, which determines MyD88 and TRIF signaling (60). This implies that one pattern recognition receptor is activated by different microenvironmental cues to generate macrophages with distinct phenotypes linked to a subset of cytokines and phosphoproteomic signaling patterns (61). In this context, our results are consistent with this finding because a lower concentration of rSIP than FLH is needed to activate TLR4. This difference is directly related to the molecular characteristics of each protein (Figure 8). Furthermore, this signaling change is supported by the start of TLR4 signaling during dimerization and the oligomerization dynamics, which determines MyD88 and TRIF signaling. Therefore, since rSIP and FLH are partial agonists, their interaction with TLR4 could also be involved in the oligomerization dynamics of this receptor.

In conclusion, these results provide further insight into the nature of TLR4 agonist protein adjuvants and their contributions to activation of the MyD88 and TRIF signaling pathways. These results are relevant since they contribute to our knowledge of how protein-based agonists of TLR4 can act as adjuvants, information that supports the use of these agonists in the development of future experimental vaccines for cancer, persistent diseases, or future

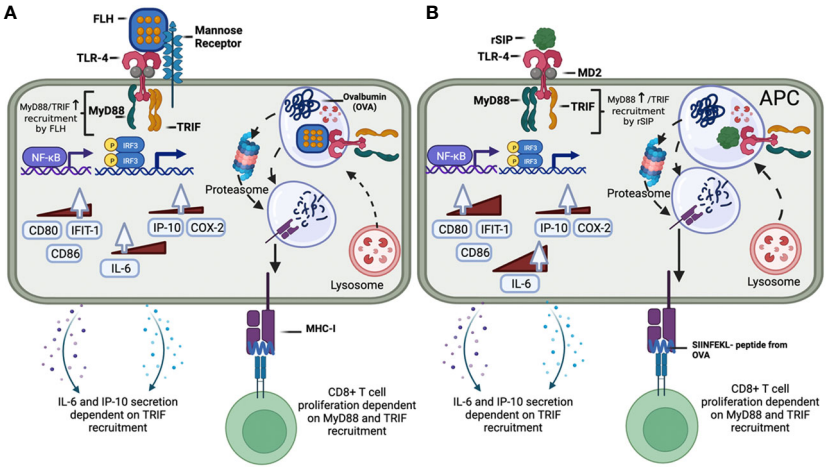


FIGURE 8
Modulation of the MyD88 and TRIF signaling pathways and downstream responses by the protein-based adjuvants rSIP and FLH in APCs. **(A)** FLH interacts with mannose receptors and activates the TLR4 signaling pathway, which recruits MyD88 and TRIF. FLH activates NF-kB and IRF. MyD88 recruitment is involved in the expression of the IL-6, CD80, IFIT-1, and CD86 mRNAs, while TRIF is involved in the expression of IL-6 and IP-10. Regarding the secretion of cytokines, only TRIF is involved in the secretion of IL-6 and IP-10. FLH then promotes OVA cross-presentation, and its effect is dependent on MyD88 and TRIF. **(B)** rSIP activates the TLR4 signaling pathway, which recruits MyD88 and TRIF. rSIP activates NF-kB and IRF. MyD88 recruitment is involved in expression of the IL-6, COX-2, CD80, IFIT-1, and CD86 mRNAs, while TRIF is involved in the expression of IL-6, COX-2, and IP-10. Regarding the secretion of cytokines, both MyD88 and TRIF are involved in the secretion of IL-6 and IP-10. rSIP then promotes OVA cross-presentation, and its effect is dependent on MyD88 and TRIF.

pandemics; an ongoing challenge related to controlling the doses of vaccine adjuvants, such as the sMLA adjuvant, which is the active component of the glucopyranosyl lipid adjuvant (GLA), exists (62, 63). Additional studies are needed to establish a preclinical model and determine the effects of these adjuvants and their contributions to the MyD88 and TRIF signaling pathways downstream of TLR4.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding authors.

Ethics statement

The animal study was reviewed and approved by Comite Institucional de Cuidado y Uso de Animales de Laboratorio (CICUAL) del Instituto de Salud Pública de Chile (ISPCh).

Author contributions

DD-D, MIB, and AEV: conception and design. DD-D, MS, DE, BC, BV, and MT: experiment execution and data acquisition. DD-D, MS, AM, FS, SL, JD, MIB, and AEV: data analysis and interpretation. DD-D, MIB, and AEV: manuscript writing. All authors contributed to the article and approved the submitted version.

Funding

This study was supported by CONICYT-CHILE FONDECYT Regular Grant 1201600 to MIB, FONDEF IDEA Grant 21I10370 and ID23I10207 to AEV and FONDAP 15130011 to SL. In addition, this work was supported by the Ministry of Sciences (Code: ANDID_FI_AVASQUEZ_216). Fellowships were awarded to DD-D (ANID N° 21200880) and to MS (ANID N 21210946). FS holds a postdoctoral fellowship from the National Commission for Scientific and Technological Research (CONICYT), Chile.

Acknowledgments

The authors are grateful to Ricardo Manzo Paredes from ISP for their valuable work in obtaining rSIP. Additionally, we thank Daniel Soto Carrasco for his work in handling laboratory animals. The authors acknowledge BioRender.com for the design of Figure 8 Finally, the authors thank Dr. Rodrigo Pacheco from Fundación Ciencia & Vida for providing the OT-I and OT-II mice used in this study.

Conflict of interest

Authors AM, FS and MIB were employed by BIOSONDA S.A.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2023.1186188/full#supplementary-material

SUPPLEMENTARY FIGURE 1
Partial and differential agonism of Toll-like receptors 4 (hTLR4) by protein-based adjuvants. **(A)** Determination of the concentrations of protein agonists needed to activate hTLR4. HEK-Blue-hTLR4 reporter cells were exposed to different concentrations of protein adjuvants. Dose–response curves were generated for cells exposed to a maximum concentration of 2,540 nM rSIP and FLH for 48 h. Data show normalized HEK-Blue mTLR4 cell responses considering LPS treatment as 100% stimulation; 100% = maximum dose plateau of the LPS agonist. **(B)** Comparison of the log EC_{50} values of protein agonists in activating mTLR4. Log (EC_{50}) values for rSIP and FLH were determined according to the relative abundance of soluble AP secreted by Hek-Blue-hTLR4 cells. Individual log (EC_{50}) values and mean values from three independent experiments are shown. The statistical significance of differences was analyzed using an unpaired t test (ns: not significant). **(C)** IRF3 activation by FLH and rSIP in BM-DCs. Phosphorylation of IRF3 induced by FLH and rSIP in BM-DCs determined by Western blotting. BM-DCs from 6 mice (pool) were used for analysis of phospho-IRF-3 (Ser396) (4D4G). As a positive control, A549 cells stimulated with poly (I:C) (C+) and without stimulation (C-) were used. The BM-DCs were stimulated for 0, 1, 2, 7, 10, 13, and 18 hours.

SUPPLEMENTARY FIGURE 2
rSIP and FLH induce classical antigen presentation. BM-DCs were stimulated for 18 h with rSIP and FLH and coadministered OVA (1 mg/mL). **(A)** The percentage and **(B)** MFI of naïve OT-II $CD4^+$ T-cell ($3x10^5$ cells) activation (CD69+) were measured after 18 h of coculture with treated BM-DCs. Data are the means ± SDs of three independent experiments. Statistical significance was determined by one-way analysis of variance (ANOVA) with Tukey's multiple comparisons test (*p<0.05; ns: statistically not significant). **(C)** Flow cytometry analyses show representative CD4+ CD69+ profiles of OT-II cells cocultured with stimulated BM-DCs.

SUPPLEMENTARY FIGURE 3
rSIP and FLH induce antigen cross-presentation through cathepsin. BM-DCs were pretreated for 1 h with DMSO, brefeldin A (1 µM), leupeptin (10 µM), and pepstatin A (40 nM) and stimulated for three hours with rSIP + OVA (1 mg/mL) and FLH + OVA (1 mg/mL). Naïve OT-I CD8+ T-cell ($2.5x10^5$ cells) proliferation was measured via CellTrace Violet staining after 3 days of coculture with treated BM-DCs. Data are the means ± SDs of three independent experiments. Statistical significance was determined using repeated measures one-way analysis of variance (ANOVA) and the *post hoc* Sidak test (*p<0.05; **p<0.01; ns: statistically not significant).

References

1. Pulendran B S, Arunachalam P, O'Hagan DT. Emerging concepts in the science of vaccine adjuvants. *Nat Rev Drug Discovery* (2021) 20(6):454–75. doi: 10.1038/s41573-021-00163-y

2. Díaz-Dinamarca DA, Salazar ML, Castillo BN, Manubens A, Vasquez AE, Salazar F, et al. Protein-based adjuvants for vaccines as immunomodulators of the innate and adaptive immune response: Current knowledge, challenges, and future opportunities. *Pharmaceutics* (2022) 14(8):1671. doi: 10.3390/pharmaceutics14081671

3. Reed SG, Tomai M, Gale MJ. New horizons in adjuvants for vaccine development. *Curr Opin Immunol* (2020) 65:97–101. doi: 10.1016/j.coi.2020.08.008

4. Wang YQ, Bazin-Lee H, Evans JT, Casella CR, Mitchell TC. MPL adjuvant contains competitive antagonists of human TLR4. *Front Immunol* (2020) 11:577823. doi: 10.3389/fimmu.2020.577823

5. Toussi D, Massari P. Immune adjuvant effect of molecularly-defined toll-like receptor ligands. *Vaccines* (2014) 2(2):323–53. doi: 10.3390/vaccines2020323

6. Steinhagen F, Kinjo T, Bode C, Klinman DM. TLR-based immune adjuvants. *Vaccine* (2011) 29(17):3341–55. doi: 10.1016/j.vaccine.2010.08.002

7. Lee W, Suresh M. Vaccine adjuvants to engage the cross-presentation pathway. *Front Immunol* (2022) 13:940047. doi: 10.3389/fimmu.2022.940047

8. Gay NJ, Symmons MF, Gangloff M, Bryant CE. Assembly and localization of Toll-like receptor signalling complexes. *Nat Rev Immunol* (2014) 14(8):546–58. doi: 10.1038/nri3713

9. Kawai T, Adachi O, Ogawa T, Takeda K, Akira S. Unresponsiveness of myD88-deficient mice to endotoxin. *Immunity* (1999) 11(1):115–22. doi: 10.1016/S1074-7613(00)80086-2

10. Yamamoto M. Role of adaptor TRIF in the myD88-independent toll-like receptor signaling pathway. *Science* (2003) 301(5633):640–3. doi: 10.1126/science.1087262

11. Zhou J, Sun T, Jin S, Guo Z, Cui J. Dual feedforward loops modulate type I interferon responses and induce selective gene expression during TLR4 activation. *iScience* (2020) 23(2):100881. doi: 10.1016/j.isci.2020.100881

12. Kolb JP, Casella CR, SenGupta S, Chilton PM, Mitchell TC. Type I interferon signaling contributes to the bias that Toll-like receptor 4 exhibits for signaling mediated by the adaptor protein TRIF. *Sci Signaling* (2014) 7(351):ra108–8. doi: 10.1126/scisignal.2005442

13. Hamdy S, Elamanchili P, Alshamsan A, Molavi O, Satou T, Samuel J. Enhanced antigen-specific primary CD4+ and CD8+ responses by codelivery of ovalbumin and toll-like receptor ligand monophosphoryl lipid A in poly(D,L-lactic-co-glycolic acid) nanoparticles. *J BioMed Mater Res* (2007) 81A(3):652–62. doi: 10.1002/jbm.a.31019

14. Borst J, Ahrends T, Bąbała N, Melief CJM, Kastenmüller W. CD4+ T cell help in cancer immunology and immunotherapy. *Nat Rev Immunol* (2018) 18(10):635–47. doi: 10.1038/s41577-018-0044-0

15. Wang C, Deng L, Hong M, Akkaraju GR, Chen ZJ. TAK1 is a ubiquitin-dependent kinase of MKK and IKK. *Nature* (2001) 412(6844):346–51. doi: 10.1038/35085597

16. Kumar S, Sunagar R, Gosselin E. Bacterial protein toll-like-receptor agonists: A novel perspective on vaccine adjuvants. *Front Immunol* (2019) 10:1144. doi: 10.3389/fimmu.2019.01144

17. Jiménez JM, Salazar ML, Arancibia S, Villar J, Salazar F, Brown GD, et al. TLR4, but neither Dectin-1 nor Dectin-2, participates in the mollusk hemocyanin-induced proinflammatory effects in antigen-presenting cells from mammals. *Front Immunol* (2019) 10:1136. doi: 10.3389/fimmu.2019.01136

18. Becker MI, Arancibia S, Salazar F, Del Campo M, De Ioannes A. Mollusk hemocyanins as natural immunostimulants in biomedical applications. In: Duc GHT, editor. *Immune response activation*. Rijeka, Croatia: InTech (2014) p. 221–42. Available at: http://www.intechopen.com/books/immune-response-activation/mollusk-hemocyanins-as-natural-immunostimulants-in-biomedical-applications.

19. De Ioannes P, Moltedo B, Oliva H, Pacheco R, Faunes F, De Ioannes AE, et al. Hemocyanin of the molluscan *Concholepas concholepas* exhibits an unusual heterodecameric array of subunits. *J Biol Chem* (2004) 279(25):26134–42. doi: 10.1074/jbc.M400903200

20. Arancibia S, Espinoza C, Salazar F, Del Campo M, Tampe R, Zhong TY, et al. A novel immunomodulatory hemocyanin from the limpet *Fissurella latimarginata* promotes potent anti-tumor activity in melanoma. *PLoS One* (2014) 9(1):e87240. doi: 10.1371/journal.pone.0087240

21. Gleisner MA, Pereda C, Tittarelli A, Navarrete M, Fuentes C, Ávalos I, et al. A heat-shocked melanoma cell lysate vaccine enhances tumor infiltration by prototypic effector T cells inhibiting tumor growth. *J Immunother Cancer* (2020) 8(2):e000999. doi: 10.1136/jitc-2020-000999

22. Reyes D, Salazar L, Espinoza E, Pereda C, Castellón E, Valdevenito R, et al. Tumour cell lysate-loaded dendritic cell vaccine induces biochemical and memory immune response in castration-resistant prostate cancer patients. *Br J Cancer* (2013) 109(6):1488–97. doi: 10.1038/bjc.2013.494

23. Salazar ML, Jiménez JM, Villar J, Rivera M, Báez M, Manubens A, et al. *N* -Glycosylation of mollusk hemocyanins contributes to their structural stability and immunomodulatory properties in mammals. *J Biol Chem* (2019) 294(51):19546–64. doi: 10.1074/jbc.RA119.009525

24. Zhong TY, Arancibia S, Born R, Tampe R, Villar J, Del Campo M, et al. Hemocyanins stimulate innate immunity by inducing different temporal patterns of proinflammatory cytokine expression in macrophages. *JI* (2016) 196(11):4650–62. doi: 10.4049/jimmunol.1501156

25. Villar J, Salazar ML, Jiménez JM, Campo MD, Manubens A, Gleisner MA, et al. C-type lectin receptors MR and DC-SIGN are involved in recognition of hemocyanins, shaping their immunostimulatory effects on human dendritic cells. *Eur J Immunol* (2021) 51(7):1715–31. doi: 10.1002/eji.202149225

26. Paccagnella M, Bologna L, Beccaro M, Mičetić I, Di Muro P, Salvato B. Structural subunit organization of molluscan hemocyanins. *Micron* (2004) 35(1–2):21–2. doi: 10.1016/j.micron.2003.10.007

27. Diaz-Dinamarca DA, Manzo RA, Soto DA, Avendaño-Valenzuela MJ, Bastias DN, Soto PI, et al. Surface immunogenic protein of *Streptococcus* group B is an agonist of toll-like receptors 2 and 4 and a potential immune adjuvant. *Vaccines* (2020) 8(1):29. doi: 10.3390/vaccines8010029

28. Diaz-Dinamarca DA, Hernandez C, Escobar DF, Soto DA, Muñoz GA, Badilla JF, et al. Mucosal vaccination with *Lactococcus lactis*-secreting surface immunological protein induces humoral and cellular immune protection against group B *Streptococcus* in a murine model. *Vaccines* (2020) 8(2):146. doi: 10.3390/vaccines8020146

29. Diaz-Dinamarca DA, Soto DA, Leyton YY, AltamIrano-Lagos MJ, Avendaño MJ, Kalergis AM, et al. Oral vaccine based on a surface immunogenic protein mixed with alum promotes a decrease in *Streptococcus agalactiae* vaginal colonization in a mouse model. *Mol Immunol* (2018) 103:63–70. doi: 10.1016/j.molimm.2018.08.028

30. Soto JA, Diaz-Dinamarca DA, Soto DA, Barrientos MJ, Carrión F, Kalergis AM, et al. Cellular immune response induced by surface immunogenic protein with AbISCO-100 adjuvant vaccination decreases group B *Streptococcus* vaginal colonization. *Mol Immunol* (2019) 111:198–204. doi: 10.1016/j.molimm.2019.04.025

31. Díaz-Dinamarca DA. The optimisation of the expression of recombinant surface immunogenic protein of group B *streptococcus* in *Escherichia coli* by response surface methodology improves humoral immunity. *Mol Biotechnol* (2018) 11:215–25. doi: 10.1007/s12033-018-0065-8

32. Lutz MB, Kukutsch N, Ogilvie ALJ, Rößner S, Koch F, ROmani N, et al. An advanced culture method for generating large quantities of highly pure dendritic cells from mouse bone marrow. *J Immunol Methods* (1999) 223(1):77–92. doi: 10.1016/S0022-1759(98)00204-X

33. Chen Q, Wang J, Zhang Q, Zhang J, Lou Y, Yang J, et al. Tumour cell-derived debris and IgG synergistically promote metastasis of pancreatic cancer by inducing inflammation via tumour-associated macrophages. *Br J Cancer* (2019) 121(9):786–95. doi: 10.1038/s41416-019-0595-2

34. Matsunaga N, Tsuchimori N, Matsumoto T, Ii M. TAK-242 (Resatorvid), a small-molecule inhibitor of toll-like receptor (TLR) 4 signaling, binds selectively to TLR4 and interferes with interactions between TLR4 and its adaptor molecules. *Mol Pharmacol* (2011) 79(1):34–41. doi: 10.1124/mol.110.068064

35. Sharma J, Boyd T, Alvarado C, Gunn E, Adams J, Ness T, et al. Reporter cell assessment of TLR4-induced NF-κB responses to cell-free hemoglobin and the influence of biliverdin. *Biomedicines* (2019) 7(2):41. doi: 10.3390/biomedicines7020041

36. Alloatti A, Kotsias F, Pauwels AM, Carpier JM, Jouve M, Timmerman E, et al. Toll-like receptor 4 engagement on dendritic cells restrains phago-lysosome fusion and promotes cross-presentation of antigens. *Immunity* (2015) 43(6):1087–100. doi: 10.1016/j.immuni.2015.11.006

37. Theisen DJ, Davidson JT, Briseño CG, Gargaro M, Lauron EJ, Wang Q, et al. WDFY4 is required for cross-presentation in response to viral and tumor antigens. *Science* (2018) 362(6415):694–9. doi: 10.1126/science.aat5030

38. Nair-Gupta P, Baccarini A, Tung N, Seyffer F, Florey O, Huang Y, et al. TLR signals induce phagosomal MHC-I delivery from the endosomal recycling compartment to allow cross-presentation. *Cell* (2014) 158(3):506–21. doi: 10.1016/j.cell.2014.04.054

39. Dingjan I, Verboogen DR, Paardekooper LM, Revelo NH, Sittig SP, Visser LJ, et al. Lipid peroxidation causes endosomal antigen release for cross-presentation. *Sci Rep* (2016) 6(1):22064. doi: 10.1038/srep22064

40. Ehlert FJ, Suga H, Griffin MT. Quantifying agonist activity at G protein-coupled receptors. *JoVE* (2011) 58):3179. doi: 10.3791/3179

41. Reed SG, Hsu FC, Carter D, Orr MT. The science of vaccine adjuvants: advances in TLR4 ligand adjuvants. *Curr Opin Immunol* (2016) 41:85–90. doi: 10.1016/j.coi.2016.06.007

42. Bonam SR, Partidos CD, Halmuthur SKM, Muller S. An overview of novel adjuvants designed for improving vaccine efficacy. *Trends Pharmacol Sci* (2017) 38 (9):771–93. doi: 10.1016/j.tips.2017.06.002

43. Zehner M, Marschall AL, Bos E, Schloetel JG, Kreer C, Fehrenschild D, et al. The translocon protein sec61 mediates antigen transport from endosomes in the cytosol for cross-presentation to CD8+ T cells. *Immunity* (2015) 42(5):850–63. doi: 10.1016/j.immuni.2015.04.008

44. Cebrian I, Croce C, Guerrero NA, Blanchard N, Mayorga LS. Rab22a controls MHC -I intracellular trafficking and antigen cross-presentation by dendritic cells. *EMBO Rep* (2016) 17(12):1753–65. doi: 10.15252/embr.201642358

45. Romerio A, Peri F. Increasing the chemical variety of small-molecule-based TLR4 modulators: An overview. *Front Immunol* (2020) 11:1210. doi: 10.3389/fimmu.2020.01210

46. Casella CR, Mitchell TC. Putting endotoxin to work for us: Monophosphoryl lipid A as a safe and effective vaccine adjuvant. *Cell Mol Life Sci* (2008) 65(20):3231–40. doi: 10.1007/s00018-008-8228-6

47. Mata-Haro V, Cekic C, Martin M, Chilton PM, Casella CR, Mitchell TC. The vaccine adjuvant monophosphoryl lipid A as a TRIF-biased agonist of TLR4. *Science* (2007) 316(5831):1628–32. doi: 10.1126/science.1138963

48. Bowen WS, Minns LA, Johnson DA, Mitchell TC, Hutton MM, Evans JT. Selective TRIF-dependent signaling by a synthetic toll-like receptor 4 agonist. *Sci Signal* (2012) 5(211):ra13-ra13. doi: 10.1126/scisignal.2001963

49. Hall RL, Wood EJ. The carbohydrate content of gastropod haemocyanins. *Biochem Soc Trans* (1976) 4(2):307–9. doi: 10.1042/bst0040307

50. Lin SC, Lo YC, Wu H. Helical assembly in the MyD88–IRAK4–IRAK2 complex in TLR/IL-1R signalling. *Nature* (2010) 465(7300):885–90. doi: 10.1038/nature09121

51. de Sousa Abreu R, Penalva LO, Marcotte EM, Vogel C. Global signatures of protein and mRNA expression levels. *Mol Biosyst* (2009) 10:1039.b908315d. doi: 10.1039/b908315d

52. Vogel C, Marcotte EM. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nat Rev Genet* (2012) 13(4):227–32. doi: 10.1038/nrg3185

53. Maier T, Güell M, Serrano L. Correlation of mRNA and protein in complex biological samples. *FEBS Letters* (2009) 583(24):3966–73. doi: 10.1016/j.febslet.2009.10.036

54. Horrevorts SK, Duinkerken S, Bloem K, Secades P, Kalay H, Musters RJ, et al. Toll-like receptor 4 triggering promotes cytosolic routing of DC-SIGN-targeted antigens for presentation on MHC class I. *Front Immunol* (2018) 9:1231. doi: 10.3389/fimmu.2018.01231

55. Loures FV, AraÃojo EF, Feriotti C, Bazan SB, Calich VLG. TLR-4 cooperates with Dectin-1 and mannose receptor to expand Th17 and Tc17 cells induced by Paracoccidioides brasiliensis stimulated dendritic cells. *Front Microbiol* (2015) 6:261/abstract. doi: 10.3389/fmicb.2015.00261/abstract

56. Wang YY, Hu CF, Li J, You X, Gao FG. Increased translocation of antigens to endosomes and TLR4 mediated endosomal recruitment of TAP contribute to nicotine augmented cross-presentation. *Oncotarget* (2016) 7(25):38451–66. doi: 10.18632/oncotarget.9498

57. Gil-Torregrosa BC, Lennon-Duménil AM, Kessler B, Guermonprez P, Ploegh HL, Fruci D, et al. Control of cross-presentation during dendritic cell maturation. *Eur J Immunol* (2004) 34(2):398–407. doi: 10.1002/eji.200324508

58. Croce C, Mayorga LS, Cebrian I. Differential requirement of Rab22a for the recruitment of ER-derived proteins to phagosomes and endosomes in dendritic cells. *Small GTPases* (2017) 29:1–9. doi: 10.1080/21541248.2017.1384088

59. Cheng Z, Taylor B, Ourthiague DR, Hoffmann A. Distinct single-cell signaling characteristics are conferred by the MyD88 and TRIF pathways during TLR4 activation. *Sci Signal* (2015) 8(385):ra69-ra69. doi: 10.1126/scisignal.aaa5208

60. Krüger CL, Zeuner MT, Cottrell GS, Widera D, Heilemann M. Quantitative single-molecule imaging of TLR4 reveals ligand-specific receptor dimerization. *Sci Signal* (2017) 10(503):eaan1308. doi: 10.1126/scisignal.aan1308

61. Piccinini AM, Zuliani-Alvarez L, Lim JMP, Midwood KS. Distinct microenvironmental cues stimulate divergent TLR4-mediated signaling pathways in macrophages. *Sci Signal* (2016) 9(443):ra86-ra86. doi: 10.1126/scisignal.aaf3596

62. Orr MT, Duthie MS, Windish HP, Lucas EA, Guderian JA, Hudson TE, et al. MyD88 and TRIF synergistic interaction is required for TH1-cell polarization with a synthetic TLR4 agonist adjuvant: Immunity to infection. *Eur J Immunol* (2013) 43 (9):2398–408. doi: 10.1002/eji.201243124

63. Gaddis DE, Michalek SM, Katz J. TLR4 Signaling via MyD88 and TRIF Differentially Shape the CD4 $^{+}$ T Cell Response to *Porphyromonas gingivalis* Hemagglutinin B. *JI* (2011) 186(10):5772–83. doi: 10.4049/jimmunol.1003192

frontiers | Frontiers in Immunology

TYPE Original Research
PUBLISHED 08 November 2023
DOI 10.3389/fimmu.2023.1259197

OPEN ACCESS

EDITED BY
Francesco Pappalardo,
University of Catania, Italy

REVIEWED BY
Elke Bergmann-Leitner,
Walter Reed Army Institute of Research, United States
Saranya Sridhar,
Sanofi Pasteur, United Kingdom

*CORRESPONDENCE
Helder I. Nakaya
helder.nakaya@einstein.br

RECEIVED 15 July 2023
ACCEPTED 23 October 2023
PUBLISHED 08 November 2023

CITATION
Gonzalez Dias Carvalho PC, Dominguez Crespo Hirata T, Mano Alves LY, Moscardini IF, do Nascimento APB, Costa-Martins AG, Sorgi S, Harandi AM, Ferreira DM, Vianello E, Haks MC, Ottenhoff THM, Santoro F, Martinez-Murillo P, Huttner A, Siegrist C-A, Medaglini D and Nakaya HI (2023) Baseline gene signatures of reactogenicity to Ebola vaccination: a machine learning approach across multiple cohorts. *Front. Immunol.* 14:1259197. doi: 10.3389/fimmu.2023.1259197

Baseline gene signatures of reactogenicity to Ebola vaccination: a machine learning approach across multiple cohorts

Patrícia Conceição Gonzalez Dias Carvalho[1,2], Thiago Dominguez Crespo Hirata[3], Leandro Yukio Mano Alves[3], Isabelle Franco Moscardini[4], Ana Paula Barbosa do Nascimento[5], André G. Costa-Martins[3,6], Sara Sorgi[7], Ali M. Harandi[8,9], Daniela M. Ferreira[1,2], Eleonora Vianello[10], Mariëlle C. Haks[10], Tom H. M. Ottenhoff[10], Francesco Santoro[7], Paola Martinez-Murillo[11], for VSV-EBOVAC Consortia, for VSV-EBOPLUS Consortia, Angela Huttner[11,12], Claire-Anne Siegrist[11], Donata Medaglini[13] and Helder I. Nakaya[14,15]*

[1]Oxford Vaccine Group, University of Oxford, Oxford, United Kingdom, [2]Department of Clinical Sciences, Liverpool School of Tropical Medicine, Liverpool, United Kingdom, [3]Department of Clinical and Toxicological Analyses, School of Pharmaceutical Sciences, University of São Paulo, São Paulo, Brazil, [4]Microbiotec Srl, Siena, Italy, [5]Division of Infectious Diseases, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, United States, [6]Artificial Intelligence and Analytics Department, Institute for Technological Research, São Paulo, Brazil, [7]Laboratory of Molecular Microbiology and Biotechnology (LAMMB), Department of Medical Biotechnologies, University of Siena, Siena, Italy, [8]Department of Microbiology and Immunology, Institute of Biomedicine, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden, [9]Vaccine Evaluation Center, BC Children's Hospital Research Institute, University of British Columbia, Vancouver, BC, Canada, [10]Department of Infectious Diseases, Leiden University Medical Center, Leiden, Netherlands, [11]Centre for Vaccinology, Faculty of Medicine, University of Geneva, Geneva, Switzerland, [12]Infectious Diseases Service, Geneva University Hospitals, Geneva, Switzerland, [13]Department of Medical Biotechnologies, University of Siena, Siena, Italy, [14]Scientific Platform Pasteur-University of São Paulo, São Paulo, Brazil, [15]Hospital Israelita Albert Einstein, São Paulo, Brazil

Introduction: The rVSVDG-ZEBOV-GP (Ervebo®) vaccine is both immunogenic and protective against Ebola. However, the vaccine can cause a broad range of transient adverse reactions, from headache to arthritis. Identifying baseline reactogenicity signatures can advance personalized vaccinology and increase our understanding of the molecular factors associated with such adverse events.

Methods: In this study, we developed a machine learning approach to integrate prevaccination gene expression data with adverse events that occurred within 14 days post-vaccination.

Results and Discussion: We analyzed the expression of 144 genes across 343 blood samples collected from participants of 4 phase I clinical trial cohorts:

Switzerland, USA, Gabon, and Kenya. Our machine learning approach revealed 22 key genes associated with adverse events such as local reactions, fatigue, headache, myalgia, fever, chills, arthralgia, nausea, and arthritis, providing insights into potential biological mechanisms linked to vaccine reactogenicity.

KEYWORDS

Ebola, rVSVDG-ZEBOV-GP vaccine, baseline gene signatures, adverse events, vaccine safety, personalized vaccinology, machine learning, data integration

1 Introduction

Ebola virus disease (EVD) is a severe and fatal infectious disease (1). rVSVΔG-ZEBOV-GP, under the name of Ervebo®, is given as a single-dose vaccine. It is a recombinant vaccine against the live and attenuated vesicular stomatitis (VSV) virus, in which the gene encoding for the VSV envelope glycoprotein has been replaced by the Ebola strain Zaire virus (ZEBOV-GP) glycoprotein gene (2). This vaccine is highly immunogenic for at least two years (3).

Live replicating VSV-based vaccines can elicit potent humoral (4–6) and strong cellular immune responses against viral (7–9). However, replication-competent vectors are frequently associated with a higher risk for adverse events (AE) (10). Although rVSVΔG-ZEBOV-GP is safe, immunogenic, and protective in human trials (11), vaccinees may report transient adverse reactions such as fever, inflammation, arthritis, dermatitis and vasculitis (11–14). Vaccine viraemia is common and associated with frequent mild-to-moderate acute inflammatory reactions and, in some vaccinees, viral dissemination, leading to arthritis and occasional dermatitis (13, 15). The occurrence of arthritis, arthralgia and other forms of joint swellings and tissue infiltration was higher in European and US vaccinees than in participants from Africa. Arthritis cases have been reported in approximately 23% (24 – 102) of vaccinees from Switzerland (11) and 4.5% (19 - 418) from USA (16), whereas a low incidence of 2.5% (1 - 40) (12) or non-incidence has been reported in Kenya and Gabon, respectively (12). The cases occurred mainly in participants aged 40 years and above, and they were self-limiting with no sequelae (15).

Although these AE did not prevent vaccine uptake (15), identifying baseline reactogenicity signatures represents an important step toward the development of personalized vaccinology and could enhance public confidence in the safety of vaccines (17). Recent studies have reported baseline predictors of post-vaccination responses for human influenza virus (18, 19), hepatitis B virus (20), as well as malaria (21) vaccination. Nonetheless, few studies focused on reactogenicity (22, 23).

In this study, we report a machine learning (ML) approach to unravel multicohort baseline transcriptional reactogenicity signatures to rVSVΔG-ZEBOV-GP. We have integrated AE reported by participants from Switzerland, USA, Gabon and Kenya clinical trials with the expression of 144 genes before the administration of rVSVΔG-ZEBOV-GP. We have identified an AE signature in which twenty-two genes and nine adverse events appear to be associated. Crucially, despite the varying baseline, the genes contribute to predicting delineated stable baseline differences across cohorts, raising the prospect of screening for AE propensity before vaccination.

2 Methods

2.1 Study design and ethics statement

The data was obtained from four clinical trials conducted for the VSV-EBOVAC and VSV-EBOPLUS Consortia on 3 different continents: North America (Phase I, randomized, double-blind, placebo-controlled, dose-response trial in the USA; Registration number NCT02314923), Europe (Phase I/II, randomized, double-blind, placebo-controlled, dose-finding trial in Geneva, Switzerland; Registration number NCT02287480) and Africa (Phase I, randomized, open-label, dose-escalation trial in Lambaréné, Gabon, and a phase I, open-label, dose-escalation trial in Kilifi, Kenya; Registration numbers PACTR201411000919191 and NCT02296983, respectively).

The trial protocols were reviewed and approved by the WHO's Ethics Committee as well as by local ethics committees (USA trial: the Chesapeake Institutional Review Boards (Columbia, MD, USA) and the Crescent City Institutional Review Board (New Orleans, LA, USA); Geneva trial: the Geneva Cantonal Ethics Commission and the Swiss Agency for Therapeutic Products (Swissmedic); Lambaréné trial: the Scientific Review Committee of Centre de Recherches Médicales de Lambaréné (CERMEL), the Institutional Ethics Committee of CERMEL, the National Ethics Committee of Gabon, and the Institutional Ethics Committee of the Universitätsklinikum Tübingen; Kilifi trial: Kilifi Ethics Committee). Placebo recipients received a normal saline injection. Information about randomization and masking and vaccine procedures were published elsewhere (16).

2.2 Available data from VSV-EBOVAC and VSV-EBOPLUS

Reactogenicity data from 782 healthy adult volunteers were collected: 512 from the United States of America (418 vaccinated and 94 placebo recipients), 115 from Geneva, Switzerland (102

vaccinated and 13 placebo recipients), 115 from Lambaréné, Gabon, and 40 from Kilifi, Kenya.

Peripheral whole blood samples were collected at several time points for transcriptomic evaluation (115 in Switzerland, 144 in the USA, 83 in Gabon, and 39 in Kenya). However, since the interest of this work was to study the host's aptitude to develop adverse reactions, we have used only expression data obtained at day 0, before immunization. Among the adults included in the study, 343/782 (43.9%) volunteers from the four cohorts had the expression of 144 genes quantified from a multiplex RT-PCR quantitative platform, which has been amplified using a two-color ligation-dependent probe called dcRT-MLPA (24), and was previously published (25).

2.3 Outcomes

We performed a predictive reactogenicity cohort evaluation within the phase 1 trials from Switzerland (randomized), Gabon (dose-escalation), Kenya, and USA. The biological and clinical outcomes of these studies have been reported elsewhere (3, 12, 16, 26).

Reactogenicity data were collected until day 14, day 28 and day 365 in the American, African, and European cohorts, respectively. For all cohorts (USA, Switzerland, Gabon and Kenya), we have selected the AEs of grade 1, 2 or 3 (mild, moderate or severe, respectively) as published previously for our VSV-EBOVAC consortia partners (3, 12, 16, 26).

Adverse event terms were standardized across all four cohorts. For the USA cohort, the terms "tenderness" and "pain in extremity" and for Switzerland, Kenya, and Gabon, the term "pain at site" were considered "any local AE". The term pyrexia in the USA and "subjective fever" in Switzerland, Kenya and Gabon were considered "fever". AE terms reported by less than 5% of the participants were removed from the next analysis. The following is the final list of adverse events in the order of frequency: "any local AE", "headache", "fatigue", "myalgia", "fever", "chills", "arthralgia", "nausea", "arthritis". The incidence within each cohort is shown in Figure 1.

2.4 Gene expression profiling

The human transcriptomic profiles of the response to the rVSVΔG-ZEBOV-GP vaccine were evaluated by the quantitative multiplex platform RT-PCR, which performs amplification using a two-color ligation-dependent probe (dcRT-MLPA). PAXgene blood RNA tubes (PreAnalytiX, Hombrechtikon, Switzerland) with 2.5 ml venous blood were collected and stored at -80°C. RNA isolation was performed using the PAXgene blood miRNA kit (PreAnalytiX) according to the manufacturer's automated protocol, including on-column DNase digestion. RNA yield was quantified using an RNA Broad Range assay Kit (ThermoFisher) with a Qubit fluorometer (ThermoFisher, Wilmington, DE, USA). The dcRT-MLPA (MLPA) assay accounts for 144 genes of critical importance whose involvement in innate and adaptive immune responses (24) is documented and used to determine the gene expression profiles of people vaccinated with rVSVΔG-ZEBOV-GP. The gene expression values thus generated were normalized according to the expression of the housekeeping gene GAPDH and transformed to log2, whereas the quality control was performed as described in previous works (25, 26). The function removeBatchEffect from limma package in R was used to remove the batch effect of the dcRT-MLPA plaque within each cohort.

2.5 Statistical analysis

GAPDH-normalized log2-transformed gene expression levels at baseline for each cohort were used for integrative analysis. Gene expression comparisons were conducted between volunteers with and without adverse events within each cohort and when combining all the cohorts. The non-parametric Wilcoxon test with Benjamini-Hochberg correction for multiple testing was applied for statistical significance. An adjusted P-value (q-value) of less than 0.05 was set as the threshold for identifying significant genes for the comparison of groups with or without adverse events. Analyses were performed with R software (version 4.0.4).

2.6 Feature selection machine-learning-based approach

The expression of the 144 immune-related genes on Day 0 (before immunization) and the information of reactogenicity obtained after immunization, for the volunteers of each cohort, were used as input files. Then, our algorithm which is a robust ML-based feature prioritization tool fully described in Figure 2 was run.

To summarize, our method first performs feature selection using three different methods: Pearson's correlation, Kbest and Recursive Feature Elimination (RFE). After generating a list of features for each method, a unified list is produced by selecting features from the intersection of 2 out the 3 methods. From this list, our approach orders the list using the Mean Decrease Gini Index (MDGI) obtained with the function 'feature_importances_' from the model trained with the Random Forest algorithm implemented in scikit-learn as "RandomForestClassifier". The features with an importance value equal to 0 are removed. Finally, it assesses the discriminatory power of the selected features and determine their effectiveness in classifying the different groups by using models trained with various machine-learning algorithms such as: Support Vector Machine (SVM), k-Nearest Neighbors (kNN), Naive Bayes and AdaBoost Classifier. Thereafter, the tool generates as output a table with the values of F1-score, area under the curve (AUC), accuracy, and precision, obtained from each model with the selected features. Using the machine-learning-based approach, we assessed the importance of genes in classifying volunteer groups with or without the selected AEs (frequency > 5%), which are "any local AE", "headache", "fatigue", "myalgia", "fever", "chills", "arthralgia", "nausea", "arthritis". The Support Vector Machine (SVM), k-Nearest Neighbors (kNN), Naive Bayes, AdaBoost Classifier and Random Forest ML algorithms were trained with a 10-fold cross-validation classification method. All the analyses were performed in

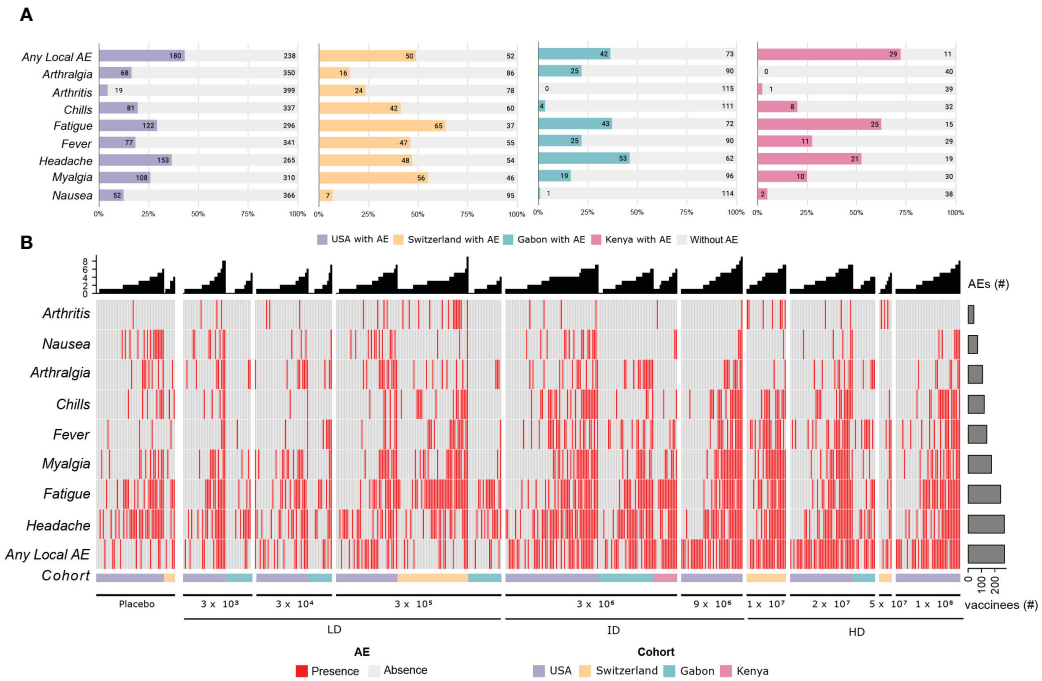


FIGURE 1
Adverse Events description for the 4 cohorts. **(A)** The stacked bar plots shows the absolute number and the frequency of the main adverse effects described in the first 14 days after vaccination with rVSVΔG-ZEBOV-GP, in the cohorts. The colored portion of each bar represents the number of participants who reported each of the adverse events, with the cohorts being represented by the colors purple (USA), yellow (Switzerland), green (Gabon), pink (Kenya) and light gray (no adverse events reported). **(B)** Heatmap showing the presence (red) and absence (light gray) of the most important adverse events. The most frequent AEs are shown at the bottom of the heatmap, and the columns are ordered per dose and cohort, as shown in the bottom annotation (USA, purple; Switzerland, yellow; Gabon, green and Kenya, pink). The number of AEs per participant is shown in the bar plot at the top, and the total number of reported adverse events is shown in the bar plot on the right.

Python (version 3.8.11). The library scikit-learn 0.24.2 was used for training the ML algorithms. All the hyperparameters were defined as default. A more detailed description of the ML-based feature prioritization tool can be found on the extended methods in the Supplementary Material.

2.7 Network construction

From the list of ranking of importance (MDGI), the top 50 features from the list were selected (Figure 3A) and their consistency across the four cohorts were evaluated. The genes with the same fold-change direction in 100% of the cohorts and shared by more than 50% of the cohorts (3 out 4, 3 out 3, 2 out 2) are kept (Figure 3B). Finally, we integrated the genes with the AEs in a network constructed using Gephi software (27) (Figure 3C).

3 Results

3.1 Reactogenicity was frequent but generally mild

The vaccine proved to be safe, even if associated with transient reactogenicity (11). We observed injection-site, systemic reactogenicity and medication use for 7 days after injection and at follow-up timepoints (days 14 and 28). We collected reactogenicity information from a total of 782 participants: 115 from the Swiss cohort (102 vaccinated and 13 placebo), 115 from Gabon, and 40 from Kenya. The remaining 512 participants were from the United States (418 vaccinated and 94 placebo). The participants included 488 males and 355 females, and the median age was 35 years (18-63), the sex and age median per cohort is shown in the Supplementary Tables 1 and 2, respectively.

Solicited and unsolicited adverse events were frequent. Majority of participants reported adverse effects in the first 14 days after vaccination, mostly mild and moderate. The side effects induced by rVSVΔG-ZEBOV-GP vaccination are self-limiting and relatively mild (Supplementary Figure 1). The most frequent side effects observed were any local AE (53.25%), fatigue (49.17%), headache (46.55%), myalgia (31.27%), fever (28.90%), chills (21.50%), arthralgia (13.67%), nausea (6.30%) and arthritis (6.39%).

Several people reported AE grade 1, including placebo recipients. Of the 782 participants, 638 (81.6%) have reported at least one adverse event, with the majority being mild or moderate. At least one adverse event (grade 1, 2 or 3) was reported by 328 of 418 (78.47%) vaccinees from the US cohort and by 96 of 102 (94.1%) from the Swiss cohort. In African cohorts, 97 of 115 (84.3%) and 37 of 40 (92.5%) participants from Gabon and Kenya, respectively, reported adverse events. Whilst among the placebo recipients, 69 of 98 (70.4%) and 11 of 13 (84.6%) participants have reported adverse events in the US and the Swiss cohorts, respectively. Grade 3 symptoms were

10.3389/fimmu.2023.1259197

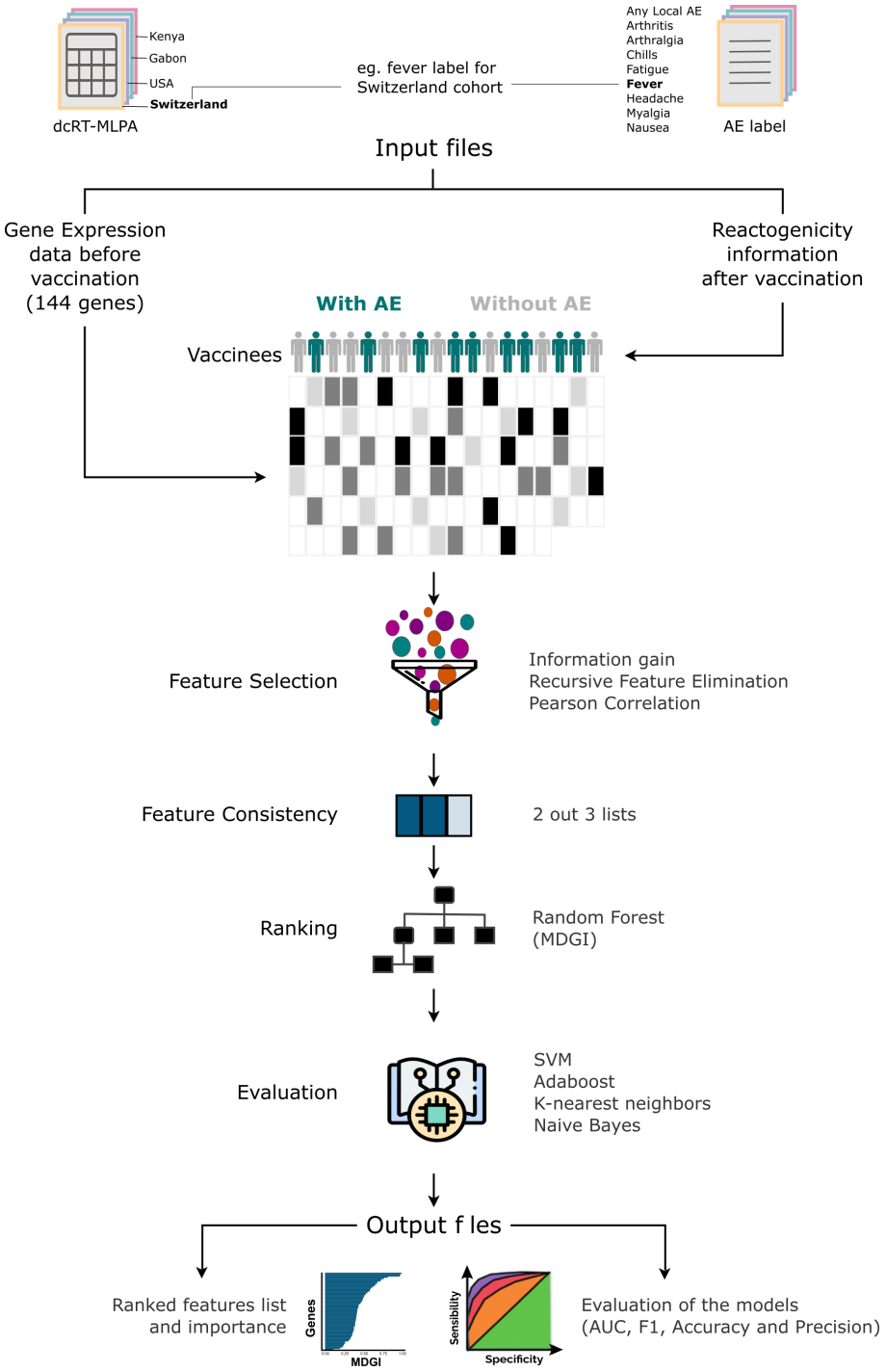


FIGURE 2
Machine-learning-based Algorithm description performed for each adverse effect per cohort. The score measures the ability of a feature to distinguish the outcome groups. First, considering that the quality of the predictive models depends on the quality of features used, the method performs the selection of features. The selection is based on the combination of 3 different methods: Pearson's correlation, Kbest and Recursive Feature Elimination (RFE). After generating a list of features for each method, a unique list is generated by selecting features from the intersection of 2 out of 3 methods. From this list, the method generates the ranking importance obtained from the Random Forest model and removes features with an importance value equal to 0. Subsequently, it evaluates the quality of gene list in discriminating the adverse events (AEs) classes in 4 machine learning models trained with the algorithms Support Vector Machine (SVM), k-Nearest Neighbors (kNN), Naive Bayes and AdaBoost Classifier. Thereafter, the tool generates a table with the F1-score, Area Under the curve (AUC), the accuracy and precision values obtained from each model with the selected features, and the median and harmonic mean calculated from all methods and metrics.

10.3389/fimmu.2023.1259197

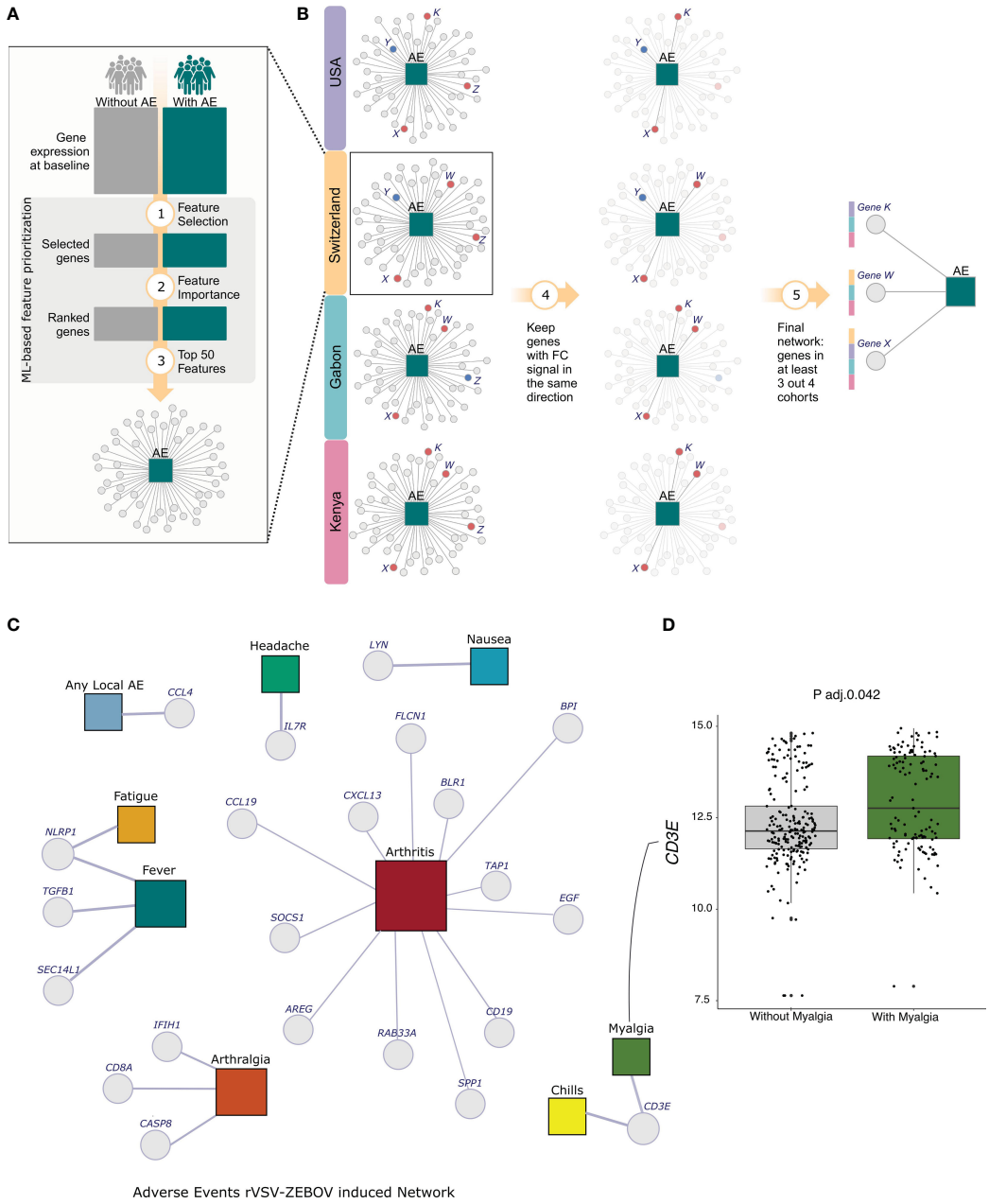


FIGURE 3
Analysis Scheme and Network of selected genes for all 9 Adverse Events. **(A)** The dcRT-MLPA with the expression of 144 genes per cohort was used to select the best features for classifying participants with or without each one of the adverse events individually. The machine-learning-based method was used for feature selection and ordering. **(B)** Among the selected ordered features, the top 50 from each cohort were chosen, and those with consistent fold-change signs across all four cohorts and shared by more than 50% of cohorts were kept. **(C)** Network Adverse Events description for the 4 cohorts. The adverse events are represented by colored squares, and the genes are represented by Light-grey circles. The squares representations are as follows: Light-blue - any local AE, Green - headache, Blue - nausea, Dark-Red - arthritis, Dark-green - myalgia, Light-yellow - chills, Orange - arthralgia, Surfie-Green - fever and Dark-yellow - fatigue. **(D)** The boxplot shows the log2 transformed expression of the gene EGF in Arthritis and non-arthritis participants.

reported by 4 of 40 (10%) vaccinees from Kenya, 23 of 418 (5.5%) from the USA, and 11 of 102 (10.8%) from Switzerland; none were reported in Gabon. Arthritis was reported in 24 participants (≃ 23%) from the Swiss cohorts, 1 (2.5%) from Kenya and 19 (≃ 4.5%) from the United States.

The complete information of adverse events, before the nomenclature combination and filtering, is shown in Supplementary Figure 1. The onset differs mainly for the grade 3 AEs (Supplementary Figure 1B). In general, the percentage of vaccinees reporting AE grade 2 or 3 increases in higher doses (Supplementary Figure 1C).

3.2 Associations between gene expression and adverse events

We collected dcRT-MLPA and reactogenicity data for a total of 343 vaccinees. Supplementary Figure 2 describes the number and proportion of volunteers who have dcRT-MLPA data (Supplementary Figure 2A), as well as the number of vaccinees who had reported the presence or absence of each of the adverse events per cohort (Supplementary Figure 2B). The percentage of participants with dcRT-MLPA available per cohort and the comparison of the ranking and frequency of adverse effects between all participants with reactogenicity data and the participants with available dcRT-MLPA data are shown in the Supplementary Tables 3 and 4, respectively.

Considering only participants with available dcRT-MLPA data, local AE is the most frequent (48.4%), followed by fatigue (48.1%), headache (47.5%), myalgia (36.7%), fever (34.4%), chills (29.9%), arthralgia (17.2%), arthritis (8.2%) and nausea (7%).

3.3 Feature selection per cohort and adverse event using our machine-learning-based approach

We integrated the reactogenicity data with the available expression data to understand the propensity of populations to AEs induced by vaccination with rVSVΔG-ZEBOV-GP. For this, we ran our ML-based feature prioritization tool described in Figure 2.

3.4 Multicohort baseline transcriptional-reactogenicity network

We kept the top fifty genes selected using our machine-learning-based approach (Figure 3A), keeping only the genes with the same fold-change signal across all cohorts. Next, we filtered out those shared by more than 50% of the cohorts (3 out 4, 3 out 3, 2 out 2) approach (Figure 3B). Finally, we integrated the genes with the AEs in a network constructed using Gephi software (27). The size of the nodes represents the degree, which denotes the number of connections in the network (Figure 3C).

After the integration, we selected a total of 22 genes for 9 adverse events. Interestingly, for six adverse events, only one gene was selected (Figure 3C). We selected the genes *CCL4* and *IL7R*, which are regulatory T-cell-associated markers, for local AE and headache, respectively. Both genes exhibited an increased expression in volunteers with adverse events, though it was not statistically significant. Nausea was associated with the gene *LYN*, which encodes a tyrosine kinase. For fatigue, we only selected the gene *NLRP1*, known to be a key mediator of programmed cell death. We selected the same gene for fever, but in combination with the genes *TGFB* and *SEC14L1*. Although the expression of the *NLRP1* gene increased in participants with both adverse events, the levels of this gene were significant only in the comparison of participants with or without fatigue (Adj. p-value = 0.0022). Similarly, we selected the gene *NLRP1* for two adverse events, and the gene *CD3E*, a T-cell marker, for chills and myalgia. However, only myalgia participants showed a significant increase in *CD3E* gene (Figure 3D, Adj. p-value 0.0409).

In the classification of participants with or without arthralgia, three genes were selected, namely *CASP8* (apoptosis-related genes), *CD8A* (Marker of lymphocyte subsets) and *IFIH1* (innate immune response related gene). This result was consistent across three cohorts since no arthralgia cases were reported in the Kenya cohort.

We identified a total of 12 genes that are associated with arthritis classification. It's important to note that the relatively high number of genes may be attributed to the fact that we only considered cohorts from Switzerland and USA since Gabon and Kenya had a lack of arthritis cases.

The arthritis-associated genes are *BLR*, G protein-coupled receptor; *RAB33A*, Small GTPases - (Rho) GTPase activating proteins; the chemokine gene *CCL19*; the cell growth associated genes *AREG* and *EGF*; the B cell marker gene *CD19*; the tumor suppressor gene *FLCN1*; *BPI*, which is associated with anti-microbial activity; *SPP1*, an epithelial-mesenchymal transition and Inflammation marker; and innate immune responses related genes, *CXCL13*, *SOCS1* and *TAP1*—the first is a myeloid associated gene, whilst the last two are IFN signaling genes.

Among the arthritis-associated genes *AREG*, *BPI*, *EGF*, *FLCN1*, *RAB33A*, *SOCS1*, *SPP1* and *TAP1* have a significant difference between arthritis and non-arthritis volunteers. Among them, *AREG*, *BPI* and *TAP1* genes showed an increased expression in arthritis participants.

4 Discussion

Although many studies describe the reactogenicity of the vaccine, only few define reactogenicity signatures. The majority of them focus on cytokines, as it has long been assumed that vaccine reactogenicity is reflected in innate responses and inflammation (28). Moreover, studies describing reactogenicity signatures using expression data are even rarer (25). To the best of our knowledge, this is the first method that integrates baseline gene expression data with several vaccine-induced reactogenicity across 4 cohorts.

The most onerous challenge in baseline data analysis is dealing with batch effects. In multi-cohort studies, data variability can be caused by inter-subject variation, technical discrepancies from sample collection or/and data acquisition and processing. In addition, the subset of participants with available expression data may not adequately represent the overall population, which raises concerns about generalization. Much like the expression data, the number of adverse events also varies within the cohort. Participants from Switzerland reported higher rates of adverse events in comparison with the other sites. Several factors may contribute to this disparity, including differences in reporting practices and clinical investigation approaches. These events, often of mild or moderate severity and not easily attributed to vaccination, may go unreported. Additionally, variations in host factors that regulate inflammatory and immune responses, such as age, sex, fitness level, physical activity, body-mass index, baseline immunity, and human leukocyte antigen types, likely differ between study populations (11, 26). Associations between vaccine dose, innate responses, and

reactogenicity was already shown by our collaborators (26). The frequency of self-reported, vaccine-induced AEs is notably lower in African settings. This same pattern was reported by Muyanja and colleagues (2014), when immunizing volunteers from Uganda and Switzerland with the yellow fever vaccine 17D (YF-17D) (29). Although self-reported vaccine-induced adverse events are notably less frequent in African settings, a study from Huttner and collaborators (2017) reveals that this reduced incidence does not correlate with weaker innate responses or higher baseline concentrations of anti-inflammatory cytokines like IL-10. Innate responses were found to be similar between European and African volunteers (26). This finding emphasizes the importance of assessing vaccine safety in the settings where they will be used (26). This is the reason we have analyzed each cohort individually before integrating the results for consistency.

Here, a baseline consistent vaccine reactogenicity signature was found across 4 different cohorts from 3 different continents. The signatures came from the gene expression analysis of 144 genes at baseline (before injection) from volunteers who had received recombinant vesicular stomatitis virus-vectored Zaire Ebola vaccine. Following which we were able to associate 22 genes with the following adverse events: any local AE, fatigue, headache, myalgia, fever, chills, arthralgia, nausea, and arthritis.

Interestingly, regulatory T-cell markers were associated with the most frequent adverse events. The genes *CCL4* and *IL7R* were associated with local AE and headache, respectively. *CCL4* and other cytokines, such as *CCL2, CCL5 and CCL8*, have been associated with the recruitment of neutrophils, eosinophils, more monocytes and DCs to the injection site in response to the activation of myeloid cells after MF59 adjuvant administration (30). The same marker levels in lymph nodes and in the muscle at the injection site has strong positive correlation in a study that evaluated mice immunized with four licensed vaccines (31). Similarly, after injection of mRNA vaccine, a strong production of chemokines (including CCL4) at the site of injection was observed by Kowalczyk and colleagues (2014) (32). Hence, the high volume of cells in the injection site can be related to local reactions.

Headache was the most frequent adverse event in a Phase Ib study that evaluated how the blockade of *IL-7* would affect immune cells and relevant clinical responses in patients with type 1 diabetes (33). Furthermore, headache was reported by 5 of the 18 healthy volunteers in a study that investigated the safety of GSK2618960, an IL-7 receptor-α subunit (CD127) monoclonal antibody (34), suggesting that dysregulation in the *IL-7* levels could be associated with headache.

The T-cell marker gene CD3E was associated with myalgia and chills in our analysis. Curiously, chills was one of the most common adverse events in participants with cutaneous T-cell lymphoma who received the Resimmune, which is a second-generation recombinant immunotoxin composed of the catalytic and translocation domains of diphtheria toxin fused to two single-chain antibody fragments reactive with the extracellular domain of CD3ϵ (35). While no direct association between CD3E gene and myalgia has been found, there is a strong indication that T cells play a key role not only in the induction but also in the suppression of pain (36, 37).

The effectiveness of rVSVΔG-ZEBOV-GP (Ervebo®) has been demonstrated in clinical studies conducted on 15,399 adults in Europe (11), Africa (AGNANDJI) (12, 13) and North America (16). In these populations, the vaccine proved to be safe and induced higher antibody titers sustained for at least 2 years in both European and African vaccines (3), but it showed transient reactogenicity (11). For this and other vaccines with similar adverse reactions, such as the ones against COVID-19 (38–41), the reactogenicity did not prevent the approval of this vaccine, since the benefits highly overcome the risks (15). Nevertheless, more studies investigating the baseline signature of vaccine-induced reactogenicity are necessary for paving the way towards precision vaccinology. This will enable us to identify who will benefit the most and who will be more vulnerable to post-immunization adverse reactions.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

Ethics statement

The trial protocols were reviewed and approved by the WHO's Ethics Committee as well as by local ethics committees (USA trial: the Chesapeake Institutional Review Boards (Columbia, MD, USA) and the Crescent City Institutional Review Board (New Orleans, LA, USA); Geneva trial: the Geneva Cantonal Ethics Commission and the Swiss Agency for Therapeutic Products (Swissmedic); Lambaréné trial: the Scientific Review Committee of Centre de Recherches Médicales de Lambaréné (CERMEL), the Institutional Ethics Committee of CERMEL, the National Ethics Committee of Gabon, and the Institutional Ethics Committee of the Universitätsklinikum Tübingen; Kilifi trial: Kilifi Ethics Committee). The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

PG-D: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. TD: Data curation, Formal Analysis, Methodology, Visualization, Writing – review & editing. LM: Formal Analysis, Software, Validation, Visualization, Writing – review & editing. IM: Formal Analysis, Software, Validation, Visualization, Writing – review & editing. AN: Formal Analysis, Software, Validation, Visualization, Writing – review & editing. AC: Data curation, Formal Analysis, Visualization, Writing – review & editing. SS: Data curation, Formal Analysis, Visualization, Writing – original draft, Writing – review & editing. AH: Conceptualization, Investigation, Writing – review &

editing. DF: Investigation, Writing – review & editing, Resources. EV: Data curation, Writing – review & editing. MH: Data curation, Writing – review & editing. TO: Data curation, Writing – review & editing. FS: Data curation, Writing – review & editing. PM-M: Data curation, Writing – review & editing. AH: Data curation, Investigation, Writing – review & editing. C-AS: Funding acquisition, Project administration, Supervision, Writing – review & editing. DM: Funding acquisition, Project administration, Supervision, Writing – review & editing. HN: Conceptualization, Data curation, Formal Analysis, Funding acquisition, Investigation, Methodology, Resources, Software, Supervision, Validation, Visualization, Writing – original draft.

Group members of the VSV-EBOVAC consortium

Selidji T. Agnandji, Rafi Ahmed, Jenna Anderson, Floriane Auderset, Philip Bejon, Luisa Borgianni, Jessica Brosnahan, Annalisa Ciabattini, Olivier Engler, Mariëlle C. Haks, Ali M. Harandi, Donald Gray Heppner, Alice Gerlini, Angela Huttner, Peter G. Kremsner, Donata Medaglini, Thomas P. Monath, Francis M. Ndungu, Patricia Njuguna, Tom H. M. Ottenhoff, David Pejoski, Mark Page, Gianni Pozzi, Francesco Santoro, and Claire-Anne Siegrist.

Group members of the VSV-EBOPLUS consortium

Selidji T. Agnandji, Luisa Borgianni, Annalisa Ciabattini, Sheri Dubey, Michael Eichberg, Olivier Engler, Alice Gerlini, Patricia Conceição Gonzalez Dias Carvalho, Mariëlle C. Haks, Ali M. Harandi, Angela Huttner, Peter G. Kremsner, Kabwende Lumeka, Donata Medaglini, Helder I. Nakaya, Sravya S. Nakka, Essone P. Ndong, Tom H. M. Ottenhoff, Gianni Pozzi, Sylvia Rothenberger, Francesco Santoro, Claire-Anne Siegrist, Suzanne van Veen, Eleonora Vianello.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by grants from the Innovative Medicines Initiative 2 Joint Undertaking under the VSV-EBOVAC (grant number 115842) and VSV-EBOPLUS (grant number 116068) projects within the Innovative Medicines Initiative Ebola+ program and also by FAPESP (grant number 18/14933-2). Development of the dcRT-MLPA probe sets was funded by GC6-74 (grant number 37772) and ADITEC (grant number 280873). Conduction of the North American trial was funded in part with Federal funds from the Department of Health and Human Services; Office of the Assistant Secretary for Preparedness and Response; Biomedical Advanced Research and Development Authority, under contract number HHSO100201500002C.

Acknowledgments

The authors thank all participants in the cohort studies.

Conflict of interest

Author IM was employed by the company Microbiotec Srl.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2023.1259197/full#supplementary-material

References

1. To KKW, Chan JFW, Tsang AKL, Cheng VCC, Yuen K-Y. Ebola virus disease: a highly fatal infectious disease reemerging in West Africa. *Microbes Infect* (2015) 17 (2):84–97. doi: 10.1016/j.micinf.2014.11.007

2. Garbutt M, Liebscher R, Wahl-Jensen V, Jones S, Möller P, Wagner R, et al. Properties of replication-competent vesicular stomatitis virus vectors expressing glycoproteins of filoviruses and arenaviruses. *J Virol* (2004) 78(10):5458–65. doi: 10.1128/JVI.78.10.5458-5465.2004

3. Huttner A, Agnandji ST, Combescure C, Fernandes JF, Bache EB, Kabwende L, et al. Determinants of antibody persistence across doses and continents after single-dose rVSV-ZEBOV vaccination for Ebola virus disease: an observational cohort study. *Lancet Infect Dis* (2018) 18(7):738–48. doi: 10.1016/S1473-3099(18)30165-8

4. Pulendran B, Oh JZ, Nakaya HI, Ravindran R, Kazmin DA. Immunity to viruses: learning from successful human vaccines. *Immunol Rev* (2013) 255(1):243–55. doi: 10.1111/imr.12099

5. Geisbert TW, Feldmann H. Recombinant vesicular stomatitis virus-based vaccines against Ebola and Marburg virus infections. *J Infect Dis* (2011) 204 Suppl 3: S1075–81. doi: 10.1093/infdis/jir349

6. Roberts A, Buonocore L, Price R, Forman J, Rose JK. Attenuated vesicular stomatitis viruses as vaccine vectors. *J Virol* (1999) 73(5):3723–32. doi: 10.1128/JVI.73.5.3723-3732.1999

7. Haglund K, Leiner I, Kerksiek K, Buonocore L, Pamer E, Rose JK. High-level primary CD8(+) T-cell response to human immunodeficiency virus type 1 gag and env generated by vaccination with recombinant vesicular stomatitis viruses. *J Virol* (2002) 76(6):2730–8. doi: 10.1128/JVI.76.6.2730-2738.2002

8. Taddeo A, Veiga IB, Devisme C, Boss R, Plattet P, Weigang S, et al. Optimized intramuscular immunization with VSV-vectored spike protein triggers a superior immune response to SARS-CoV-2. *NPJ Vaccines* (2022) 7(1):82. doi: 10.1038/s41541-022-00508-7

9. Poetsch JH, Dahlke C, Zinser ME, Kasonta R, Lunemann S, Rechtien A, et al. Detectable vesicular stomatitis virus (VSV)-specific humoral and cellular immune responses following VSV-ebola virus vaccination in humans. *J Infect Dis* (2019) 219 (4):556–61. doi: 10.1093/infdis/jiy565

10. Fathi A, Dahlke C, Addo MM. Recombinant vesicular stomatitis virus vector vaccines for WHO blueprint priority pathogens. *Hum Vaccin Immunother* (2019) 15 (10):2269–85. doi: 10.1080/21645515.2019.1649532

11. Huttner A, Dayer J-A, Yerly S, Combescure C, Auderset F, Desmeules J, et al. The effect of dose on the safety and immunogenicity of the VSV Ebola candidate vaccine: a randomised double-blind, placebo-controlled phase 1/2 trial. *Lancet Infect Dis* (2015) 15(10):1156–66. doi: 10.1016/S1473-3099(15)00154-1

12. Agnandji ST, Huttner A, Zinser ME, Njuguna P, Dahlke C, Fernandes JF, et al. Phase 1 trials of rVSV Ebola vaccine in Africa and Europe. *N Engl J Med* (2016) 374 (17):1647–60. doi: 10.1056/NEJMoa1502924

13. Regules JA, Beigel JH, Paolino KM, Voell J, Castellano AR, Hu Z, et al. A recombinant vesicular stomatitis virus ebola vaccine. *N Engl J Med* (2017) 376(4):330–41. doi: 10.1056/NEJMoa1414216

14. Henao-Restrepo AM, Longini IM, Egger M, Dean NE, Edmunds WJ, Camacho A, et al. Efficacy and effectiveness of an rVSV-vectored vaccine expressing Ebola surface glycoprotein: interim results from the Guinea ring vaccination cluster-randomised trial. *Lancet* (2015) 386(9996):857–66. doi: 10.1016/S0140-6736(15)61117-5

15. Bache BE, Grobusch MP, Agnandji ST. Safety, immunogenicity and risk-benefit analysis of rVSV-ΔG-ZEBOV-GP (V920) Ebola vaccine in Phase I-III clinical trials across regions. *Future Microbiol* (2020) 15:85–106. doi: 10.2217/fmb-2019-0237

16. Heppner DG, Kemp TL, Martin BK, Ramsey WJ, Nichols R, Dasen EJ, et al. Safety and immunogenicity of the rVSVΔG-ZEBOV-GP Ebola virus vaccine candidate in healthy adults: a phase 1b randomised, multicentre, double-blind, placebo-controlled, dose-response study. *Lancet Infect Dis* (2017) 17(8):854–66. doi: 10.1016/S1473-3099(17)30313-4

17. Pellegrino P, Falvella FS, Perrone V, Carnovale C, Brusadelli T, Pozzi M, et al. The first steps towards the era of personalised vaccinology: predicting adverse reactions. *Pharmacogenomics J* (2015) 15(3):284–7. doi: 10.1038/tpj.2014.57

18. Tsang JS, Schwartzberg PL, Kotliarov Y, Biancotto A, Xie Z, Germain RN, et al. Global analyses of human immune variation reveal baseline predictors of postvaccination responses. *Cell* (2014) 157(2):499–513. doi: 10.1016/j.cell.2014.03.031

19. HIPC-CHI Signatures Project Team and HIPC-I Consortium. Multicohort analysis reveals baseline transcriptional predictors of influenza vaccination responses. *Sci Immunol* (2017) 2(14). doi: 10.1126/sciimmunol.aal4656

20. Bartholomeus E, De Neuter N, Meysman P, Suls A, Keersmaekers N, Elias G, et al. Transcriptome profiling in blood before and after hepatitis B vaccination shows significant differences in gene expression between responders and non-responders. *Vaccine* (2018) 36(42):6282–9. doi: 10.1016/j.vaccine.2018.09.001

21. Warimwe GM, Fletcher HA, Olotu A, Agnandji ST, Hill AVS, Marsh K, et al. Peripheral blood monocyte-to-lymphocyte ratio at study enrollment predicts efficacy of the RTS,S malaria vaccine: analysis of pooled phase II clinical trial data. *BMC Med* (2013) 11:184. doi: 10.1186/1741-7015-11-184

22. O'Connor D, Pinto MV, Sheerin D, Tomic A, Drury RE, Channon-Wells S, et al. Gene expression profiling reveals insights into infant immunological and febrile responses to group B meningococcal vaccine. *Mol Syst Biol* (2020) 16(11):e9888. doi: 10.15252/msb.20209888

23. Syenina A, Gan ES, Toh JZN, de Alwis R, Lin LZ, Tham CYL, et al. Adverse effects following anti-COVID-19 vaccination with mRNA-based BNT162b2 are alleviated by altering the route of administration and correlate with baseline enrichment of T and NK cell genes. *PloS Biol* (2022) 20(5):e3001643. doi: 10.1371/journal.pbio.3001643

24. Haks MC, Goeman JJ, Magis-Escurra C, Ottenhoff THM. Focused human gene expression profiling using dual-color reverse transcriptase multiplex ligation-dependent probe amplification. *Vaccine* (2015) 33(40):5282–8. doi: 10.1016/j.vaccine.2015.04.054

25. Vianello E, Gonzalez-Dias P, van Veen S, Engele CG, Quinten E, Monath TP, et al. Transcriptomic signatures induced by the Ebola virus vaccine rVSVΔG-ZEBOV-GP in adult cohorts in Europe, Africa, and North America: a molecular biomarker study. *Lancet Microbe* (2022) 3(2):e113–23. doi: 10.1016/S2666-5247(21)00235-4

26. Huttner A, Combescure C, Grillet S, Haks MC, Quinten E, Modoux C, et al. A dose-dependent plasma signature of the safety and immunogenicity of the rVSV-Ebola vaccine in Europe and Africa. *Sci Transl Med* (2017) 9(385). doi: 10.1126/scitranslmed.aaj1701

27. Bastian M, Heymann S, Jacomy M. Gephi: an open source software for exploring and manipulating networks. *Third Int AAAI Conf Weblogs Soc Media* (2009) 3(1):361–2. doi: 10.1609/icwsm.v3i1.13937

28. Hervé C, Laupèze B, Del Giudice G, Didierlaurent AM, Tavares Da Silva F. The how's and what's of vaccine reactogenicity. *NPJ Vaccines* (2019) 4:39. doi: 10.1038/s41541-019-0132-6

29. Muyanja E, Ssemaganda A, Ngauv P, Cubas R, Perrin H, Srinivasan D, et al. Immune activation alters cellular and humoral responses to yellow fever 17D vaccine. *J Clin Invest* (2014) 124(7):3147–58. doi: 10.1172/JCI75429

30. Pulendran B, Arunachalam P S, O'Hagan DT. Emerging concepts in the science of vaccine adjuvants. *Nat Rev Drug Discov* (2021) 20(6):454–75. doi: 10.1038/s41573-021-00163-y

31. McKay PF, Cizmeci D, Aldon Y, Maertzdorf J, Weiner J, Kaufmann SH, et al. Identification of potential biomarkers of vaccine inflammation in mice. *eLife* (2019) 8. doi: 10.7554/eLife.46149

32. Kowalczyk A, Döner F, Jasny E, Noth J, Scheel B, Koch SD, et al. Self-adjuvanted RNActive® vaccine induces local immune responses at the injection site leading to potent adaptive immunity in mice and humans. *J Immunother Cancer* (2014) 2(Suppl 3):P172. doi: 10.1186/2051-1426-2-S3-P172

33. Herold KC, Bucktrout SL, Wang X, Bode BW, Gitelman SE, Gottlieb PA, et al. Immunomodulatory activity of humanized anti–IL-7R monoclonal antibody RN168 in subjects with type 1 diabetes. *JCI Insight* (2019) 4(24):e12605. doi: 10.1172/jci.insight.126054

34. Ellis J, van Maurik A, Fortunato L, Gisbert S, Chen K, Schwartz A, et al. Anti-IL-7 receptor α monoclonal antibody (GSK2618960) in healthy subjects - a randomized, double-blind, placebo-controlled study. *Br J Clin Pharmacol* (2019) 85(2):304–15. doi: 10.1111/bcp.13748

35. Frankel AE, Woo JH, Ahn C, Foss FM, Duvic M, Neville PH, et al. Resimmune, an anti-CD3ϵ recombinant immunotoxin, induces durable remissions in patients with cutaneous T-cell lymphoma. *Haematologica* (2015) 100(6):794–800. doi: 10.3324/haematol.2015.123711

36. Kavelaars A, Heijnen CJ. T cells as guardians of pain resolution. *Trends Mol Med* (2021) 27(4):302–13. doi: 10.1016/j.molmed.2020.12.007

37. Deyhle MR, Hyldahl RD. The role of T lymphocytes in skeletal muscle repair from traumatic and contraction-induced injury. *Front Physiol* (2018) 9:768. doi: 10.3389/fphys.2018.00768

38. Polack FP, Thomas SJ, Kitchin N, Absalon J, Gurtman A, Lockhart S, et al. Safety and efficacy of the BNT162b2 mRNA Covid-19 vaccine. *N Engl J Med* (2020) 383 (27):2603–15. doi: 10.1056/NEJMoa2034577

39. Ramasamy MN, Minassian AM, Ewer KJ, Flaxman AL, Folegatti PM, Owens DR, et al. Safety and immunogenicity of ChAdOx1 nCoV-19 vaccine administered in a prime-boost regimen in young and old adults (COV002): a single-blind, randomised, controlled, phase 2/3 trial. *Lancet* (2021) 396(10267):1979–93. doi: 10.1016/S0140-6736(20)32466-1

40. Baden LR, El Sahly HM, Essink B, Kotloff K, Frey S, Novak R, et al. Efficacy and safety of the mRNA-1273 SARS-CoV-2 vaccine. *N Engl J Med* (2021) 384(5):403–16. doi: 10.1056/NEJMoa2035389

41. Arunachalam PS, Scott MKD, Hagan T, Li C, Feng Y, Wimmers F, et al. Systems vaccinology of the BNT162b2 mRNA vaccine in humans. *Nature* (2021) 596 (7872):410–6. doi: 10.1038/s41586-021-03791-x

frontiers | Frontiers in Immunology

TYPE Original Research
PUBLISHED 28 November 2023
DOI 10.3389/fimmu.2023.1284366

OPEN ACCESS

EDITED BY
Joe Hou,
Fred Hutchinson Cancer Center,
United States

REVIEWED BY
Anuj Kumar,
ICMR-National Institute of Cancer Prevention and Research, India
Junjie Yue,
Beijing Institute of Biotechnology, China

*CORRESPONDENCE
Abdulaziz Alamri
abalamri@ksu.edu.sa

†These authors have contributed equally to this work

RECEIVED 04 September 2023
ACCEPTED 01 November 2023
PUBLISHED 28 November 2023

CITATION
Ali SL, Ali A, Alamri A, Baiduissenova A, Dusmagambetov M and Abduldayeva A (2023) Genomic annotation for vaccine target identification and immunoinformatics-guided multi-epitope-based vaccine design against Songling virus through screening its whole genome encoded proteins. *Front. Immunol.* 14:1284366. doi: 10.3389/fimmu.2023.1284366

Genomic annotation for vaccine target identification and immunoinformatics-guided multi-epitope-based vaccine design against Songling virus through screening its whole genome encoded proteins

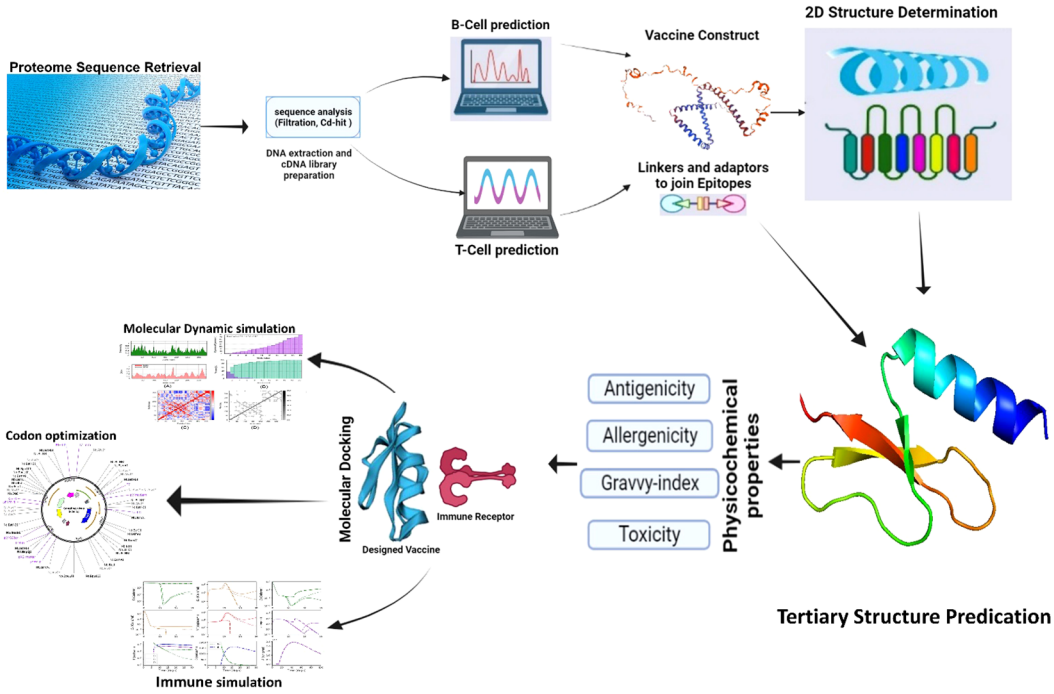
S. Luqman Ali[1†], Awais Ali[1†], Abdulaziz Alamri[2*], Aliya Baiduissenova[3], Marat Dusmagambetov[3] and Aigul Abduldayeva[4]

[1]Department of Biochemistry, Abdul Wali Khan University Mardan, Mardan, Pakistan, [2]Department of Biochemistry, College of Science, King Saud University, Riyadh, Saudi Arabia, [3]Department of Microbiology and Virology, Astana Medical University, Astana, Kazakhstan, [4]Preventive Medicine, Astana Medical University, Astana, Kazakhstan

Songling virus (SGLV), a newly discovered tick-borne orthonairovirus, was recently identified in human spleen tissue. It exhibits cytopathic effects in human hepatoma cells and is associated with clinical symptoms including headache, fever, depression, fatigue, and dizziness, but no treatments or vaccines exist for this pathogenic virus. In the current study, immunoinformatics techniques were employed to identify potential vaccine targets within SGLV by comprehensively analyzing SGLV proteins. Four proteins were chosen based on specific thresholds to identify B-cell and T-cell epitopes, validated through IFN-γ epitopes. Six overlap MHC-I, MHC-II, and B cell epitopes were chosen to design a comprehensive vaccine candidate, ensuring 100% global coverage. These structures were paired with different adjuvants for broader protection against international strains. Vaccine constructions' 3D models were high-quality and validated by structural analysis. After molecular docking, SGLV-V4 was selected for further research due to its lowest binding energy (-66.26 kcal/mol) and its suitable immunological and physiochemical properties. The vaccine gene is expressed significantly in *E. coli bacteria* through *in silico* cloning. Immunological research and MD simulations supported its molecular stability and robust immune response within the host cell. These findings can potentially be used in designing safer and more effective experimental SGLV-V4 vaccines.

KEYWORDS

immunoinformatics, *Songling virus*, reverse vaccinology, molecular docking, MD simulation



GRAPHICAL ABSTRACT

1 Introduction

The recent discovered *Songling virus* (SGLV) in China marked a significant moment in pathogenic viruses (1). Its genomic configuration exhibited profound structural homologies with reputable orthonairoviruses, spanning sequence similarities from 46.5% to 65.7%. Phylogenetic analyses situated SGLV distinctly within the Tamdy orthonairovirus group, encapsulating its place within the broader framework of the Nairoviridae family. Microscopic scrutiny validated SGLV's morphological congruence with the hallmark attributes of orthonairoviruses (2). Notably, it's essential to mention that SGLV is a single-stranded RNA (ssRNA) virus, contributing to its classification within this viral group.

Functionally, isolated SGLV strains sourced from patients exhibited the capacity to induce prominent cytopathic effects in human hepatoma cells, accentuating its potential for pathogenesis. Between 2017 and 2018, SGLV's impact on human health materialized, materializing as symptoms encompassing headaches, fever, depression, fatigue, and dizziness. Serological investigations illuminated a pivotal facet: a significant 69% of patients exhibited the development of virus-specific antibody responses during the acute phase (3).

Remarkably, the absence of discernible SGLV viral RNA and the conspicuous scarcity of specific antibodies within healthy cohorts underscored its nuanced selectivity in its interaction with human physiology. Beyond human hosts, SGLV found ecological footing within ticks such as Ixodes crenulatus, Haemaphysalis longicornis, Haemaphysalis concinna, and Ixodes persulcatus within the northeastern precincts of China. Significantly, the viral L segments of SGLV came to the fore, manifesting in 2.2% of spleen samples from great gerbils. BLASTn alignments divulged a compelling narrative of genetic resonance, as the SGLV in great gerbils demonstrated a remarkable alignment of 93.7% (236/252 nucleotides) and 94.0% (78/83 amino acids) with its counterpart detected in human patients with a history of tick encounters within the confines of northeastern China (1). This multifaceted interplay of genetics, ecology, and clinical ramifications has woven a comprehensive tapestry, enriching our comprehension of SGLV's influence on human health within the geographical landscape of northeastern China (1, 4).

Reverse vaccinology is a cutting-edge strategy that has been widely applied to the introduction of new vaccinations. The strategy aims to combine immunogenomics and immunogenetics with bioinformatics for the development of novel vaccine targets (5). With the recent advancements in genome or protein sequence databases, this rapid *in silico* method has gained significant appeal (6). An innovative vaccine fuses CTL and HTL segments with specialized linkers and adjuvants, showcasing high antigenicity, non-allergenicity, and stability. Molecular docking finds strong binding energy to TLRs, promising robust immunogenic responses. Immune simulation employed to simulates a natural immune response, where the top candidate activates essential immune components (IgG, IgM, T-cells, B-cells, and cytokines), promising protection against the Songling virus (7). Further investigations, including molecular dynamics and computational cloning for efficient E. coli expression, solidify the vaccine's prospects. Vaccine may recognize and boost immunity against infection in the body. Therefore, predicting allergenicity is a crucial stage in the creation of a neuropeptide vaccine. Immunoinformatics techniques and tools were utilized to design

a non-allergic, immunogenic, and thermostable recombinant vaccine against the *Songling virus*, and we expect wet lab researchers to confirm our prediction.

2 Methodology

The systematic methodology employed in this study to design a multi-epitope vaccine construct targeting the *Songling virus* (SGLV) (Figure 1).

2.1 Protein sequence retrieval and filtration

Songling virus proteome data was downloaded from the National Centre for Biotechnology Information (NCBI) database under reference taxonomy ID: 2795181, and sequences were verified from Virus Pathogen Resource (ViPR) (8, 9). The protein sequence data for SGLV was in FASTA format and submitted to NCBI on May 6, 2023. Different parameters of CD-hit were used for obtaining 85% sequence similarity and removing redundancy to acquire non-paralogous sequences of proteins (10). BLASTp of

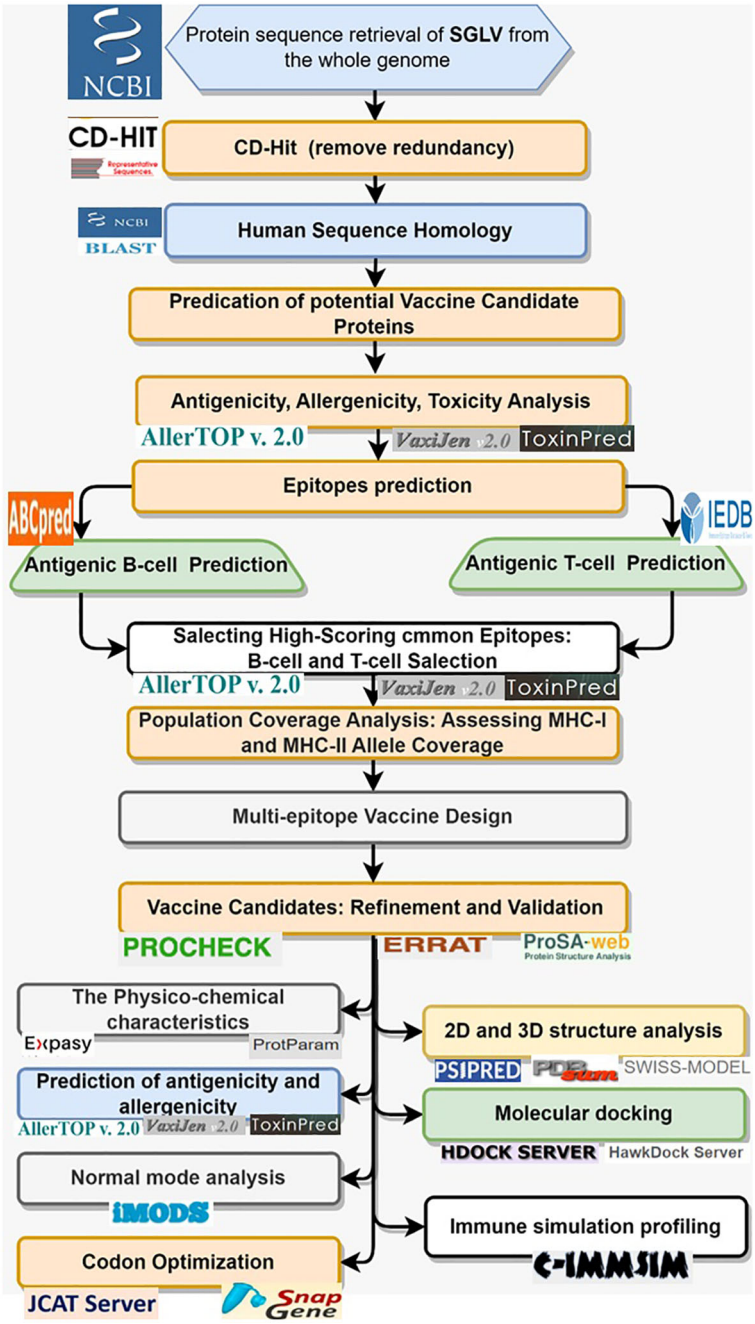


FIGURE 1
Systematic flow chart of *in-silico* based multi-epitope vaccine design.

NCBI was used for sequence homology to human proteins with thresholds of percent identity < 35, query coverage 75, e-value 10-4, bit score <100, and others used as defaults (11). For identification of the allergenicity of SGLV proteins, the AllerTop online server was used (12). The antigenicity was determined by setting the 0.4 parameter in VaxiJen online server (13). For the identification of the toxicity of SGLV proteins, the Toxinpred 3.0 online server was implemented (14).

2.2 Prediction of T-cell and B-cell

MHC-I epitopes were predicted using the IEDB MHC-I stickiness predictions program (15) accessed on June 23, 2022. The prediction algorithm used the SMM approach, and sequences were provided in FASTA format. It has been determined that humans will be the host species. The output format was set to XHTML tables, and all other options and parameters were left as default. Similarly, the IEDB MHC-II binding prediction tool (16) accessed on June 25, 2023, was used to predict the MHC-II epitopes by selecting the SMM prediction method. Data was provided in FASTA format. The HLA-DR was chosen as the species/locus couple, and then alleles were chosen using the typical length values associated with each species/locus (17). The other variables were kept at their default settings, and the final result format was set to XHTML table.

The B cell is an important element of the body's defense system. It is responsible for secreting antibodies that provide long-term immunity (18). For the detection of a continuously growing 12-mer long B-cell lymphoid (BCL) for the selected amino acids with a threshold number of 8.0, ABCPred (http://crdd.osdd.net/raghava/abcpred/) tool was used for this analysis (18). linear B-Cell and MHCI & II overlapped epitopes are selected based on physiochemical properties by employing VaxiJen v2.0, AllerTOP v.2.0, & ToxinPred 3.0 tools (12, 13, 19).

2.3 Population coverage

The IEDB population coverage assessment tool (https://tools.iedb.org/population/) was utilized to check the designed vaccine had successfully covered the entire world population (20). Populations research in China, eastern the People's Republic, southern the Caribbean, the Southeast Asian region, and the ocean was conducted to understand the global nature of the *Songling virus* pandemic. Default values were used to evaluate coverage for MHC class I and class II HLA binding alleles (15). This strategy takes advantage of the worldwide distribution of HLA-binding genotypes to calculate the abundance of certain epitopes.

2.4 Multi-epitope vaccine construct design

Effective vaccine construct design and proper epitope separation depend on all candidate epitopes being connected together through linkers and adjuvants. The B-cell epitope was linked to the CTL targets using the EAAAK, AAY linkers, and the HTL epitopes were linked to the CTL targets using the GPGPG linker (Figure 2). To facilitate future glycosylation with a carrier protein, a cysteine residue was included in the N-terminal of the multi-epitope vaccine construct (21). The antigenicity, allergenicity, toxicities, and physicochemical features of the vaccine construct were analyzed further using the ProtParam tool (https://web.expasy.org/protparam/) (19).

2.5 2D and 3D structure modeling and validation

PDBsum (http://www.ebi.ac.uk/thornton-srv/databases/pdbsum/) was used to determine the secondary structure of the new vaccine construct, which was then validated through the PSIPRED server

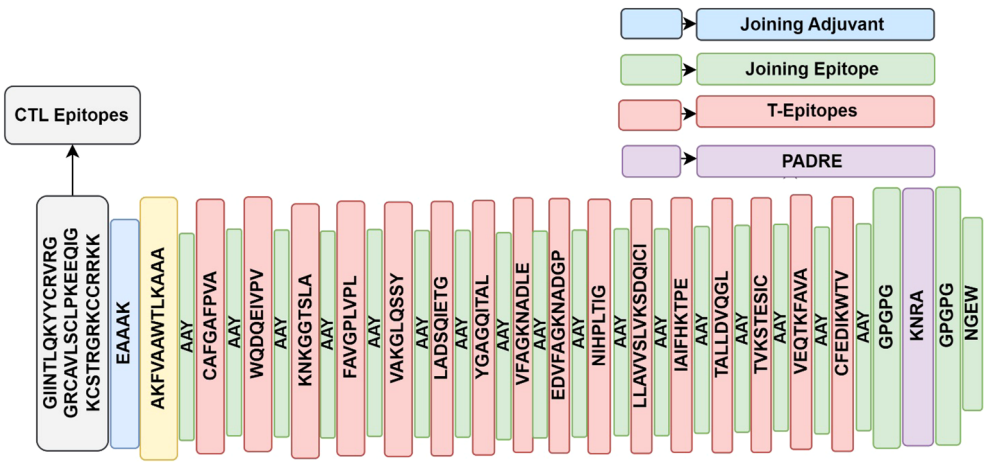


FIGURE 2
Graphical illustration of multi-epitope vaccine construct, showing CTL, and T-cell epitopes joined by appropriate links and Adjuvant.

(http://bioinf.cs.ucl.ac.uk/psipred/) (22–24). SWISS Model (https://swissmodel.expasy.org/) was used to model proteins, and then ProSAweb Server (https://prosa.services.came.sbg.ac.at/prosa.php) was used to analyze protein structure and validate the model (25, 26). Godard plot analysis utilizing the ERRAT server (https://servicesn.mbi.ucla.edu/ERRAT/) and RAMPAGE server (http://mordred.bioc.cam.ac.uk/rapper/rampage.php) (27).

2.6 Molecular docking with TLRs receptors

Employing TLRs alongside their precisely designed synthetic ligands in vaccines can initiate a potent chain of cytokines, crucial for robust immune reactions (28). Understanding the pattern of interactions among design vaccines with TLR3, TLR4, and TLR8 immune cell receptors is crucial for efficiently inducing immunological responses. The vaccine constructs were docked into the human TLR3 receptor (PDB ID: 2a0z) (29), TLR4 receptor (PDB ID: 4G8A) and TLR8 receptor [PDB ID: 3w3m (29–31)], using web servers Hdock (http://hdock.phys.hust.edu.cn/) and HawkDock (http://cadd.zju.edu.cn/hawkdock/), in order to evaluate the chemical reactions between immune receptors (TLR3, TLR4, and TLR8) and vaccine constructs (V1, V2, V3, and V4) (32). Another webserver HADDOCK (High ambiguity driven protein–protein Docking server) was utilized to generates informative visual representations of the docking outcome, facilitating a comprehensive analysis of the results. These graphical plots allow for easy comparison of the top-docked structure with the complete set of generated structures, providing insights into key parameters such as docking score and RMSD (root mean square deviation) (33).

2.7 MD simulation

The best docking results for the SGLV-V4-TLR4 molecule were used to study a chemical dynamics (MD) research simulation. MD simulations, energy efficiency, and protein flexibility were all calculated using the iMODS web server (https://imods.iqfr.csic.es/) (34). iMODS is based on normal mode analysis (NMA) in the internal (dihedral) coordinates of macromolecules that naturally reproduce the collective functional movements of biological macromolecules. Using these modes, iMODS builds pathways for functional transitions between two proteins with homologous structures. The server can simulate potential with several coarse-grained atomic representations and provides an enhanced arrow model based on an affine model to describe the complicated domain dynamics of macromolecules. The service analyses the dynamic molecular structure and the docked protein structure with other ligands as an amino acid of interest in order to deliver elastic network-related data according to NMA, which is equal for the particular instance of deform Eigenvalues, which changes the B-factor (mobility profiles), along with a variation map. The SGLV-V4-TLR4 complex docked PDB file was uploaded to the iMODS service, and results were obtained with all parameters set to their default values (34).

2.8 Immune simulation

The C-ImmSim webserver (http://www.cbs.dtu.dk/services/C-ImmSim-10.1/) was used to model computational immunological simulation of our prioritized vaccine design. This platform employs a potent combination of predictive modeling techniques, including Position-Specific Scoring Matrices (PSSM) and a variety of cutting-edge machine-learning algorithms, to assess and predict the cellular and humoral responses elicited by our antigenic vaccine candidate (35). The application leverages antigenic peptide sequences and lymphocyte receptors to replicate the intricate dynamics of immunogenic responses. Throughout our investigation, we precisely observed to a standard clinical protocol, administering two doses of the vaccine with a four-week interval to assess immune responses (36). Our focus lay on six specific human leukocyte antigens: HLA-A0101, HLA-A0201, HLA-B0702, HLA-B3901, HLA-DRB10101, and HLA-DRB10401, each monitored at time intervals of 1, 84, and 168 hours. Immune simulation was executed using the application's default settings, encompassing 1000 iterative steps (37).

2.9 Codon optimization

The JAVA Codon optimization Tool (Jcat) (http://www.jcat.de) was used to optimize the coding and execute a reverse translation of the sequence of amino acid sequences for the suggested immunization (38). After that, an E. coli production gene vector called pET28 was used alongside Snapgene version 5.2 to digitally clone the genetic code sequence (39).

2.10 Prediction of the vaccine mRNA secondary structure

The Transcription and Translation web based Tool (http://biomodel.uah.es/en/lab/cybertory/analysis/trans.htm) was employed to acquire the mRNA sequence of the vaccine. To predict the secondary structure of the vaccine mRNA, two web-based servers, Mfold v2.3 (http://www.unafold.org/mfold/applications/rna-folding-form-v2.php) (40) and RNAfold (http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi) (41), were utilized. The primary outcome of interest centered on the minimum free energy (expressed in units of Kcal/mol), with lower values indicating a greater degree of stability within the mRNA's folding structure.

3 Results

3.1 Proteins prediction of SGLV vaccine candidate

The complete proteome of the SGLV strain, containing 40 proteins from different strains across the world, was extracted from NCBI in FASTA format (Supplementary File S1). Utilizing CD-hit,

redundancy was minimized, and human blast yielded four distinct proteins (Supplementary File S2). Subsequently, we screened these proteins for allergenicity, antigenicity, and toxicity, identifying optimal candidates with high antigenicity, non-allergen, and non-toxic properties for epitope prediction (Supplementary Table 1).

3.2 Prediction of T-cell and B-cell and population coverage analysis

For further analysis, Four proteins are selected to recognize the lead epitopes for producing a chimeric vaccine construct against SGLV. T-cell (major histocompatibility complex class I and class II) epitopes were determined for the selected proteins using the IEDB server, with an IC_{50} threshold of 50 nM. The ABCpred scores reached greater than 0.8, and the specificities of the estimated ubiquitous B-cell epitopes were 75%. Vaccines were developed based on predictions of twelve (13) overlapping lead regions for each prioritized protein. the top 12 epitopes based on their antigenicity, IFN positivity, toxicity, and allergenicity (Table 1). The main goal was to recognize lead epitopes with the potential to induce humoral and cell-mediated immunogenic responses and host interferons.

The epitopes chosen exhibited 100% coverage across the global population (Supplementary Table 2). Analysis from the IEDB database indicated a notably high population coverage for these predicted epitopes, particularly in regions significantly impacted by SGLV, such as east Asia, Europe and south east Asia (Figure 3).

3.3 Multi-epitope vaccine design

To generate a multi-epitope vaccine EAAAK, CTL, HTL, and GGGS/HEYGAEALERAG linkers and adjuvants were used. When administered intramuscularly, vaccine containing linkers offer superior protection against each epitope (42). Increased immunogenic responses were achieved by coupling the epitopes to various adjuvants, such as HBHA protein molecules, beta-defensin, 50S ribosome enzyme L7/L12 adjuvants, and N-terminally abbreviated HBHA comparable amino acid sequences. Immunization strategies have used the PADRE peptide sequence to protect against difficulties caused by local variations in HLA-DR. Previous studies have shown that vaccine formulations including PADRE provide enhanced immune protection and high cytotoxic T lymphocyte (CTL) responses, as shown in Supplementary Table 3.

3.4 Immunological and physiochemical properties

Based on their immunological features, none of the immunization strategies were found to be harmful or allergic. Each one of the multi-epitope vaccine formulations has a substantial antigenic property, as demonstrated by antigenicity scores > 0.8 as estimated by the VaxiJen 2.0 server. Using cross-validation on the peptide sequence based on established datasets, VaxiJen 2.0 determines the antigenicity of viral sequences and identifies their protective properties. Each structure VaxiJen 2.0 score fell between 0.4333 and 0.5197, which is the same as the default threshold for viruses (13). Using the ProtParam service, we were able to determine the physiochemical parameters of the vaccine compositions and found that the molecular weights of each epitope in these innovations ranged from 30 kDa to 55 kDa. The selected vaccine designs have GRAVY values around -0.128 and -0.310, indicating that they are hydrophilic. The numerical pI values fell between 8.87 and 10.07. The thermostability of these structures was demonstrated by aliphatic index values between 69.09 and 82.50. The stability of these constructs at different temperatures was projected to be shown by their unpredictability index values, which ranged between 30.93 and 41.81 (Table 2). The

TABLE 1 IFN-g epitope prediction, investigation of allergenicity and toxicity, and prediction of overlapping T- and B-cell epitopes.

Protein IDs	MHC-I	IC50	MHC-II	B-cell Epitopes	ABCpred score	IFNepitope score	Allergenicity	Toxicity
YP_010840762.1	CAFGAFPVA	17	SDMVCAFGAFPVAEP	RICSDMVCAFGAFPVA	387	-0.1401593	non-allergen	Non-toxic
	WQDQEIVPV	74	WQDQEIVPVEHMLHQ	SGWQDQEIVPVEHMLH	444	-0.6228511	non-allergen	Non-toxic
	KNKGGTSLA	4.3	GSWTKKNKGGTSLAV	WGSWTKKNKGGTSLAV	215	2	non-allergen	Non-toxic
YP_010840761.1	FAVGPLVPL	10	FAVGPLVPLESAQKV	TKFAVGPLVPLESAQK	988	-0.2917396	non-allergen	Non-toxic
	VAKGLQSSY	42	AIKVEAVAKGLQSSY	EEIQQYLNDCSKGLLN	1261	-0.110257	non-allergen	Non-toxic
	LADSQIETG	28	RNIILADSQIETGTT	SEELLAFVDSQYVLTI	307	-0.0578239	non-allergen	Non-toxic
YP_010840760.1	YGAGQITAL	69	RPSYGAGQITALLDV	GRPSYGAGQITALLDV	1100	-1	non-allergen	Non-toxic
	IAIFHKTPE	13	IAIFHKTPERDLFDL	DIAIFHKTPERDLFDL	963	-0.5001685	non-allergen	Non-toxic
	TALLDVQGL	53	AGQITALLDVQGLLL	GAGQITALLDVQGLLL	1105	-0.1869516	non-allergen	Non-toxic
UWI48350.1	TVKSTESIC	9	EDIKWTVKSTESICE	FEDIKWTVKSTESICE	44	-1	non-allergen	Non-toxic
	VEQTKFAVA	51	ADWVEQTKFAVAPLV	ADWVEQTKFAVAPLVP	4	-0.394556	non-allergen	Non-toxic
	CFEDIKWTV	15	CRYRGCFEDIKWTVK	ECCRYRGCFEDIKWTV	36	-1	non-allergen	Non-toxic

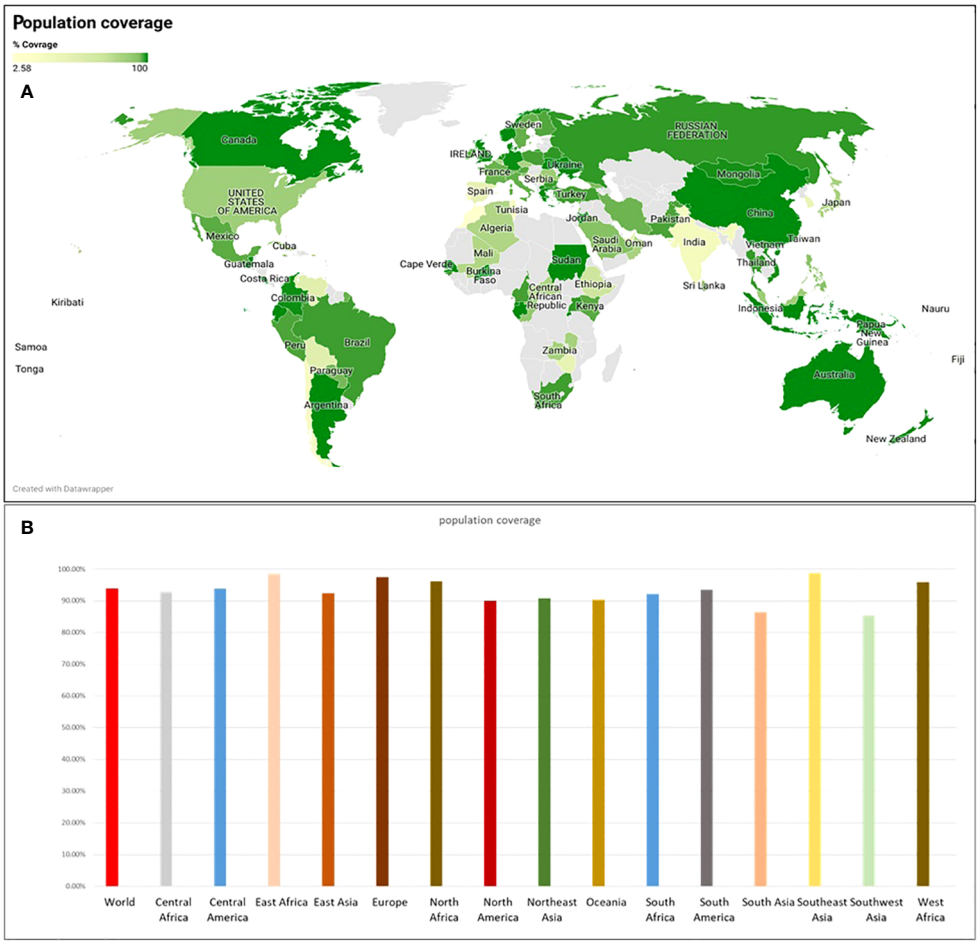


FIGURE 3
The population coverage was determined using the IEDB webserver, **(A)** population coverage across the world's countries and **(B)** population coverage across different ethnicities.

only real difference among the designs was the adjuvant; therefore, none of them changed much in terms of their physicochemical qualities. The vaccine designs' ability to elicit robust immunogenic reactions in the human host was deduced from an examination of their immunogenic and physiochemical properties (43). More experimental work is needed to confirm the reliability of these results.

3.5 Structures modeling, validation, and refinement

Computational approaches that anticipated secondary structure components were used to analyze the structural characteristics of the vaccine constructs. The PDBsum server displays proteins in their residue conservation and 2D structure. It shows which parts of the protein are not the same (colored blue) and which parts are very similar (colored red). Figure 4A. Similarly, the PSIPRED 4.0 server, which uses position-specific scoring matrices (PSSM), was utilized to predict transmembrane helices and topology within the peptide sequence, as well as identify fold and domain regions, as shown in Supplementary Figure 4. A stable and functional 3D structure of a vaccine is crucial for studying its molecular interactions with immune receptor proteins. Figure 4B displays the vaccine constructions predicted by the homology modeling techniques implemented in the Swiss Modelling server; these constructs were further refined by the DeepRefiner web server and submitted to a physical validation study. The binding energy of the JSmol structure is shown in Figure 4C. 73.7% of the V1 construct, 82.8% of the V2, 91.5% of the V3, and 97.4% of the V4 acids remained in the plots' favorable region (Figure 4D), indicating that the vaccine constructions were highly stable. The improved vaccine designs had ERRAT quality ratings between 58% and 97%. The ProSA-web server found that the Z score of all vaccine constructions might range from -0.88 to -4.71 (Figure 4E). Table 3 displays the 3D structural validation of vaccine constructs.

3.6 Molecular docking

Molecular docking is used for predicting the suitable binding between multi-epitope vaccine (MEV) and receptor molecules.

TABLE 2 Physiochemical properties of the vaccine constructs using ProtParam server and JCAT server.

Vaccine constructs	No of Amino Acids	Molecular weight (Da)	Instability index	Theoretical PI	Grand average of hydropathicity (GRAVY)*	GC content	CAI (0.85-1.0)	Aliphatic index
Con#1 adjuvant = HBHA adjuvant	427	43117	41.32 protein as stable	10.07	-0.282	52.22	1.0	69.09
Con#2 adjuvant = Beta definsin adjuvant	512	51396	35.65 protein as stable	9.51	-0.128	51.43	1.0	78.07
Con#3 adjuvant= HBHA conserved adjuvant	541	55585	41.80 protein as unstable	9.55	-0.310	53.23	1.0	75.71
Con#4 adjuvant = Ribosomal protein adjuvant	292	30649	30.93 protein as stable	8.87	0.129	52.73	1.0	82.50

Human surface TLR3, TLR4, and TLR8 immune system receptors were used to dock MEV with the help of Hdock Server (a blind docking technique) and Hawkdock Server. In the result, only one structure is prioritized for each dock based on its high score and lowest binding energy (Table 4). And the 3D structure of each docking is shown in Supplementary Figures 1–3. In this study, the binding energy of the V4 was found to be lower with TLR4 area by a significant margin (-66.26 kcal/mol) as compared to docking with TLR3 and TLR8 receptors. The prioritized complex TRL4-V4 was then assigned to HADDOCK to evaluate various parameters in our analysis, including HADDOCK scores, cluster size, van der Waals energy, electrostatic energy, desolvation energy, restraints violation energy, buried surface area, and Z score (Figure 5). Notably, Cluster 6 exhibited exceptional characteristics with a Z score of -2.0, HADDOCK scores of -9.5, a cluster size of 7, van der Waals energy of -26.2, electrostatic energy of -309.1, desolvation energy of -2.5, restraints violation energy of 2998.6, and a substantial buried surface area of 1708.2. Consequently, we selected the most promising structure from Cluster 6 for molecular dynamics simulation. The docking investigation reveals that the vaccine designs have strong binding capabilities to the TLR4 protein.

3.7 Molecular dynamic simulation

The TLR4 receptor was chosen due to its lowest binding energy with the SGLV-V4 construct. To comprehensively evaluate the stability of proteins and the enthalpy efficiency within SGLV-V4-TLR4 complexes, molecular dynamics (MD) simulations were employed. In parallel, the iMODS platform facilitated an in-depth analysis of atomic and molecular movements within the vaccine's biological context, elucidating macromolecular mobility via the normal mode analysis (NMA) methodology. For a more detailed understanding, Figure 6 provides a visual representation of the outcomes stemming from MD simulations and NMA conducted on the SGLV-V4 and TLR4 docked complexes. Drawing from the work of Ichiye and Karplus in 1991 (44), we utilized Equation 2 in conjunction with C Cartesian coordinates to compute the correlation matrix, thereby revealing the intricate interplay of atoms through an elastic network model. Each point on the graph symbolizes a spring connecting specific atom pairs, with varying shades of grey denoting differing levels of stiffness (as seen in Figure 6A). The complexity of molecular interactions within the system is further elucidated by the covariance map of the complex. By utilizing covariance analysis, this map highlights correlated (red), uncorrelated (white), or anti-correlated (blue) atomic movements, thus providing valuable insights into the dynamics of the complex molecule (Figure 6B). Additionally, eigenvalues, reflecting the stiffness of motion, hold a direct proportionality to the energy required for structural deformations. A lower eigenvalue indicates greater ease of deformation for the carbon alpha atoms. Notably, the SGLV-V4-TLR4 complex exhibited an eigenvalue of 2.395982e-05, signifying its stability (as observed in Figure 6C).

Furthermore, NMA-derived B-factor analysis was instrumental in portraying the relative amplitude of atomic displacements within the molecular complex. Figure 6D, displaying the B-factor graph, illustrates the correlation between the mobility identified in the docked complex NMA and the PDB scores. In this context, RMSD minimization based on local and global structure superposition enabled iterative deformation of the input structure, modeling potential transitions. Meanwhile, the total atomic displacements across all modes of residues at individual atomic sites provide an insightful measure of main-chain deformability. The complex's deformability graph, illustrated in Figure 6E, identifies peak regions representing the protein's more flexible areas, while inflexible sections exhibit lower values. Additionally, the variance graph, inversely linked to the eigenvalue (as demonstrated in Figure 6F), is connected to each normal mode of the complex, elucidating both individual and cumulative variances for a comprehensive depiction of the system's dynamics.

3.8 Immune response simulation

The focused MEV significantly boosted secondary responses, as predicted by immune modeling. In principle, this sequence can help

10.3389/fimmu.2023.1284366

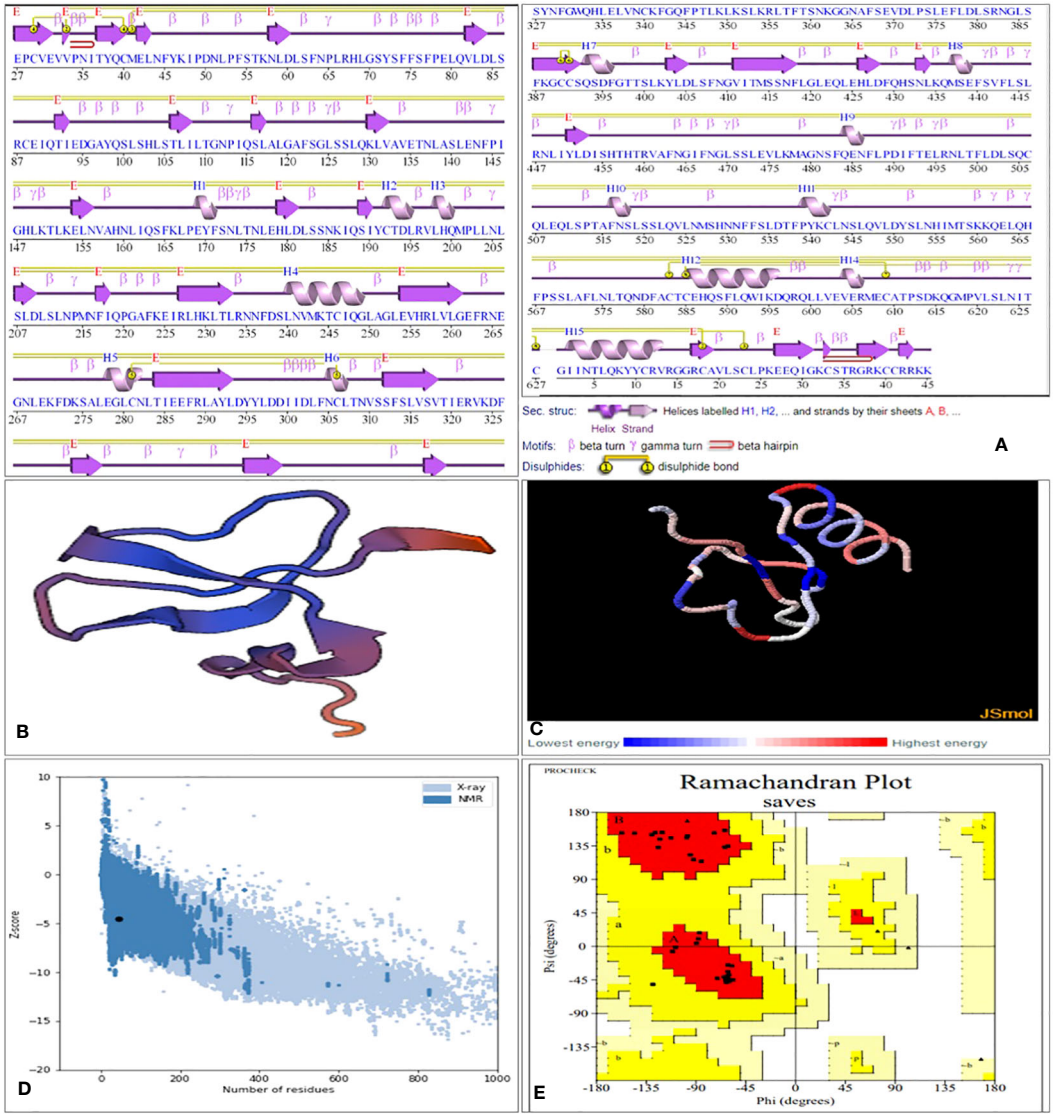


FIGURE 4
SGLV-V4 three-dimensional structural analysis, refinement, and validation **(A)** Protein's Secondary Structure with Graceful Elements: strands (elegant pink arrows), helices (royal purple springs), and captivating motifs in shades of red (-hairpins, mesmerizing -turns, and more), **(B)** The Swiss Model designed a 3D model of the multi-epitope vaccination using a homology modeling method. **(C)** Binding energy of the JSmol structure **(D)** ProSA-web yields a Z-score of -4.52. **(E)** Ramachandran plot analysis reveals 90% of the residues, 20% are in the allowed region, and 1% are in the prohibited portion of the plot.

the immune system quickly respond to threats. High levels of IgM were the prime simulated response. The simulated secondary and tertiary responses revealed considerable increases in B-cell populations as well as high concentrations of IgG1 + IgG2, IgM, and IgM + IgG antibodies. However, there was a decrease in the antigen levels (Figures 7A, B). The increased level of memory B-cell population and isotype switching indicate the formation of immunological memory in this case. Following the subsequent exposure to chimeric antigens, this caused a fast antigen reduction (Figure 7C). After further antigen exposure, it was hypothesized that both cytotoxic (TC) and helper (TH) cell subsets would form a similar memory (Figures 7D, E). High levels of activity in the macrophage, flare cell, and natural killer cell populations were also sustained during the vaccination period (Figures 7F–H). Higher levels of cytokines like interleukins IL-2 and IFN-y were also present (Figure 7I). These results provide support for the research showing that the anticipated vaccine formulation induced successful immune reactions against SGLV.

TABLE 3 3D structural validation of vaccine constructs via ERRAT, PROCHECK (Ramachandran plot favored region), and ProSA-Web Server.

Vaccine Construct	ERRAT (%)	PROCHECK (%)	ProSA (Z-score)
SGLV -V1	58.3333	73.7	-4.71
SGLV -V2	99	82.8	-4.11
SGLV -V3	97.3154	91.5	-0.88
SGLV -V4	93.75	97.4	-4.52

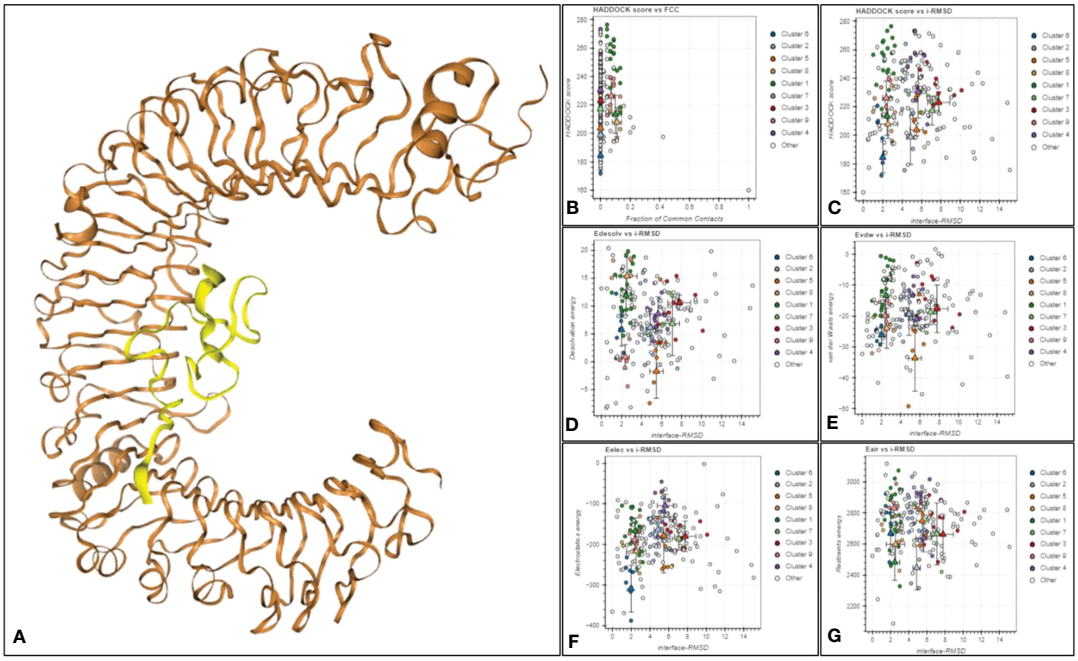


FIGURE 5
(A) The docked complex of TRL4-V4. The TLR4 receptor is depicted in brown, while yellow represents the SGLV-V4 vaccine construct. **(B)** Haddock score against a fraction of frequent contacts. **(C)** Haddock score against ligand RMSD. **(D)** Electrostatic Solvation Energy (EDESOLV) against Initial-RMSD in Molecular Simulations (I-RMSD), **(E)** van der Waals energy against interface I-RMSD, **(F)** Electrostatic energy (Eelec) of docked molecule against interface-RMSD, **(G)** (Ensemble-Averaged Interaction-Reweighted Simulation) EAIR outperforms I-RMSD in predicting the structure of receptor-ligand complexes.

3.9 Molecular cloning and codon optimization

Designing a vaccine with an appropriate expression system is the initial stage in evaluating a vaccination candidate, which requires a serological study. Prior to *in vitro* expression, a similar strategy was employed in earlier experiments for *in silico*-designed vaccines. The bacterial cell expresser E. coli was selected. Cloning and transcription are greatly facilitated by the Java Codon Adaptation Test (JCAT), which makes the E. coli K12 strain a great host organism. The estimated GC content of the improved sequence was 52.7%, which is significantly higher than the value of 50.73 found in E. coli. The modified sequence had a CAI (codon adaptation index) of 1.0. The multi-epitope vaccine (MEV) vector's codon usage curve is shown in Figure 8. Finally, SnapGene software was used to create a recombinant plasmid sequence by inserting the final vaccine construct V4 modified codon sequence into the plasmid vector pET28a (+), ensuring heterologous cloning and expression in the E. coli system (Figure 8).

3.10 Secondary structure of vaccine mRNA

The RNAfold server predicts the vaccine mRNA's secondary structure with a minimum free energy of -268.90 kcal/mol (Figure 9A), while the centroid secondary structure shows -229.37 kcal/mol (Figure 9B). mFold v2.3 server calculates the optimal secondary structure's minimum free energy at -283.12 kcal/mol (Figure 9C). A lower minimal free energy suggests greater stability for the vaccine mRNA post-expression *in vivo*.

TABLE 4 Docking scores and Binding energies of multi-epitope vaccine constructs and TLRs.

Constructs	TLR3 (2a0z)			TLR4 (4G8A)			TLR8 (3w3m)		
	Docking Score		Binding Energy	Docking Score		Binding Energy	Docking score		Binding Energy
	Hdock	Hawkdock		Hdock score	Hawkdock		Hdock score	Hawkdock	
V1	-274.53	-5209.48	-41.79	-298.62	-5097.4	-43.37	-301.63	-4694.39	-43.47
V2	-240.85	-5463.77	-0.26	-200.17	-4634.95	12.82	-240.85	-4098.19	-16.02
V3	-385.59	-4092.04	-0.35	-312.43	-2352.47	-5.81	-368.66	-3037.87	-14.13
V4	-278.24	-4788.78	-35.78	-266.51	-6853.68	-66.26	-274.59	-4318.53	-28.92

10.3389/fimmu.2023.1284366

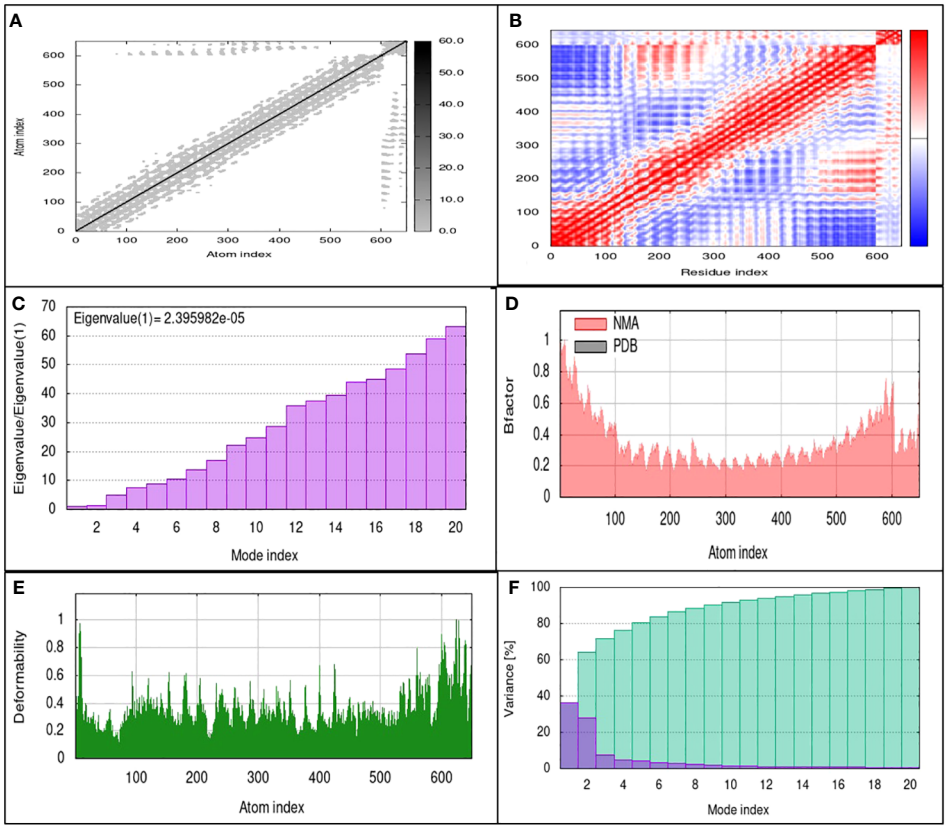


FIGURE 6
MD simulation results of SGLV-V4 with TLR4 **(A)** The elastic network model uses springs between atoms, indicated by colored dots for stiffness. **(B)** Covariance matrix Shoes paired residue mobility motions, i.e., uncorrelated (white), correlated (red), and anti-correlated (blue). **(C)** Eigenvalues, **(D)** Averaged RMS indicated by B-factor, **(E)** Deformability, and **(F)** Shows variances in Colored bars (purple) represent individuals, and cumulative is represented by green.

4 Discussion

In response to the growing concern over the flow of Songling virus (SGLV) cases worldwide and the absence of available vaccines, this study investigated the challenge of preventing future Songling virus epidemics. Employing cutting-edge immunoinformatics techniques, we embarked on designing innovative multi-epitope Songling virus vaccine constructs by examining the proteome of the Songling virus to pinpoint targets for a potential vaccine. Using strict standards, they forecasted epitopes for B-cells, MHC-I, and MHC-II. These epitopes play a crucial role by sparking a protective immune response that blocks viruses and establishes long-term defense (45).

The overlapped epitopes are prioritized from MHCI & II, and B-cell which are highly antigenic, non-allergen, and produce humoral response, that combats infections by eliminating infected cells or releasing antiviral substances to establish lasting immunity (46). To enhance this response, a novel vaccine was created using various CTL and HTL segments combined with specialized suitable linkers and adjuvants. Additionally, the vaccine's design incorporates EAAAK, AAY, and GPGPG linkers and adjuvants, which improve the structure and stability of the vaccine. Four vaccine constructs were designed from selected epitopes. These vaccine models displayed impressive traits: high antigenicity, non-allergenicity, and non-toxicity. Analysis of the vaccine's physiochemical characteristics indicated its robustness, alkaline nature, and hydrophobic properties, all of which indicate its potential to induce potent and targeted immunogenic responses in infected individuals.

Molecular docking analysis was then employed to explore the interaction between the vaccine constructs and the crucial immune cell receptors, i.e., TLRs. TLRs are known for their pivotal role in immune cell activation and the recognition of viral peptide structures (46). The results revealed strong binding affinities of SGLV-V4 toward TLR4, suggesting that the designed vaccine constructs have the capacity to generate robust immunogenic responses upon exposure. The C-ImmSim server, an immune response evaluation tool, was used to assess a newly designed vaccine's ability to induce an immunological response. This method simulates key components of the mammalian immune system and tracks how various immune cells respond to the vaccine (7). The goal is to design a vaccine that not only offers immediate protection but also triggers a long-lasting immune response,

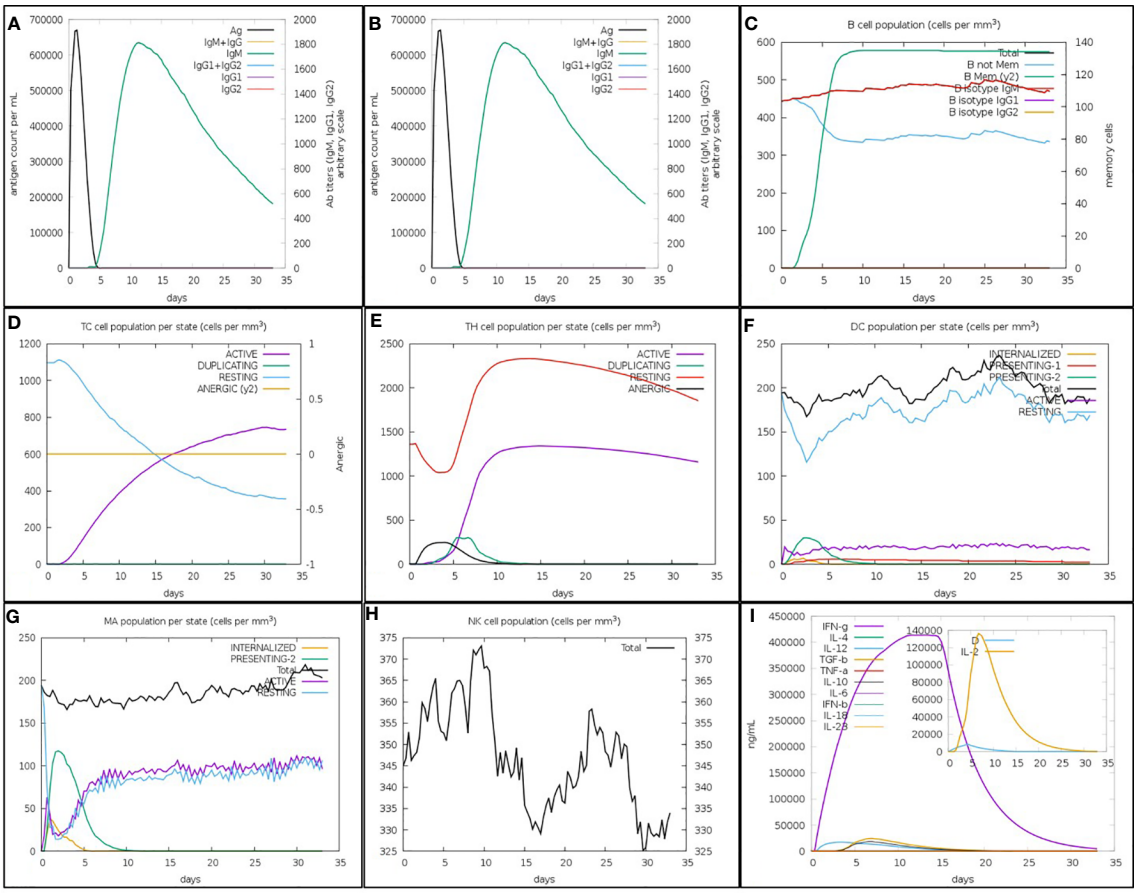


FIGURE 7
The *in silico* inflamed simulation used by the C-ImmSim Servers allows for an estimation of the SGLV-V4 recombinant peptide vaccination's immunological potential. **(A)** Vaccines cause an increase in immunoglobin antibodies and a decrease in antigen levels. As seen in **(B)**, B-cell numbers increase and antigen titers fall after immunization. Increased B-cell counts as a result of repeated antigen exposure **(C)**. T-cytotoxic and T-helper cell counts rise **(D, E)** after repeated antigen exposure. Dendritic cells, macrophages, and natural killer cells all grew in number during the vaccination window **(F–H)**. Increased antigen exposure leads to increased cytokine and interleukin **(I)** production. This danger signal is depicted alongside leukocytes and the rate of expansion factor IL-2 in the inset graphic.

simulating natural immunity. The top-ranked vaccine candidate, SGLV-V4, activated essential immune components, including antibodies (IgG and IgM), T-cells, B-cells, and cytokines (Figure 7) (47). This multi-epitope-based subunit vaccine shows suitable in protecting against the Songling virus. For further investigation to assess the stability and biomolecular process of SGLV-V4, the molecular dynamic simulation and NMA were performed. To enhance vaccine expression, computational cloning was performed on a pET28a (+) vector after codon optimization with the JCAT web service. The optimized codon adaptation index (CAI) and GC content fall within acceptable ranges, ensuring efficient expression in E. coli (strain K12). This optimization is crucial for successful vaccine production.

As we progress toward the next critical steps, we anticipate *in vitro* immunological assays to be conducted to confirm and validate the immunogenicity of the designed vaccine. Subsequently, a challenge-protection preclinical trial will be initiated, presenting a crucial opportunity to rigorously evaluate and substantiate the efficacy and safety of the SGLV-V4-TLR4 vaccine construct. These endeavors aim to provide a comprehensive framework to combat Songling virus infections effectively, potentially mitigating their impact and safeguarding public health against this evolving threat. However, to determine the vaccine's safety and efficacy, further experimental validation is required, which may involve the production of vaccine proteins with thorough *in vivo* and *in vitro* tests. However, the current research relies entirely on the results of computational approaches for technical equipment.

5 Conclusion

In our study on Songling virus (SGLV), a tick-borne pathogen lacking treatment or vaccines, we employed immunoinformatics to identify four potential vaccine target proteins. Designing a comprehensive vaccine candidate with broad global coverage, we combined B-cell and T-cell epitopes and validated them through IFN-γ epitopes. SGLV-V4, selected for its strong performance in molecular docking and favorable properties, was efficiently

10.3389/fimmu.2023.1284366

GGCATTATTAACACCCTGCAGAAATATTATTGCCGCGTGCGCGGCGGCCGCTGCGCGGTGCTGAGCTGCCTGCCGAAAGAA
GAACAGATTGGCAAATGCAGCACCCGCGGCCGCAAATGCTGCCGCCGCAAAAAAGAAGCGGCGGCGAAAGCGAAATTTGT
GGCGGCGTGGACCCTGAAAGCGGCGGCGGCGGCGTATTGCGCGTTTGGCGCGTTTCCGGTGGCGGCGGCGTATTGGCAG
GATCAGGAAATTGTGCCGGTGGCGGCGTATAAAAACAAAGGCGGCACCAGCCTGGCGGCGGCGTATTTTGCGGTGGGCCC
GCTGGTGCCGCTGGCGGCGTATGTGGCGAAAGGCCTGCAGAGCAGCTATGCGGCGTATCTGGCGGATAGCCAGATTGAAA
CCGGCGCGGCGTATTATGGCGCGGGCCAGATTACCGCGCTGGCGGCGTATGTGTTTGCGGGCAAAAACGCGGATCTGGAA
GCGGCGTATGAAGATGTGTTTGCGGGCAAAAACGCGGATGGCCCGGCGGCGTATGGCCCGGGCAACAGCAGCATGCCGTT
TGCGGCGTATCATAACATTCATCCGCTGACCATTGGCGCGGCGTATATTGCGATTTTTCATAAAACCCCGGAAGCGGCGTAT
ACCGCGCTGCTGGATGTGCAGGGCCTGGCGGCGTATACCGTGAAAAGCACCGAAAGCATTTGCGCGGCGTATGTGGAACA
GACCAAATTTGCGGTGGCGGCGGCGTATTGCTTTGAAGATCTGCTGGCGGTGGTGAGCCTGGTGAAAAGCGATCAGATTT
GCATTATTAAATGGACCGTGGGCCCGGGCCCGGGCAAAAACCGCGCGGGCCCGGGCCCGGGCAACGGCGAATGG

A

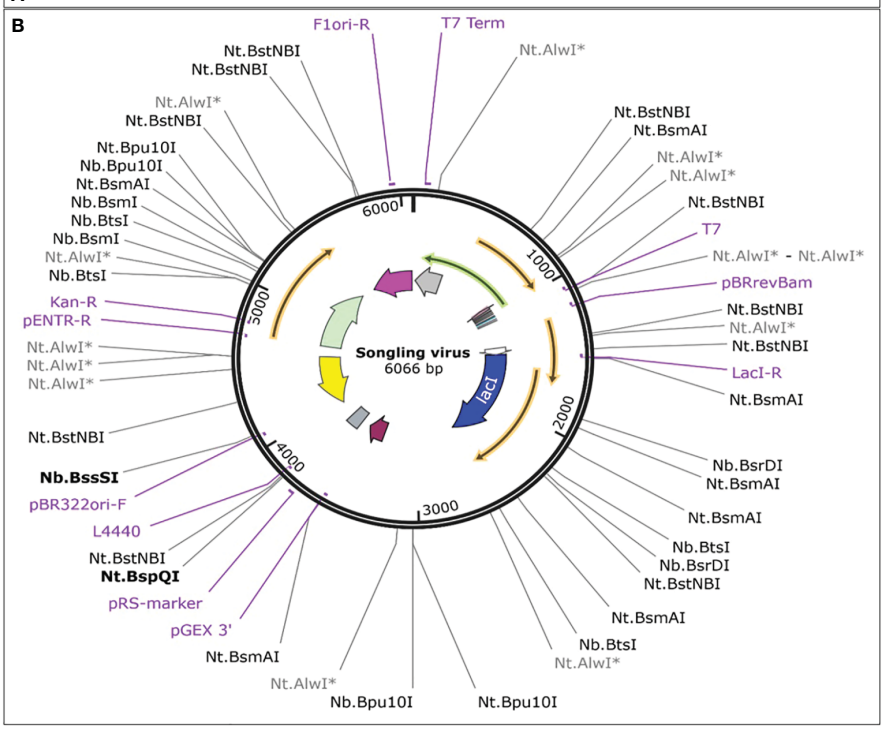


FIGURE 8
(A) Reverse-translated primary DNA sequence of the SGLV-V4, **(B)** SGLV-V4 cloned specifically into the E. coli expression vector [pET28a(+)].

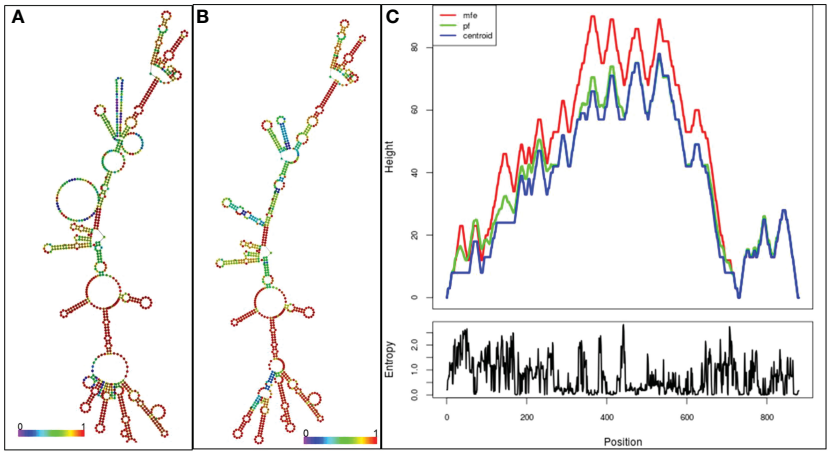


FIGURE 9
(A) Optimal secondary structure of the vaccine mRNA **(B)** Central secondary configuration of the vaccine mRNA **(C)** thermodynamic ensemble of mRNA structure, and the centroid structure are vividly depicted in the mountain plot representation. Additionally, the positional entropy plot unveils the intricacies of each position.

expressed in E. coli bacteria. Immunological research and simulations confirmed its stability and robust immune response, offering a promising avenue for safer and more effective SGLV-V4 vaccine development.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding author.

Author contributions

SA: Writing – original draft, Writing – review & editing, Conceptualization, Data curation, Formal Analysis, Investigation, Methodology. AAli: Writing – original draft, Writing – review & editing, Investigation, Software, Visualization. AAla: Funding acquisition, Validation, Writing – review & editing. AB: Validation, Writing – review & editing. MD: Writing – review & editing. AAb: Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This research received funding form King Saud University project number: RSPD2023R552.

Acknowledgments

Authors are thankful for Researchers Supporting Project number (RSPD2023R552), King Saud University, Riyadh, Saudi Arabia.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2023.1284366/full#supplementary-material

References

1. Ma J, Lv X-L, Zhang X, Han S-Z, Wang Z-D, Li L, et al. Identification of a new orthonairovirus associated with human febrile illness in China. *Nat Med* (2021) 27:434–9. doi: 10.1038/s41591-020-01228-y

2. Cotmore SF, Agbandje-McKenna M, Canuti M, Chiorini JA, Eis-Hubinger A-M, Hughes J, et al. ICTV virus taxonomy profile: Parvoviridae. *J Gen Virol* (2019) 100:367–8. doi: 10.1099/jgv.0.001212

3. Ji N, Wang N, Liu G, Zhao S, Liu Z, Tan W, et al. Tacheng tick virus 1 and songling virus infection in great gerbils (Rhombomys opimus) in Northwestern China. *J Wildl Dis* (2023) 59:138–42. doi: 10.7589/JWD-D-21-00137

4. Cai X, Cai X, Xu Y, Shao Y, Fu L, Men X, et al. Virome analysis of ticks and tick-borne viruses in Heilongjiang and Jilin Provinces, China. *Virus Res* (2023) 323:199006. doi: 10.1016/j.virusres.2022.199006

5. Poland GA, Ovsyannikova IG, Jacobson RM. Application of pharmacogenomics to vaccines. *Pharmacogenomics* (2009) 10(5). doi: 10.2217/pgs.09.25

6. Flower DR. *Bioinformatics for vaccinology*. Wiley: John Wiley & Sons (2008).

7. Suleman M, Rashid F, Ali S, Sher H, Luo S, Xie L, et al. Immunoinformatic-based design of immune-boosting multiepitope subunit vaccines against monkeypox virus and validation through molecular dynamics and immune simulation. *Front Immunol* (2022) 13:1042997. doi: 10.3389/fimmu.2022.1042997

8. Pruitt KD, Tatusova T, Maglott DR. NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res* (2005) 33:D501–4. doi: 10.1093/nar/gki025

9. Wheeler DL, Barrett T, Benson DA, Bryant SH, Canese K, Chetvernin V, et al. Database resources of the national center for biotechnology information. *Nucleic Acids Res* (2007) 35:D5–D12. doi: 10.1093/nar/gkl1031

10. Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* (2012) 28:3150–2. doi: 10.1093/bioinformatics/bts565

11. Mahram A, Herbordt MC. NCBI BLASTP on high-performance reconfigurable computing systems. *ACM Trans Reconfigurable Technol Syst* (2015) 7:1–20. doi: 10.1145/2629691

12. Dimitrov I, Flower DR, Doytchinova I. AllerTOP-a server for *in silico* prediction of allergens. *BMC Bioinf* (2013) 14: 1–9. doi: 10.1186/1471-2105-14-S6-S4

13. Doytchinova IA, Flower DR. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinf* (2007) 8:1–7. doi: 10.1186/1471-2105-8-4

14. Rathore AS, Arora A, Choudhury SPS, Tijare P, Raghava GPS. *ToxinPred 3.0: An improved method for predicting the toxicity of peptides*. bioRxiv (2023) p. 2008–23.

15. Dhanda SK, Mahajan S, Paul S, Yan Z, Kim H, Jespersen MC, et al. IEDB-AR: immune epitope database—analysis resource in 2019. *Nucleic Acids Res* (2019) 47: W502–6. doi: 10.1093/nar/gkz452

16. Zhang Q, Wang P, Kim Y, Haste-Andersen P, Beaver J, Bourne PE, et al. Immune epitope database analysis resource (IEDB-AR). *Nucleic Acids Res* (2008) 36: W513–8. doi: 10.1093/nar/gkn254

17. Roewer L, Nagy M, Schmidt P, Epplen JT, Herzog-Schröder G. Microsatellite and HLA class II oligonucleotide typing in a population of Yanomami Indians. *DNA fingerprinting State Sci* (1993), 221–30. doi: 10.1007/978-3-0348-8583-6_18

18. Malik AA, Ojha SC, SChaduangrat N, Nantasenamat C. ABCpred: a webserver for the discovery of acetyl-and butyryl-cholinesterase inhibitors. *Mol Divers* (2022) 18:1–21. doi: 10.1007/s11030-021-10292-6

19. ExPASy BRP. *ProtParam tool. SIB bioinforma resour portal* . Available at: http// web expasy org/protparam/ (Accessed Oct 2016).

20. Bui H-H, Sidney J, Dinh K, Southwood S, Newman MJ, Sette A. Predicting population coverage of T-cell epitope-based diagnostics and vaccines. *BMC Bioinf* (2006) 7:1–5. doi: 10.1186/1471-2105-7-153

21. Bandyopadhyay A, Cambray S, Gao J. Fast and selective labeling of N-terminal cysteines at neutral pH *via* thiazolidino boronate formation. *Chem Sci* (2016) 7:4589–93. doi: 10.1039/C6SC00172F

22. McGuffin LJ, Bryson K, Jones DT. The PSIPRED protein structure prediction server. *Bioinformatics* (2000) 16:404–5. doi: 10.1093/bioinformatics/16.4.404

23. Jones DT. Protein secondary structure prediction based on position-specific scoring matrices. *J Mol Biol* (1999) 292:195–202. doi: 10.1006/jmbi.1999.3091

24. Laskowski RA. PDBsum: summaries and analyses of PDB structures. *Nucleic Acids Res* (2001) 29:221–2. doi: 10.1093/nar/29.1.221

25. Biasini M, Bienert S, Waterhouse A, Arnold K, Studer G, Schmidt T, et al. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res* (2014) 42:W252–8. doi: 10.1093/nar/gku340

26. Wiederstein M, Sippl MJ. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res* (2007) 35: W407–10. doi: 10.1093/nar/gkm290

27. Lovell SC, Davis IW, Arendall WBIII, de Bakker PI, Word JM, Prisant MG, et al. Structure validation by Calpha geometry: phi, psi and Cbeta deviation, Proteins50, 437–450. *Google Sch There is no Corresp Rec this Ref* (2003) 50(3). doi: 10.1002/prot.10286

28. Koganty RR. Vaccines: exploiting the role of toll-like receptors. *Expert Rev Vaccines* (2002) 1:123–4. doi: 10.1586/14760584.1.2.123

29. Matsumoto M, Seya T. TLR3: interferon induction by double-stranded RNA including poly (I: C). *Adv Drug Delivery Rev* (2008) 60:805–12. doi: 10.1016/j.addr.2007.11.005

30. Ciesielska A, Matyjek M, Kwiatkowska K. TLR4 and CD14 trafficking and its influence on LPS-induced pro-inflammatory signaling. *Cell Mol Life Sci* (2021) 78:1233–61. doi: 10.1007/s00018-020-03656-y

31. Cervantes JL, Weinerman B, Basole C, Salazar JC. TLR8: the forgotten relative revindicated. *Cell Mol Immunol* (2012) 9:434–8. doi: 10.1038/cmi.2012.38

32. Yan Y, Zhang D, Zhou P, Li B, Huang S-Y. HDOCK: a web server for protein–protein and protein–DNA/RNA docking based on a hybrid strategy. *Nucleic Acids Res* (2017) 45:W365–73. doi: 10.1093/nar/gkx407

33. De Vries SJ, Van Dijk M, Bonvin AMJJ. The HADDOCK web server for data-driven biomolecular docking. *Nat Protoc* (2010) 5:883–97. doi: 10.1038/nprot.2010.32

34. López-Blanco JR, Aliaga JI, Quintana-Ortí ES, Chacón P. iMODS: internal coordinates normal mode analysis server. *Nucleic Acids Res* (2014) 42:W271–6. doi: 10.1093/nar/gku339

35. Nain Z, Abdulla F, Rahman MM, Karim MM, Khan MSA, Bin SS, et al. Proteome-wide screening for designing a multi-epitope vaccine against emerging pathogen Elizabethkingia anophelis using immunoinformatic approaches. *J Biomol Struct Dyn* (2020) 38:4850–67. doi: 10.1080/07391102.2019.1692072

36. Kroger A, Bahta L, Long S, Sanchez P. *General best practice guidelines for immunization: best practices guidance of the Advisory Committee on Immunization Practices (ACIP).* (2023).

37. Aiman S, Ali F, Zia A, Aslam M, Han Z, Shams S, et al. Core genome mediated potential vaccine targets prioritization against Clostridium difficile via reverse vaccinology-an immuno-informatics approach. *J Biol Res* (2022) 29. doi: 10.26262/jbrt.v29i0.8481

38. Grote A, Hiller K, Scheer M, Münch R, Nörtemann B, Hempel DC, et al. JCat: a novel tool to adapt codon usage of a target gene to its potential expression host. *Nucleic Acids Res* (2005) 33:W526–31. doi: 10.1093/nar/gki376

39. Biotech G. *Snapgene viewer*. Glick B, editor (2020), p. 3.

40. Zuker M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* (2003) 31:3406–15. doi: 10.1093/nar/gkg595

41. Gruber AR, Lorenz R, Bernhart SH, Neuböck R, Hofacker IL. The vienna RNA websuite. *Nucleic Acids Res* (2008) 36:W70–4. doi: 10.1093/nar/gkn188

42. Rahman N, Ali F, Basharat Z, Shehroz M, Khan MK, Jeandet P, et al. Vaccine design from the ensemble of surface glycoprotein epitopes of SARS-CoV-2: an immunoinformatics approach. *Vaccines* (2020) 8:423. doi: 10.3390/vaccines8030423

43. Mehmood A, Kaushik AC, Wei D. Prediction and validation of potent peptides against herpes simplex virus type 1 via immunoinformatic and systems biology approach. *Chem Biol Drug Des* (2019) 94:1868–83. doi: 10.1111/cbdd.13602

44. Ichiye T, Karplus M. Collective motions in proteins: a covariance analysis of atomic fluctuations in molecular dynamics and normal mode simulations. *Proteins Struct Funct Bioinforma* (1991) 11:205–17. doi: 10.1002/prot.340110305

45. Bacchetta R, Gregori S, Roncarolo M-G. CD4+ regulatory T cells: mechanisms of induction and effector function. *Autoimmun Rev* (2005) 4:491–6. doi: 10.1016/j.autrev.2005.04.005

46. Vaure C, Liu Y. A comparative review of toll-like receptor 4 expression and functionality in different animal species. *Front Immunol* (2014) 5:316. doi: 10.3389/fimmu.2014.00316

47. Chen R. Bacterial expression systems for recombinant protein production: E. *coli beyond. Biotechnol Adv* (2012) 30:1102–7. doi: 10.1016/j.biotechadv.2011.09.013

frontiers | Frontiers in Immunology

TYPE Original Research
PUBLISHED 06 December 2023
DOI 10.3389/fimmu.2023.1278534

OPEN ACCESS

EDITED BY
Joe Hou,
Fred Hutchinson Cancer Center,
United States

REVIEWED BY
Sumeet Patiyal,
National Cancer Institute (NIH),
United States
Kyle O'Donnell,
National Institutes of Health (NIH),
United States

*CORRESPONDENCE
Eduard Porta-Pardo
eduard.porta@bsc.es
Antoni Torres
atorres@clinic.cat
Laia Fernández-Barat
lfernan1@recerca.clinic.cat

†These authors share first authorship

‡These authors share senior authorship

RECEIVED 16 August 2023
ACCEPTED 14 November 2023
PUBLISHED 06 December 2023

CITATION
Farriol-Duran R, López-Aladid R, Porta-Pardo E, Torres A and Fernández-Barat L (2023) Brewpitopes: a pipeline to refine B-cell epitope predictions during public health emergencies.
Front. Immunol. 14:1278534.
doi: 10.3389/fimmu.2023.1278534

Brewpitopes: a pipeline to refine B-cell epitope predictions during public health emergencies

Roc Farriol-Duran[1†], Ruben López-Aladid[2,3†], Eduard Porta-Pardo[1,4*‡], Antoni Torres[2,3*‡] and Laia Fernández-Barat[2,3*‡]

[1]Barcelona Supercomputing Center (BSC), Barcelona, Spain, [2]CELLEX Research Laboratories, CibeRes (Centro de Investigación Biomédica en Red de Enfermedades Respiratorias ,Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Barcelona, Spain, [3]Pneumology Department, Hospital Clínic, Barcelona, Spain, [4]Josep Carreras Leukaemia Research Institute (IJC), Badalona, Spain

The application of B-cell epitope identification to develop therapeutic antibodies and vaccine candidates is well established. However, the validation of epitopes is time-consuming and resource-intensive. To alleviate this, in recent years, multiple computational predictors have been developed in the immunoinformatics community. Brewpitopes is a pipeline that curates bioinformatic B-cell epitope predictions obtained by integrating different state-of-the-art tools. We used additional computational predictors to account for subcellular location, glycosylation status, and surface accessibility of the predicted epitopes. The implementation of these sets of rational filters optimizes *in vivo* antibody recognition properties of the candidate epitopes. To validate Brewpitopes, we performed a proteome-wide analysis of SARS-CoV-2 with a particular focus on S protein and its variants of concern. In the S protein, we obtained a fivefold enrichment in terms of predicted neutralization versus the epitopes identified by individual tools. We analyzed epitope landscape changes caused by mutations in the S protein of new viral variants that were linked to observed immune escape evidence in specific strains. In addition, we identified a set of epitopes with neutralizing potential in four SARS-CoV-2 proteins (R1AB, R1A, AP3A, and ORF9C). These epitopes and antigenic proteins are conserved targets for viral neutralization studies. In summary, Brewpitopes is a powerful pipeline that refines B-cell epitope bioinformatic predictions during public health emergencies in a high-throughput capacity to facilitate the optimization of experimental validation of therapeutic antibodies and candidate vaccines.

KEYWORDS

bioinformatics and computational biology, immunology and infectious diseases, vaccine development, antibody therapeutics, epitope prediction and antigenicity prediction

Abbreviations: S protein, Spike protein of SARS-CoV-2; VOCs, Variants of Concern from SARS-CoV-2; RSA, Relative Solvent Accessibility; IEDB, Immune Epitope Database.

Introduction

Neutralizing antibodies play a major role in the adaptive immune response against pathogens (1). Hence, the prediction of the protein regions driving pathogen neutralization is key to guide the understanding of their mechanism of action (1). These protein regions, termed neutralizing B-cell epitopes, have the potential to spread through the entire proteome of the target pathogen. Such a wide distribution requires high-throughput techniques to unravel the full epitope landscape. In this context, the bioinformatic prediction of B-cell epitopes has become a necessary exploration to prioritize which candidates should be selected for experimental validation (Table 1). For instance, in the race against the SARS-CoV-2 pandemic, accurate bioinformatic B-cell epitope predictors significantly contributed to the success of COVID-19 preventive and therapeutic strategies (22) (Table 1). For this reason, many groups dedicated their efforts to the identification of SARS-CoV-2 antibody binding regions using different bioinformatic approaches as a first step to later characterize neutralizing antibodies or to design immunogens for vaccines (Table 1) (22, 23).

B-cell epitope predictors recommended by the Immune Epitope Database (IEDB) (24) such as Bepipred (2), or Discotope (8), and other existing SOTA methods (Table 1) (5, 7, 9–13) are tools able to identify candidate continuous and discontinuous B-cell epitopes in a minute scale. However, even state-of-the-art B-cell epitope prediction tools frequently output lists of predicted epitopes that are excessively large to validate experimentally (25). Moreover, many of the predicted epitopes will not necessarily function *in vivo* (25). Hence, the development of new predictive tools that will refine the available computational B-cell epitope predictions is a priority. Such tools will provide a rapid and accurate reaction in case of emergency situations such as the COVID-19 pandemic or the appearance of new variants of concern (VOCs) that escape the immune response of vaccinated subjects (3, 26).

To this end, we have designed Brewpitopes, a new predictive pipeline that integrates additional important features of known epitopes, such as glycosylation or structural accessibility using specific computational methods. To curate B-cell epitope predictions for neutralizing antibody recognition, Brewpitopes outputs curated lists of refined epitopes with an increased likelihood to be functional *in vivo*. To validate Brewpitopes, the pipeline was implemented to predict B-cell epitopes in antibody binding regions on the entire the proteome of SARS-CoV-2, with a special focus on the S protein and its VOCs.

Materials and methods

All three-dimensional protein figures have been generated with PyMol 2.5 and Chimera X. All statistical analyses have been performed using R statistical software (R version 3.6.3). All data and software can be obtained from public sources for academic use.

Dataset curation

The SARS-CoV-2 proteome in UniprotKB consists of 16 reviewed proteins (27). We used the corresponding FASTA

TABLE 1 Biophysical features included in state-of-the-art B-cell epitope predictors and the strategies followed for epitope landscape determination during early SARS-CoV-2 pandemics.

A							
Method	Type of Epitope	Subcellular Location	Glycosylation	Surface Accessibility	Based on	Comment	Reference
Bepipred 2.0	Linear	No	No	No	Antibody–antigen structures	Buried residues are predicted within epitopes	(2)
ABCpred	Linear	No	No	No	BCIPEP	High numbers of epitopes due to different window lengths	(3, 4)
EpitopeVec	Linear	No	No	No	IEDB+ BCIPEP	–	(4–6)
SVMTRIP	Linear	No	No	No	IEDB	–	(6, 7)
Discotope 2.0	Conformational	No	No	No	Antibody–antigen structures	–	(6, 8)
SeRenDIP-CE	Conformational	No	No	No	SAbDab	–	(9, 10)
SEPPA3.0	Conformational	Yes	Yes	No	Antibody–antigen structures	Predefined subcellular location. Does not account for surface accessibility or O-glycosylations.	(11)
Ellipro	Conformational	No	No	No	Antigen structures	–	(12)
Epitope3D	Conformational	No	No	No	Antibody–antigen structures	–	(13)

(*Continued*)

TABLE 1 Continued

A							
Method	Type of Epitope	Subcellular Location	Glycosylation	Surface Accessibility	Based on	Comment	Reference
Brewpitopes	Linear + Conformational	Yes	Yes	Yes	Antibody–antigen structures	–	–

B.								
Authors	Type of Epitope	Epitope Predictor	Subcellular Location	Glycosylation	Surface Accessibility	Based on	Comment	Reference
Almoftl et al.	Linear + Conformational	IEDB	No	No	No	IEDB		(14)
EzaJ et al.	Linear + Conformational	IEDB	No	No	No	Molecular dynamics		(4)
Sikora et al.	Conformational	–	No	Yes	No	Homology modeling		(15)
Khare et al.	Conformational	–	No	No	Yes	–	Integrates sequence conservation and functional domains.	(16)
Smith et al.	Linear	–	No	Yes	Yes	Epitope mapping		(17)
Li et al.	Linear	–	No	No	No	Epitope mapping		(18)
Schwarz et al.	Linear	–	No	No	No	Homology modeling		(19)
Grifoni et al.	Linear	IEDB	No	No	Yes	Bacterial display libraries		(20)
Haynes et al.	Conformational	SERA	No	No	No	Ab–Antigen structures		(21)
Farriol-Duran et al.	Linear + Conformational	Multiple	Yes	Yes	Yes			

(A) Comparison of the state-of-the-art B-cell epitope predictors. Classification of the different tools according to the inclusion of subcellular location, glycosylation status, and surface accessibility. Training datasets are annotated in the "Based on" column.
(B) Comparison of the B-cell epitope studies of SARS-CoV-2. Collection of B-cell epitope prediction approaches analyzing the epitope landscape of SARS-CoV-2. Strategies were classified according to the aforementioned biophysical constraints. B-cell epitope predictors and validation techniques are annotated in the "Based on" column.

10.3389/fimmu.2023.1278534

sequences as starting data for linear epitope predictions. To perform structural epitope predictions, when available, we obtained the PDB structures from the Protein Data Bank database selecting the structures with the best resolution and more protein sequence coverage (28). For those proteins with no available structure in PDB, we used Alphafold2.0 (29) or Modeller (30) to model their 3D structure.

Linear epitope predictions

To predict linear epitopes on protein sequences, we used ABCpred (31) and Bepipred 2.0 (2). We used ABCpred (31), an artificial neural network trained on B-cell epitopes from the Bcipep database (32), to predict linear epitopes given a FASTA sequence. The identification threshold was set to 0.5 as indicated by default (accuracy 65.9%) and all the window lengths were used for prediction (10–20mers). Additionally, we kept the overlapping filter on. To further augment the specificity of the predictions, we increased the ABCpred score to 0.8.

In addition, we used Bepipred 2.0 (2), a random forest algorithm trained on epitopes annotated from antibody–antigen complexes, as a second source to predict linear epitopes. The epitope identification threshold was set to ≥0.55 leading to a specificity of 0.81 and a sensitivity of 0.29 (32).

Structural epitope predictions

We used PDBrenum (33) to map the PDB residue numbers to their original positions at the UniprotKB FASTA sequence. The reason behind this step was that factors such as the inclusion of mutations to stabilize the crystal may lead to discordances between the residue numbers in the PDB and FASTA sequence from the same protein.

In order to model those SARS-CoV-2 proteins with missing structures in PDB, we used Alphafold 2.0 (29). We then refined the models by restraining our analysis to those regions with a pDLLT threshold of 0.7 to only assess highly confident regions. The proteins that required Alphafold modeling were M, NS6, ORF9C, ORF3D, ORF3C, NS7B, and ORF3B.

To predict conformational or structural B-cell epitopes, we used Discotope 2.0, a method based on surface accessibility and a novel epitope propensity score (8). The epitope identification threshold was set to −3.7, as specified by default, which determined a sensitivity of 0.47 and a specificity of 0.75.

Epitope extraction and integration

Bepipred 2.0 (2), ABCpred (31), and Discotope 2.0 (8) predictions resulted in different tabular outputs. To extract and curate the predicted epitopes, we created a suite of computational tools in R statistical programming language and Python, available at https://github.com/rocfd/brewpitopes.

Subcellular location predictions

When publicly available, the protein topology information was retrieved from the subcellular location section in UniprotKB (27). For those proteins with unavailable topology, we predicted their extracellular regions using Constrained Consensus TOPology prediction (CCTOP) (6), a consensus method based on the integration of HMMTOP (34), Membrain (35), Memsat-SVM (36), Octopus (37), Philius (38), Phobius (39), Pro and Prodiv (40), Scampi (41), and TMHMM (42). The.xml output of CCTOP was parsed using an in-house R script (xml_parser.R) and then the extracted topology served as reference to select epitopes located in extracellular regions using the script Epitopology.R.

Glycosylation predictions

To investigate *in silico* which residues would be glycosylated, we used NetNGlyc 1.0 (43) for N-glycosylation and Net-O-Glyc 4.0 (44) for O-glycosylations. NetNglyc uses an artificial neural network to examine the sequences of human proteins in the context of Asn-Xaa-Ser/Thr sequons. NetOglyc produces neural network predictions of mucin type GalNAc O-glycosylation sites in mammalian proteins. We parsed the corresponding outputs using tailored R scripts and then we extracted the glycosylated positions to filter out those epitopes containing glycosylated residues using Epiglycan.py.

Accessibility predictions

To predict the accessibility of epitopes within their parental protein structure, we computed the relative solvent accessibility (RSA) values using ICM browser from Molsoft (45). We used an in-house IEC browser script (Compute_ASA.icm) to compute RSA and we considered buried those residues with RSA threshold less than 0.20. Then, the ICM-browser output was parsed to extract the buried positions, which then served as a filter to discard epitopes containing inaccessible or buried residues using Episurf.py.

Variants of concern analysis

The mutations accumulated by the VOCs Alpha, Beta, Delta, Gamma, and Omicron in the S protein were obtained from the CoVariants webpage (4), which is empowered by GISAID data (46). A fasta sequence embedding each variant's mutations was generated using fasta_mutator.R.

Results

Brewpitopes, a pipeline to curate B-cell epitope predictions based on determinant features for *in vivo* antibody recognition

While there are some tools available to predict the presence of B-cell epitopes in a protein sequence or structure, these tools are

mainly based on machine learning methods trained with experimentally validated epitopes (Table 1). However, these methods sometimes do not account for other factors that might affect the antigenicity or the potential of a protein region to be recognized specifically by antibodies.

Brewpitopes was designed as a streamlined pipeline that generates a consensus between linear and conformational epitope predictions and curates them following the *in vivo* antibody recognition constraints (Figure 1). To this end, a suite of computational tools was created to integrate the output of different SOTA B-cell epitope predictor and to filter the candidates using predictions of the aforementioned biophysical features (Figures 1, 2).

In Brewpitopes, we included predictions of linear epitopes, which are continual stretches of residues located at the surface of proteins, and predictions of conformational epitopes, which are discontinuous residues recognized by antibodies due to their structural disposition. For both cases, state-of-the-art predictors exist (Table 1). To start with, in Brewpitopes, we have predicted linear epitopes using Bepipred2.0 (2) and ABCpred (31) and we have searched for conformational epitopes using Discotope2.0 (8). Once predicted, we have extracted the epitopes using tailored R scripts named Epixtractor and then integrated the results using Epimerger.

Once the predictions are integrated, we propose a set of serial biophysical filters organized in a pipeline. First, since neutralizing antibodies only inspect the external surface of cells or viral particles, we propose that those epitopes predicted in intracellular and transmembrane regions of viral proteins cannot be targets for antibody neutralization (Figure 1). Hence, the subcellular location of an epitope is a recognition constraint (47), which our pipeline uses to prioritize epitopes located on extracellular protein regions while discarding those located in intracellular and transmembrane regions. To predict the subcellular location of a protein region, we used protein topology information. For some proteins, the topology is already available at UniProtKB (27); however, for some others, topology is not described. In such cases, the alternative is to predict the topology of the target protein. In Brewpitopes, there is a module to upload experimentally described protein topologies. Complementarily, for undescribed proteins, we used CCTOP to predict their transmembrane, intracellular, and extracellular regions (6). Once we had obtained or predicted the extracellular regions, we labeled the epitopes using Epitopology.

Glycan coverage can limit the surface accessibility of predicted B-cell epitopes that contain glycosylated residues, thus reducing their *in vivo* antibody recognition potential (Figure 1) (48). For this reason, our pipeline uses *in silico* tools to predict glycosylated sites on protein sequences. Concretely, we have used NetNglyc1.0 (43) and NetOglyc4.0 (44), for the prediction of N-glycosylations and O-glycosylations, respectively. These methods are based on artificial neural networks trained on glycosylation patterns by which they can predict glycosylation sites *ab initio* given a protein sequence. With this information, Brewpitopes discards all the epitopes that include glycosylated residues using Epiglycan.

As the third filter, we include the accessibility of the epitope within the antigenic protein structure as another antibody recognition constraint (49) (Figure 1). Accordingly, our pipeline calculates the relative solvent accessibility (RSA) values of all the residues in the target protein and filters out those epitopes containing at least one buried residue (RSA < 0.2). To compute the RSA values based on crystal structures, we have used Molsoft (45) and the in-house script compute_asa.icm.

The last step of the Brewpitopes pipeline is Epifilter, which uses the annotations of the previous steps to filter out those epitopes predicted as intracellular, glycosylated, or buried. Additionally, a length filter was used to discard epitopes SHORTER than five amino acids in length, which were considered unspecific. Therefore, the final candidates refined using Brewpitopes are extracellular, non-glycosylated, and accessible, properties that enhance the antibody recognition *in vivo*.

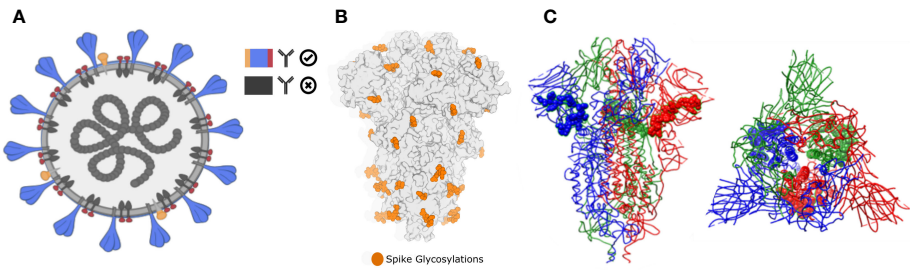


FIGURE 1
Biophysical constraints for *in vivo* antibody recognition. **(A)** Recognition of extracellular or extra-viral protein regions. Neutralizing antibodies only inspect the external surface of viral particles. Therefore, predicted epitopes located in intracellular or transmembrane epitopes will not be recognized. In Brewpitopes, we used protein topology-annotated information and topology predictors to assess the subcellular location of the target protein regions with predicted epitopes. Exclusively, candidates located on extracellular protein regions were selected. **(B)** Glycosylation coverage prevents *in vivo* antibody recognition of neutralizing epitopes. Predicted epitopes that contain glycosylation motifs are likely covered by glycans supporting the selection of predicted epitopes without glycosylated residues. In Brewpitopes, we predicted the glycosylation profiles of target proteins using Net-N-glyc and Net-O-glyc for N- and O-glycosylations, respectively. Only predicted epitopes without glycosylated residues pass this filter. **(C)** Epitope accessibility on parental protein surface. Predicted epitopes that contain buried residues will be less accessible for *in vivo* antibody recognition. Left: structure of S protein of SARS-CoV-2 highlighting a fully accessible predicted epitope. Right: structure of the S protein displaying a highly buried predicted epitope. In Brewpitopes, to assess epitope accessibility, we calculated the Residue Solvent Accessibility (RSA) of the predicted epitope sequences using crystal or structural models. Once predicted, fully accessible epitopes (all residues RSA $\geq$ 0.2) were selected and buried candidates were discarded (at least one residue RSA < 0.2).

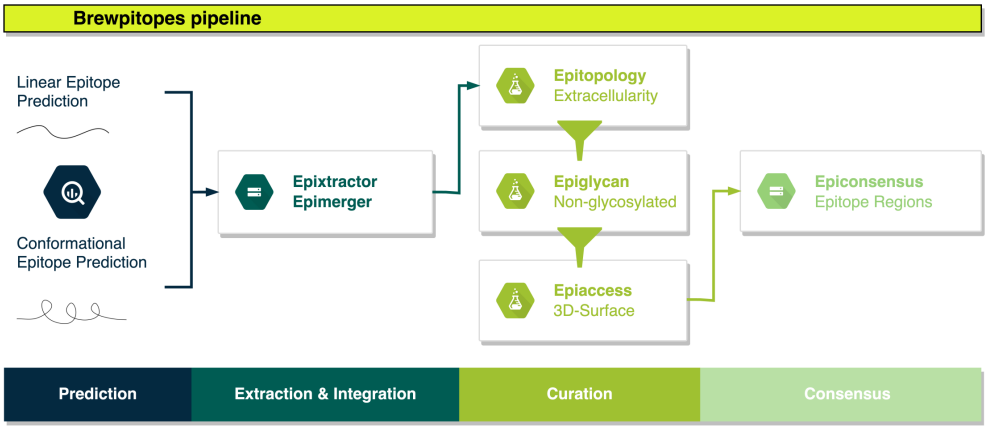


FIGURE 2
Brewpitopes pipeline. Linear and conformational epitope predictions are performed using Bepipred2.0, ABCpred, and Discotope2.0. Epitope extraction is customized in each tool's output using Epixtractor. Extracted epitopes are standardized using Epimerger. Subsequently, Brewpitopes implements three *in silico* predictors of biophysical constraints for *in vivo* antibody recognition: subcellular location, glycosylation coverage, and surface accessibility. Protein topology information to determine subcellular location can be uploaded into Brewpitopes using annotated data or via CCTOP predictions (.xml output) using Epitopology. Predicted epitopes located in extracellular regions are selected. Intracellular and transmembrane epitopes are discarded. Glycosylation patterns of target proteins are predicted with Net-N-Glyc and Net-O-Glyc and the output is used by Epiglycan to label all predicted epitopes containing one glycosylated residue as "glycosylated" and candidates not containing glycosylated positions as "non-glycosylated". Epitope accessibility on the 3D surface of the parental protein structure is computed via *compute_asa.icm* (Molsoft - ICM Browser) and a PDB file obtained from a crystal structure or a computational model. Predicted RSA values are used by Epiaccess to label fully accessible epitopes as "accessible" (all residues RSA $\geq$ 0.2) and candidates containing at least one buried residue as "buried" (RSA < 0.2). The filtering of the candidate epitopes according to the predicted biophysical constraints (labeled as "extracellular", "non-glycosylated", and "accessible") is performed by Epifilter. Curated candidates predicted by different tools will result in overlapping epitopes that are merged into epitope regions using Epiconsensus.

The final list of curated epitopes derives from the different tools integrated at the initial step of Brewpitopes. Thus, frequently epitopes with overlapping positions will be encountered. To prevent the prioritization of different but redundant candidates, Brewpitopes merges overlapping epitopes into epitope regions with the aim to generate a consensus between B-cell epitope predictors. Complementarily, the selection of a short sequence length threshold was useful to integrate epitopes predicted by different tools into larger epitope regions. To this end, we designed Epiconsensus, a tool that not only merges overlapping epitopes but also enables the scoring of the merged epitope regions, setting a prioritized order of the initial B-cell epitope predictor scores.

Bioinformatic validation of Brewpitopes in the proteome of SARS-CoV-2

Brewpitopes can be implemented to any target protein or organism, but due to the pandemic context and the interest in B-cell epitopes and neutralizing antibodies against SARS-CoV-2, to validate the pipeline, we analyzed the proteome of this virus. Within SARS-CoV-2, we specially focused on the S protein due to its importance in vaccine and therapeutic antibody design plus the known role of Spike for immune evasion (50). Our results confirm the neutralizing potential of the S protein but additionally identify other SARS-CoV-2 proteins containing epitopes of interest.

Focusing on the S protein, linear epitope predictions resulted in 213 epitopes and structural predictions in 6. Once integrated, 10 epitopes were discarded due to their intraviral location. Next, since it had been established that S protein is heavily glycosylated (26), 52 epitopes were filtered out due to their likelihood to include glycosylated residues. Lastly, 143 epitopes were discarded because they contained at least one residue buried within the 3D structure of the S protein. As a result, 14 epitopes derived from S were curated for optimized antibody recognition (Figure 3). Compared to the initial state-of-the-art epitope predictions, our results show that only a 5.5% of the predicted epitopes for the S protein will have high antibody recognition *in vivo* potential due to the recognition constraints analyzed with Brewpitopes (Figure 4). Furthermore, to generate a consensus between linear and conformational predictions from different tools, the overlapping epitopes were merged into epitope regions. In the case of S protein, the 14 candidates were merged into seven epitope regions (Figure 3).

As an external control, the epitope regions identified in the S protein were cross-validated with the epitopes reported at the IEDB database (51). Notably, the regions identified in our pipeline were all encountered among IEDB annotated epitopes, which confirms the validity of our predictions. However, our epitope regions represented less than 1% of the epitopes for the S protein listed in the IEDB. Compared to the initial output from the computational tools, the final list of prioritized epitopes from our pipeline was enriched fivefold in validated epitopes from IEDB (p < 2e-4). This confirms the power of Brewpitopes to refine B-cell epitope computational predictions to a reduced set of epitopes with greater probability for *in vivo* antibody recognition (Figure 3).

To extend our proteome-wide analysis of SARS-CoV-2, we used Brewpitopes to search for other epitopes and antigenic viral proteins with antibody recognition potential. Overall, 4/15 of the

10.3389/fimmu.2023.1278534

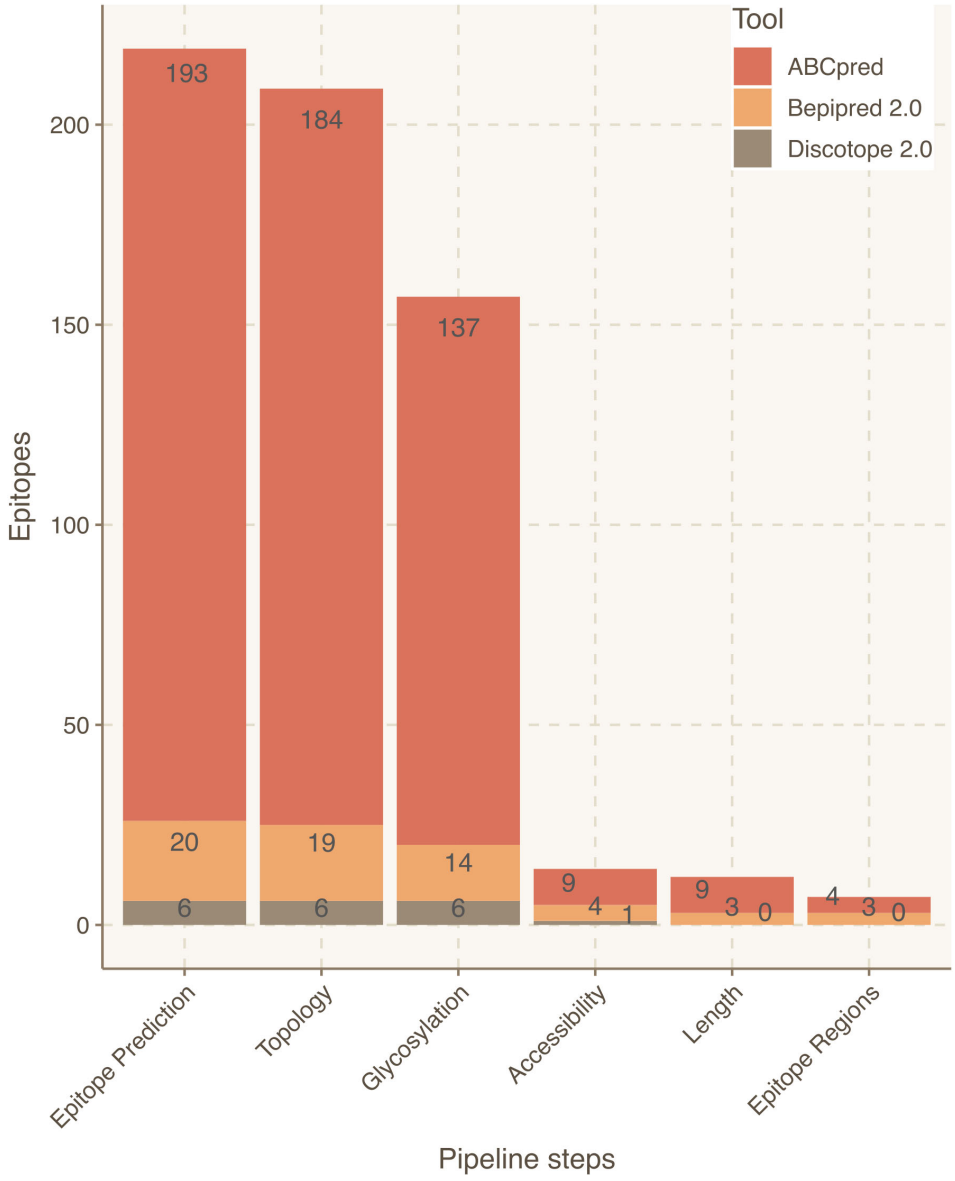


FIGURE 3
Epitope refinement for SARS-CoV-2 Wuhan S protein. The *x*-axis represents the filtering steps of the pipeline. The *y*-axis displays the number of epitopes refined by each filtering step of Brewpitopes.

remaining proteins contained candidate epitopes for neutralizing antibodies (R1AB, R1A, AP3A, and ORF9C) (Table 2). The remaining proteins (11/15) did not contain epitopes due to their major intraviral location (NS7A, NS7B, ORF3D, ORF3C, ORF9B, ORF3B, NS8, NS6, M, E, and N) and the absence of predicted epitopes in their short extracellular regions.

Within the proteins that contained curated epitopes, R1AB and R1A stood out, including 479 and 348 epitopes, respectively. The large numbers of epitopes predicted in these proteins is mainly explained by their long sequences, 7,096 and 4,405 amino acids, respectively. Remarkably, R1A corresponds to the N-terminal region of R1AB explaining the high degree of shared predictions. R1AB is a complex polyprotein cleaved into 15 chains. In this analysis, all the chains were analyzed together using the standard R1AB UniProt sequence. On the other hand, we could also identify epitopes located in shorter proteins as ORF9C and AP3A. Accordingly, these presented a lower number of predicted candidates. In terms of epitope regions, R1AB contains 62 regions; R1A, 46 regions; ORF9C, 2 regions; and AP3A, 1 region. Altogether, these results corroborate that four SARS-CoV-2 proteins other than S have at least one candidate epitope region with *in vivo* antibody recognition potential.

Analysis of epitope conservation in the S protein of variants of concern

We studied the effect of mutations accumulated in the S protein of the VOCs (Alpha, Beta, Delta, Gamma, and Omicron) of SARS-CoV-2 in the development of immune escape mechanisms

10.3389/fimmu.2023.1278534

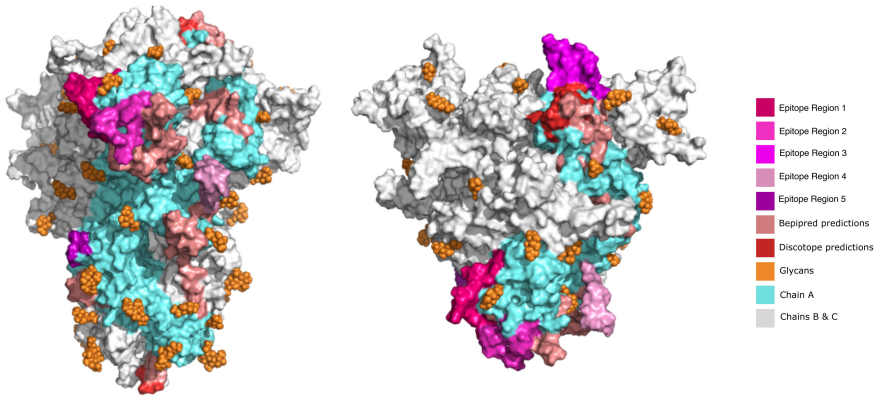


FIGURE 4
Visualization of predicted epitope location on the 3D structure of SARS-CoV-2 S protein to compare the initially predicted epitopes versus the epitopes refined by Brewpitopes. This representation depicts the shrinkage of the region to be explored and experimentally validated since unrefined predictions represent a much larger surface than the epitopes refined by Brewpitopes. Left: Front view of the S protein 3D structure. Right: Top view. All the epitopes were only labeled on the chain A of the S protein for visualization purposes (blue). The epitope regions 6 and 7 were not displayed because they escaped the limits of the represented structure. Owing to the large number candidates predicted by ABCpred, only the best scored candidates of this software were included in the 3D representation.

implementing Brewpitopes on the S protein sequences of the different variants (Table S1; Figure 5). We generated tailored FASTA files including the mutations of each variant and we retrieved the structures from PDB when available. For the Omicron variant, we modeled its structure using Modeller (30). Once we had run Brewpitopes, we compared the final number of epitopes with neutralizing potential identified in each variant with the epitopes generated by our analysis of the Wuhan S protein, considered the wild type. Concretely, we aimed at identifying epitope losses due to the presence of mutations, the appearance of new glycosylation sites and structures changed, leading to new buried positions. Additionally, we accounted for newly predicted epitopes generated by unique mutations of each variant. To compare epitope regions in WT versus those of the VOCs, the length of these epitope regions was added and divided by the total length of the S protein to obtain a protein-wide epitope coverage

TABLE 2 Epitope refinement on SARS-CoV-2 proteome.

Protein	UniProt ID	Predicted Epitopes	Curated Epitopes	Epitope Refinement (%)	Epitopic Regions
Spike	P0DTC2	219	12	5.5	7
E-protein	P0DTC4	10	0	0	0
N-protein	P0DTC9	115	0	0	0
M-protein	P0DTC5	22	0	0	0
R1AB	P0DTD1	1,111	479	43.1	62
R1A	P0DTC1	668	348	52.1	46
AP3A	P0DTC3	17	2	11.8	1
NS6	P0DTC6	13	0	0	0
NS7A	P0DTC7	15	0	0	0
NS7B	P0DTD8	2	0	0	0
NS8	P0DTC8	14	0	0	0
ORF3B	P0DTF1	0	0	0	0
ORF3C	P0DTG1	1	0	0	0
ORF3D	P0DTG0	12	0	0	0
ORF9B	P0DTD2	12	0	0	0
ORF9C	P0DTD3	9	4	44.44	2

Predicted epitopes correspond to the number of epitopes obtained using individual linear and structural predictors. Curated epitopes refer to refined epitopes obtained using Brewpitopes. Epitope refinement is the percentage of curated epitopes over the initial number of predicted epitopes obtained using individual state-of-the-art tools. Epitope regions result from the integration of overlapping predictions by different tools.

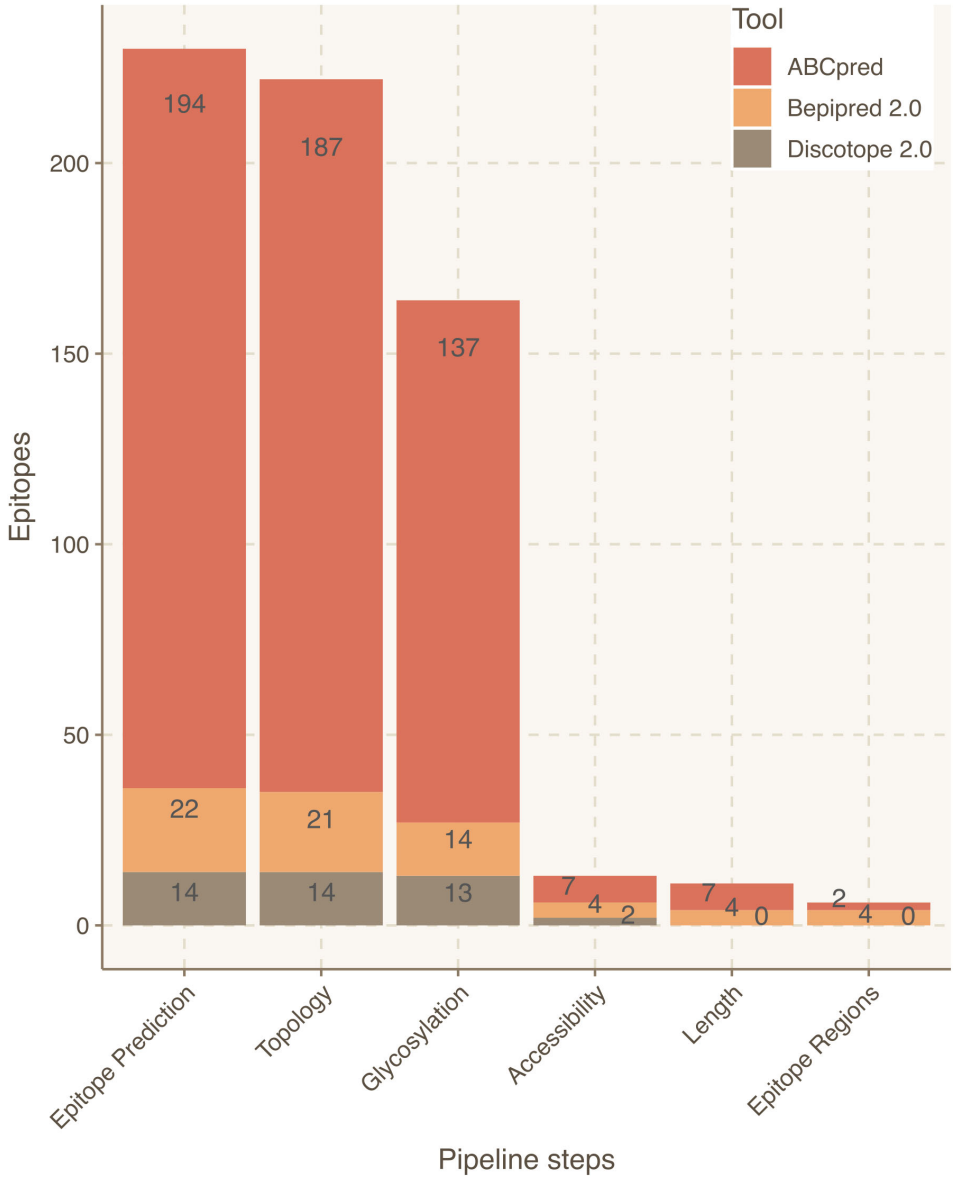


FIGURE 5
Epitope refinement for the S protein of the Omicron variant. The *x*-axis represents the steps of the Brewpitopes pipeline and the *y*-axis denotes the number of epitopes selected by each filtering step of Brewpitopes (Figure 2). Omicron's epitope yield obtained with Brewpitopes (six epitope regions) is lower than Wuhan WT's yield (seven epitope regions).

metric. In other words, this metric is the fraction of the protein sequence that is covered by predicted epitopes. This analysis predicted an epitope coverage of 9.43% for the WT S variant.

To visualize the accumulation of mutations in the VOC's S protein, we calculated the intersections of shared mutations between variants (Table S1; Figure S1). Accordingly, the UpSet plot shows how the Omicron variant accumulates the largest number of mutations (4), of which 28 are exclusive. Gamma accumulates eight unique mutations; Delta, seven mutations; Beta, six mutations; and Alpha, four mutations. Also, the degree of shared mutations between variants is low, with Alpha and Omicron being the variants that share more mutations, with four. The other VOC's pairs share a single mutation while the intersection of all variants also points to a single foundational mutation. This high diversity in the mutations accumulated in S protein across variants points towards separate evolutionary paths. This phenomenon can derive into variant-specific immune evasion mechanisms such as decreased antibody recognition. The fact that Omicron accumulates more than three times more mutations at S than the remaining VOCs indicates a greater potential for epitope disruption.

The accumulation of more variant-specific mutations in the S protein than shared mutations (Figure S1; Table S1) implies a potential development of specific epitope landscape in each variant (Tables 3, 4). Additionally, these variant landscapes are likely to differ from the patterns observed in the WT Wuhan variant. Considering epitope region conservation against the wild-type virus, the Alpha variant loses ER7; the Beta variant loses ER4 and ER7 but gains an epitope region at 828–845; the Gamma

TABLE 3 Epitope refinement on S protein in Wuhan and Alpha, Beta, Delta, Gamma, and Omicron variants.

Variant	ID	Predicted Epitopes	Curated Epitopes	Epitopic Regions	Epitope Refinement (%)	Epitope Conservation (%)	Epitopic Region Conservation (%)
Wuhan-2	WT	219	12	7	5.5	100	100
Alpha	B.1.1.7	206	13	6	6.3	108.3	85.7
Beta	B.1.351	225	11	6	4.9	91.7	85.7
Delta	P.1	213	15	7	7	125	100
Gamma	B.1.617.2	214	6	5	2.8	50	71.4
Omicron	B.1.1.529	230	11	6	4.8	91.7	85.7

Predicted epitopes correspond to the number of epitopes obtained using individual linear and structural predictors. Curated epitopes refer to refined epitopes obtained using Brewpitopes. Epitope refinement is the percentage of curated epitopes over the initial number of predicted epitopes obtained using individual state-of-the-art tools. Epitope regions result from the integration of overlapping predictions by different tools.
Epitope conservation refers to the percentage of refined epitopes shared between each variant and the WT S protein. Epitope region conservation refers to the percentage of epitope regions shared between each variant and the WT S protein.

variant loses ER2, ER3, and ER4; the Delta variant loses ER3, ER4, and ER6 but gains ER1, ER5, and ER8; and the Omicron variant loses ER2, ER3, and ER4 partially and ER7 entirely (Table 4; Figure 5).

In terms of epitope coverage, the major loss is prediction on Gamma (4%) and Omicron (2%) variants while Alpha and Beta loss is less than 1.5%. Differently, Delta gains 0.5% in epitope coverage in respect to WT due to the prediction of a large epitope. The differences in variant epitope landscape can be attributed to partial losses in antibody recognition. However, using Brewpitopes, a core of epitope regions conserved across variants could be identified (Table 5).

Discussion

In vivo antibody recognition is constrained by molecular features not frequently integrated in state-of-the-art B-cell epitope predictors. These include extracellular location of the epitope, absence of glycosylation coverage, and surface accessibility on the parental protein (Table 1). In Brewpitopes, we have implemented these features as filters to refine bioinformatic B-cell epitope predictions. Thus, Brewpitopes optimizes *in vivo* antibody recognition properties of predicted epitopes. The proteome-wide SARS-CoV-2 analysis demonstrates the obtainment of a refined set of epitopes with neutralizing potential in S protein and its conservation in VOCs (Alpha, Beta, Delta, Gamma, and Omicron). Additionally, we identified four proteins with candidate epitope regions for neutralization studies. As exemplified in this study, Brewpitopes is a ready-to-use tool to enhance the accuracy and response rates of bioinformatic B-cell epitope predictions for future public health emergencies such as the appearance of vaccine-resistant SARS-CoV-2 variants and other pathogenic threads.

Profiling of the B-cell epitope landscape in SARS-CoV-2 has been a research-intensive topic since the start of the COVID-19 pandemic for its implications in vaccine and therapeutic antibody development (Table 1) (14–22, 52, 53). However, none of the proposed strategies jointly integrates the prediction of subcellular location, glycosylation status, or 3D accessibility of the epitope as factors influencing antibody recognition. For this reason, Brewpitopes is a first-in-class pipeline thanks to a streamlined implementation of *in silico* predictors of biophysical constraints. Furthermore, the available methods can only predict linear or conformational epitopes separately, whereas with Brewpitopes, we propose an integration of both types of predictions into linear epitope regions using the Epiconsensus tool.

The filters implemented in Brewpitopes are based on computational predictions, such as CCTOP for subcellular location of protein regions or Net-N-glyc and Net-O-glyc for glycosylations. The usage of bioinformatic tools expands the applicability of Brewpitopes enabling *ab initio* predictions on the proteome of understudied organisms or new pathogens. These tools preclude the requirement of previous protein topology, glycosylation, and accessibility of experimental determinations. Thus, Brewpitopes can be implemented rapidly and without large resource requirements. However, relying on bioinformatic predictions inevitably implies at least a minimal degree of false positives and false negatives among the curated and discarded candidates.

In the case of glycosylation predictions, the dynamics of this type of PTM or its effects on neighboring epitopes cannot be assessed *in silico* using a sequence-based approach as Brewpitopes. In terms of structural accessibility, many candidates predicted by individual tools used in this study contained buried residues. This can limit the recognition of the candidates as compared to fully accessible epitopes (47). To minimize this effect, in Brewpitopes, we discard all epitopes containing a single buried residue (RSA <0.2). This criterion is the most stringent filter of the pipeline. In the case of S protein, it downsized the number of candidates from 137 to 14 (Figure 3; Table 2). As expected, after the implementation of this stringent filter, a proportion of epitopes discarded may still have antigenic activity. Still, since the objective of the pipeline is to

TABLE 4 Epitope regions identified in the WT S protein using Brewpitopes compared to the epitope regions of the variants of concern.

Variant	Wuhan_2	Alpha	Mutations	Glycosilations	Buried
Epitope Region 1	NA	NA	NA	NA	NA
Epitope Region 2	168-FEYVSQPFLMDLEGKQGN-185	164-TFEYVSQPFLMDLEGKQGNFK-184	NA	NA	NA
Epitope Region 3	244-LHRSYLTPGDSSSGWTA-260	248-PGDSSSGWT-256	NA	NA	NA
Epitope Region 4	470-TEIYQAGSTPCNGVEGFNCYFP-491	NA	NA	NA	472, 475, 487, 488, 491
Epitope Region 5	NA	NA	NA	NA	NA
Epitope Region 6	621-PVAIHADQLTPTWRVYSTGS-640	620-AIHADQLTPTWRVYSTGSNVFQT-642	NA	NA	NA
Epitope Region 7	809-PSKPS-813	NA	NA	NA	NA
Epitope Region 8	NA	828-AGFIKQYGDCLGDIAARD-845	NA	NA	NA
Epitope Region 9	1155-YFKNHTSPDVDLGDISGINASV-1176	1152-YFKNHTSPDVDLGDISGINASVVNIQKE-1179	NA	NA	NA
Epitope Region 10	1195-ESLIDLQELGKYEQYI-1210	1192-ESLIDLQELGKYEQYI-1207	NA	NA	NA
Variant	Wuhan_2	Beta	Mutations	Glycosilation	Buried
Epitope Region 1	NA	NA	NA	NA	
Epitope Region 2	168-FEYVSQPFLMDLEGKQGN-185	176-LMDLEGKQGNFK-187	NA	NA	168
Epitope Region 3	244-LHRSYLTPGDSSSGWTA-260	249-GDSSSGW-255	NA	NA	241*
Epitope Region 4	470-TEIYQAGSTPCNGVEGFNCYFP-491	NA	E484K	NA	472, 475, 480, 487, 488, 491
Epitope Region 5	NA	NA	NA	NA	NA
Epitope Region 6	621-PVAIHADQLTPTWRVYSTGS-640	620-AIHADQLTPTWRVYSTGSNVFQT-642	NA	NA	NA
Epitope Region 7	809-PSKPS-813	NA	NA	NA	806*
Epitope Region 8	NA	828-AGFIKQYGDCLGDIAARD-845	NA	NA	NA
Epitope Region 9	1155-YFKNHTSPDVDLGDISGINASV-1176	1152-YFKNHTSPDVDLGDISGINASVVNIQKE-1179	NA	NA	NA
Epitope Region 10	1195-ESLIDLQELGKYEQYI-1210	1192-ESLIDLQELGKYEQYI-1207	NA	NA	NA
Variant	Wuhan_2	Gamma	Mutations	Glycosilations	Buried
Epitope Region 1	NA	NA	L18F, T20N	17	NA

(Continued)

10.3389/fimmu.2023.1278534

TABLE 4 Continued

Variant	Wuhan_2	Gamma	Mutations	Glycosilations	Buried
Epitope Region 2	168-FEYVSQPFLMDLEGKQGN-185	NA	NA	NA	168
Epitope Region 3	244-LHRSYLTPGDSSSGWTA-260	NA	NA	NA	244, 246, 258
Epitope Region 4	470-TEIYQAGSTPCNGVEGFNCYFP-491	NA	E484K	NA	473, 475, 476, 487, 488, 489, 491
Epitope Region 5	NA	NA	NA	NA	NA
Epitope Region 6	621-PVAIHADQLTPTWRVYSTGS-640	621-PVAIHADQLTPTWRVYSTGS-640	NA	NA	NA
Epitope Region 7	809-PSKPS-813	809-PSKPS-813	NA	NA	NA
Epitope Region 8	NA	NA	NA	NA	NA
Epitope Region 9	1155-YFKNHTSPDVDLGDISGINASV-1176	1141-LQPELD-1146//1155-YFKNHTSPDVDLGDISGINASF-1176	NA	NA	NA
Epitope Region 10	1195-ESLIDLQELGKYEQYI-1210	1195-ESLIDLQELGKYEQYI-1210	NA	NA	NA
Variant	Wuhan_2	Delta	Mutations	Glycosilations	Buried
Epitope Region 1	NA	14-QCVNLRTRTQ-23	T19R	NA	NA
Epitope Region 2	168-FEYVSQPFLMDLEGKQGN-185	NA	NA	NA	173
Epitope Region 3	244-LHRSYLTPGDSSSGWTA-260	243-HRSYLTPGDSSSGWTA-258	NA	NA	NA
Epitope Region 4	470-TEIYQAGSTPCNGVEGFNCYFP-491	NA	T478K	NA	478
Epitope Region 5	NA	496-QPTNG-500	NA	NA	NA
Epitope Region 6	621-PVAIHADQLTPTWRVYSTGS-640	NA	NA	NA	631
Epitope Region 7	809-PSKPS-813	807-PSKPS-811	NA	NA	NA
Epitope Region 8	NA	827-ADAGFIKQYGDCLGDIAA-844	NA	NA	NA
Epitope Region 9	1155-YFKNHTSPDVDLGDISGINASV-1176	1136-YDPLQPELDSFKEELDKYFKNHTSPDVDLGDISGINASVVNIQKEIDRLNEVAKN-1190	NA	NA	NA
Epitope Region 10	1195-ESLIDLQELGKYEQYI-1210	1193-ESLIDLQELGKYEQYIKWPW-1212	NA	NA	NA
Variant	Wuhan_2	Omicron	Muts	Glycosilations	Buried
Epitope Region 1	NA	NA	NA	N17	NA
Epitope Region 2	168-FEYVSQPFLMDLEGKQGN-185	178-QGNFK-182	NA	NA	172

(Continued)

10.3389/fimmu.2023.1278534

10.3389/fimmu.2023.1278534

TABLE 4 Continued

Variant	Wuhan_2	Omicron	Muts	Glycosilations	Buried
Epitope Region 3	244-LHRSYLTPGDSSSGWTA-260	248-PGDSSSGWT-256	NA	NA	243*
Epitope Region 4	470-TEIYQAGSTPCNGVEGFNCYFP-491	467-TEIYQAGNKPCNGVAGFNCYFPL-489	S477N, T478K, E484A	NA	491
Epitope Region 5	NA	NA	G496S, Q498R	NA	495, 497
Epitope Region 6	621-PVAIHADQLTPTWRVYSTGS-640	627-TPTWRVYSTGSNVFQT-642	NA	NA	620*
Epitope Region 7	809-PSKPS-813	NA	NA	NA	815, 816*
Epitope Region 8	NA	NA	NA	NA	(…)
Epitope Region 9	1155-YFKNHTSPDVDLGDISGINASV-1176	1152-YFKNHTSPDVDLGDISGINASVVNIQKE-1179	NA	NA	NA
Epitope Region 10	1195-ESLIDLQELGKYEQYI-1210	1192-ESLIDLQELGKYEQYI-1207	NA	NA	NA

"Mutations", "Glycosylations", and "Buried" refer to the residues that cause the disruption of epitope regions in each viral variant.

TABLE 5 Epitope coverage of the WT S protein versus variants of concern.

Variant	Epitope Coverage (%)
Wuhan_2	9.43
Alpha	9.03
Beta	8.17
Delta	10.13
Gamma	5.42
Omicron	7.62

Epitope coverage is the percentage of the total protein sequence (the S protein) that is covered by curated epitope regions predicted using Brewpitopes. It estimates the antigenicity potential of a protein. The loss of epitope coverage in variants of concern is a proxy to estimate their immune escape potential due to the loss of in vivo antibody neutralization.

obtain the greatest immunogenicity enrichment in the refined candidates; we consider that this filter strongly serves this purpose. Complementarily, accessibility predictions depend on optimal structural resolution, which is difficult to obtain for highly flexible protein regions. To circumvent this, we labeled these regions as unmodeled, but due to their high flexibility, these were included as exposed regions and epitopes predicted within these passed the accessibility filter.

In terms of software flexibility, Brewpitopes is built upon Discotope2.0, and Bepipred2.0, which, during the pipeline development and SARS-CoV-2 analysis, were considered state of the art by the IEDB analysis resource tool (51). ABCpred was also included in the analysis, but it can no longer be considered a cutting-edge method. Accordingly, Brewpitopes succeeds in discarding a major quantity of candidates predicted by this tool. In addition, Brewpitopes' design flexibility enables a straightforward integration of new state-of-the-art methods and can be easily maintained to keep up with the fast evolution pace of the field.

While Brewpitopes can be applied to any protein or organism, given the wealth of SARS-CoV-2 data and biomedical interest, we focused on the analysis of this virus. We performed a proteome-wide analysis of the epitope landscape in SARS-CoV-2 to obtain a curated list of epitopes with neutralizing potential. To study the immune evasion mechanisms by SARS-CoV-2, we predicted the epitope profiles of WT S protein and we assessed how these were affected by variant-specific mutations. This comparison led to the discovery of six epitope regions conserved across variants, which could explain the conserved protection of vaccinated patients against new variants (54). In this line, the restrictive nature of Brewpitopes' filtering criteria led to a significant reduction of predicted epitopes on the S protein to be validated. This study serves as an example of the value of the pipeline in terms of experimental resource optimization.

The identification of potentially neutralizing epitopes in R1AB, R1A, AP3A, and ORF9C highlights the importance of studying proteome regions with low variability. Despite the fact that these proteins are not considered key for viral survival and cellular entry, the presence of extracellular regions accessible for antibody recognition supports their neutralizing potential. The

restricted viral evolution of these proteins can limit the advantage of variants in terms of antigen drift and immune escape while leading to greater vaccine protection rates.

Despite losses in epitope coverage observed in S protein variants, Brewpitopes could identify several epitope regions shared across variants. This finding has beneficial implications for vaccine efficacy versus new VOCs. Brewpitopes reported a lower epitope coverage loss for Omicron than for the Gamma variant. The epitope coverage loss predicted in Omicron versus Wuhan could partially explain the large loss of neutralization against this variant reported by previous studies (55). Discordances between neutralization studies (55) and the results of Brewpitopes can be explained by relevant differences between *in vitro* and *in silico* methods. As aforementioned, Brewpitopes' stringency could discard a proportion of truly antigenic epitopes and thus underrepresent the neutralization loss observed in Omicron.

Brewpitopes is a pipeline that refines bioinformatic B-cell epitope predictions straightforwardly for use against any target protein or organism's proteome. The integration of multiple state-of-the-art B-cell epitope algorithms coupled with the addition of *ab initio* predictions of important features for *in vivo* antibody recognition is a relevant advantage over existing pipelines and individual predictors. Furthermore, implementing Brewpitopes to the proteome of SARS-CoV-2 Wuhan WT variant versus VOCs, we have identified an epitope core in S protein conserved across variants and new antigenic regions in four SARS-CoV-2 proteins less prone to immune escape due to lower immune pressure and antigenic drift rates.

In conclusion, Brewpitopes is a streamlined pipeline that assesses biophysical properties not accounted for in state-of-the-art B-cell epitope predictors. The usage of *in silico* predictors of subcellular location, glycosylation status, and surface accessibility has been demonstrated as crucial to enrich the neutralization potential of predicted epitopes in SARS-CoV-2.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding authors.

Author contributions

RFD: Investigation, Methodology, Writing – original draft, visualization, conceptualization and investigation. RL-A: Investigation, Writing – original draft. EPP: Conceptualization, Supervision, Writing – review & editing, funding acquisition, methodology, validation and visualization. AT: Funding acquisition, Supervision, Writing – review & editing. LF-B: Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Validation, Visualization, Writing – original draft.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. RF-D received support by a La Caixa Junior Leader Fellowship (LCF/BQ/PI18/11630003) from Fundación La Caixa. EP-P received support by a La Caixa Junior Leader Fellowship (LCF/BQ/PI18/11630003) from Fundación La Caixa and a Ramon y Cajal fellowship from the Spanish Ministry of Science (RYC2019-026415-I). LF-B and RL-A received support by Direcció General de Recerca i Inovació en Salut (DGRIS) and BIOCAT (https://www.biocat.cat/ca) (Code: BIOCAT_DGRIS_COVID19) awarded to AT and LF-B; ISCIII-FOS (FI19/00090) grant awarded to RL-A, CB 06/06/0028/CIBER de enfermedades respiratorias (Ciberes), Ciberes is an initiative of ISCIII. ICREA Academy/Institució Catalana de Recerca i Estudis Avançats awarded to AT; 2.603/IDIBAPS, SGR/Generalitat de Catalunya awarded to AT. Funders did not play any role in project design, data collection, data analysis, interpretation, or writing of the paper.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2023.1278534/full#supplementary-material

SUPPLEMENTARY FIGURE 1
Mutations accumulated in the protein S of the Variants of Concern Alpha, Beta, Delta, Gamma and Omicron. Representation of unique and shared mutations of each variant. Total mutations per each variant are displayed in the lower barplot. The accumulation of mutations in the S protein of viral variants can be linked to a greater potential of immune escape due to the potential disruption of epitopes caused by changes in the sequence. Omicron stands out accumulating the 3 times more mutations than other variants.

SUPPLEMENTARY TABLE 1
Amino acid changes and sequence position of mutations in the S protein of variants of concern.

SUPPLEMENTARY TABLE 2
Comparison of glycosylation sites in S protein determined by MS or predicted using computational tools. Experimental sources include mass spectrometry glycosylation studies. *In silico* glycosylation sites were predicted using computational tools (Net-N-Glyc and Net-O-Glyc).

References

1. Cyster JG, Allen CDC. B cell responses: cell interaction dynamics and decisions. *Cell* (2019) 177:524–40. doi: 10.1016/j.cell.2019.03.016

2. Jespersen MC, Peters B, Nielsen M, Marcatili P. BepiPred-2.0: Improving sequence-based B-cell epitope prediction using conformational epitopes. *Nucleic Acids Res* (2017) 45(W1):W24–9. doi: 10.1093/nar/gkx346

3. Zhang Y, Banga Ndzouboukou JL, Gan M, Lin X, Fan X. Immune evasive effects of SARS-CoV-2 variants to COVID-19 emergency used vaccines. *Front Immunol* (2021) 12. doi: 10.3389/fimmu.2021.771242

4. *CoVariants*. Available at: https://covariants.org/.

5. Bahai A, Asgari E, Mofrad MRK, Kloetgen A, McHardy AC. EpitopeVec: Linear epitope prediction using deep protein sequence embeddings. *Bioinformatics* (2021) 37 (23):4517–25. doi: 10.1093/bioinformatics/btab467

6. Dobson L, Reményi I, Tusnády GE. CCTOP: A Consensus Constrained TOPology prediction web server. *Nucleic Acids Res* (2015) 43(W1):W408–12. doi: 10.1093/nar/gkv451

7. Walker JM. *Methods in molecular biology*. Available at: http://www.springer.com/series/7651.

8. Kringelum JV, Lundegaard C, Lund O, Nielsen M. Reliable B cell epitope predictions: impacts of method development and improved benchmarking. *PloS Comput Biol* (2012) 8(12). doi: 10.1371/journal.pcbi.1002829

9. Blasse C, Saalfeld S, Etournay R, Sagner A, Eaton S, Myers EW. PreMosa: Extracting 2D surfaces from 3D microscopy mosaics. *Bioinformatics* (2017) 33 (16):1–7. doi: 10.1093/bioinformatics/btx195

10. Raybould MIJ, Marks C, Lewis AP, Shi J, Bujotzek A, Taddese B, et al. Thera-SAbDab: the therapeutic structural antibody database. *Nucleic Acids Res* (2020) 48(D1): D383–8. doi: 10.1093/nar/gkz827

11. Zhou C, Chen Z, Zhang L, Yan D, Mao T, Tang K, et al. SEPPA 3.0 - enhanced spatial epitope prediction enabling glycoprotein antigens. *Nucleic Acids Res* (2019) 47 (W1):W388–94. doi: 10.1093/nar/gkz413

12. Ponomarenko J, Bui HH, Li W, Fusseder N, Bourne PE, Sette A, et al. ElliPro: A new structure-based tool for the prediction of antibody epitopes. *BMC Bioinf* (2008) 9. doi: 10.1186/1471-2105-9-514

13. Da Silva BM, Myung Y, Ascher DB, Pires DEV. Epitope3D: A machine learning method for conformational B-cell epitope prediction. *Brief Bioinform* (2022) 23(1). doi: 10.1093/bib/bbab423

14. Schwarz T, Heiss K, Mahendran Y, Casilag F, Kurth F, Sander LE, et al. SARS-CoV-2 proteome-wide analysis revealed significant epitope signatures in COVID-19 patients. *Front Immunol* (2021) 12. doi: 10.3389/fimmu.2021.629185

15. Cromer D, Juno JA, Khoury D, Reynaldi A, Wheatley AK, Kent SJ, et al. Prospects for durable immune control of SARS-CoV-2 and prevention of reinfection. *Nat Rev Immunol Nat Res* (2021) 21:395–404. doi: 10.1038/s41577-021-00550-x

16. Almofti YA, Abd-elrahman KA, Eltilib EEM. Vaccinomic approach for novel multi epitopes vaccine against severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2). *BMC Immunol* (2021) 22(1). doi: 10.1186/s12865-021-00412-0

17. Ezaj MMA, Junaid M, Akter Y, Nahrin A, Siddika A, Afrose SS, et al. Whole proteome screening and identification of potential epitopes of SARS-CoV-2 for vaccine design-an immunoinformatic, molecular docking and molecular dynamics simulation accelerated robust strategy. *J Biomol Struct Dyn* (2022) 40(14):6477–502. doi: 10.1080/07391102.2021.1886171

18. Sikora M, von Bülow S, Blanc FEC, Gecht M, Covino R, Hummer G. Computational epitope map of SARS-CoV-2 spike protein. *PloS Comput Biol* (2021) 17(4). doi: 10.1371/journal.pcbi.1008790

19. Khare S, Azevedo M, Parajuli P, Gokulan K. Conformational changes of the receptor binding domain of SARS-CoV-2 spike protein and prediction of a B-cell antigenic epitope using structural data. *Front Artif Intell* (2021) 4. doi: 10.3389/frai.2021.630955

20. VanBlargan LA, Adams LJ, Liu Z, Chen RE, Gilchuk P, Raju S, et al. A potently neutralizing SARS-CoV-2 antibody inhibits variants of concern by utilizing unique binding residues in a highly conserved epitope. *Immunity* (2021) 54(10):2399–2416.e6. doi: 10.1016/j.immuni.2021.08.016

21. Wang C, Li W, Drabek D, Okba NMA, van Haperen R, Osterhaus ADME, et al. A human monoclonal antibody blocking SARS-CoV-2 infection. *Nat Commun* (2020) 11(1). doi: 10.1038/s41467-020-16256-y

22. Corti D, Purcell LA, Snell G, Veesler D. Tackling COVID-19 with neutralizing monoclonal antibodies. *Cell* (2021) 184:3086–108. doi: 10.1016/j.cell.2021.05.005

23. Stoddard CI, Galloway J, Chu HY, Shipley MM, Sung K, Itell HL, et al. Epitope profiling reveals binding signatures of SARS-CoV-2 immune response in natural infection and cross-reactivity with endemic human CoVs. *Cell Rep* (2021) 35(8). doi: 10.1016/j.celrep.2021.109164

24. Fleri W, Paul S, Dhanda SK, Mahajan S, Xu X, Peters B, et al. The immune epitope database and analysis resource in epitope discovery and synthetic vaccine design. *Front Immunol* (2017) 8. doi: 10.3389/fimmu.2017.00278

25. Caoili SEC. Benchmarking B-cell epitope prediction for the design of peptide-based vaccines: Problems and prospects. *J Biomed Biotechnol* (2010) 2010. doi: 10.1155/2010/910524

26. Khan WH, Hashmi Z, Goel A, Ahmad R, Gupta K, Khan N, et al. COVID-19 pandemic and vaccines update on challenges and resolutions. *Front Cell Infect Microbiol* (2021) 11. doi: 10.3389/fcimb.2021.690621

27. Apweiler R, Bairoch A, Wu CH, Barker WC, Boeckmann B, Ferro S, et al. UniProt: The universal protein knowledgebase. *Nucleic Acids Res* (2004) 32 (DATABASE ISS.). doi: 10.1093/nar/gkh131

28. RCSB PDB. *Homepage*. Available at: https://www.rcsb.org/.

29. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. *Nature* (2021) 596(7873):583–9. doi: 10.1038/s41586-021-03819-2

30. Webb B, Sali A. Comparative protein structure modeling using MODELLER. *Curr Protoc Bioinf* (2016) 2016:5.6.1–5.6.37. doi: 10.1002/cpbi.3

31. Saha S, Raghava GPS. Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. *Proteins: Structure Funct Genet* (2006) 65(1):40–8. doi: 10.1002/prot.21078

32. Saha S, Bhasin M, Raghava GPS. Bcipep: A database of B-cell epitopes. *BMC Genomics* (2005) 6. doi: 10.1186/1471-2164-6-79

33. Faezov B, Dunbrack RL. PDBrenum: A webserver and program providing Protein Data Bank files renumbered according to their UniProt sequences. *PloS One* (2021) 16(7 July). doi: 10.1371/journal.pone.0253411

34. Tusnády E, Tusnády T, Istv´ I, Simon I. The HMMTOP transmembrane topology prediction server. *Bioinf Appl NOTE* (2001) 17:849–50. doi: 10.1093/bioinformatics/17.9.849

35. Shen H, Chou JJ. Membrain: Improving the accuracy of predicting transmembrane helices. *PloS One* (2008) 3(6). doi: 10.1371/journal.pone.0002399

36. Nugent T, Jones DT. Detecting pore-lining regions in transmembrane protein sequences. *BMC Bioinf* (2012) 3(1). doi: 10.1186/1471-2105-13-169

37. Viklund H, Elofsson A. OCTOPUS: Improving topology prediction by two-track ANN-based preference scores and an extended topological grammar. *Bioinformatics* (2008) 24(15):1662–8. doi: 10.1093/bioinformatics/btn221

38. Reynolds SM, Käll L, Riffle ME, Bilmes JA, Noble WS. Transmembrane topology and signal peptide prediction using dynamic Bayesian networks. *PloS Comput Biol* (2008) 4(11). doi: 10.1371/journal.pcbi.1000213

39. Käll L, Krogh A, Sonnhammer ELL. Advantages of combined transmembrane topology and signal peptide prediction-the Phobius web server. *Nucleic Acids Res* (2007) 35(SUPPL.2). doi: 10.1093/nar/gkm256

40. Viklund H, Elofsson A. Best α-helical transmembrane protein topology predictions are achieved using hidden Markov models and evolutionary information. *Protein Sci* (2004) 13(7):1908–17. doi: 10.1110/ps.04625404

41. Bernsel A, Viklund H, Falk J, Lindahl E, Von Heijne G, Elofsson A. *Prediction of membrane-protein topology from first principles* (2008). Available at: www.pnas.org/cgi/content/full/.

42. Kahsay RY, Gao G, Liao L. An improved hidden Markov model for transmembrane protein detection and topology prediction and its applications to complete genomes. *Bioinformatics* (2005) 21(9):1853–8. doi: 10.1093/bioinformatics/bti303

43. *NetNGlyc 1.0 - DTU health tech - bioinformatic services*. Available at: https://services.healthtech.dtu.dk/services/NetNGlyc-1.0/.

44. Steentoft C, Vakhrushev SY, Joshi HJ, Kong Y, Vester-Christensen MB, Schjoldager KTBG, et al. Precision mapping of the human O-GalNAc glycoproteome through SimpleCell technology. *EMBO J* (2013) 32(10):1478–88. doi: 10.1038/emboj.2013.79

45. Molsoft L.L.C. *ICM-browser* (2023). Available at: https://www.molsoft.com/icm_browser.html.

46. *GISAID - gisaid.org*. Available at: https://gisaid.org/.

47. Xu Z, Kulp DW. Protein engineering and particulate display of B-cell epitopes to facilitate development of novel vaccines. *Curr Opin Immunol* (2019) 59:49–56. doi: 10.1016/j.coi.2019.03.003

48. Wintjens R, Bifani AM, Bifani P. Impact of glycan cloud on the B-cell epitope prediction of SARS-CoV-2 Spike protein. *NPJ Vaccines* (2020) 5(1). doi: 10.1038/s41541-020-00237-9

49. Zobayer N, Hossain AA, Rahman M. A combined view of B-cell epitope features in antigens. *Bioinformation* (2019) 15(7):530–4. doi: 10.6026/97320630015530

50. Smith CC, Olsen KS, Gentry KM, Sambade M, Beck W, Garness J, et al. Landscape and selection of vaccine epitopes in SARS-CoV-2. *Genome Med* (2021) 13 (1). doi: 10.1186/s13073-021-00910-1

51. Vita R, Mahajan S, Overton JA, Dhanda SK, Martini S, Cantrell JR, et al. The immune epitope database (IEDB): 2018 update. *Nucleic Acids Res* (2019) 47(D1):D339–43. doi: 10.1093/nar/gky1006

52. Grifoni A, Sidney J, Zhang Y, Scheuermann RH, Peters B, Sette A. A sequence homology and bioinformatic approach can predict candidate targets for immune responses to SARS-CoV-2. *Cell Host Microbe* (2020) 27(4):671–680.e2. doi: 10.1016/j.chom.2020.03.002

53. Haynes WA, Kamath K, Bozekowski J, Baum-Jones E, Campbell M, Casanovas-Massana A, et al. High-resolution epitope mapping and characterization of SARS-CoV-2 antibodies in large cohorts of subjects with COVID-19. *Commun Biol* (2021) 22(1). doi: 10.1101/2020.11.23.20235002

54. Yi C, Sun X, Lin Y, Gu C, Ding L, Lu X, et al. Comprehensive mapping of binding hot spots of SARS-CoV-2 RBD-specific neutralizing antibodies for tracking immune escape variants. *Genome Med* (2021) 13(1). doi: 10.1186/s13073-021-00985-w

55. Dejnirattisai W, Shaw RH, Supasa P, Liu C, Stuart AS, Pollard AJ, et al. Reduced neutralisation of SARS-CoV-2 omicron B.1.1.529 variant by post-immunisation serum. *Lancet Elsevier B.V.* (2022) 399:234–6. doi: 10.1016/S0140-6736(21)02844-0

frontiers | Frontiers in Immunology

TYPE Original Research
PUBLISHED 08 December 2023
DOI 10.3389/fimmu.2023.1285899

OPEN ACCESS

EDITED BY
Helder Nakaya,
University of São Paulo, Brazil

REVIEWED BY
Ravi Kumar,
Central University of Haryana, India
Sudeep Kumar Maurya,
University of Pittsburgh Medical Center, United States

*CORRESPONDENCE
Li C. Xue
me.lixue@gmail.com
Mohammad Hossein Karimi-Jafari
mhkarimijafari@ut.ac.ir

†These authors have contributed equally to this work

RECEIVED 30 August 2023
ACCEPTED 17 November 2023
PUBLISHED 08 December 2023

CITATION
Parizi FM, Marzella DF, Ramakrishnan G, 't Hoen PAC, Karimi-Jafari MH and Xue LC (2023) PANDORA v2.0: Benchmarking peptide-MHC II models and software improvements.
Front. Immunol. 14:1285899.
doi: 10.3389/fimmu.2023.1285899

PANDORA v2.0: Benchmarking peptide-MHC II models and software improvements

Farzaneh M. Parizi[1,2†], Dario F. Marzella[1†], Gayatri Ramakrishnan[1], Peter A. C. 't Hoen[1], Mohammad Hossein Karimi-Jafari[2*] and Li C. Xue[1*]

[1]Medical BioSciences Department, Radboud University Medical Center, Nijmegen, Netherlands, [2]Department of Bioinformatics, Institute of Biochemistry and Biophysics, University of Tehran, Tehran, Iran

T-cell specificity to differentiate between self and non-self relies on T-cell receptor (TCR) recognition of peptides presented by the Major Histocompatibility Complex (MHC). Investigations into the three-dimensional (3D) structures of peptide:MHC (pMHC) complexes have provided valuable insights of MHC functions. Given the limited availability of experimental pMHC structures and considerable diversity of peptides and MHC alleles, it calls for the development of efficient and reliable computational approaches for modeling pMHC structures. Here we present an update of PANDORA and the systematic evaluation of its performance in modelling 3D structures of pMHC class II complexes (pMHC-II), which play a key role in the cancer immune response. PANDORA is a modelling software that can build low-energy models in a few minutes by restraining peptide residues inside the MHC-II binding groove. We benchmarked PANDORA on 136 experimentally determined pMHC-II structures covering 44 unique αβ chain pairs. Our pipeline achieves a median backbone Ligand-Root Mean Squared Deviation (L-RMSD) of 0.42 Å on the binding core and 0.88 Å on the whole peptide for the benchmark dataset. We incorporated software improvements to make PANDORA a pan-allele framework and improved the user interface and software quality. Its computational efficiency allows enriching the wealth of pMHC binding affinity and mass spectrometry data with 3D models. These models can be used as a starting point for molecular dynamics simulations or structure-boosted deep learning algorithms to identify MHC-binding peptides. PANDORA is available as a Python package through Conda or as a source installation at https://github.com/X-lab-3D/PANDORA.

KEYWORDS

peptide:MHC, MHC class II, peptide binding, 3D structures, large-scale 3D modelling

1 Introduction

The ability of T-cells to recognize and eliminate infected or transformed cells relies on their ability to distinguish between self and non-self peptides presented by the Major Histocompatibility Complex (MHC) on the surface of these cells. Upon recognition of a non-self peptide by T-cell receptors (TCR), T-cells activate and initiate an immune response. MHC class I (MHC-I) molecules typically present intracellular antigens to cytotoxic CD8+ T-cells, which eliminate the cell presenting the antigen. MHC-II molecules present extracellular antigens to helper CD4+ T-cells, which assist other immune cells by releasing cytokines and orchestrating the immune response (1, 2). To unravel the mechanisms of peptide presentation to T-cells and immune response, it is essential to investigate how peptides bind to MHC molecules.

Understanding the mechanism of peptide-MHC (pMHC) binding raises an intriguing research question regarding how MHC molecules effectively bind to a wide range of peptides while maintaining strong binding and specificity. Previous research focusing on the structural aspects of pMHC complexes has provided valuable insights into our understanding of antigen presentation specificity (3) and peptide binding dynamics (4, 5). Allele-specific residues at anchor positions and complementary pockets in the MHC molecule play a significant role in determining the promiscuity and specificity of peptide recognition by MHC molecules (6, 7). Notably, the presence of hydrophobic anchors and the formation of hydrogen bonds have been discovered to stabilize the pMHC-II interaction (8, 9). Similarly, in the case of MHC class I, peptide-dependent stability is achieved through the establishment of conserved hydrogen bonds at the N and C termini of peptides, along with anchor residues that fit into pockets of MHC class I (10, 11). Furthermore, structural investigations have provided insights into other mechanisms, such as the molecular basis of autoimmune diseases (10) and T-cell recognition (11, 12). The knowledge gained from structural studies has also facilitated the design of novel therapies and can help the development of effective vaccine strategies (13, 14). Therefore, access to structural information on pMHC is crucial for these advancements.

This work focuses on pMHC-II binding. MHC-II is crucial in antigen presentation, particularly for extracellular antigens. Additionally, MHC-II mediated CD4+ T-cell responses are reported to account for the predominant immune responses following cancer vaccine treatment (15–18). The MHC-II complex consists of two membrane-anchored chains: an α- and β-chain (Figure 1), and it can bind peptides up to 25 residues in length (20, 21). The binding groove of MHC-II can hold a 9-mer core (22). The residues outside the groove form the Peptide Flanking Regions (PFR), namely the left (N-terminus) and right (C-terminus) PFRs. A peptide is kept in place within the groove by three or four main conserved binding pockets: Pockets 1, 4, 6, and 9 (Figure 1A, and alongside these, there are smaller auxiliary anchor pockets (23, 24).

To accommodate a diverse range of antigens within the MHC groove, the MHC locus stands out as the most polymorphic region in the human genome (2, 25). With over 10,754 alleles for MHC class II, there is a significant variation in MHC-II alleles and the peptides they can bind (26). Unfortunately, only a few pMHC-II structures have been experimentally resolved [about 240 entries in the PDB, the Protein Data Bank (27)]. This necessitates the development of fast, structure-based computational modeling methods to overcome the scarcity of available pMHC-II structures. However, only a few modeling methods have been explicitly developed for pMHC-II complexes.

Most existing pMHC-II modelling methods rely on grid-based docking, including pDock and EpiDock (28–33). Among them, pDock has demonstrated improved performance in generating peptide core conformations bound to MHC-II. The pDock's approach involves receptor modeling followed by flexible peptide docking into the binding groove while retaining its starting conformation using loose restraints. Current pMHC-II modeling approaches are often limited in terms of usability due to: 1) long computation times; 2) the use of closed-source software; 3) limited

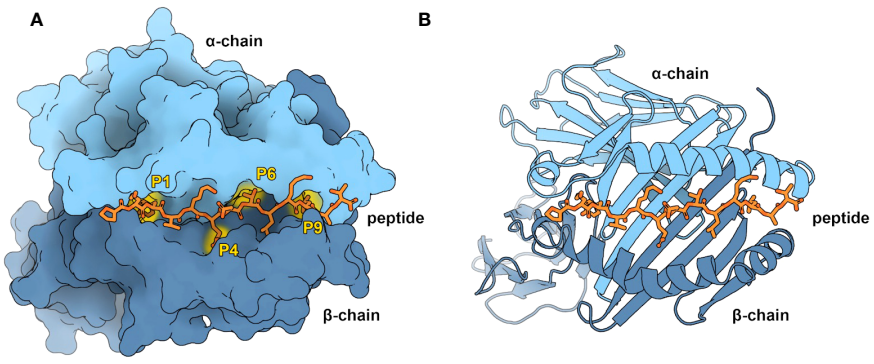


FIGURE 1
Overview of the pMHC-II complex **(A)** Representation of an MHC-II molecule by its accessible surface area, visualized with Protein Imager (19). MHC-II consists of an α-chain (light blue) and a β-chain (dark blue). Shown are the four characteristic pockets in the binding groove (P1, P4, P6, and P9), occupied by the corresponding peptide (orange) anchor residues. As modeling restraints (yellow), PANDORA uses the atomic contacts between the peptide anchor residues and the MHC-II pockets. **(B)** Cartoon representation of pMHC-II. The peptide binding groove consists of two α-helices on a floor of β-sheets, in which the peptide resides (PDB ID: 1DLH).

coverage of diverse MHC alleles; and 4) uncertainty regarding the quality of PFR conformations. Additionally, structural modelling of pMHC-II complexes is fraught with challenges. It is not always clear which region of a peptide forms the core and is directly anchored to the MHC-II receptors (34, 35). Existing methods, such as NetMHCIIpan-4.0 (36), can provide reasonably accurate predictions for the binding core. Furthermore, the flexibility of PFRs poses additional hurdles. To address these challenges, the development of fast and pan-allelic pMHC-II modelling software is required to integrate prediction of the binding core and generation of plausible conformations for the entire peptide bound to MHC-II.

We present here the utility and performance of our pMHC modelling software, PANDORA v2.0, for pMHC-II modeling and its new version updates. We have earlier demonstrated PANDORA's reliable performance for modeling pMHC-I complexes (37, 38). PANDORA leverages two pieces of domain knowledge: 1) the high conservation of MHC structures and 2) the anchoring of peptides to the main pockets of MHC molecules (Figure 2). We benchmarked PANDORA on 136 experimentally resolved pMHC-II structures, including mouse alleles. When compared with an existing pMHC-II modelling technique, pDock (32), and also with AlphaFold (39), we show that PANDORA outperforms these methods in terms of generated model quality and computational efficiency. Additionally, we evaluate the effectiveness of the anchor prediction tool used in our approach (NetMHCIIpan-4.0). PANDORA's quality and speed show the potential for boosting structure-based Deep Learning (DL) algorithms, making it a valuable tool in developing effective vaccine designs. We also discuss the existing limitations of anchor predictions and propose the integration of a structural and physics-based anchor predictor as a potential solution. Furthermore, we highlight the importance of further research in the modeling of post-translational modifications (PTMs) on peptide-MHC interactions.

2 Materials and methods

2.1 Structural template set building

Building the template dataset is similar to our previous work (37) and is expanded to make it suitable for pMHC-II. PANDORA retrieved structures of pMHC-II complexes from IMGT/3Dstructure-DB (40) and filters for those with peptides of lengths between 7 and 25 residues. Structures including the DM chaperone and the CLIP peptide, both known to affect the MHC-II conformation, are discarded (21, 41). The MHC-II alpha chain is renamed as chain "M" and the beta chain is renamed as chain "N" to make a distinction from MHC-I β2-microglobulin which is renamed as chain "B". The peptide chain is renamed as chain "P". For the benchmark experiment presented in this work, the parsing resulted in a total of 136 pMHC-II templates, spanning over 32 α chain alleles, 81 β chain alleles, and a total of 44 unique MHC-II αβ pairs (see details in Supplementary Table S1).

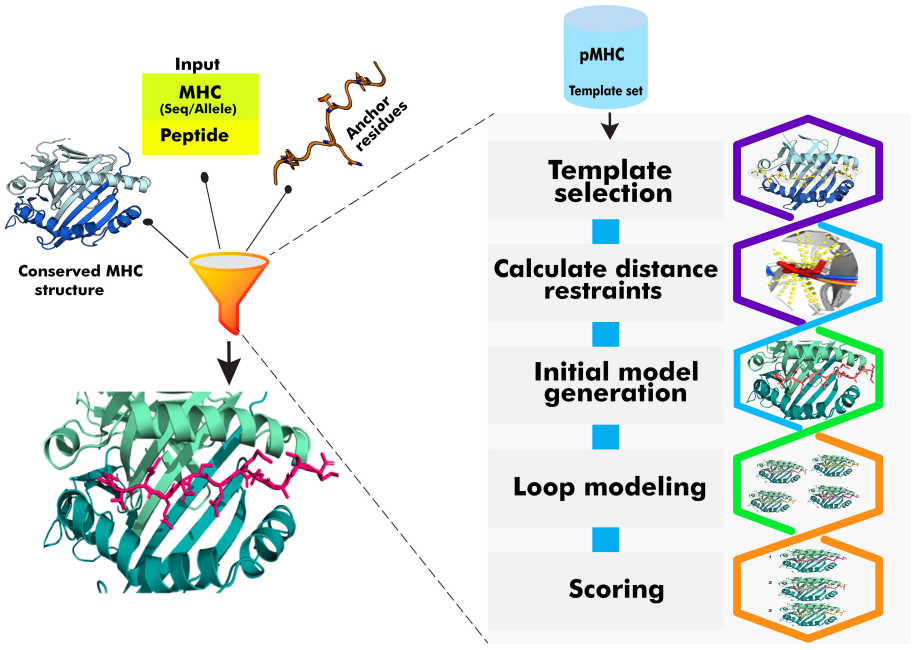


FIGURE 2
Overview of the PANDORA pMHC modelling framework. PANDORA as an integrative modelling protocol, leverages two domain knowledge aspects: the highly conserved nature of MHC structures and the binding of peptides to MHC pockets with anchor residues. PANDORA takes the sequence information of a target peptide and MHC as input and selects a template pMHC structure from a template set based on sequence similarity. The target peptide core is superposed onto the template peptide core. In flexible mode, it applies distance restraints for anchor residues. The framework performs loop modeling of flexible regions and energy minimization of pMHC-II conformations. Conformations are ranked to select those resembling the near-native conformation.

2.2 BLAST databases generation

BLAST (v2.10) is used to assign allele names (needed by NetMHCIIpan-4.0 for predicting binding cores) to the MHC sequences provided by the user and, independently, for the template selection step. The current version of PANDORA uses two BLAST databases. The first one (BLAST-DB1) is generated from the manually curated MHC sequences taken from https://www.ebi.ac.uk/ipd/, and it is used to assign the allele name to any MHC sequence provided by the user. This allele name will later be used as input for NetMHCIIpan4.0 to predict the binding core (see Template selection). The second one (BLAST-DB2) is generated from the template set sequences extracted by the PDB files retrieved as described above, and it is used for the template selection step.

2.3 Template selection

The template selection step has been updated from the first version of PANDORA (which used allele type names to identify templates) to a BLAST-based template selection. First, the target MHC sequences are queried against the BLAST-DB2 database with default parameters, and the results are ranked by percentage sequence identities. Templates sharing the highest sequence identity with the target sequences are selected and further ranked by peptide alignment score. Our peptide alignment method includes alignment of the binding core of the peptides followed by the addition of gaps at both their termini to account for different peptide lengths. The binding cores of the templates are derived from their corresponding structures. The binding core for the query peptide is predicted by NetMHCIIpan4.0. The peptides' alignments are then scored using a PAM30 substitution matrix. The highest-ranking template is then selected for modeling.

2.4 Modeling

We perform 3D modeling as described previously. For MHC-II, we restrain four anchor positions (P1, 4, 6, and 9) while keeping the peptide flanking regions flexible during the modeling step. In the default mode for pMHC-II cases, the whole peptide core is kept fixed as the template conformation. PANDORA v2.0 also supports restraints-flexible modelling mode for the peptide core, where users can provide anchors' restraints standard deviation, thereby specifying the extent of deviation of restraints from those in the templates in Angstroms. By default, 20 (adjustable) 3D models are produced, which are ranked by MODELLER's (42) internal molpdf score.

2.5 L-RMSD calculation

The L-RMSD is calculated as described in (43) as the backbone L-RMSD (including only the backbone atoms N, Cα, C, and O). We calculate "Core L-RMSD" for the binding core residues of the peptide, "Flanking L-RMSD" for the flanking regions of the peptide (i.e., the residues at the N-terminal of the first anchor and at the C-terminal of the fourth anchor), and "Whole L-RMSD" for all the residues of the peptide. The lower the L-RMSD, the better a model is.

3 Results

3.1 Modeling performance on the benchmark set

We benchmarked PANDORA's performance in reproducing X-ray crystal structures of pMHC-II complexes from the template set (n = 136). We carried out a leave-one-out validation approach where we iteratively removed a structure from the template database and allowed PANDORA to predict the pMHC-II complex using sequence and anchor information. To rule out the impact of anchor predictions, the anchor positions provided to PANDORA in this experiment were obtained from the target experimental structure to assess the modelling quality (see discussion on anchor prediction effects in the "NetMHCIIpan's anchor prediction" section).

We analyzed the distribution of the best model (i.e, the model with the lowest L-RMSD) conformations obtained for the whole and core peptide regions (Figures 3A, C, E; detailed information on different RMSD values is reported in Supplementary Table S2). The results demonstrate that for 91.1% (125 out of 136) cases, PANDORA was able to sample at least one high-quality model (whole peptide L-RMSD < 2 Å) with an overall mean L-RMSD of 1.11 ± 0.86 Å (i.e, Figure 3B). A small number of cases (11 out of 136) showed a relatively higher whole peptide L-RMSD of > 2 Å (see Figures 3E, F, and "The PFR Conformation Evaluation" section). We investigated the distribution of whole and core L-RMSDs over various peptide lengths, as illustrated in Figure 3A. Our analysis reveals a correlation between peptide lengths and the L-RMSD values, with longer peptides exhibiting higher L-RMSD values (Supplementary Figures S1A, B).

Furthermore, in terms of model ranking, we examined the performance of PANDORA by reporting L-RMSD for the top-ranked model (i.e., the conformation ranked as the top model using molpdf scoring function) (Figures 3D, F; for details, see Supplementary Figure S2). Our results show that PANDORA achieved an 85% success rate (L-RMSD < 2 Å) for the top 5 ranked models in the entire template set.

3.2 PANDORA generates low-energy conformations for the binding core

With four anchor positions in the binding groove, the structure of pMHC-II is well-suited for a restraint-based modelling approach. With the default mode (see Modelings in Methods), PANDORA demonstrates high accuracy in reproducing high-quality core conformations, with an average core L-RMSD of 0.49 ± 0.27 Å

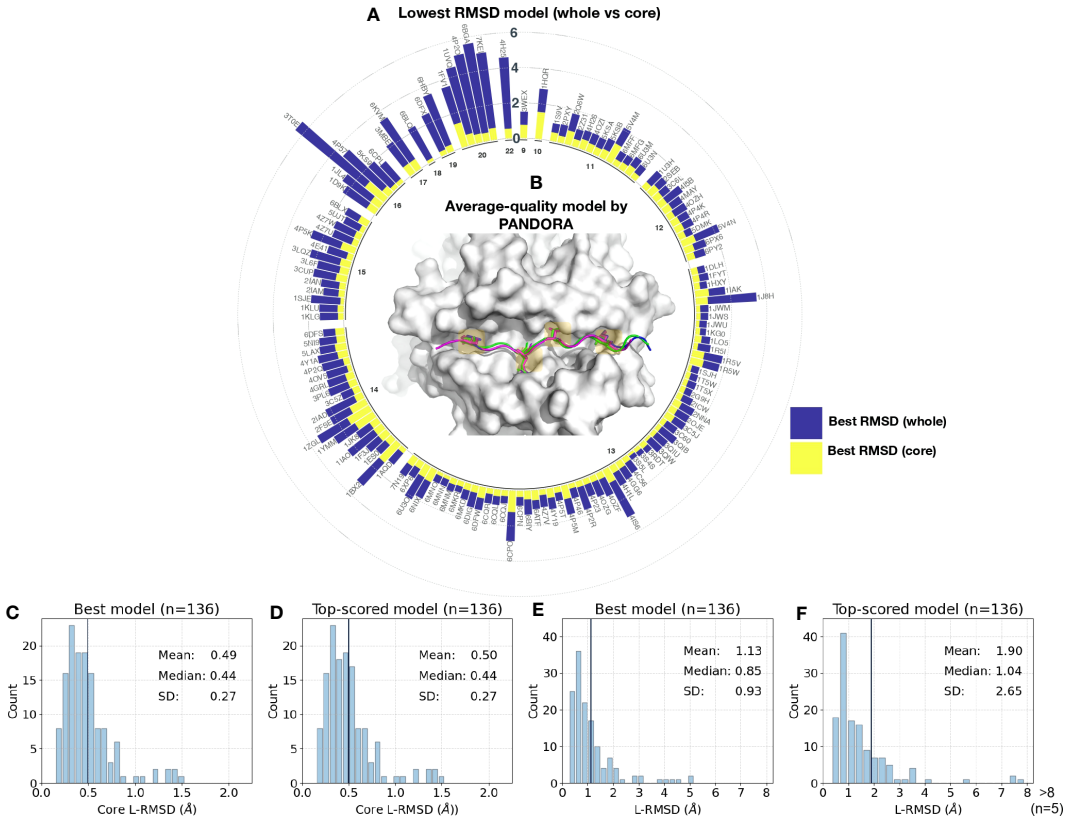


FIGURE 3
Benchmark results on reproducing 136 pMHC-II complexes with X-ray structures. **(A)** Sampling performance of the PANDORA benchmark experiment. The conformation with the lowest RMSD was chosen as the best RMSD model. A circular bar plot grouped based on peptide length (represented by the numbers in the inner circle) reports the lowest backbone L-RMSD (Y-axis) for the whole peptide (navy) and binding core (yellow). **(B)** An example of an average-quality 3D model generated by PANDORA. The target peptide (PDB ID: 4I5B) is marked in green; the template structure (PDB ID: 2OJE) is marked in magenta; and the PANDORA model structure (best conformation among the top 5 ranked) is marked in darkblue. **(C, D)** Histogram of the lowest backbone L-RMSD models in the peptide binding core vs. the whole peptide. **(E, F)** Complete performance of PANDORA (modeling + scoring). Histogram for the top-ranked models by PANDORA in terms of backbone L-RMSD on the peptide binding core vs. the whole peptide.

(93.38% of the cases having an L-RMSD < 1 Å) (Figures 3C, E). The fully-flexible mode, which allows for flexibility in the peptides' binding core, yielded an average core L-RMSD of 0.47 ± 0.2 Å (Supplementary Figure S3). However, the restraints-flexible mode increases the computational time by 90%, while marginally enhancing the overall quality (~ 6.46 min/case in the fully flexible mode vs. 3.75 min/case in the default mode).

3.3 Comparisons with AlphaFold and pDock

We compared PANDORA's performance against existing approaches, such as pDOCK (32) and AlphaFold (39). To assess the general performance of the pipeline, we used NetMHCIIpan's predicted anchor positions for this comparison.

pDock uses the ICM (Internal Coordinate Mechanics) algorithm to perform a flexible peptide docking into the MHC binding groove. During docking, the position of the peptide is only loosely constrained so that it retains a conformation close to its initial structure. For comparisons against pDock, we modeled pMHC-II complexes using PANDORA for the cases reported by Khan and Ranganathan (32). We obtained a mean L-RMSD of 0.27 ± 0.07 Å for Cα core while pDock achieved 0.59 ± 0.24 Å (Table 1). pDock retained RMSD estimates by redocking experimental pMHC X-ray structures; thus, the core residues are referred to as a priori. PANDORA automatically predicts anchor residues (using NetMHCIIpan-4.0 (36)) and a suitable template, generating higher-quality peptide core conformations. We did not use pDock to perform cross-docking on our template set since pDock is not publicly available for download and usage.

We also compared PANDORA with one of the best AI methods available for protein structure predictions, i.e., AlphaFold. AlphaFold is an advanced deep neural network approach that achieves unprecedented accuracy in protein folding predictions (44). However, since AlphaFold relies on sequence conservation information, it performs poorly on proteins where such information is absent, such as antibody-antigens and peptides (e.g., synthetic peptides or frame-shift mutated peptides) (45). For an objective comparison, we chose to use a version of AlphaFold that also uses templates to predict MHC structures (colabfold (44)). Our comparison shows that not all Alphafold-generated pMHC-II conformations have the correct anchor positions. Out of four randomly selected cases (Figures 4A–D), in two cases (PDB ID:

TABLE 1 Comparison of PANDORA and pDock in pMHC-II modelling.

PDB	PANDORA's best core Cα L-RMSD (Å)	pDOCK Cα core L-RMSD(Å)*
1FYT	0.38	0.35
1KLU	0.30	0.59
1T5W	0.24	0.65
1PYW	NA	0.32
1SJE	0.21	0.37
1AQD	0.24	1.01

NA, not available. At this stage, PANDORA could not model one case as the peptide sequence includes two non-canonical residues not handled by MODELLER.
The calculated core Cα L-RMSD (Å) on modeling 5 pMHC-II complexes using integrative homology modeling and a grid-based docking method. PANDORA's best model quality is compared to pDock as no pDock scoring function was disclosed in pDock so it seems that pDock reported the best RMSDs in their paper. *Data extracted from Khan & Ranganathan (32).

3C5Z and 6PX6), AlphaFold was unsuccessful in predicting the peptide's conformation with the correct anchor residues (Figures 4A, C). Notably, they were both part of AlphaFold's training set. This is mainly because PANDORA correctly identified the binding core for the four cases using NetMHCIIpan binding core predictions (see "NetMHCIIpan4.0 performance" in the next section).

Additionally, considering computational cost, PANDORA outperforms AlphaFold (regarding resources) and pDock. PANDORA is much more efficient considering template selection, anchor prediction, and modeling require ~3-4 minutes (from 3.75 to 6.46 minutes per case, depending on the mode, with shorter times for pMHC-I) on one core from an Intel(R) Xeon(R) Gold 6142 CPU @ 2.60 GHz. While pDock reported requiring 10 minutes for modeling on 2 CPUs 3.20 GHz (without homology modelling). Also, AlphaFold requires a significant amount of computation power-up to 18 GB of GPU power and 20 minutes to model a single pMHC case.

3.4 The impacts of binding core prediction on PANDORA model quality

The interaction between the peptide binding core and the MHC binding groove directly impacts the quality of a model; therefore, choosing the correct binding core is critical. In the absence of user-

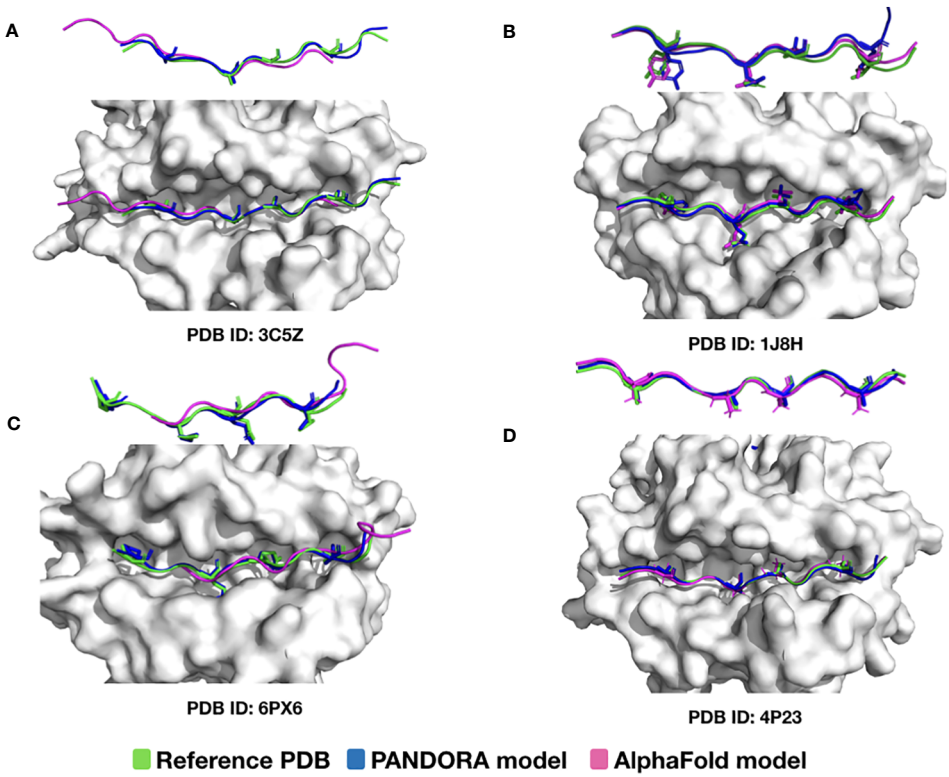


FIGURE 4
Comparison of PANDORA and AlphaFold in reproducing pMHC-II complexes. The peptide conformations are colored as follows: reference PDB in green, AlphaFold model in magenta, and PANDORA model in blue. The AlphaFold models were generated by Colab-fold using template-based modelling (default 5 top models generated). The presented model for AlphaFold and PANDORA is the best model (lowest L-RMSD model) among the top-5 ranked models. Two overviews of the peptide backbone conformation from two different angles are shown for each case (anchor residues are shown with side-chains). Overall, AlphaFold models have quite good backbone predictions (probably due to the usage of templates), but in 2 out of 4 cases, the peptide core conformations are shifted. **(A)** 24-mer peptide binding to H2-AB1*01 (PDB ID: 3C5Z); however, the predicted core conformation generated by AlphaFold is shifted by 3 residues; **(B)** 13-mer peptide binding to HLA-DRA*0101, HLA-DRB1*0401 (PDB ID: 1J8H) the binding core is accurately identified by AlphaFold **(C)** 12-mer peptide binding to HLA-DQA1*0201, HLA-DQB1*0201 (PDB ID: 6PX6); however, the predicted AlphaFold binding core conformation is shifted by 4 residues; **(D)** 13-mer peptide binding to H2-AB1*01 (PDB ID: 4P23) and the binding core is accurately identified by AlphaFold. Considering that **(A, B, D)** were already in AlphaFold training, it is noteworthy that A is predicted with an incorrect binding core.

defined anchor residues, PANDORA uses NetMHCIIpan to predict the binding core. Hence, we evaluate NetMHCIIpan's binding core prediction accuracy by comparing its predictions to known cores from experimental PDB structures. Our results show that the anchors were incorrectly predicted in 33 of the 136 cases in the benchmark dataset. In most cases, the observed shifts were by one or two residues (26 of 33), but misalignments of up to 8 residues were also observed (Supplementary Figure S4).

3.5 PANDORA as a pan-allele modelling method

Owing to the high structural similarity across MHC-II alleles, it is possible to model pMHC-II complexes using different MHC-II alleles as templates. Our results show that even when a template with the same MHC allele type for either of the chains was not available in the template set (25% of cases), PANDORA was still able to provide models with a mean L-RMSD of 0.86 Å for the best-RMSD models and 1.05 Å for the top 10 ranked models (Figure 5).

3.6 Software improvements

PANDORA v2.0 includes major improvements from the first release:

Frontend (User side):

- Capability to use MHC-sequence as input instead of only allele name, leading to much broader allele coverage than version 1.0.
- Addition of command-line interface for easier accessibility and bash integration.
- Addition of restraints-flexible modelling mode to avoid small clashes caused by rigid restraints (see Materials and Methods).
- Improvements in the python user interface.
- Easier software and database installation.
- Addition of an option to remove or keep beta2-microglobulin in the generated models, as Beta2-microglobulin can be crucial or not, depending on what the models will be used for (MD, AI, manual exploration, etc.).

Backend (internal software side):

- BLAST-based template selection instead of allele-name based template selection.
- Addition of a reference sequence database for allele names and MHC sequence automatic retrieval.
- The allele name is now automatically retrieved with BLAST when only the sequence is provided.
- Improvement in the MHC-II template parsing to prevent multiple structures from being discarded or from missing the allele name.
- Addition of parallelization and minor optimization improvements for the template set generation, drastically increasing its speed.

4 Discussion

PANDORA is a 3D modelling software for both pMHC-I and -II. Here we evaluated the performance of PANDORA on reliably generating peptide conformations binding to MHC class II complexes alongside the software improvements. We applied homology-driven restraint-based modelling to reduce the computational time during sampling (3.75 min/case on one CPU core). The proposed method was tested on 136 complexes, making it the largest modeling effort of pMHC-II complexes to date. Our results show that PANDORA was able to effectively model these complexes, achieving an 85% success rate (L-RMSD < 2 Å) for the top 5 ranked models in the entire template set and generating particularly high-quality peptide core conformations.

PANDORA outperforms pDock (32) and AlphaFold (39) regarding computational time and core L-RMSD values. PANDORA incorporates domain knowledge into the modeling. In contrast, AlphaFold is a general protein structure prediction method that relies on sequence conservation information, and conservation on the peptide side has little or no bearing on this binding. Our comparison shows that not all AlphaFold-generated pMHC-II conformations have the correct anchor positions. AlphaFold's higher computational cost is a major impediment to model millions of pMHCs, whereas PANDORA is a more practical choice.

PANDORA has the following unique features: 1) Fast: enabling high-throughput modeling of 3D pMHC-II complexes; 2) Reliable: generating low-energy models; 3) Efficient: With the use of anchor distance restraints, it to work on both MHC-I and MHC-II; 4) Template availability: providing an extensively cleaned template database of pMHC complexes, valuable for reliable homology modeling; 5) Highly Modular: It is easy to customize or extend; 6) Pan-allele: User may include MHCs from different species.

PANDORA has a user-friendly interface allowing users to incorporate new configurations such as 1) more extensive sampling (especially with longer peptides); 2) specification of secondary structure restraints (23% of benchmark cases formed beta-strand PFR, Supplementary Figure S5D); 3) fully fixed mode vs. flexible mode for the core conformation; 4) Manually defining the anchor residues; 5) Possibility of changing to other anchor predictor software. Its highly modular framework (Supplementary Figure S7) facilitates future community-wide development.

Knowledge of the peptide binding core is required to generate the pMHC-II complex structure. When the user doesn't input the anchor residues' position, PANDORA currently relies on NetMHCIIpan-4.0 as an anchor predictor (36). This software has a limited, yet large, set of available MHC alleles to utilize, and it can sometimes fail to predict the correct binding core (Supplementary Figure S4). Using an anchor predictor relying on structural and physics-based data could overcome these limitations for the pMHC anchor prediction, allowing for more accurate, pan-allelic anchor predictions.

PFR can influence TCR interactions (46–49); introducing a modeling program to generate credible PFR conformations is an important step forward. It is important to note that a singular X-ray structure exclusively depicts only one snapshot of the complex conformation. This implies that a method could generate a possible

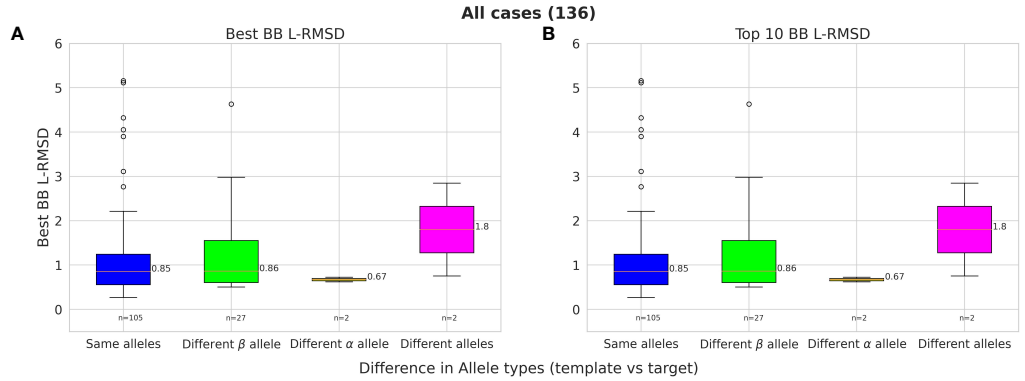


FIGURE 5
The effect of modelling with a different template allele-type and its effect on performance. Given two allele-type for each α- and β-chains of the template and target pMHC-II, 4 different scenarios are compared in each box-plot column; 1) Both allele-type is the same for template and target (blue); 2) Only the Alpha allele-type the same (green); 3) Only Beta allele-type the same (orange); and 4) both chain allele-types are different (pink). **(A)** Best L-RMSD models and **(B)** Top-ranked models using scoring.

PFR conformation that is not currently cataloged in the PDB but holds biological significance. To address this issue, PANDORA generates an ensemble of near-native conformations (top N-ranked conformations).

Further work is needed to model the post-translational modifications (PTM) in peptides binding to MHC, which have been shown to modulate antigen presentation and recognition (50, 51) and, moreover, PTMs on peptides increase the vast number of possible pMHC combinations. PTMs have a structural impact on the stability of pMHC complexes and the consequent modulations of immune responses (52). Although it has not yet been extensively evaluated within our framework, we recognize its potential benefits for the field and remain committed to conducting additional research and possibly incorporating this method into our future research.

While PANDORA provides energy scores, its primary focus is on 3D modeling rather than predicting binding affinity, it might be possible to utilize the energy scores from PANDORA models or running molecular dynamics on PANDORA models to gain insights into MHC binding specificities. In addition, PANDORA can potentially contribute to advancing our understanding of cancer biology, particularly in unraveling the impact of peptide mutations on MHC binding and the exposure of peptide side chains to T-cells or (see * marked cases in Supplementary Tables S1, S2). Although not intended for neoantigen identification, PANDORA was used to evaluate the effects of point mutations on a melanoma tumor antigen. PANDORA accurately modeled both peptides' side chains (see Supplementary Figure S6), resulting in high-affinity energy scores and a slight improvement in mutant binding.

In conclusion, the ability of PANDORA to generate high-quality peptide conformations within the MHC-II binding groove lends great reliability to the models employed for analyzing molecular interactions at the atomic level. Due to PANDORA's computational efficiency, initial conformations for molecular dynamics simulations can be quickly built.

It is now feasible to enrich the actively accumulating wealth of pMHC binding affinity and mass spectrometry data with physics-based PANDORA models and aid structure-boosted artificial intelligence algorithms in identifying antigenic peptides (for example, by training the deep learning framework DeepRank on these 3D models). As such, it can be leveraged to identify cancer neoantigens or viral antigenic peptides that hold promise as vaccine candidates. It will therefore pave the way for developing novel cancer immunotherapies.

Data availability statement

The template database and package are available at DOI: 10.5281/zenodo.6373630 PANDORA is available as a Python package on conda at https://anaconda.org/csb-nijmegen/csbpandora. The package source code and the documentation are available on GitHub at https://github.com/X-lab3D/PANDORA. Further inquiries can be directed to the corresponding authors.

Author contributions

FP: Conceptualization, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. DM: Conceptualization, Data curation, Investigation, Methodology, Software, Writing – original draft, Writing – review & editing. GR: Formal Analysis, Validation, Visualization, Writing – review & editing. PH: Conceptualization, Funding acquisition, Project administration, Resources, Supervision, Writing – review & editing. MK-J: Formal Analysis, Supervision, Visualization, Writing – review & editing. LX: Conceptualization, Formal Analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Validation, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. LX, FP, and DM received support from the Hypatia Fellowship from Radboudumc (Rv819.52706). GR is supported by Europees Fonds voor Regionale Ontwikkeling (EFRO) (R0005582).

Acknowledgments

The authors would like to thank Giulia Crocioni, Derek van Tilborg, and Shahiel Maassen for their help with the software development. We also thank Hassan Rasouli and Yannick J.M. Aarts for their assistance in plot visualization.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2023.1285899/full#supplementary-material

References

1. Janeway C. *Immunobiology: the immune system in health and disease. 5. ed.* New York, NY: Garland Publ (2001). p. 732.

2. Blum JS, Wearsch PA, Cresswell P. Pathways of antigen processing. *Annu Rev Immunol* (2013) 31(1):443–73. doi: 10.1146/annurev-immunol-032712-095910

3. Madden DR. The three-dimensional structure of peptide-MHC complexes. *Annu Rev Immunol* (1995) 13(1):587–622. doi: 10.1146/annurev.iy.13.040195.003103

4. Abualrous ET, Stolzenberg S, Sticht J, Wieczorek M, Roske Y, Günther M, et al. MHC-II dynamics are maintained in HLA-DR allotypes to ensure catalyzed peptide exchange. *Nat Chem Biol* (2023) 19:1–9. doi: 10.1038/s41589-023-01316-3

5. Thomas C, Tampé R. Structure of the TAPBPR–MHC I complex defines the mechanism of peptide loading and editing. *Science* (2017) 358(6366):1060–4. doi: 10.1126/science.aao6001

6. Zhang C, Anderson A, DeLisi C. Structural principles that govern the peptide-binding motifs of class I MHC molecules. *JMB* (1998) 281(5):929–47. doi: 10.1006/jmbi.1998.1982

7. Sinigaglia F, Hammer J. Defining rules for the peptide-MHC class II interaction. *Curr Opin Immunol* (1994) 6(1):52–6. doi: 10.1016/0952-7915(94)90033-7

8. Ferrante A, Gorski J. Cooperativity of hydrophobic anchor interactions: evidence for epitope selection by MHC class II as a folding process. *J Immunol* (2007) 178 (11):7181–9. doi: 10.4049/jimmunol.178.11.7181

9. Murthy VL, Stern LJ. The class II MHC protein HLA-DR1 in complex with an endogenous peptide: implications for the structural basis of the specificity of peptide binding. *Structure* (1997) 5(10):1385–96. doi: 10.1016/S0969-2126(97)00288-8

10. Wucherpfennig KW. Insights into autoimmunity gained from structural analysis of MHC–peptide complexes. *Curr Opin Immunol* (2001) 13(6):650–6. doi: 10.1016/S0952-7915(01)00274-6

11. Madura F, Rizkallah PJ, Holland CJ, Fuller A, Bulek A, Godkin AJ, et al. Structural basis for ineffective T-cell responses to MHC anchor residue-improved "heteroclitic" peptides: Molecular immunology. *Eur J Immunol* (2015) 45(2):584–91. doi: 10.1002/eji.201445114

12. Smith AR, Alonso JA, Ayres CM, Singh NK, Hellman LM, Baker BM. Structurally silent peptide anchor modifications allosterically modulate T cell recognition in a receptor-dependent manner. *Proc Natl Acad Sci USA* (2021) 118(4): e2018125118. doi: 10.1073/pnas.2018125118

13. Saotome K, Dudgeon D, Colotti K, Moore MJ, Jones J, Zhou Y, et al. Structural analysis of cancer-relevant TCR-CD3 and peptide-MHC complexes by cryoEM. *Nat Commun* (2023) 14(1):2401. doi: 10.1038/s41467-023-37532-7

14. Sušac L, Vuong MT, Thomas C, Von Bülow S, O'Brien-Ball C, Santos AM, et al. Structure of a fully assembled tumor-specific T cell receptor ligated by pMHC. *Cell* (2022) 185(17):3201–3213.e19. doi: 10.1016/j.cell.2022.07.010

15. Sahin U, Derhovanessian E, Miller M, Kloke BP, Simon P, Löwer M, et al. Personalized RNA mutanome vaccines mobilize poly-specific therapeutic immunity against cancer. *Nature* (2017) 547(7662):222–6. doi: 10.1038/nature23003

16. Ott PA, Hu Z, Keskin DB, Shukla SA, Sun J, Bozym DJ, et al. An immunogenic personal neoantigen vaccine for patients with melanoma. *Nature* (2017) 547 (7662):217–21. doi: 10.1038/nature22991

17. Kreiter S, Vormehr M, van de Roemer N, Diken M, Löwer M, Diekmann J, et al. Mutant MHC class II epitopes drive therapeutic immune responses to cancer. *Nature* (2015) 520(7549):692–6. doi: 10.1038/nature14426

18. Alspach E, Lussier DM, Miceli AP, Kizhvatov I, DuPage M, Luoma AM, et al. MHC-II neoantigens shape tumour immunity and response to immunotherapy. *Nature* (2019) 574(7780):696–701. doi: 10.1038/s41586-019-1671-8

19. Tomasello G, Armenia I, Molla G. The Protein Imager: a full-featured online molecular viewer interface with server-side HQ-rendering capabilities. *Bioinformatics* (2020) 36(9):2909–11. doi: 10.1093/bioinformatics/btaa009

20. Jardetzky TS, Brown JH, Gorga JC, Stern LJ, Urban RG, Strominger JL, et al. Crystallographic analysis of endogenous peptides associated with HLA-DR1 suggests a common, polyproline II-like conformation for bound peptides. *Proc Natl Acad Sci USA* (1996) 93(2):734–8. doi: 10.1073/pnas.93.2.734

21. Wieczorek M, Abualrous ET, Sticht J, Álvaro-Benito M, Stolzenberg S, Noé F, et al. Major histocompatibility complex (MHC) class I and MHC class II proteins: conformational plasticity in antigen presentation. *Front Immunol* (2017) 8:292. doi: 10.3389/fimmu.2017.00292

22. Hammer J, Bono E, Gallazzi F, Belunis C, Nagy Z, Sinigaglia F. Precise prediction of major histocompatibility complex class II-peptide interaction based on peptide side chain scanning. *J Exp Med* (1994) 180(6):2353–8. doi: 10.1084/jem.180.6.2353

23. Stern LJ, Brown JH, Jardetzky TS, Gorga JC, Urban RG, Strominger JL, et al. Crystal structure of the human class II MHC protein HLA-DR1 complexed with an influenza virus peptide. *Nature* (1994) 368(6468):215–21. doi: 10.1038/368215a0

24. Brown JH, Jardetzky TS, Gorga JC, Stern LJ, Urban RG, Strominger JL, et al. Three-dimensional structure of the human class II histocompatibility antigen HLA-DR1. *Nature* (1993) 364(6432):33–9. doi: 10.1038/364033a0

25. Kulski JK, Suzuki S, Shiina T. Human leukocyte antigen super-locus: nexus of genomic supergenes, SNPs, indels, transcripts, and haplotypes. *Hum Genome Var* (2022) 9(1):49. doi: 10.1038/s41439-022-00226-5

26. Barker DJ, Maccari G, Georgiou X, Cooper MA, Flicek P, Robinson J, et al. The IPD-IMGT/HLA database. *Nucleic Acids Res* (2023) 51(D1):D1053–60. doi: 10.1093/nar/gkac1011

27. Berman HM. The protein data bank. *Nucleic Acids Res* (2000) 28(1):235–42. doi: 10.1093/nar/28.1.235

28. Patronov A, Dimitrov I, Flower DR, Doytchinova I. Peptide binding prediction for the human class II MHC allele HLA-DP2: a molecular docking approach. *BMC Struct Biol* (2011) 11(1):32. doi: 10.1186/1472-6807-11-32

29. Patronov A, Salamanova E, Dimitrov I, Flower D, Doytchinova I. Histidine Hydrogen Bonding in MHC at pH 5 and pH 7 Modeled by Molecular Docking and Molecular Dynamics Simulations. *CAD* (2014) 10(1):41–9. doi: 10.2174/15734099113096660050

30. Bordner AJ. Towards universal structure-based prediction of class II MHC epitopes for diverse allotypes. *PloS One* (2010) 5(12):e14383. doi: 10.1371/journal.pone.0014383

31. Tong JC, Tan TW, Ranganathan S. Modeling the structure of bound peptide ligands to major histocompatibility complex. *Protein Sci* (2004) 13(9):2523–32. doi: 10.1110/ps.04631204

32. Khan J, Ranganathan S. pDOCK: a new technique for rapid and accurate docking of peptide ligands to Major Histocompatibility Complexes. *Immunome Res* (2010) 6. doi: 10.1186/1745-7580-6-S1-S2

33. Atanasova M, Patronov A, Dimitrov I, Flower DR, Doytchinova I. EpiDOCK: a molecular docking-based tool for MHC class II binding prediction. *PEDS* (2013) 26 (10):631–4. doi: 10.1093/protein/gzt018

34. Racle J, Michaux J, Rockinger GA, Arnaud M, Bobisse S, Chong C, et al. Robust

prediction of HLA class II epitopes by deep motif deconvolution of immunopeptidomes. *Nat Biotechnol* (2019) 37(11):1283–6. doi: 10.1038/s41587-019-0289-6

35. Liu Z, Jin J, Cui Y, Xiong Z, Nasiri A, Zhao Y, et al. DeepSeqPanII: an interpretable recurrent neural network model with attention mechanism for peptide-HLA class II binding prediction. *IEEE/ACM Trans Comput Biol Bioinf.* (2022) 19 (4):2188–96. doi: 10.1109/TCBB.2021.3074927

36. Reynisson B, Alvarez B, Paul S, Peters B, Nielsen M. NetMHCpan-4.1 and NetMHCIIpan-4.0: improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data. *NAR* (2020) 48 (W1):W449–54. doi: 10.1093/nar/gkaa379

37. Marzella DF, Parizi FM, Tilborg Dv, Renaud N, Sybrandi D, Buzatu R, et al. PANDORA: A fast, anchor-restrained modelling protocol for peptide: MHC complexes. *Front Immunol* (2022) 13:878762. doi: 10.3389/fimmu.2022.878762

38. Marzella DF, Crocioni G, Parizi FM, Xue LC. The PANDORA software for anchor-restrained peptide:MHC modeling. *Methods Mol Biol* (2023) 2673:251–71. doi: 10.1007/978-1-0716-3239-0_18

39. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. *Nature* (2021) 596:1–11. doi: 10.1038/s41586-021-03819-2

40. Ehrenmann F, Kaas Q, Lefranc MP. IMGT/3Dstructure-DB and IMGT/DomainGapAlign: a database and a tool for immunoglobulins or antibodies, T cell receptors, MHC, IgSF and MhcSF. *Nucleic Acids Res* (2010) 38:D301–7. doi: 10.1093/nar/gkp946

41. Neefjes J, Jongsma MLM, Paul P, Bakke O. Towards a systems understanding of MHC class I and MHC class II antigen presentation. *Nat Rev Immunol* (2011) 11 (12):823–36. doi: 10.1038/nri3084

42. Webb B, Sali A. Comparative protein structure modeling using MODELLER. *Curr Protoc Bioinf* (2016) 54(5):1–5. doi: 10.1002/prot.21804

43. Lensink MF, Méndez R, Wodak SJ. Docking and scoring protein complexes: CAPRI 3rd Edition. *Proteins* (2007) 69(4):704–18. doi: 10.1002/prot.21804

44. Mirdita M, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M. ColabFold: making protein folding accessible to all. *Nat Methods* (2022) 19(6):679–82. doi: 10.1038/s41592-022-01488-1

45. Evans R, O'Neill M, Pritzel A, Antropova N, Senior A, Green T, et al. Protein complex prediction with AlphaFold-Multimer. *Bioinformatics;* (2021) 4:2021–10. doi: 10.1101/2021.10.04.463034

46. Carson RT, Vignali KM, Woodland DL, Vignali DAA. T cell receptor recognition of MHC class II–bound peptide flanking residues enhances immunogenicity and results in altered TCR V region usage. *Immunity* (1997) 7 (3):387–99. doi: 10.1016/S1074-7613(00)80360-X

47. Vignali DA, Strominger JL. Amino acid residues that flank core peptide epitopes and the extracellular domains of CD4 modulate differential signaling through the T cell receptor. *JEM* (1994) 179(6):1945–56. doi: 10.1084/jem.179.6.1945

48. Muller CP, Ammerlaan W, Fleckenstein B, Krauss S, Kalbacher H, Schneider F, et al. Activation of T cells by the ragged tail of MHC class II-presented peptides of the measles virus fusion protein. *Int Immunol* (1996) 8(4):445–56. doi: 10.1093/intimm/8.4.445

49. Zavala-Ruiz Z, Strug I, Walker BD, Norris PJ, Stern LJ. A hairpin turn in a class II MHC-bound peptide orients residues outside the binding groove for T cell recognition. *Proc Natl Acad Sci USA* (2004) 101(36):13279–84. doi: 10.1073/pnas.0403371101

50. Levy R, Alter Regev T, Paes W, Gumpert N, Cohen Shvefel S, Bartok O, et al. Large-scale immunopeptidome analysis reveals recurrent posttranslational splicing of cancer- and immune-associated genes. *MCP* (2023) 22(4):100519. doi: 10.1016/j.mcpro.2023.100519

51. Bloodworth N, Barbaro NR, Moretti R, Harrison DG, Meiler J. Rosetta FlexPepDock to predict peptide-MHC binding: An approach for non-canonical amino acids. *PloS One* (2022) 17(12):e0275759. doi: 10.1371/journal.pone.0275759

52. Sandalova T, Sala BM, Achour A. Structural aspects of chemical modifications in the MHC-restricted immunopeptidome; Implications for immune recognition. *Front Chem* (2022) 10:861609. doi: 10.3389/fchem.2022.861609

frontiers | Frontiers in Immunology

TYPE Original Research
PUBLISHED 03 January 2024
DOI 10.3389/fimmu.2023.1279003

OPEN ACCESS

EDITED BY
Joe Hou,
Fred Hutchinson Cancer Center, United States

REVIEWED BY
Morgan Erin Brisse,
National Institute of Allergy and Infectious Diseases (NIH), United States
Marija Zaric,
International AIDS Vaccine Initiative Inc, United States
Bronwyn Gunn,
Washington State University, United States

*CORRESPONDENCE
Paola Andrea Martinez-Murillo
paola.martinez@unige.ch

†These authors have contributed equally to this work

RECEIVED 17 August 2023
ACCEPTED 07 December 2023
PUBLISHED 03 January 2024

CITATION
Martinez-Murillo PA, Huttner A, Lemeille S, Medaglini D, Ottenhoff THM, Harandi AM, Didierlaurent AM and Siegrist C-A (2024) Refined innate plasma signature after rVSVΔG-ZEBOV-GP immunization is shared among adult cohorts in Europe and North America. *Front. Immunol.* 14:1279003. doi: 10.3389/fimmu.2023.1279003

Refined innate plasma signature after rVSVΔG-ZEBOV-GP immunization is shared among adult cohorts in Europe and North America

Paola Andrea Martinez-Murillo[1*], Angela Huttner[2,3,4,5], Sylvain Lemeille[1], Donata Medaglini[6], Tom H. M. Ottenhoff[7], Ali M. Harandi[8,9], Arnaud M. Didierlaurent[1†] and Claire-Anne Siegrist[1,2†] for the VEBCON, VSV-EBOVAC and VSV-EBOPLUS Consortia

[1]Center of Vaccinology, Department of Pathology and Immunology, Faculty of Medicine, University of Geneva, Geneva, Switzerland, [2]Center for Vaccinology, Geneva University Hospitals, Geneva, Switzerland, [3]Division of Infectious Diseases, Geneva University Hospitals, Geneva, Switzerland, [4]Faculty of Medicine, University of Geneva, Geneva, Switzerland, [5]Center for Clinical Research, Geneva University Hospitals, Geneva, Switzerland, [6]Laboratory of Molecular Microbiology and Biotechnology, Department of Medical Biotechnologies, University of Siena, Siena, Italy, [7]Department of Infectious Diseases, Leiden University Medical Center, Leiden, Netherlands, [8]Department of Microbiology and Immunology, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden, [9]Vaccine Evaluation Centre, BC Children's Hospital Research Institute, University of British Columbia, Vancouver, BC, Canada

Background: During the last decade Ebola virus has caused several outbreaks in Africa. The recombinant vesicular stomatitis virus-vectored Zaire Ebola (rVSVΔG-ZEBOV-GP) vaccine has proved safe and immunogenic but is reactogenic. We previously identified the first innate plasma signature response after vaccination in Geneva as composed of five monocyte-related biomarkers peaking at day 1 post-immunization that correlates with adverse events, biological outcomes (haematological changes and viremia) and antibody titers. In this follow-up study, we sought to identify additional biomarkers in the same Geneva cohort and validate those identified markers in a US cohort.

Methods: Additional biomarkers were identified using multiplexed protein biomarker platform O-link and confirmed by Luminex. Principal component analysis (PCA) evaluated if these markers could explain a higher variability of the vaccine response (and thereby refined the initial signature). Multivariable and linear regression models evaluated the correlations of the main components with adverse events, biological outcomes, and antibody titers. External validation of the refined signature was conducted in a second cohort of US vaccinees (n=142).

Results: Eleven additional biomarkers peaked at day 1 post-immunization: MCP2, MCP3, MCP4, CXCL10, OSM, CX3CL1, MCSF, CXCL11, TRAIL, RANKL and IL15. PCA analysis retained three principal components (PC) that accounted for 79% of the vaccine response variability. PC1 and PC2 were very robust and had different biomarkers that contributed to their variability. PC1 better discriminated different doses, better defined the risk of fever and

myalgia, while PC2 better defined the risk of headache. We also found new biomarkers that correlated with reactogenicity, including transient arthritis (MCP-2, CXCL10, CXCL11, CX3CL1, MCSF, IL-15, OSM). Several innate biomarkers are associated with antibody levels one and six months after vaccination. Refined PC1 correlated strongly in both data sets (Geneva: r = 0.97, P < 0.001; US: r = 0.99, P< 0.001).

Conclusion: Eleven additional biomarkers refined the previously found 5-biomarker Geneva signature. The refined signature better discriminated between different doses, was strongly associated with the risk of adverse events and with antibody responses and was validated in a separate cohort.

KEYWORDS

innate plasma signature, rVSVΔG-ZEBOV-GP, biomarkers, adverse events, immunogenicity

Introduction

Since the identification of the ebolaviruses in 1976, several outbreaks of Ebola disease have been identified in sub-Saharan Africa. Ebola virus disease (EVD) induces a high mortality rate (50-90%) and can result in uncontrolled epidemics, as witnessed in 2014-16 during the largest Ebola outbreak ever reported (1). The international response to this outbreak supported international collaborations to test EVD vaccine candidates. rVSVΔG-ZEBOV-GP, the most advanced candidate at that time, is a live-attenuated vaccine whose vesicular stomatitis virus glycoprotein-encoding gene has been deleted (VSVΔG) and replaced with the Zaire Ebola virus (ZEBOV-GP) glycoprotein. This vaccine induced 100% protection against EVD in challenged non-human primates (NHP) (2–4).

rVSVΔG-ZEBOV-GP proved safe and immunogenic in different clinical trials held in the USA, Europe and Africa (5–11), but induces transient reactogenicity (12). It was shown to be effective within days in the ring vaccination trial held in 2015 in Guinea (10) and during the 2018–19 outbreak in the Democratic Republic of Congo (13). All these findings supported fast tracked vaccine licensure, resulting in a prequalification by WHO for rVSVΔG-ZEBOV-GP to be used in countries at high risk in 2019 (14), and to its license under the name of Ervebo® by the FDA (15) and by the EMA (16).

Although rVSVΔG-ZEBOV-GP is highly effective against EVD, only a few studies have explored its principal innate and adaptive induced immune mechanisms and its ability to induce early protection. Studies in NHP models have demonstrated that antibodies and $CD4^+$ T-cells are necessary for rVSV-EBOV-mediated protection against lethal infection, while $CD8^+$ T-cells play a minor role (17). Interestingly, rVSVΔG-ZEBOV-GP induced partial and total protection in NHP as early as 3 and 7 days after challenge, in absence of detectable antigen-specific IgG and low IgM-specific serum antibodies (18), suggesting a role of innate responses in mediating early protection.

rVSVΔG-ZEBOV-GP induces a robust innate immune response characterized by the mobilization of monocytes and natural killer (NK) cell in humans, and NK cell activation and CXCL10 levels correlates with antigen-specific antibody responses (8, 19). Similarly, other rVSV-based vaccines evaluated in NHPs induce the secretion of cytokines/chemokines and NK cell activation [VSV-MARV (20, 21)] and the transcription of genes involved in NK and innate immune pathways [rVSVΔG-LASV-GPC (22)]. We showed in Geneva vaccinees that this mobilization and activation of circulating NK cells was rapid and dose-dependent (23). We also identified the first innate plasma signature response to rVSVΔG-ZEBOV-GP in healthy vaccinees, derived in a European cohort (Geneva, Switzerland) and validated in an African cohort (Lambaréné, Gabón) (24). Among the six monocyte-related cytokines/chemokines which peaked at day 1 post-immunization, five (MCP-1, IL-1Ra, TNF-α, IL-10 and IL-6) defined a signature that was vaccine dose-dependent and correlated with viremia, biological outcomes and adverse events, including transient arthritis (24). Here, we aimed to identify additional markers in Geneva vaccinees that could refine the previous signature and to validate this refined signature in a US cohort.

Methods

Study design, population, and key previous outcomes

We used plasma samples obtained from two clinical trials conducted in Europe (phase 1/2, randomized, double-blind, placebo-controlled, dose-finding trial in Geneva, Switzerland [November 2014, to January 2015; NCT02287480]) (12) and in

North America (phase 1b, randomized, double-blind, placebo-controlled, dose-response trial in the USA [Dec 5, 2014, to June 23, 2015; NCT02314923]) (25). The trial protocols were reviewed and approved by the WHO's Ethics Committee as well as by local ethics committees (USA trial: the Chesapeake Institutional Review Boards (Columbia, MD, USA) and the Crescent City Institutional Review Board (New Orleans, LA, USA); Geneva trial: the Geneva Cantonal Ethics Commission and the Swiss Agency for Therapeutic Products (Swissmedic). All participants had provided written informed consent to participate in those studies (12, 25).

As genetic and environmental factors may influence vaccine response, we used the Geneva trial as the derivation cohort (n=115) and the US trial as the validation cohort (see Supplementary Figure 1). As a wider range of vaccine doses were tested in this US trial (7, 9), we randomly choose a subset of individuals (n=130) grouped to best match Geneva low dose (n=48), high dose (n=60) and placebo (n=22) recipients (Supplementary Figure 1).

Pilot high-throughput screening in plasma from Geneva vaccinees

O-link (OLINK AB, Uppsala) is a semi-quantitative assay based on Proximity Extension Assay (PEA) technology with no cross reactivity. It measures proteins via an antibody-mediated detection system linked to synthetic DNA. The method has been described previously (26). Briefly, paired oligonucleotide-coupled antibodies with overlapping sequences are allowed to bind to proteins in the sample. When paired antibodies are brought in proximity to one another through binding to their target, their oligonucleotide sequences overlap to form a PCR target, which can be semi-quantified with real-time PCR. We used three O-link panels (inflammation, immune and metabolic panels, each panel detecting 92 proteins) to screen for 276 markers. Inflammatory panel was tested first, and we evaluated days 0, 1, 3 and 7. Immune and metabolism panels were used later, and we evaluated only day 0 and 1. Following data pre-processing, including quality control, the relative level (NPX) of each of the 276 proteins was assessed. Proteins with more than 30% of samples with NPX values below the limit of detection (n=53) were excluded from further analysis.

In this pilot screening, we selected a subgroup of participants of the Geneva cohort (n=49), including all participants that reported transient arthritis and matched the samples by dose, sex and age (Figure 1A), with the aim to identify potential arthritis-associated biomarkers. We first assessed the number of markers peaking at D1, D3 and D7 (Figure 1B). Subsequently, the identified biomarkers were confirmed and quantified by Luminex in each participant of the Geneva cohort (n=115).

Quantification of biomarkers by Luminex assay

A customized Luminex assay (Magnetic Luminex assay, R&D Systems) was used to measure the plasma concentration of most of the markers identified by O-link, as some were not available for testing with the Luminex technology. Assays were performed according to the supplier's instructions using) the Luminex xMAP Technology (Luminex Corporation) and read on the Bio-Plex 200 array reader (Bio-Rad Laboratories). Five-parameter logistic regression curve (Bio-Plex Manager 6.0) was used to calculate sample concentrations. In addition to previously reported biomarkers (IL-1Ra, MCP1, IL-6, IL-10, MIP1b, and TNF-α) (24), additional markers from the O-Link analysis were MCP2, MCP3, MCP4, CXCL10, OSM, CX3CL1, MCSF, CXCL11, TRAIL, RANKL and IL15 were measured in both Geneva and US cohort. All data below thresholds (last point of the standard curve) were set to half the value of the corresponding threshold.

ZEBOV-GP-binding antibodies

We used the data generated in studies performed in Geneva, reported in (12) and in the US, reported in (25). For the present study, we refer to measurements performed at day 28 and 180. Briefly, quantification of ZEBOV-GP-specific antibodies for the Geneva cohort was done at the US Army Medical Research Institute for Infectious Diseases (USAMRIID) in Frederick, Maryland, USA in the Diagnostic Systems Division using USAMRIID's standard operating procedure (SOP AP-03-35; USAMRIID ELISA) (8, 12, 27) by the Filovirus Animal Non-Clinical Group (FANG). For the US cohort, ZEBOV-GP-specific antibodies were tested in Focus Diagnostics, San Juan Capistrano, CA, based on the assay developed by FANG. The homologous Zaire–Kikwit strain GP was used as specified in the SOP. The log10 transformed ELISA units per mL was used for correlation analysis in the present study.

Identification of the Geneva and US signatures

We applied the same methods as previously (24) to identify signatures of the vaccine response. PCA was done for all participants of each cohort and for all 17 identified markers for which we used the log10 D1/D0 ratio to normalize the data. To build the model, the normalized data were standardized so that the means and the SD equalled 0. PCA components with eigen values greater than 1 were retained. Because of the number of variables introduced in the PCA (n=17) and the number of vaccinees (Geneva cohort: n=100; USA cohort: n=113), a risk of overfitting was suspected, thus a bootstrap procedure was used to check the robustness of the number of retained principal components. For this, 50,000 re-samplings with replacements were done: for each resampling, the same PCA was conducted. Cronbach's alpha values were used to indicate whether the variation of markers upregulated between days 0 and 1 was based on a single trait. The Kaiser-Meyer-Olkin was used to measure the adequacy of the data to factor analysis (28). Our validation cohort was the US cohort, and we used the same approach to calculate the signature by PCA. The score for

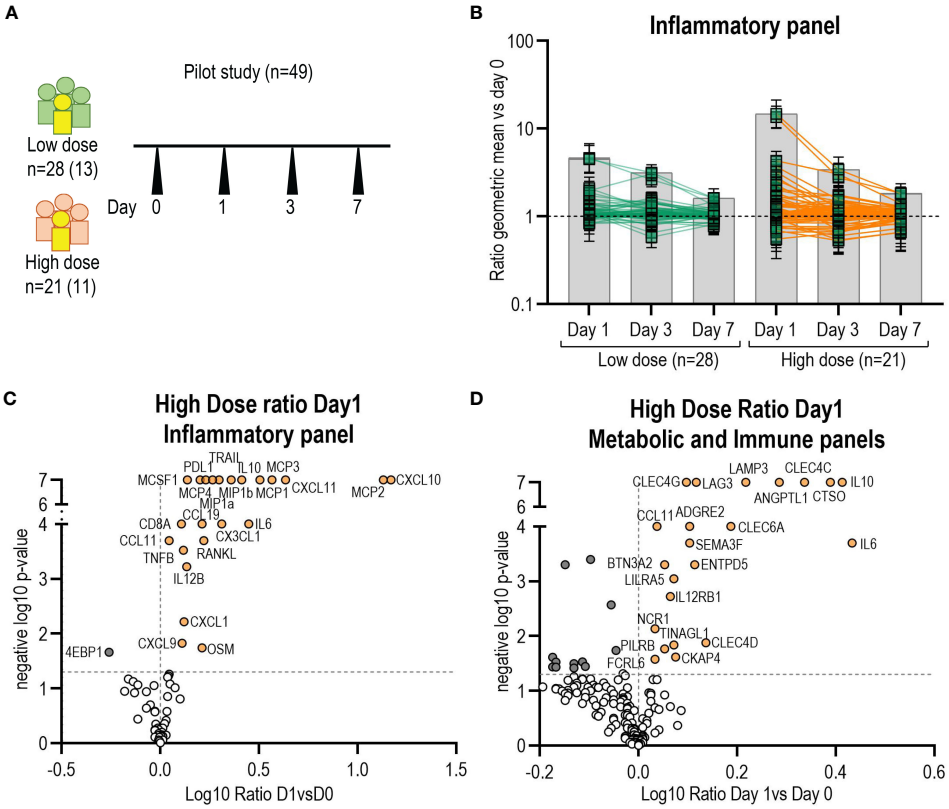


FIGURE 1
Identification of additional biomarkers by O-link. **(A)** Schematic of the pilot study samples used to screen for new markers (n=49). In yellow and in parenthesis number of participants with arthritis. **(B)** Kinetics of biomarkers from O-link inflammatory panel (96 markers) expressed as the ratio of the mean at day1, day3 and day7 versus day 0. Each square represents the mean for a single marker and confidence interval is included. Volcano plots from O-link inflammatory **(C)** and metabolic panels **(D)** of the high dose group displaying the log10 fold change (x axis) against the t test-derived negative log10 statistical P value (y axis) for all proteins differentially secreted between day1 and day0. Thresholds (dotted grey line), p-value cut-off was fixed at 0.05 (1,3 negative log10) and fold change cut-offs was 1 (0 in the log10 scale). P-value of zero was set up as 0,0000001 (7 neg log10). Open circles represent all proteins below the p-value and in dark grey all proteins below fold change cut-offs. Proteins above the fold-change cut off are labelled as orange circles.

each observation was calculated by applying the equations of each component, which then was used to evaluate the correlation with adverse events and biological outcomes.

Statistical methods

Biomarkers were reported by vaccine dose and timepoint using log10 geometric mean concentrations (GMCs). GMCs were compared between independent groups using t-tests or ANOVA (with Scheffe's correction for multiplicity of tests and *post hoc* analyses) and over time using linear regression models with mixed effects to account for repeated measures. The association between the signature and biological outcomes/AEs was assessed using linear and logistic regression models with adjustment for the dose. The type I error level was 0.05, and all statistical tests were two-sided. AUCs of the previous and refined signature were compared by using Delong's non-parametric test for paired ROC curves. Analyses were conducted in R 3.2.2 (R Foundation for Statistical Computing, version 2.15.2) and STATA 14.0 IC (StataCorp LP).

Results

Identification of additional biomarkers of innate responses to rVSVΔG-ZEBOV

We set up a pilot experiment using an O-link approach that can measure up to 276 analytes to identify additional plasma markers associated with the vaccine response compared to our previous study (Figure 1A). Markers significantly peaked at day 1 in both the high and low dose groups, but not at day 3 or 7 (Figure 1B). Therefore, we subsequently only analysed the ratio of D1/D0. In the high-dose (HD) vaccinees group, 18 new additional proteins from the inflammatory panel were significantly elevated and one protein (4EBP1) showed a significant decrease (Figures 1B, C). In the low-dose (LD) vaccinees group, 18 new proteins were significantly elevated (16 were shared with HD vaccinees) and one (MMP1) was significantly decreased (Supplementary Figure 2A). The analysis of the metabolic and immune panels showed that in the HD group 17 new proteins were significantly increased, and 13 were significantly decreased on day 1 compared to day 0 (Figure 1D), whereas in the LD group four new proteins were significantly elevated and eight were

significantly decreased (Supplementary Figure 2B) (no new markers were shared with HD vaccinees). We observed that all the proteins identified in our previous study (24) had significantly increased on day 1, confirming our previous findings, and supporting the use of O-link as an adequate screening tool. Secreted proteins with a D1/D0 ratio greater than 1 but without statistical significance are shown in Supplementary Figures 2C, D. We did not find statistically significant differences in biomarkers levels between arthritis and non-arthritis in this subset of patients in the inflammatory panel and metabolic panel analysed (Supplementary Table 1).

In conclusion, use of O-link screening in a subset of the Geneva cohort (n=49) allowed us to identify 18 additional proteins significantly secreted at higher levels on day 1 in both high and low dose groups.

Confirmation and quantification of the biomarker signature

Out of the 18 additional markers found by O-link, eleven were available for measurement by Luminex and were quantified on days 0, 1, 3, 7 in plasma samples of the entire Geneva cohort (n=115). The eleven markers included chemokines: monocyte chemoattractant protein 2 (MCP2/CCL8), monocyte chemoattractant protein 3 (MCP3/CCL7), monocyte chemoattractant protein 4 (MCP4/CCL13), chemokine C-X3-C motif ligand 1 (CX3CL1/Fractalkine), interferon gamma-induced protein 10 (IP10/CXCL10), interferon-gamma-inducible protein 9 (IP-9/CXCL11); cytokines: Interleukin 15 (IL-15), Oncostatin M (OSM) and macrophage colony-stimulating factor (M-CSF); and ligands: Tumor necrosis factor ligand superfamily member 10 (TRAIL/TNFSF10), Tumor necrosis factor ligand superfamily member 11 (RANKL/TNFSF11).

We calculated the geometric mean concentrations (GMCs) for each marker and the ratio of D1/D0. As expected, in the placebo control group, no marker significantly increased with time, except for CXCL10 that showed a significant decline at day 1 (Table 1). We confirmed that all eleven additional markers significantly peaked at day 1 in the Geneva cohort (Figure 2), with the largest fold increases reported in HD for CXCL11 [21.0 (95% CI, 15.1 to 29.2)], CXCL10 [14.2 (95% CI, 11 to 18.4)] and MCP2 [13.3 (95% CI, 11 to 16.1)] (Table 1). HD vaccinees showed significantly higher increases in GMCs than LD vaccinees for all markers except RANKL (Figure 2).

We found that all additional markers except RANKL were significantly correlated between each other and with the previously reported markers, irrespective of the vaccine dose (Supplementary Figure 3) The strongest associations were observed between CXCL10 and CXCL11 at both doses (Spearman's correlation coefficient r = 0.92, p <0.001; r=0.88, p<0.001) and between MCP1 and MCP2 (Spearman's correlation coefficient r = 0.61, p<0.001; r=0.82, p<0.001 at the two doses respectively) (Supplementary Figure 3).

TABLE 1 Ratio day 1/day 0 of the geometric mean (GM) of the additional identified markers measured in the plasma of Geneva participants.

	Placebo (n=13)			Low Dose (n=51)			High Dose (n=51)		
Marker	Ratio GM	Confidence Interval	*p-value*	Ratio GM	Confidence Interval	*p-value*	Ratio GM	Confidence Interval	*p-value*
CXCL11	0,85	(0,68 - 1,07)	0,150	2,65	(1,98 - 3,54)	**<0,001**	21	(15,14 - 29,23)	**<0,001**
CXCL10	0,81	(0,69 - 0,96)	**0,019**	3,08	(2,39 - 3,97)	**<0,001**	14,2	(10,99 - 18,35)	**<0,001**
MCP2	1,12	(0,89 - 1,41)	0,298	3,95	(2,94 - 5,29)	**<0,001**	13,3	(10,95 - 16,14)	**<0,001**
MCSF	0,95	(0,56 - 1,61)	0,831	2,07	(1,58 - 2,72)	**<0,001**	7,41	(5,59 - 9,82)	**<0,001**
MCP3	1,35	(0,83 - 2,22)	0,207	1,71	(1,27 - 2,3)	**<0,001**	6,18	(4,54 - 8,43)	**<0,001**
OSM	0,93	(0,65- 1,33)	0,660	2,01	(1,72 - 2,34)	**<0,001**	4,78	(3,83 - 5,97)	**<0,001**
TRAIL	0,92	(0,76 - 1,12)	0,368	1,72	(1,5 - 1,98)	**<0,001**	4,15	(3,62 -4,76)	**<0,001**
CX3CL1	0,99	(0,77 - 1,28)	0,925	1,27	(1,12 - 1,44)	**<0,001**	3,83	(2,95 - 4,98)	**<0,001**
IL15	1,34	(0,91 - 1,96)	0,123	1,4	(1,19 - 1,65)	**<0,001**	3,15	(2,64 - 3,77)	**<0,001**
RANKL	1,21	(0,74 - 1,96)	0,415	1,36	(1,2 - 1,54)	**<0,001**	2,13	(1,69 - 2,68)	**<0,001**
MCP4	0,9	(0,73- 1,11)	0,295	1,13	(1,06 - 1,21)	**<0,001**	1,64	(1,48 - 1,82)	**<0,001**
IL1-Rα	0,97	(0,78 - 1,21)	0,81	1,77	(1,39 - 2,26)	**<0,001**	10,6	(8,41 - 13,37)	**<0,001**
IL-6	0,7	(0,35 - 1,38)	0,31	1,82	(1,18 - 2,79)	**0,007**	13,5	(8,29 - 21,91)	**<0,001**
IL-10	0,74	(0,39 - 1,38)	0,35	2,11	(1,19 - 3,75)	**0,011**	7,08	(4,68 - 10,70)	**<0,001**
TNF-α	1,3	(0,59 - 2,87)	0,51	1,33	(0,78 - 2,27)	0,3	3,98	(2,43 - 6,51)	**<0,001**
MCP-1	0,89	(0,78 - 1,02)	0,11	1,4	(1,22 - 1,62)	**0,011**	3,35	(2,97 - 3,78)	**<0,001**
MIP-1β	0,96	(0,81 - 1,15)	0,64	1,33	(1,14 -1,55)	**0,011**	2,31	(2,09 - 2,56)	**<0,001**

Ratio of GM: log10 base ratio Day 1/Day 0. Significant difference between day 1 and day 0 are represented by P-values highlighted in bold. Markers are presented according to the ratio GM levels in the high dose. Previous signature biomarkers reported in Huttner et al., 2017 (24) are shaded in grey.

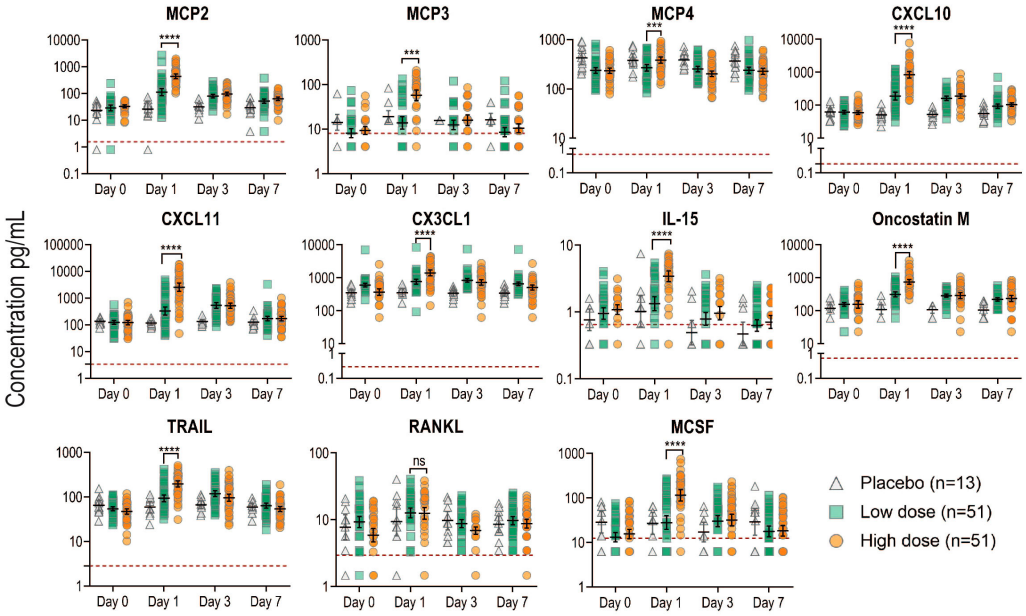


FIGURE 2
Kinetics of newly identified biomarkers measured in the plasma of all Geneva participants. Plasma concentration in pg/ml for each marker measured by Luminex was plotted at each time point in the different groups: placebo (gray), low dose (green) and high dose (orange). Each dot represents a participant (n=115). Black lines represent the geometric mean concentrations with the CI. Red dotted lines indicate the limit of detection for each marker. Samples below the limit of detection were assigned a value corresponding to 50% of the last standard dilution value. P values less than 0.001 are summarized with three asterisks, and P values less than 0.0001 are summarized with four asterisks.

In summary, we found eleven additional markers at day 1 after vaccination that correlated with the previously identified signature in the Geneva cohort.

Refinement of the innate plasma signature

PCA was conducted for the 17 markers described above (6 previously reported and the 11 additional reported here). PCA showed that the new refined signature accounted for 77.8% of the variability of the day 1 immune response versus baseline and three components were retained (PC1: 63.2%, PC2 8.5% and PC3 6.1% of the variance; Figure 3A). The bootstrap analysis confirmed the robustness of the first three components. The frequency of the number of retained components (Eigen value > 1) over the 50'000 re-sampling was PC1: n=50000/50000 (100%); PC2: n=49849/50000 (99.7%); PC3: n=34580/50000 (69.16%); PC4: n=113/50000 (0.23%); PC5: n=0/50000 (0%). Cronbach's alpha values (LD:0.94, HD: 0.94) indicated that the variability in the markers induced by the vaccine was highly reliable and mostly based on a common trait. The overall measure of adequacy was 0.9, considered by Kaiser et al. (28) as very robust data for factor analysis.

After normalization and standardization, the equation of the first component (PC1) was defined by "$0.083\times IL1Ra^{STD} + 0.067xIL6^{STD} + 0.057xTNFa^{STD} + 0.06xIL10^{STD} + 0.083xMCP1^{STD} + 0.07xMIP1b^{STD} + 0.076xMCP3^{STD} + 0.086xCXCL10^{STD} + 0.068xOSM^{STD} + 0.076xMCP4^{STD} + 0.075xCX3CL1^{STD} + 0.075xMCSF^{STD} + 0.088xCXCL11^{STD} + 0.084xTRAIL^{STD} + 0.084xMCP2^{STD} + 0.03xRANKL^{STD} + 0.074xIL15^{STD}$", i.e., 17 biomarkers. PC2 equation is reported in Supplementary Table 2.

The biomarkers contributing to component 1 were all positively correlated, while the ones contributing to component 2 showed both a positive and negative correlations (Figure 3A). In the component 1, eleven biomarkers were above the expected average contribution, six of them strongly contributing to the component variability (CXCL11, CXCL10, MCP-2, TRAIL, IL1Ra, MCP-1; Figure 3B), while for the component 2, four biomarkers strongly contributed to component variability (IL-10, TNFA, MP1b, IL-6; Figure 3C).

We next found that the refined signature discriminated better than the previous signature between placebo recipients and LD vaccinees [AUC: 0.87 (95% CI, 0.75 to 0.99) vs 0.79 (95% CI, 0.69 to 0.91); p=0.37], and between low- and HD vaccinees [0.91 (95% CI, 0.85 to 0.97) vs 0.88 (95% CI, 0.81 to 0.95); p=0.059]. Both signatures discriminated almost perfectly placebo recipients and HD vaccinees with area under ROC curves close to 1 (Figure 3D). Altogether, these results show that the addition of eleven markers refined the previous plasma signature as it explained a higher percentage of the variability in the response and improved the discrimination between the two vaccine doses.

Additional biomarkers are associated with vaccine-related adverse events

We next performed a multivariable analysis to assess whether the refined signature was associated with the risk of adverse events

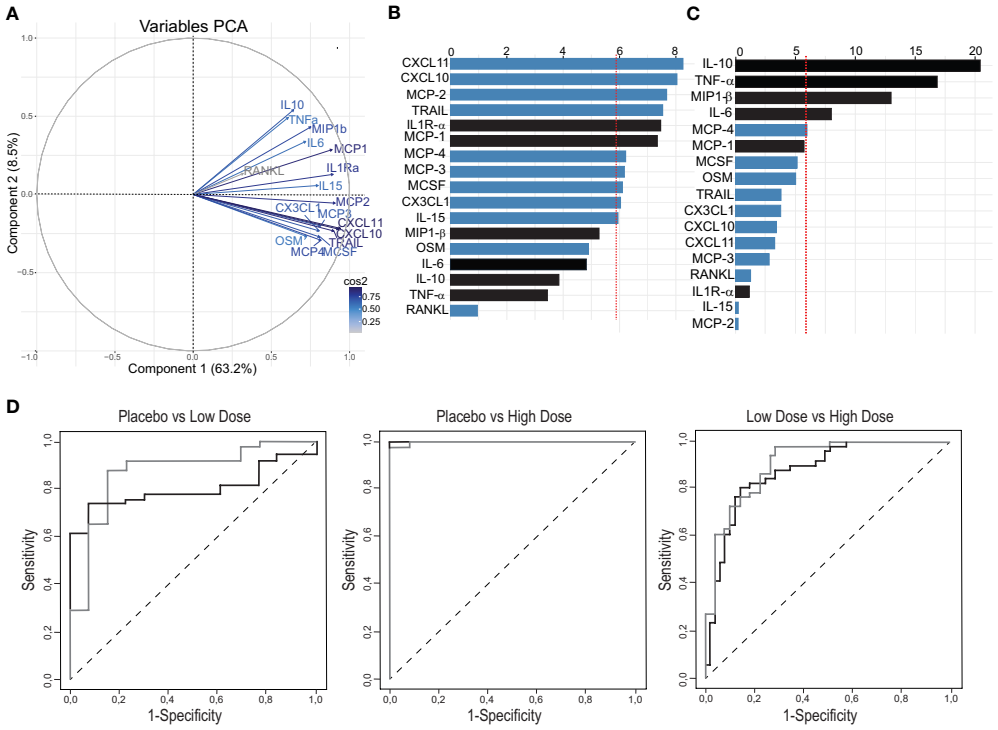


FIGURE 3
Definition of a refined signature by PCA after rVSVΔG-ZEBOV-GP vaccination in the Geneva cohort. **(A)** A variable correlation plot shows the magnitude (length of the arrow) and direction of the correlations of each marker (n=17) to each of the two principal components. Cos2 values indicate how well represented the marker is on the principal component and are shown in a gradient of colours shown in the legend. **(B, C)** Graphs showing the percentage of the contribution of each marker to the variability on component 1 **(B)** and component 2 **(C)**. Red dashed line indicates the average contribution. Blue bars indicate additional markers and bars in black indicate previous markers. **(D)** Comparison Area Under the Curve (AUC) between previous signature (black line) and refined signature (grey line).

following vaccination, as previously described (24). Similarly, we showed that a score higher than one of the Components 1 and 2 of the refined signature increased the risk of injection-site pain, subjective fever and chills in HD vaccinees, (Table 2). In contrast to our previous report, only Component 1 of the refined signature was associated with a higher risk of objective fever and myalgia, while Component 2 was associated with higher risk of headache in HD vaccinees. Because adverse events (AEs) were reported mainly in HD vaccinees (97%), which corresponds to the vaccine dose used in Ervebo®, we focused on this group for further analyses. Headache was associated with significant increase in CXCL10, CXCL11, MCSF, MCP-2 and TNF-alpha, while fatigue was associated with significant increases in CXCL10, MCP-4 and TNF-α (Figure 4A). Increase in MCP-2 was specifically associated with subjective fever and chills, while CX3CL1 and TNF-α were associated with objective fever and myalgia. In contrast, a significant decrease of the anti-inflammatory cytokine IL-10 was associated with arthralgia. No identified biomarker was associated with local pain. Overall, TNF-α and MCP-2 were key biomarkers associated with most systemic AEs.

Twenty-four percent (24%) of participants reported transient vaccine-induced arthritis in the Geneva cohort (12), which was previously associated with lower day 1 signature scores only in HD vaccinees (24). Here, we report a similar finding, Component 1 was significantly lower in HD vaccinees with transient arthritis (GM non-arthritis 0,93 (0,7-0,17) vs GM arthritis 0,34 (-0,05-0,73) p: 0,011) and levels of seven innate plasma biomarkers were also significantly lower (MCP-2, CXCL10, CXCL11, CX3CL1, MCSF, IL-15, OSM), complementary to the four previous biomarkers reported (IL-6, TNF-a, MCP-1 and MIP-1b) (Figure 4B).

Of note, the refined signature showed little to no association with age but was associated with gender (lower scores of Component 1 in females [−0.22 versus 0.19, p=0.029)], confirming what was reported for the previous signature (24).

Overall, the refined signature can thus better predict the risk of objective fever, myalgia and headache and several additional biomarkers were found to be significantly associated with specific systemic adverse events including transient arthritis.

The refined signature and the additional markers are differentially associated with hematological, virological and immunological outcomes

rVSVΔG-ZEBOV-GP immunization triggers a transient, dose-dependent viremia and hematological changes (8, 12). We observed a significant positive association between Component 1 of the refined signature and viremia mainly in LD vaccinees (Supplementary Table 3) that was ruled by IL-15, RANKL and MCSF (Supplementary Table 4). We found a negative correlation for both

TABLE 2 Multivariable analyses of the determinants of clinical outcomes of the refined innate signature in Geneva vaccinees (n=100).

			1st component		2nd component	
Adverse Event	Predictor		Adjusted OR (95%CI)	p-value	Adjusted OR (95%CI)	p-value
Objective fever	Dose	Low dose	Ref		Ref	
		High dose	15.99 (2.3 to 331.34)	**0,017**	16.31 (3.03 to 303.15)	**0,009**
	Signature	<0	Ref		Ref	
		>=0	1.05 (0.23 to 5.8)	0,956	0.64 (0.18 to 2.14)	0,472
Subjective fever	Dose	Low dose	Ref		Ref	
		High dose	3.73 (1.36 to 10.78)	**0,012**	5.07 (2.19 to 12.31)	**<0.001**
	Signature	<0	Ref		Ref	
		>=0	1.72 (0.6 to 4.78)	0,302	0.69 (0.29 to 1.61)	0,388
Headache	Dose	Low dose	Ref		Ref	
		High dose	2.14 (0.79 to 5.93)	0,133	2.67 (1.19 to 6.14)	**0,018**
	Signature	<0	Ref		Ref	
		>=0	1.47 (0.53 to 4.01)	0,446	0.63 (0.28 to 1.43)	0,272
Fatigue	Dose	Low dose	Ref		Ref	
		High dose	1.22 (0.43 to 3.58)	0,706	0.75 (0.33 to 1.7)	0,495
	Signature	<0	Ref		Ref	
		>=0	0.44 (0.15 to 1.24)	0,129	1 (0.44 to 2.28)	0,996
Myalgia	Dose	Low dose	Ref		Ref	
		High dose	2.81 (1.04 to 7.98)	**0,045**	3.15 (1.4 to 7.31)	**0,006**
	Signature	<0	Ref		Ref	
		>=0	1.22 (0.43 to 3.31)	0,702	0.61 (0.26 to 1.39)	0,242
Chills	Dose	Low dose	Ref		Ref	
		High dose	3.2 (1.15 to 9.63)	**0,030**	3.1 (1.36 to 7.35)	**0,008**
	Signature	<0	Ref		Ref	
		>=0	0.96 (0.32 to 2.69)	0,935	0.85 (0.37 to 1.95)	0,694
Arthralgia	Dose	Low dose	Ref		Ref	
		High dose	0.97 (0.25 to 3.78)	0,960	1.28 (0.44 to 3.88)	0,655
	Signature	<0	Ref		Ref	
		>=0	1.62 (0.42 to 6.58)	0,486	0.88 (0.3 to 2.6)	0,807
Pain	Dose	Low dose	Ref		Ref	
		High dose	19.64 (5.81 to 91.45)	**<0.001**	17.81 (6.59 to 55.78)	**<0.001**
	Signature	<0	Ref		Ref	
		>=0	0.62 (0.13 to 2.13)	0,484	2.92 (1.06 to 8.98)	0,046

Multivariable analyses were performed to assess the association between the refine innate signature components 1 and 2, and adverse events (AEs) adjusting for the vaccine dose. Logistic regression models were used. The reported adjusted odds ratios (ORs) capture the increase in risk of an AE compared with the reference category (denoted "Ref"). In grey, results that were similar between previous and refined signature.
Significant difference against the reference in the Doses or in the Signature component is represented by P-values highlighted in bold.

doses between component 1 of the refined signature and day 1 lymphopenia and thrombopenia, which was maintained until day 3 only for HD vaccinees. These negative associations of both doses with lymphopenia were correlated with all additional biomarkers while the negative correlation with thrombopenia was related to different biomarkers (Supplementary Table 3). Component 1 was differently associated with neutropenia according to the vaccine dose. Early (day 1) neutropenia was positively associated in HD vaccinees and was influenced mainly by MCP-3, while delayed neutropenia was negatively associated with LD vaccination.

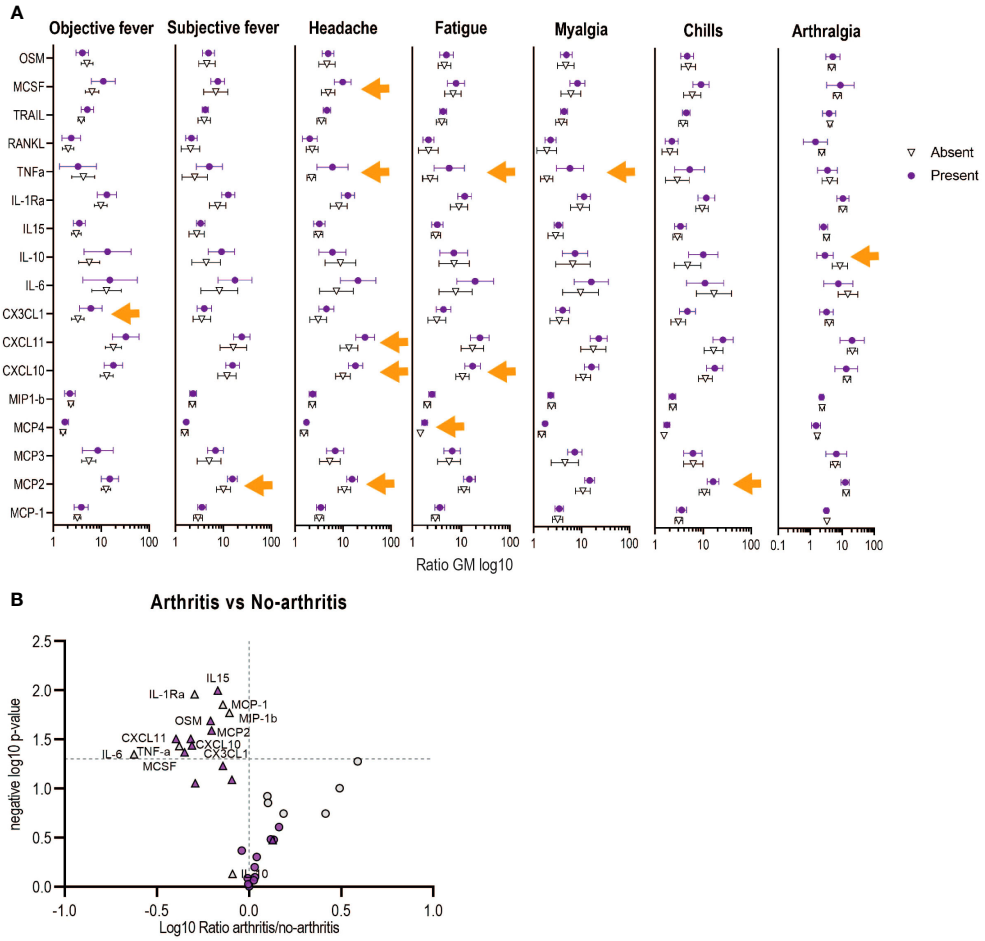


FIGURE 4
Associations between the refined signature biomarkers with early adverse events (AEs) in Geneva vaccinees receiving high vaccine dose. **(A)** Each symbol represents the ratio of geometric mean (log10 of day 1/day 0) for each biomarker. Bars shown mean and 95% CI. Orange arrows shows significance difference between having or not the indicated AE. **(B)** Volcano plots from the ratio of the plasma markers measured in those with arthritis vs those with no-arthritis in log10 fold change (x axis) against the t-test-derived negative log10 statistical P-value (y axis) for the additional (purple) and previous (grey) biomarkers of the refined signature. Thresholds (dotted grey line): p-value cut-off was fixed at 0.05 (1,3 negative log10) and fold change cut-offs is 1 (0 in the log10 scale). The two vaccine doses are shown with circles (low dose) or triangles (high dose).

Finally, in this analysis, we found limited correlation of the two PCs with antibody response except for Component 2 in HD vaccinees that positively correlated with antibody levels 180 days after vaccination (Supplementary Table 3). Others have reported correlation between the antibody levels at day 28 with day 3 CXCL10 levels when considering all vaccinees irrespective of the vaccine dose (19). A similar univariate analysis grouping the LD and HD groups showed that the antibody levels at day 28 positively correlated with the ratio D1/D0 (or actual concentrations at day 1) of several cytokines and chemokines, including CXCL-10 (Supplementary Figure 4). This correlation was limited to a more limited set of cytokines at day 3. Antibody response at day 180 was associated with the D1/D0 ratio of IL-10, MCP-1 and MIP-1b, and in HD only with IL-10 that drives the positive association found with Component 2 in HD vaccinees. In line with the multivariate analysis, there were fewer correlations between the antibody levels at day 28 with innate plasma biomarkers when considering each dose group separately, limited to positive correlation with D1/D0 ratio of MCP-1 and MIP-1b levels (LD group) and negative correlation with CXCL10 level (HD group; Supplementary Figure 4).

In summary, component 1 of the refined signature differentially correlated with LD viremia (positive) and hematological (negative) outcomes, several innate plasma biomarkers including CXCL10 were associated with antibody titers one month after vaccination but fewer with long-term specific antibody response.

Validation of the refined signature in an independent US cohort

The kinetics of the response of the 17 biomarkers in the US cohort was similar to the ones observed in the Geneva participants, although some differences were noted in the magnitude of the response (Supplementary Figure 5). In US HD vaccinees, the largest fold increases were observed for IL10 [58.1 (95% CI, 43 to 78)], CXCL10 [57.8 (95% CI, 43 to 79)] and CXL11 [28.6 (95% CI, 20 to

40)]. Although weaker in magnitude, the same markers including MCP-2 showed the largest fold increase in LD vaccinees (Supplementary Table 5). At baseline, most biomarkers were significantly lower in the US cohort, while the D1/D0 ratio showed similar responses in both cohorts, CXCL10, CXCL11, IL-10 and MCP-2 being the biomarkers with the highest ratio in both cohorts (Table 1; Supplementary Table 5).

To evaluate whether the signature defined using the Geneva cohort could predict rVSVΔG-ZEBOV-GP responses elicited in a different cohort, we applied an independent PCA to the US data. Similar to what was found in Geneva, three components explained 75.9% of the variability of the D1/D0 ratios (PC1 explained 63.6% of the variance, PC2: 6.4% and PC3: 5.9%) (Figure 5A). The bootstrap showed that the first three components were robust (PC1: n=50000/50000 (100%), PC2: n=49333/50000 (98.67%), PC3: n=27657/50000 (55.31%). The overall measure of adequacy was 0.93. Thus, the PCA model in the US samples was adequate and behaved very similarly as for the Geneva samples.

Comparable to what was observed in Geneva cohort, the first component also discriminates well between LD and HD (Supplementary Figure 6) and had a similar equation for component 1: "$0.085\times IL1Ra^{STD} + 0.07xIL6^{STD} + 0.08xTNFa^{STD} + 0.085xIL10^{STD} + 0.078xMCP1^{STD} + 0.064xMIP1b^{STD} + 0.056xMCP3^{STD} + 0.085xCXCL10^{STD} + 0.077xOSM^{STD} + 0.052xMCP4^{STD} + 0.08xCX3CL1^{STD} + 0.079xMCSF^{STD} + 0.085xCXCL11^{STD} + 0.078xTRAIL^{STD} + 0.067xMCP2^{STD} + 0.027xRANKL^{STD} + 0.08xIL15^{STD}$". Component 2's equation is shown in Supplementary Table 2. In addition, biomarkers contributing to component 1 were positively correlated, while the ones contributing to component 2 had both a positive and negative correlations (Figure 5A). In component 1, eleven biomarkers were above the expected average contribution with CXCL11, CXCL10, IL-10 and ILR-α being the highest, while for the component 2, three biomarkers contributed to the component variability, with RANKL representing 58% of the contribution (Figure 5B).

We next asked whether applying the Geneva first two components to the US data and vice versa would generate comparable results. Only the first component correlated strongly in both data sets, using Geneva data (r = 0.97, P < 0.001) and using US data (r = 0.99, P=0) (Figure 5C), and discriminated well the participants receiving the LD and the HD in both cohorts (Supplementary Figure 6).

The validation confirms that Component 1 of the refined signature accurately predicts the variability in response to the rVSVΔG-ZEBOV-GP vaccine.

GP-specific antibody levels also correlate with biomarkers in the US cohort

Similar to Geneva cohort, when considering all vaccinees irrespective of the vaccine dose the antibody levels at day 28 in all vaccinees positively correlated with the D1/D0 ratio of several cytokines, such as IL1RA, IL-10, MCP-1, CXCL10, MIP1b, CX3CL1, MCSF, CXCL11, TRAIL, IL-15. The correlation was also mostly lost when considering day 3 cytokine ratio and when

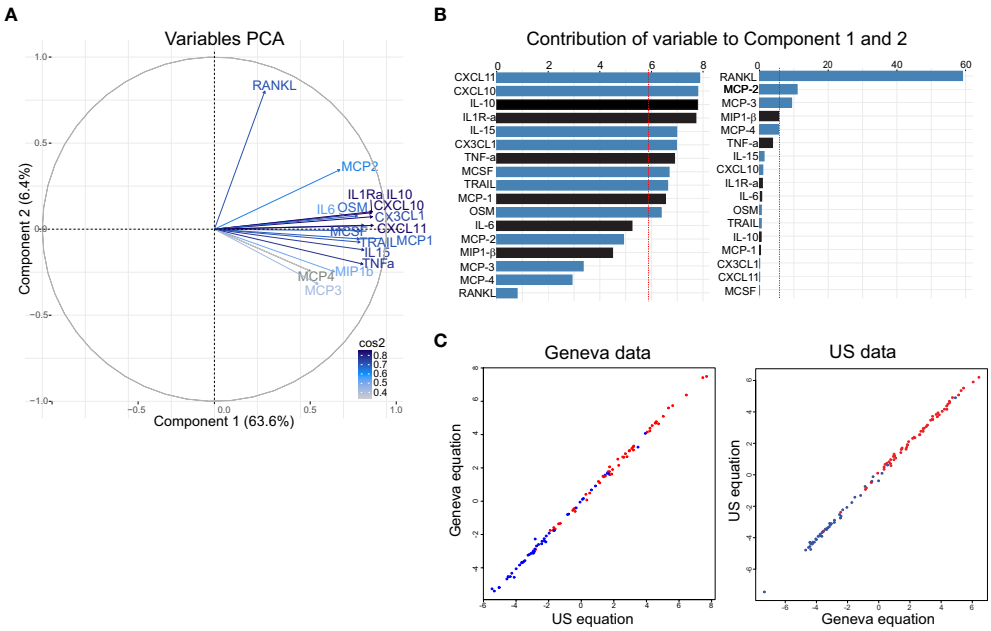


FIGURE 5
Analysis of the signature in the US cohort and validation of the refine signature defined in the Geneva cohort. **(A)** Variable correlation plot shows the magnitude (length of the arrow) and direction of the correlations of each marker (n=17) to each of the two principal components. Cos2 values indicates how well represented the marker is on the principal component and are shown in a gradient of colours: grey represent low values, light blue represents mid values, dark blue represents high values. **(B)** Percentage of contribution to the variability of each marker in the component 1 (left) and component 2 (right). Red dashed line indicates the expected average contribution. Bars in blue indicate additional markers and in black previous markers **(C)** Correlation between Geneva equation and US equation using Geneva data (left) and US data (right).

splitting by dose (Supplementary Figure 4). Unlike in the Geneva cohort, most of these correlations were maintained until day 180 after vaccination (Supplementary Figure 4).

Overall, US cohort innate plasma signature biomarkers also correlate with antibody levels at day 28 and 180 after vaccination with rVSVΔG-ZEBOV-GP.

Discussion

We showed that the inclusion of additional biomarkers refined the first plasma signature identified previously in Geneva. The refined signature, which now includes 17 markers, better discriminated between vaccine doses as it performed better at capturing the variability of the vaccine responses, and better defined the risk of fever, myalgia and headache. We also found new biomarkers that correlated with reactogenicity and transient arthritis, and that were associated with antibody levels one and six months after vaccination. Finally, the results were cross validated in a separate cohort.

We used O-link to screen for additional markers: of the many markers screened, only 18 were significantly higher in both HD and LD vaccinees. These markers are related to monocytes recruitment as well as to biological processes involved in vaccine responses such as pro-inflammatory cytokines, chemokine-signaling pathways, chemotaxis of different immune populations (monocytes, neutrophils, eosinophils and lymphocytes) and cellular response to interferon gamma. CXCL10, CXCL11, MCP-2, IL1R-α were the markers with the highest D1/D0 ratio as well as the ones with the greatest contribution to the variability of Component 1. CXCL10 and CXCL11 are IFN-dependent cytokines and plays an important role in the chemotaxis of monocytes, T-cells, NK cells and dendritic cells. They are secreted by monocytes, endothelial cells and fibroblasts, and their secretion is enhanced in the presence of TNF-α (29). This is in line with the positive correlation that we observed between CXCL10 and CXCL11 with TNF-α. Previous transcriptomic analysis from blood samples of the same cohorts have shown that interferon signaling genes (ISGs) were upregulated at day 1 post-vaccination and, consistent with our results, CXCL10 was upregulated at day 1 (30, 31). Similarly, the replication incompetent Ebola vaccine Ad26.ZEBOV increases the expression of IFN-stimulated genes (CXCL9, CXCL11, and CXCL10), and those associated with monocyte and lymphocyte recruitment such as CCL2 (MCP-1),CCL8 (MCP-2), and CCL7 (MCP-3) (32). However, compared to rVSVΔG-ZEBOV-GP, Ad26-ZEBOV combined with MVA-BN-Filo (Zabdeno/Mvabea) as well as another adenovirus-based Ebola vaccine cAd3-EBOZ is less immunogenic with less persisting antibody response, requiring higher doses to reach the same level of immunogenicity (33, 34).

Of note, we did not detect an increase in plasma IFN protein level (similar to previous reports (19)) but CXCL10 and CXCL11 increase may result from a transient and earlier IFN response before day 1. This discrepancy between gene expression and the protein level of IFN in blood might reflect rapid migration of cells to secondary lymphoid organs (33), rapid kinetics of the IFNs secretion (32) and/or a sub-optimal sensitivity of the assay used to detect these proteins. The innate vaccine response induced by the live attenuated rVSVΔG-ZEBOV-GP it is mainly related to monocyte recruitment and activation, whereas live-attenuated yellow fever mainly induces a dendritic-cell (DC) innate signature (35, 36) and the adjuvanted influenza-H1N vaccine induces a lymphoid gene-expression signature (37). More recently, SARSCoV-2 infection as well as mRNA vaccination were shown to induce a monocyte and DC innate signature with enhanced serum levels of IFN-α (38) and IFN-gamma, respectively (39).

Compared with the first signature reported previously (24), the refined signature presented herein explains a higher proportion of the variability of the D1/D0 ratios. Components 1 and 2 were both very robust and included different biomarkers that contributed to their variability, which can explain the different associations observed with dose, adverse events and biological outcomes. For instance, in contrast with the previous signature, component 1 was associated with risk of objective fever and myalgia, while component 2 (which represented only 8.5% of the variability) was the only one significantly associated with a risk of headache and with the GP-specific antibody response six months after vaccination.

Another important distinction with the previous signature was that several specific biomarkers were associated with the presence of systemic adverse events in HD vaccinees. Most of these associations were with single markers, for example high levels of CX3CL1 and MCP-2 were associated with the presence of objective fever and subjective fever, respectively. Increase in CX3CL1 plasma level has been associated with Hanta virus fever (40). CX3CL1 shedding can be induced by MCP-1 via p38 signaling (41). This is in line with the positive correlation we saw between plasma levels of CX3CL1 and MCP-1, suggesting that MCP-1 could induce shedding of CX3CL1. In addition, the correlation between fatigue and headache with TNF-α plasma levels found in the previous signature is now extended to several additional biomarkers including CXCL10.

Similarly, the risk of transient arthritis after vaccination is associated with the reduction of various additional biomarkers mainly in HD vaccinees. After rVSVΔG-ZEBOV-GP vaccination, 24% of Geneva trial reported transient arthritis and the virus was isolated in the synovial fluid (12). While in US trial the frequency of reported transient arthritis was 5%, the cases were dispersed across multiple doses including placebo (7, 9), likely confounding a direct comparison. In agreement with the previous signature (24), we found in Geneva cohort that Component 1 was significantly lower in HD vaccines who developed arthritis, this was ruled by 12 out of 17 biomarkers constituting the signature that had significantly lower plasma levels in HD vaccinees with arthritis. The topmost differentially expressed markers were IL-6, CXCL10, CXCL11, TNF-α and MCSF. Although the roles of IL-6 and TNF-α in rheumatoid arthritis (42, 43) and in chronic chikungunya arthritis (44) are well established, we saw a reduction during the acute phase. However, it is also well established that a robust cytokine response during the acute phase of viral infection is vital for clearance and control of viral dissemination, and prevention of chronic chikungunya arthritis (45). Our results suggest that individuals who developed arthritis after a HD vaccine (which in close to the dose currently in use in the field $72x10^6$pfu/dose)had a lower level of inflammatory response and therefore, we hypothesize have a less effective early control of viral dissemination, which may in turn leads to viral presence in privileged

sites such as joints, and thus could enhance the risk of vaccine-induced viral arthritis (8, 12, 46). The lack of association with bone resorption markers such as RANKL (47) is in line with the absence of bone resorption lesions in our arthritis patients (12), in contrast to chikungunya arthritis (46). Recently, transcriptomic analysis of the same Geneva cohort identified an early five-gene signature associated with the risk of arthritis that included T-cell subset genes CD4 and CCR7, IFN-regulatory sign gene FCGR1A, myeloid-associated gene IL12A, and Th2-associated gene GATA3 (30). Taken together, we hypothesized that the loss of T-cell homeostasis, a weak innate response during the acute phase (in HD vaccinees) and age at the time of vaccination (in LD vaccinees) are associated with transient arthritis after rVSVΔG-ZEBOV-GP vaccination.

We did not analyze the impact of baseline in the incidence of the adverse events observed after vaccination, but this was evaluated using machine learning in other paper by members of the consortium using the same cohorts as in the present study. In this study, 22 genes at baseline were associated with fatigue, headache, myalgia, fever, chills, arthralgia, nausea and arthritis (48).

Others have reported a correlation between the early innate response and specific-GP antibody levels one month after rVSVΔG-ZEBOV-GP vaccination that involved upregulation of ISGs such as IFI6 gene at day 7 (30) and CXCL10 protein levels at day 3 (19). We also found a positive correlation in both cohorts between specific-GP antibody titers one month after vaccination in all vaccinees and D1/D0 ratio of several innate plasma signature biomarkers including CXCL10 and IL-15. IL-15 and IFN-g have been reported to correlate with antibody response after the second dose of BTN162b2 mRNA COVID-19 vaccine (49). We also saw that in US cohort more innate biomarkers correlate with the antibody levels compared to the Geneva cohort; we can not exclude that this could be due to a difference in the vaccine dose in the two countrioes, since for the HD groups participants in the US received $100x10^6$pfu/dose (n=30) and $20x10^6$pfu/mL (n=30), while in Geneva, participants received $10x10^6$pfu/dose (n=35) $50x10^6$pfu/dose (n=16).These results highlight the key role of early activation of interferon-dependent responses at the transcriptional and protein level in the generation of high antibody levels, as reported for other vaccines (49–53).

We validated this refined signature in a US cohort. Although baseline levels of IL-10 were higher in the US than in the Geneva cohort, the kinetics of the biomarkers as well as the components of rVSVΔG-ZEBOV-GP early response were remarkably comparable. This implies that innate responses induced after rVSVΔG-ZEBOV-GP vaccination were very robust, likely independent of genetic and environmental background. The biomarkers that contributed to the US Component 1 variability were similar to the ones in Geneva's, except for IL-10, which was significantly higher for the US Component 1. In contrast, Component 2 in the two cohorts have different sets of markers that contribute to the variability. For instance, in Geneva IL-10, TNF-a, MIP-1b and IL-6 are the main contributors, whereas for the US cohort the main contributors are RANKL, MCP-2 and MCP-3.

The study identified certain markers by O-link, with CLEC4G/4C/4D/6A showing significant increases in the high-dose (HD) group, while only CLEC4C increased in the low-dose (LD) group. These markers belong to C-type lectin ligands receptors (CLRs), recognized as pattern recognition receptors (PRRs) and are crucial for initiating innate immune responses. CLEC4G, known as LSECtin, serves as an attachment factor for Ebola and SARS viruses, (54) and plays an important role in Ebola GP-mediated inflammatory responses in human DCs by inducing TNF-α and IL-6 secretion (55). CLEC4C is found exclusively on plasmacytoid dendritic cells (pDCs) and can bind various cells and viruses, including HIV-1 and hepatitis C virus (56, 57, Florentin et al., 2011). CLEC6A (Dectin-2) is an FcRγ-coupled receptor on macrophages and dendritic cells, proposed as a potential attachment factor for Ebola (56). CLEC4D (MCL) is a macrophage C-type lectin implicated in the upregulation of innate genes post-vaccination. Altogether, this suggest that CLEC proteins that increased after rVSVΔG-ZEBOV-GP vaccination may have the potential to bind to the Ebola glycoprotein. This binding could lead to the activation of monocytes, macrophages, and dendritic cells. However, further research is required to fully understand the role of these CLEC proteins in the context of vaccination.Our study has limitations, we were not able to quantify all the markers that were found with the initial O-link screening because they are not available within the Luminex technology, a technique that we had to use to allow comparison with our previous study. Binding antibody responses were assessed at different labs on samples from two different cohorts. This may have also led to some variability in correlation analysis. It would also be interesting to conduct *in vitro* studies to define which cells produce these biomarkers associated with AEs upon rVSVΔG-ZEBOV-GP exposure, in particular cells from the joint, skin or vascular.

In conclusion, we refined the early plasma innate signature induced by rVSVΔG-ZEBOV-GP vaccine, which now better correlates with the presence of AEs, hematological changes, viremia and antibody titer in Geneva cohort. This refined signature was validated in an independent US cohort and showed strong correlation between cohorts, demonstrating its robustness and potential for broad applicability. This innate refined plasma signature highlights the importance of the innate response, especially of monocytes, in the development of rVSV-vaccine responses, and its potential role in controlling vaccine dissemination to prevent arthritis. Altogether, these results provide new insights into early blood biomarkers of immunogenicity and reactogenicity of the rVSVΔG-ZEBOV-GP vaccine.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The study involved humans and were approved by Ethics committees of the WHO for both cohorts. Local Ethics committees for Geneva trial: Canton of Geneva and and Swiss Agency for Therapeutic Products (Swissmedic). Local committees for USA trial: the Chesapeake Institutional Review Boards (Columbia, MD, USA) and the Crescent City Institutional Review Board (New Orleans, LA, USA). The studies were conducted in accordance with the local

legislation and institutional requirements. The human samples used in this study were acquired from previous studies for which ethical approval was obtained as mentioned before and informed consent for participation was signed. For this study extra written informed consent for participation was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and institutional requirements.

Author contributions

PM-M: Conceptualization, Investigation, Methodology, Project administration, Validation, Visualization, Writing – original draft, Writing – review & editing. AH: Conceptualization, Writing – review & editing, Supervision. SL: Data curation, Formal analysis, Methodology, Visualization, Writing – review & editing. DM: Conceptualization, Funding acquisition, Resources, Writing – review & editing. TO: Conceptualization, Funding acquisition, Resources, Writing – review & editing. AH: Conceptualization, Funding acquisition, Resources, Writing – review & editing. AD: Supervision, Visualization, Writing – original draft, Writing – review & editing. C-AS: Conceptualization, Funding acquisition, Methodology, Project administration, Resources, Supervision, Writing – original draft, Writing – review & editing.

Group members of VEBCON Consortium

Gabon Selidji Todagbe Agnandji, Sanjeev Krishna, Peter G. Kremsner, Jessica S. Brosnahan (Centre de Recherches Médicales de Lambaréné). *Germany* Selidji Todagbe Agnandji, Sanjeev Krishna, Peter G. Kremsner, Jessica S. Brosnahan (Institut für Tropenmedizin, Universitätsklinikum Tübingen); Marylyn M. Addo (University Medical Center Hamburg); Stephan Becker, Verena Krähling (Institute of Virology, Marburg). *UK* Sanjeev Krishna (St. George's University of London). *Kenya* Philip Bejon, Patricia Njuguna (Kenya Medical Research Institute, Kilifi). S*witzerland* Claire-Anne Siegrist, Angela Huttner (Geneva University Hospitals, Geneva); and Marie-Paule Kieny, Vasee Moorthy, Patricia Fast, Barbara Savarese, Olivier Lapujade (World Health Organization, Geneva).

Group members of VSV-EBOVAC Consortium

Selidji Todagbe Agnandji, Rafi Ahmed, Sravya S. Nakka, Floriane Auderset, Philip Bejon, Luisa Borgianni, Jessica Brosnahan, Simone Lucchesi, Olivier Engler, Mariëlle C. Haks, Ali M. Harandi, Donald Gray Heppner, Alice Gerlini, Angela Huttner, Peter Gottfried Kremsner, Donata Medaglini, Thomas Monath, Francis Ndungu, Patricia Njuguna, Tom H. M. Ottenhoff, David Pejoski, Mark Page, Gianni Pozzi, Francesco Santoro, Claire-Anne Siegrist.

Group members of VSV-EBOPLUS Consortium

Selidji Todagbe Agnandji, Sravya S. Nakka, Luisa Borgianni, Annalisa Ciabattini, Sheri Dubey, Michael Eichberg, Olivier Engler, Patrícia Gonzalez-Dias, Peter Gottfried Kremsner, Ali M. Harandi, Alice Gerlini, Angela Huttner, Donata Medaglini, Helder Nakaya, Tom H. M. Ottenhoff, Gianni Pozzi, Sylvia Rothenberger, Francesco Santoro, Eleonora Vianello, Claire-Anne Siegrist.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This project has received funding from the Innovative Medicines Initiative 2 Joint Undertaking under grant agreement No 116068 (VSV-EBOPLUS project) and No 115842 (VSV-EBOVAC project). This Joint Undertaking receives support from the European Union's Horizon 2020 research and innovation program and EFPIA. Conduction of the North American trial was funded in part with Federal funds from the Department of Health and Human Services; Office of the Assistant Secretary for Preparedness and Response; Biomedical Advanced Research and Development Authority, under contract number HHSO100201500002C.

Acknowledgments

The authors thank all participants in the cohort studies. We also thank Christophe Combescure for his statistical expertise advice during the data analysis of the refined signature. We also express gratitude to Michael Eichberg and Sheriy Dubey for granting us access to the US cohort plasma samples used as validation in this study.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2023.1279003/full#supplementary-material

References

1. Sivanandy P, Jun PH, Man LW, Wei NS, Mun NFK, Yii CAJ, et al. A systematic review of ebola virus disease outbreaks and an analysis of the efficacy and safety of newer drugs approved for the treatment of ebola virus disease by the US food and drug administration from 2016 to 2020. *J Infect Public Health* (2022) 15(3):285–92. doi: 10.1016/j.jiph.2022.01.005

2. Geisbert TW, Daddario-Dicaprio KM, Geisbert JB, Reed DS, Feldmann F, Grolla A, et al. Vesicular stomatitis virus-based vaccines protect nonhuman primates against aerosol challenge with ebola and marburg viruses. *Vaccine* (2008) 26(52):6894–900. doi: 10.1016/j.vaccine.2008.09.082

3. Geisbert TW, Daddario-Dicaprio KM, Lewis MG, Geisbert JB, Grolla A, Leung A, et al. Vesicular stomatitis virus-based ebola vaccine is well-tolerated and protects immunocompromised nonhuman primates. *PloS Pathog* (2008) 4(11):e1000225. doi: 10.1371/journal.ppat.1000225

4. Jones SM, Feldmann H, Stroher U, Geisbert JB, Fernando L, Grolla A, et al. Live attenuated recombinant vaccine protects nonhuman primates against ebola and marburg viruses. *Nat Med* (2005) 11(7):786–90. doi: 10.1038/nm1258

5. Regules JA, Beigel JH, Paolino KM, Voell J, Castellano AR, Hu Z, et al. A recombinant vesicular stomatitis virus ebola vaccine. *N Engl J Med* (2015) 376(4):330–41. doi: 10.1056/NEJMoa1414216

6. ElSherif MS, Brown C, MacKinnon-Cameron D, Li L, Racine T, Alimonti J, et al. Assessing the safety and immunogenicity of recombinant vesicular stomatitis virus ebola vaccine in healthy adults: A randomized clinical trial. *CMAJ Can Med Assoc J = J l'Association medicale Can* (2017) 189(24):E819–e27. doi: 10.1503/cmaj.170074

7. Heppner DG Jr., Kemp TL, Martin BK, Ramsey WJ, Nichols R, Dasen EJ, et al. Safety and immunogenicity of the rvsvg-zebov-gp ebola virus vaccine candidate in healthy adults: A phase 1b randomised, multicentre, double-blind, placebo-controlled, dose-response study. *Lancet Infect Dis* (2017) 17(8):854–66. doi: 10.1016/S1473-3099(17)30313-4

8. Agnandji ST, Huttner A, Zinser ME, Njuguna P, Dahlke C, Fernandes JF, et al. Phase 1 trials of rvsv ebola vaccine in Africa and Europe. *N Engl J Med* (2015) 374 (17):1647–60. doi: 10.1056/NEJMoa1502924

9. Halperin SA, Arribas JR, Rupp R, Andrews CP, Chu L, Das R, et al. Six-month safety data of recombinant vesicular stomatitis virus–zaire ebola virus envelope glycoprotein vaccine in a phase 3 double-blind, placebo-controlled randomized study in healthy adults. *J Infect Dis* (2017) 215(12):1789–98. doi: 10.1093/infdis/jix189

10. Henao-Restrepo AM, Camacho A, Longini IM, Watson CH, Edmunds WJ, Egger M, et al. Efficacy and effectiveness of an rvsv-vectored vaccine in preventing ebola virus disease: final results from the Guinea ring vaccination, open-label, cluster-randomised trial (Ebola ça suffit!). *Lancet (London England)* (2017) 389(10068):505–18. doi: 10.1016/s0140-6736(16)32621-6

11. Kennedy SB, Bolay F, Kieh M, Grandits G, Badio M, Ballou R, et al. Phase 2 placebo-controlled trial of two vaccines to prevent ebola in Liberia. *N Engl J Med* (2017) 377(15):1438–47. doi: 10.1056/NEJMoa1614067

12. Huttner A, Dayer JA, Yerly S, Combescure C, Auderset F, Desmeules J, et al. The effect of dose on the safety and immunogenicity of the vsv ebola candidate vaccine: A randomised double-blind, placebo-controlled phase 1/2 trial. *Lancet Infect Dis* (2015) 15(10):1156–66. doi: 10.1016/S1473-3099(15)00154-1

13. Wells CR, Pandey A, Parpia AS, Fitzpatrick MC, Meyers LA, Singer BH, et al. Ebola vaccination in the democratic republic of the Congo. *Proc Natl Acad Sci USA* (2019) 116(20):10178–83. doi: 10.1073/pnas.1817329116

14. WHO. *Who prequalifies ebola vaccine, paving the way for its use in high-risk countries* Vol. 12. . Geneva, Switzerland: WHO (2019). p. 2020. Available at: https://wwwwhoint/news-room/detail/12-11-2019-who-prequalifiesebola-vaccine-paving-the-way-for-its-use-in-high-risk-countries.

15. FDA. Drug Administration. *First FDA-approved vaccine for the prevention of ebola virus disease, marking a critical milestone in public health preparedness and response. FDA* Vol. 19. FDA (2019). Available at: https://www.fda.gov/news-events/press-announcements/first-fda-approved-vaccine-prevention-ebola-virus-disease-marking-critical-milestone-public-health.

16. EMA. *First vaccine to protect against ebola.* (2019), *EMA/CHMP/565403/2019*. Available at: https://www.ema.europa.eu/en/news/first-vaccine-protect-against-ebola.

17. Marzi A, Engelmann F, Feldmann F, Haberthur K, Shupert WL, Brining D, et al. Antibodies are necessary for rvsv/zebov-gp-mediated protection against lethal ebola virus challenge in nonhuman primates. *Proc Natl Acad Sci USA* (2013) 110(5):1893–8. doi: 10.1073/pnas.1209591110

18. Marzi A, Robertson SJ, Haddock E, Feldmann F, Hanley PW, Scott DP, et al. Vsv-ebov rapidly protects macaques against infection with the 2014/15 ebola virus outbreak strain. *Science* (2015) 349(6249):739–42. doi: 10.1126/science.aab3920

19. Rechtien A, Richert L, Lorenzo H, Martrus G, Hejblum B, Dahlke C, et al. Systems vaccinology identifies an early innate immune signature as a correlate of antibody responses to the ebola vaccine rvsv-zebov. *Cell Rep* (2017) 20(9):2251–61. doi: 10.1016/j.celrep.2017.08.023

20. Marzi A, Jankeel A, Menicucci AR, Callison J, O'Donnell KL, Feldmann F, et al. Single dose of a vsv-based vaccine rapidly protects macaques from marburg virus disease. *Front Immunol* (2021) 12:774026. doi: 10.3389/fimmu.2021.774026

21. O'Donnell KL, Feldmann F, Kaza B, Clancy CS, Hanley PW, Fletcher P, et al. Rapid protection of nonhuman primates against marburg virus disease using a single low-dose vsv-based vaccine. *EBioMedicine* (2023) 89:104463. doi: 10.1016/j.ebiom.2023.104463

22. Cross RW, Woolsey C, Prasad AN, Borisevich V, Agans KN, Deer DJ, et al. A recombinant vsv-vectored vaccine rapidly protects nonhuman primates against heterologous lethal lassa fever. *Cell Rep* (2022) 40(3):111094. doi: 10.1016/j.celrep.2022.111094

23. Pejoski D, de Rham C, Martinez-Murillo P, Santoro F, Auderset F, Medaglini D, et al. Rapid dose-dependent natural killer (Nk) cell modulation and cytokine responses following human rvsv-zebov ebolavirus vaccination. *NPJ Vaccines* (2020) 5(1):32. doi: 10.1038/s41541-020-0179-4

24. Huttner A, Combescure C, Grillet S, Haks MC, Quinten E, Modoux C, et al. A dose-dependent plasma signature of the safety and immunogenicity of the rvsv-ebola vaccine in Europe and Africa. *Sci Transl Med* (2017) 9(385). doi: 10.1126/scitranslmed.aaj1701

25. Heppner DG Jr., Kemp TL, Martin BK, Ramsey WJ, Nichols R, Dasen EJ, et al. Safety and immunogenicity of the rvsvΔG-zebov-gp ebola virus vaccine candidate in healthy adults: A phase 1b randomised, multicentre, double-blind, placebo-controlled, dose-response study. *Lancet Infect Dis* (2017) 17(8):854–66. doi: 10.1016/s1473-3099(17)30313-4

26. Assarsson E, Lundberg M, Holmquist G, Björkesten J, Thorsen SB, Ekman D, et al. Homogenous 96-plex PEA immunoassay exhibiting high sensitivity, specificity, and excellent scalability. *PLoS One* (2014) 9(4):e95192. doi: 10.1371/journal.pone.0095192

27. Rudge TL Jr., Sankovich KA, Niemuth NA, Anderson MS, Badorrek CS, Skomrock ND, et al. Development, qualification, and validation of the filovirus animal nonclinical group anti-ebola virus glycoprotein immunoglobulin G enzyme-linked immunosorbent assay for human serum samples. *PloS One* (2019) 14(4): e0215457. doi: 10.1371/journal.pone.0215457

28. Kaiser HF. An index of factorial simplicity. *Psychometrika* (1974) 39(1):31–6. doi: 10.1007/BF02291575

29. Tokunaga R, Zhang W, Naseem M, Puccini A, Berger MD, Soni S, et al. Cxcl9, cxcl10, cxcl11/cxcr3 axis for immune activation - a target for novel cancer therapy. *Cancer Treat Rev* (2018) 63:40–7. doi: 10.1016/j.ctrv.2017.11.007

30. Vianello E, Gonzalez-Dias P, van Veen S, Engele CG, Quinten E, Monath TP, et al. Transcriptomic signatures induced by the ebola virus vaccine rvsvδg-zebov-gp in adult cohorts in Europe, Africa, and North America: A molecular biomarker study. *Lancet Microbe* (2022) 3(2):e113–e23. doi: 10.1016/S2666-5247(21)00235-4

31. Santoro F, Donato A, Lucchesi S, Sorgi S, Gerlini A, Haks MC, et al. Human transcriptomic response to the vsv-vectored ebola vaccine. *Vaccines (Basel)* (2021) 9(2). doi: 10.3390/vaccines9020067

32. Blengio F, Hocini H, Richert L, Lefebvre C, Durand M, Hejblum B, et al. Identification of early gene expression profiles associated with long-lasting antibody responses to the ebola vaccine ad26.Zebov/mva-bn-filo. *Cell Rep* (2023) 42(9):113101. doi: 10.1016/j.celrep.2023.113101

33. Woolsey C, Geisbert TW. Current state of ebola virus vaccines: A snapshot. *PloS Pathog* (2021) 17(12):e1010078. doi: 10.1371/journal.ppat.1010078

34. Malik S, Kishore S, Nag S, Dhasmana A, Preetam S, Mitra O, et al. Ebola virus disease vaccines: development, current perspectives & Challenges. *Vaccines (Basel)* (2023) 11(2). doi: 10.3390/vaccines11020268

35. Pulendran B, Oh JZ, Nakaya HI, Ravindran R, Kazmin DA. Immunity to viruses: learning from successful human vaccines. *Immunol Rev* (2013) 255(1):243–55. doi: 10.1111/imr.12099

36. Querec T, Bennouna S, Alkan S, Laouar Y, Gorden K, Flavell R, et al. Yellow fever vaccine yf-17d activates multiple dendritic cell subsets via tlr2, 7, 8, and 9 to stimulate polyvalent immunity. *J Exp Med* (2006) 203(2):413–24. doi: 10.1084/jem.20051720

37. Sobolev O, Binda E, O'Farrell S, Lorenc A, Pradines J, Huang Y, et al. Adjuvanted influenza-H1n1 vaccination reveals lymphoid signatures of age-dependent early responses and of clinical adverse events. *Nat Immunol* (2016) 17(2):204–13. doi: 10.1038/ni.3328

38. Vono M, Huttner A, Lemeille S, Martinez-Murillo P, Meyer B, Baggio S, et al. Robust innate responses to sars-cov-2 in children resolve faster than in adults without compromising adaptive immunity. *Cell Rep* (2021) 37(1):109773. doi: 10.1016/j.celrep.2021.109773

39. Li C, Lee A, Grigoryan L, Arunachalam PS, Scott MKD, Trisal M, et al. Mechanisms of innate and adaptive immunity to the pfizer-biontech bnt162b2 vaccine. *Nat Immunol* (2022) 23(4):543–55. doi: 10.1038/s41590-022-01163-9

40. Zhang C, Tang K, Zhang Y, Ma Y, Du H, Zheng X, et al. Elevated plasma fractalkine level is associated with the severity of hemorrhagic fever with renal syndrome in humans. *Viral Immunol* (2021) 34(7):491–9. doi: 10.1089/vim.2020.0244

41. Green SR, Han KH, Chen Y, Almazan F, Charo IF, Miller YI, et al. The cc chemokine mcp-1 stimulates surface expression of cx3cr1 and enhances the adhesion of monocytes to fractalkine/cx3cl1 via P38 mapk. *J Immunol (Baltimore Md 1950)* (2006) 176(12):7412–20. doi: 10.4049/jimmunol.176.12.7412

42. Radner H, Aletaha D. Anti-tnf in rheumatoid arthritis: an overview. *Wiener medizinische Wochenschrift (1946)* (2015) 165(1-2):3–9. doi: 10.1007/s10354-015-0344-y

43. Favalli EG. Understanding the role of interleukin-6 (Il-6) in the joint and beyond: A comprehensive review of il-6 inhibition for the management of rheumatoid arthritis. *Rheumatol Ther* (2020) 7(3):473–516. doi: 10.1007/s40744-020-00219-2

44. Chow A, Her Z, Ong EK, Chen JM, Dimatatac F, Kwek DJ, et al. Persistent arthralgia induced by chikungunya virus infection is associated with interleukin-6 and granulocyte macrophage colony-stimulating factor. *J Infect Dis* (2011) 203(2):149–57. doi: 10.1093/infdis/jiq042

45. Chang AY, Tritsch S, Reid SP, Martins K, Encinales L, Pacheco N, et al. The cytokine profile in acute chikungunya infection is predictive of chronic arthritis 20 months post infection. *Dis (Basel Switzerland)* (2018) 6(4). doi: 10.3390/diseases6040095

46. Chen W, Foo SS, Sims NA, Herrero LJ, Walsh NC, Mahalingam S. Arthritogenic alphaviruses: new insights into arthritis and bone pathology. *Trends Microbiol* (2015) 23(1):35–43. doi: 10.1016/j.tim.2014.09.005

47. Cao X. Rankl-rank signaling regulates osteoblast differentiation and bone formation. *Bone Res* (2018) 6(1):35. doi: 10.1038/s41413-018-0040-9

48. Gonzalez Dias Carvalho PC, Dominguez Crespo Hirata T, Mano Alves LY, Moscardini IF, do Nascimento APB, Costa-Martins AG, et al. Baseline gene signatures of reactogenicity to ebola vaccination: A machine learning approach across multiple cohorts. *Front Immunol* (2023) 14:1259197. doi: 10.3389/fimmu.2023.1259197

49. Bergamaschi C, Terpos E, Rosati M, Angel M, Bear J, Stellas D, et al. Systemic il-15, ifn-Γ, and ip-10/cxcl10 signature associated with effective immune response to sars-cov-2 in bnt162b2 mrna vaccine recipients. *Cell Rep* (2021) 36(6):109504. doi: 10.1016/j.celrep.2021.109504

50. Li S, Rouphael N, Duraisingham S, Romero-Steiner S, Presnell S, Davis C, et al. Molecular signatures of antibody responses derived from a systems biology study of five human vaccines. *Nat Immunol* (2014) 15(2):195–204. doi: 10.1038/ni.2789

51. Nakaya HI, Wrammert J, Lee EK, Racioppi L, Marie-Kunze S, Haining WN, et al. Systems biology of vaccination for seasonal influenza in humans. *Nat Immunol* (2011) 12(8):786–95. doi: 10.1038/ni.2067

52. Pulendran B, Ahmed R. Immunological mechanisms of vaccination. *Nat Immunol* (2011) 12(6):509–17. doi: 10.1038/ni.2039

53. Li S, Sullivan NL, Rouphael N, Yu T, Banton S, Maddur MS, et al. Metabolic phenotypes of response to vaccination in humans. *Cell* (2017) 169(5):862–77 e17. doi: 10.1016/j.cell.2017.04.026

54. Gramberg T, Hofmann H, Möller P, Lalor PF, Marzi A, Geier M, et al. LSECtin interacts with filovirus glycoproteins and the spike protein of SARS coronavirus. *Virology* (2005) 340(2):224–36. doi: 10.1016/j.virol.2005.06.026

55. Zhao D, Han X, Zheng X, Wang H, Yang Z, Liu D, et al. The Myeloid LSECtin is a DAP12-coupled receptor that is crucial for inflammatory response induced by Ebola virus glycoprotein. *PLoS Pathog* (2016) 12(3):e1005487. doi: 10.1371/journal.ppat.1005487

56. Kerscher B, Willment JA, Brown GD. The Dectin-2 family of C-type lectin-like receptors: an update. *Int Immunol* (2013) 25(5):271–7. doi: 10.1093/intimm/dxt006

57. Riboldi E, Daniele R, Parola C, Inforzato A, Arnold PL, Bosisio D, et al. Human C-type lectin domain family 4, member C (CLEC4C/BDCA-2/CD303) is a receptor for asialo-galactosyl-oligosaccharides. *J Biol Chem* (2011) 286(41):35329–33. doi: 10.1074/jbc.C111.290494

frontiers | Frontiers in Immunology

TYPE Technology and Code
PUBLISHED 08 February 2024
DOI 10.3389/fimmu.2024.1322712

OPEN ACCESS

EDITED BY
Helder Nakaya,
University of São Paulo, Brazil

REVIEWED BY
Qingfei Pan,
St. Jude Children's Research Hospital, United States
Jan Zaucha,
AstraZeneca, United Kingdom

*CORRESPONDENCE
Morten Nielsen
morni@dtu.dk

†These authors have contributed equally to this work

RECEIVED 16 October 2023
ACCEPTED 08 January 2024
PUBLISHED 08 February 2024

CITATION
Høie MH, Gade FS, Johansen JM, Würtzen C, Winther O, Nielsen M and Marcatili P (2024) DiscoTope-3.0: improved B-cell epitope prediction using inverse folding latent representations. *Front. Immunol.* 15:1322712. doi: 10.3389/fimmu.2024.1322712

DiscoTope-3.0: improved B-cell epitope prediction using inverse folding latent representations

Magnus Haraldson Høie[1], Frederik Steensgaard Gade[1], Julie Maria Johansen[1], Charlotte Würtzen[1], Ole Winther[2,3,4], Morten Nielsen[1*†] and Paolo Marcatili[1†]

[1]Department of Health Technology, Section for Bioinformatics, Technical University of Denmark (DTU), Kgs. Lyngby, Denmark, [2]Section for Cognitive Systems, DTU Compute, Technical University of Denmark (DTU), Kgs. Lyngby, Denmark, [3]Center for Genomic Medicine, Rigshospitalet (Copenhagen University Hospital), Copenhagen, Denmark, [4]Department of Biology, Bioinformatics Centre, University of Copenhagen, Copenhagen, Denmark

Accurate computational identification of B-cell epitopes is crucial for the development of vaccines, therapies, and diagnostic tools. However, current structure-based prediction methods face limitations due to the dependency on experimentally solved structures. Here, we introduce DiscoTope-3.0, a markedly improved B-cell epitope prediction tool that innovatively employs inverse folding structure representations and a positive-unlabelled learning strategy, and is adapted for both solved and predicted structures. Our tool demonstrates a considerable improvement in performance over existing methods, accurately predicting linear and conformational epitopes across multiple independent datasets. Most notably, DiscoTope-3.0 maintains high predictive performance across solved, relaxed and predicted structures, alleviating the need for experimental structures and extending the general applicability of accurate B-cell epitope prediction by 3 orders of magnitude. DiscoTope-3.0 is made widely accessible on two web servers, processing over 100 structures per submission, and as a downloadable package. In addition, the servers interface with RCSB and AlphaFoldDB, facilitating large-scale prediction across over 200 million cataloged proteins. DiscoTope-3.0 is available at: https://services.healthtech.dtu.dk/service.php?DiscoTope-3.0.

KEYWORDS

structure-based, B cell epitope prediction, inverse-folding, antibody epitope prediction, ESM-IF1, immunogenicity prediction, vaccine design

1 Introduction

A key mechanism in humoral immunity is the precise binding of B-cell receptors and antibodies to their molecular targets, named antigens. The antigen regions that are involved in the binding are known as B-cell epitopes. B-cell epitopes are found on the surface of antigens, and in the case of proteins they can be classified as linear if the epitope residues are

sequentially arranged along the antigen sequence, or discontinuous if they are only proximal in the antigen tertiary structure, but not in the primary structure. Identification of B-cell epitopes has large biotechnological applications, including rational development of vaccines and immunotherapeutics. However, experimental mapping of epitopes remains expensive and resource intensive. Computational tools for B-cell epitope prediction offer a viable and large-scale alternative to experiments. However, prediction of B-cell epitopes remains a challenging problem (1, 2). Historically, in-silico prediction methods have been either antigen sequence- or structure-based. Sequence-based methods such as BepiPred-2.0 (3) are attractive given the high availability of protein sequences. BepiPred-2.0 utilizes a random forest trained on structural features predicted from the antigen sequence, but has limited accuracy and struggles to predict conformational or non-linear epitopes (4). In a recent work, BepiPred-3.0 (5) further improves the method, demonstrating large gains by exploiting sequence representations from the protein language model ESM-2 (6). It was shown to outperform previous sequence based tools, including Seppa-3.0 (7), ElliPro (8), BeTop (9) and EPSVR/EPMeta (10).

Structure-based methods should benefit from having direct access to the antigen tertiary structure, and in particular, its surface topology. DiscoTope-2.0 (11) was published in 2012, and it estimates epitope propensity from the local geometry of each residue, taking into consideration both its solvent accessibility and the direction of its side chain. Older structure-based methods like DiscoTope-2.0 and the newer epitope3D (12) are still outperformed by the sequence-based BepiPred-3.0 (5). However novel methods such as the inverse-folding based SEMA (13) and the geometric deep-learning network ScanNet (14) have shown promising advances. Recently, ScanNet demonstrated improved performance by explicitly considering geometric details at both the resolution of individual atoms and amino-acids. However, while structure-based prediction tools may demonstrate improved performance, they are limited by the availability of antigen structures.

Data scarcity affects the accuracy of prediction tools in different ways. Firstly, they constrain the amount of data on which such tools can be trained. As of January 2023, less than 5500 antibody structures in complex with an antigen are available in the antibody-specific structural database SabDab (15). After filtering this dataset for redundancy, one may be left with less than 1500 structures for training, which limits the complexity of the models that can be reliably trained without incurring in overfitting (5).

Secondly, the available data is a biased sampling of the possible antibody-antigen complexes. We find that most antigens are found only once in the dataset, while others, likely due to medical or biological interest, have been resolved in complex with as many as 43 (15) different antibodies. This means that one cannot confidently annotate negative residues; they might be part of antibody-antigen complexes yet to be solved.

Lastly, undersampling of epitopes will also result in an imprecise assessment of the tools' accuracy; predicted epitopes that appear as false positives may just be in antibody bound regions yet to be identified. The last two points (bias and undersampling) are typical of a class of problems known as Positive-Unlabeled (PU) learning. In this scenario, we are only confident of positive epitopes, while all remaining (surface) residues should be treated as unlabeled. Several approaches have been proposed for increasing the accuracy of B-cell epitope prediction methods and their estimated metrics in such cases (12, 16, 17). A simple yet effective strategy is to train ensemble predictors based on bootstrapping of samples in the Unlabelled class (18), also known as PU bagging, which is the approach that we employ in this work.

With recent advances in protein structure prediction, AlphaFold2 (19) has enabled accurate prediction of protein structures directly from sequences. Currently, over 200 million pre-computed structures are available in AlphaFold DB (20), covering every currently cataloged protein in UniProt (21). The three-dimensional coordinates of the proteins, together with the local quality reported as pLDDT scores, are readily accessible from the database.

To truly harness the remarkable progress in generating accurate structural models, we must develop robust and informative numerical representations of both predicted and resolved structures. This is especially crucial for deep-learning methods, which thrive on such tasks. The ESM-IF1 inverse folding model is an equivariant graph neural network pre-trained to recover native protein sequences from protein backbones structures (Ca, C and N atoms). The structure-based representations which may be extracted from this model have been shown to outperform sequence-based representations on tasks such as predicting binding affinity and change in protein stability (22). Crucially, ESM-IF1 is explicitly trained on both solved and AlphaFold predicted structures, enabling large-scale application of its representations even when solved structures are unavailable.

In this work, we train DiscoTope-3.0, a structure-based B-cell epitope prediction tool exploiting inverse folding representations generated from either AlphaFold predicted or solved structures. DiscoTope-3.0 is trained on both predicted and solved antigen structures using a positive-unlabelled learning ensemble approach, enabling large-scale prediction of epitopes even when solved structures are unavailable. We compare its performance versus previous tools and the impact in performance when using predicted structures versus solved structures, in both cases showing substantially improved accuracy. DiscoTope-3.0 is implemented as a web server and downloadable package interfacing with both RCSB and AlphaFoldDB.

2 Results

The positive-unlabelled ensemble training strategy for DiscoTope-3.0 is shown in Figure 1. First, epitopes from solved antibody-antigen complexes are mapped onto the antigen sequences (1). Using sequences as input, antigen structures are predicted using AlphaFold2 (2). Next, per-residue structural representations, for both solved and predicted structures, are extracted using the ESM-IF1 protein inverse folding model (3 and 4). During training, random subsets of epitopes and unlabelled residues are sampled across the dataset (5), before finally training an ensemble of XGBoost models on the individual data subsets (6). The final DiscoTope-3.0 score is given as the average score from the ensemble models (7). More details on the training procedure are available below and in the Methods section.

10.3389/fimmu.2024.1322712

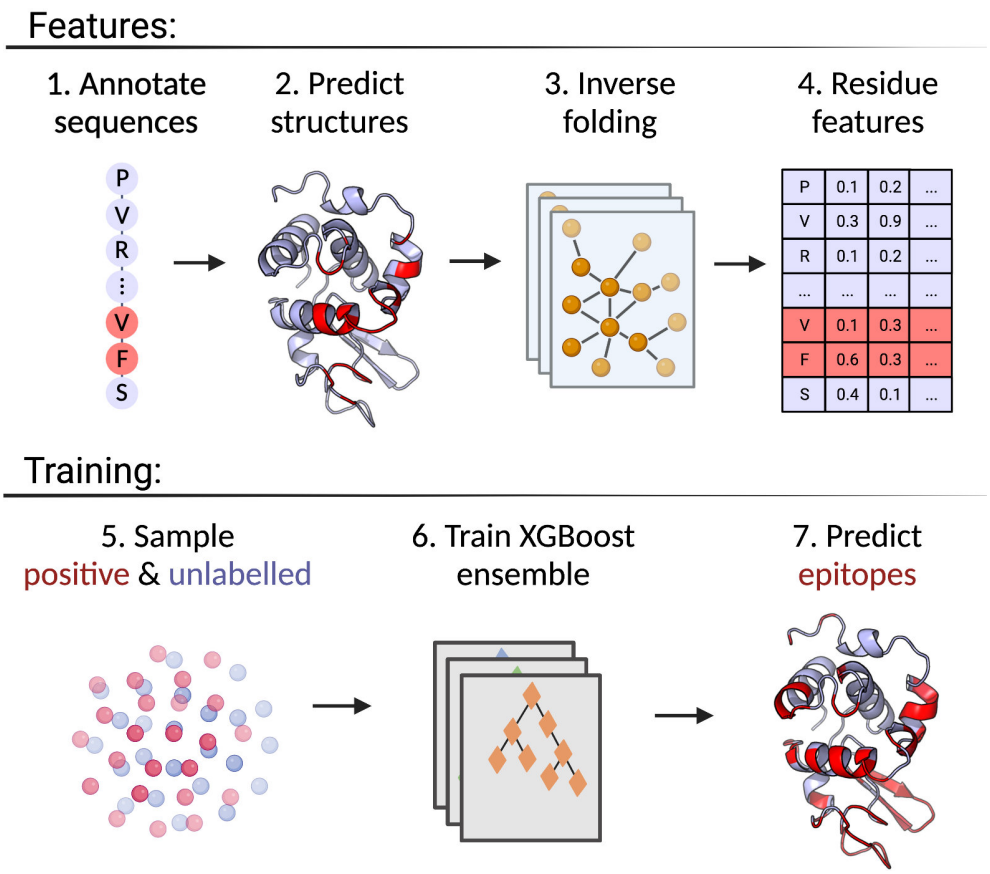


FIGURE 1
Overview of the DiscoTope-3.0 method. 1) Antigen sequences are extracted from solved antibody-antigen complexes with mapped epitopes. 2) Antigen structures are either provided as solved, or predicted with AlphaFold2. 3) Structures are embedded with ESM-IF1. 4) Per-reside features are extracted. 5) Labelled epitope residues and non-labelled residues are extracted. 6) An XGBoost ensemble is trained on the resulting dataset. 7) The final model predicts epitope-propensity on input solved or predicted structures. Created with BioRender.com.

Here, we present a quick overview of the dataset and feature pre-processing procedure. DiscoTope-3.0 training and validation is based on the BepiPred-3.0 dataset of 582 antibody-antigens complexes, covering a total of 1466 antigen chains IEDB (23). Epitopes are defined as the set of residues within 4 Å of any antibody heavy atom (see Methods). The training and hyperparameter tuning is based on 2 different datasets: Training and Validation, while evaluation is performed on the Validation and external test sets. The external test set consists of 24 antigens collected from SAbDab (15) and PDB (24) on October 20, 2022. These antigens share at most 20% similarity to both our own, BepiPred and ScanNet's training datasets (see Methods).

In addition to using experimentally solved antigens for training, structures for the individual antigen chains were additionally predicted using AlphaFold2. Both the solved and predicted chains were then embedded with ESM-IF1. Further we extract for each residue its relative surface accessibility (RSA), AlphaFold local quality score (pLDDT) as well as the antigen length and a one-hot encoding for the antigen sequence (see Methods and Table 1). These structural features (or subsets) were used to train an ensemble of XGBoost models and the ensemble average is used as the final prediction score.

We chose to use XGBoost for our architecture due to their robustness to outliers and noise, minimal need to adjust model hyperparameters (25), and enabling combination of multiple "weak learners" in our PU (positive and unlabelled) learning ensemble (26, 27), to produce a robust final prediction.

Structure-based representations have been shown to be a powerful representation in different downstream tasks. To see if this is also the case for B-cell epitope prediction, we evaluated the results obtained using different feature encoding schemes on our validation set of AlphaFold structures (for details on this dataset refer to Methods and Table 1). First, we assess whether training a single XGBoost model using structure representations from predicted structures outperforms a similar model based on the sequence representations from ESM-2 (Figure 2). Here, we observe a marginal but consistent epitope prediction performance using the structure (AUC-ROC 0.767 ± 0.003) vs sequence representations (AUC-ROC 0.751 ± 0.003) ($p < 0.0001$).

As explained in the introduction, the B-cell epitope prediction problem can be categorized in the broad class of PU training. Incorrectly labeled negative examples can negatively affect the training, by introducing frustration in the learning process (28). We can observe that, by using an ensemble learning strategy with a dataset bagging approach based on previous works (28–30) (see Methods), we can further improve performance (AUC-ROC 0.791 ± 0.001) and generalization.

10.3389/fimmu.2024.1322712

TABLE 1 Feature overview.

Feature	Dimensions	Description
ESM-IF1 embeddings	512	Inverse folding representations from input antigen structures
Antigen sequence	20	Amino-acid sequence, one-hot encoded
Antigen length	1	Length of sequence
AlphaFold quality score	1	Residue pLDDT score, as predicted by AlphaFold2
Relative surface accessibility	1	Calculated by Shrake-Rupley algorithm
Total:	535	L x 535

Input features for the XGBoost model architecture.

2.1 Effect of using predicted versus solved structures

One of the risks in training models on either exclusively solved or AlphaFold structures is that the methods might over-specialize to one source and perform worse on the other, or even be affected by data leakage. For example, a model may overfit on conformational changes present in the side chains of epitope residues in solved antibody-antigen complexes.

By training on both predicted and solved structures, we obtain a final model which performs well on both structure types, with an AUC-ROC 0.799 ± 0.001 for predicted structures (Figure 2), and 0.807 ± 0.001 when predicting solved structures (Supplementary Figure S1). We note that training separate models, namely using only solved or only predicted structures, does indeed improve performance slightly when tested on the same class (AUC-ROC 0.813 and 0.805 respectively), but comes at added complexity. To simplify comparison with other tools, we therefore chose the DiscoTope-3.0 version trained on both structure types for further analysis.

2.2 Benchmark comparison to state-of-the-art methods

To further test the effect of using predicted versus solved structures, we used the external test set of 24 antigens. These antigens share at most 20% sequence similarity to both our own, BepiPred and ScanNet's training datasets (see Methods). We benchmark against the structure-based tools ScanNet and SEMA, while including BepiPred-3.0, as a purely sequence-based and independent of the different structural variations, and a naïve predictor using relative surface accessibility as a score. We note that all benchmarked tools use the same definition for epitope residues, thus ensuring a fair comparison.

The precision and recall scores of the tools were calculated on this test set. The results of this evaluation are displayed in Figure 3. Here, DiscoTope-3.0 outperforms all other tools, for both predicted (AUC-PR 0.232 ± 0.02 vs closest 0.177 ± 0.02 BepiPred-3.0) and solved structures (0.223 ± 0.02 vs closest 0.185 ± 0.02 SEMA) (see Supplementary Figure S2 and Table 2 for more performance metrics). We note that DiscoTope-3.0 here strongly outperforms BepiPred-3.0, and point to the BepiPred-3.0 publication for an extensive benchmark demonstrating BepiPred-3.0 again outperforming epitope3D, Seppa-3.0, Ellipro and the previous version of DiscoTope-2.0.

We introduce a novel metric, the epitope rank score, primarily due to the need for a fair and normalized comparison among different tools, that operate with varying score scales. To put it simply, to calculate the epitope rank scores, we rank-normalize the scores for a given antigen, then find the mean rank for all observed epitopes in that antigen. For instance, a mean epitope rank score of 70% signifies that, on average, epitopes score in the upper 70th percentile of residue scores (see

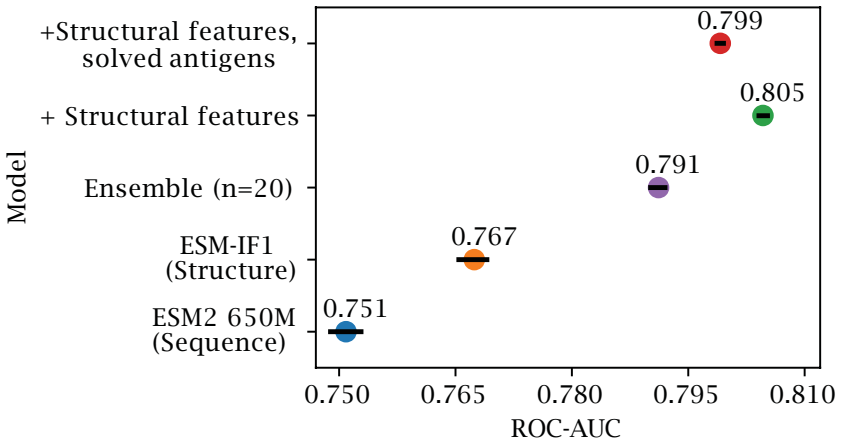


FIGURE 2
Effects of inverse folding and bagging. Ablation results on the validation set of AlphaFold predicted structures, with less than 50% sequence similarity to the training set. The plot reports the AUC-ROC for a single XGboost model trained on representations based on ESM-2 650M parameter (blue) and on ESM-IF1 (orange), for an ensemble of 20 XGboost models based on bootstraps of ESM-IF1 representations (purple), for models where additional structural features are included (see Methods) tested on both AlphaFold models (green) and on the corresponding solved structures (red) (see Methods). Error bars indicate 95% confidence interval.

10.3389/fimmu.2024.1322712

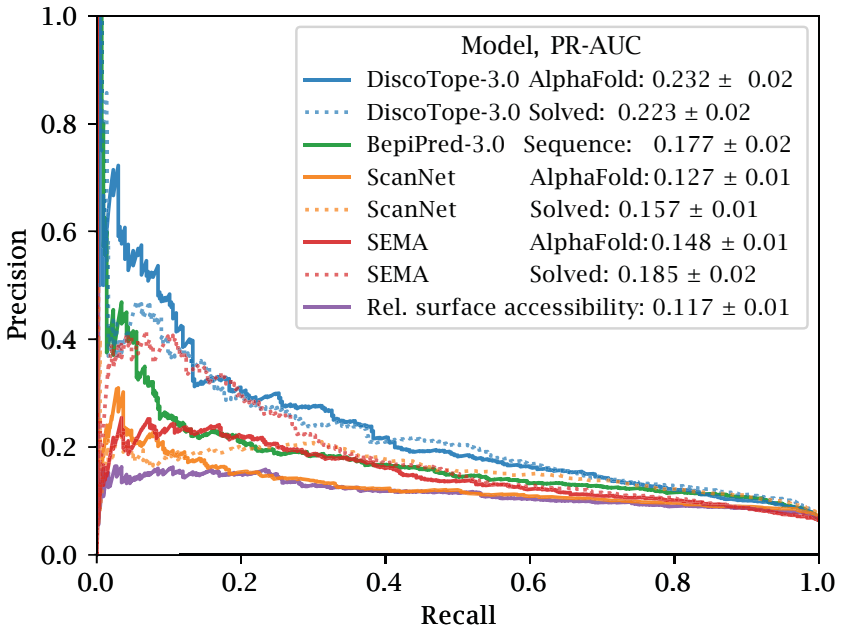


FIGURE 3
Improved performance on solved and predicted structures. AUC-PR curve plots on the external test set of 24 antigen chains, at most 20% similar to the training set of all models. Structures provided as AlphaFold predicted, experimentally solved, or sequence in the case of BepiPred-3.0. Standard deviation calculated from bootstrapping 1000 times (see Methods). See Supplementary Figure S2 and Table 2 for additional performance metrics. Please see BepiPred-3.0 publication (5) for its improved performance versus DiscoTope-2.0, epitope3D, Seppa-3.0 and ElliPro.

Methods). A typical real-case scenario for this metric, would be for users to submit individual antigens, and then to analyze the top scoring epitope residues, regardless of their specific scores. Using this metric, DiscoTope-3.0 consistently outperforms ScanNet, SEMA and BepiPred-3.0 in the case of predicted structures, and is only matched in performance by ScanNet on solved structures (Supplementary Figure S3).

2.3 Robustness to relaxation and predicted structures

We note that DiscoTope-3.0's performance is largely unaffected by the type of structures used for prediction. To further test the robustness of the tools to minor differences in the antigen structures, we performed an energy minimization on the solved structures using the software FoldX (31). This minimization only impacts the side chain, thus leaving the backbone of the native structure unaltered. The ESM-IF model does not use the side chain atoms in its structure representations, and consequently DiscoTope-3.0 should not be affected by the relaxation process.

We observe that after side-chain relaxation in solved structures, ScanNet's epitope rank scores are reduced by $\sim$ 3.1 percentile points, while swapping solved for predicted structures leads to a loss of $\sim$ 7.5 percentile points (see Methods). In contrast to this, DiscoTope-3.0 only loses $\sim$ 0.1 and $\sim$ 0.6 percentile points respectively, again indicating robustness to the modeling process (Supplementary Figure S4).

These observations can be attributed to the different ways the two models process structural features. ScanNet uses side-chain atomic coordinates explicitly, whereas DiscoTope-3.0 relies solely on the accuracy of backbone modeling. This difference suggests that some models, like ScanNet, might overfit to the specific orientations of side-chains present only in bound antibody-antigen complexes, information which would not be useful in predicting novel epitopes. By training models on both predicted and relaxed, solved structures, we can potentially avoid this overfitting and increase the generalizability of the models.

TABLE 2 Performance on benchmarking datasets.

Metric	Dataset		
	Lysozyme (AlphaFold)	External test set (AlphaFold)	External test set (Solved)
AUC-PR	0.722	0.232	0.223
AUC-ROC	0.809	0.783	0.795
MCC	0.521	0.227	0.214
Total residues	129	6788	6788
Observed epitope residues	55 (223 before collapsing)	436	436
# antigen structures	1 (12 before collapsing)	24	24

Overview of external test set and lysozyme test sets for solved and AlphaFold predicted antigens. Matthew correlation coefficient (MCC) calculated at optimal sensitivity-specificity threshold using the Youden-index.

2.4 Improved prediction on exposed and non-linear epitopes

We also investigated if the structural information available to DiscoTope-3.0, ScanNet and SEMA affects the prediction of different types of epitopes. To this aim, epitopes were split into different sub-categories (Exposed, Buried, Linear and Non-linear). Exposed and Buried epitope residues are defined depending on whether their relative surface accessibility was above or below 20%, respectively. Linear epitopes are defined as any group of 3 or more epitope residues found sequentially along the antigen sequence, allowing for a possible gap of up to 1 unlabeled residue in between. Finally non-linear epitopes were defined as epitopes not satisfying the conditions of the linear group.

The result of this performance evaluation in the external test set reveals improved performance of DiscoTope-3.0 across all epitope subsets (Figure 4). DiscoTope-3.0 performance is remarkably good for non-linear epitopes. In the case of buried epitopes (relative surface accessibility < 20%), all models score poorly in the 30-37th percentile (not shown). This low performance is likely an artifact of the epitope labeling definition (shared between all tools), where inaccessible residues in proximity to the bound antibody are included in an epitope patch, despite not directly being involved in molecular interactions with the antibody.

2.5 Effect of predicted structural quality

Next, we investigate how the quality of the AlphaFold predicted structures affects the prediction of exposed epitopes. Overall, lower structural quality leads to small decrease in predictive performance (Figure 5), with high quality structures (pLDDT 95-100) having a mean epitope rank score of 84.2%, and moderate quality structures (pLDDT 85-95) having a non-significant decrease in mean epitope rank score of 81.2%. Only the group of antigens in the lowest quality pLDDT 60-85 group ($\sim$ 9% of antigens) perform significantly worse, with a score of 75.5% ($p < 0.005$). Fitting a linear model, the epitope rank score on average lowers by about 5 percentile points for every 10 point decrease in structural quality or pLDDT score (Figure 5B).

2.6 Calibrating scores for antigen length and surface area

We note that DiscoTope, BepiPred and SEMA exhibit a bias towards lower scores for longer antigen lengths (Pearson correlation -0.74, -0.71 and -0.51 respectively on external test set, not shown). If using a fixed threshold for binary epitope prediction, this results in most residues in shorter antigens being assigned as positives, while longer antigens may have all residues assigned as negatives.

To correct for this length bias, we calibrate antigen scores based on a predicted μ and standard deviation value, calculated from the antigen length and its mean surface score (see Methods and Supplementary Figure S8). The calibrated scores demonstrate independence towards the antigen length, and clear separation of buried and exposed residues across antigens in the validation set S6. Furthermore, we find that calibrating the scores enables setting a fixed threshold that provides reliable epitope recall across shorter and longer antigens. For example, if we choose the 50th percentile calibrated score for exposed epitopes (in the validation set), and use this for binary epitope prediction on the lysozyme case study (see

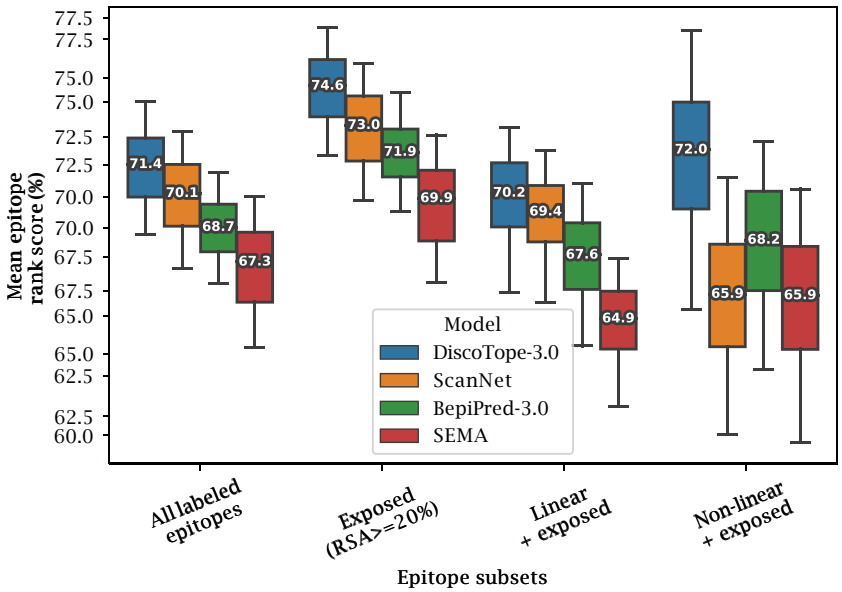


FIGURE 4
Improved performance on linear and non-linear epitopes. Mean epitope rank scores across antigens in the external test set, for the following epitope subsets: All labeled epitopes, Exposed (relative surface accessibility < 20%), Exposed Linear epitopes and exposed Non-linear epitopes (see text and Methods). Mean values calculated after bootstrapping 1000 times, with whiskers showing 95% distribution range.

10.3389/fimmu.2024.1322712

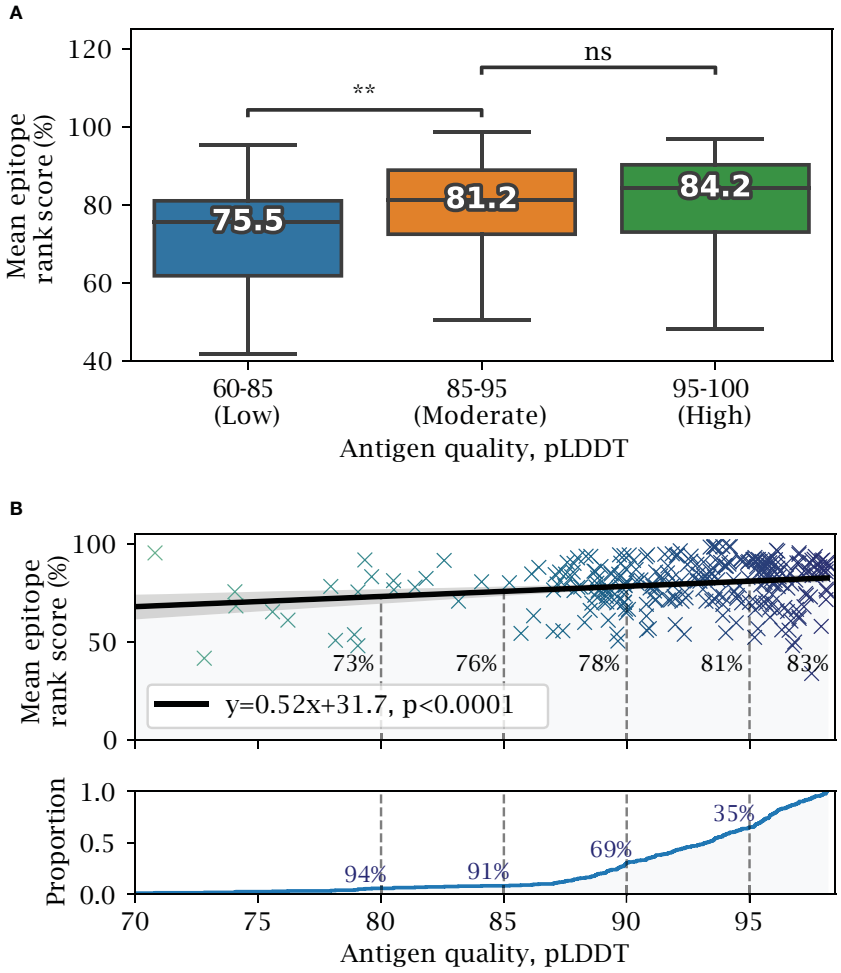


FIGURE 5
Effects of predicted structural quality. Validation set performance on AlphaFold predicted antigens dependent on predicted structural quality, excluding buried epitopes. **(A)** Epitope rank score distribution for antigens split into increasing quality bins of mean antigen pLDDT 60-85, 85-95 and 95-100. Median value for each distribution is shown, with paired one-tailed t-test comparison (** = $p < 0.005$). **(B)** Mean antigen pLDDT versus mean epitope rank score, with a fitted linear model shown in black. Below, cumulative distribution of mean antigen pLDDT, with a 91% proportion exceeding a pLDDT of 85, and 35% exceeding 95 respectively.

next section), we achieve an expected ~ 50% epitope recall (see Methods and Supplementary Figure S7).

2.7 Lysozyme case study with collapsed epitopes

As a noteworthy test case, we evaluate the performance of DiscoTope-3.0 on lysozyme, a well-studied antigen extensively mapped against different antibodies. First, we identified 12 lysozyme chains with mapped epitopes at 90% similarity to the chain C of the PDB structure 1A2Y. Next, DiscoTope-3.0 was re-trained excluding these chains. Next, we calculated an antibody hit rate, a ratio of on the number of times a given epitope residue was observed as an epitope across all of the 12 structures. Here, a score of 90% means the same residue was observed as an epitope in 11 out of 12 of the chains, which is the case for 5 out of 129 residues.

Overall, we find that calibrated DiscoTope-3.0 scores correlate with the observed epitope count or antibody hit rate with a Spearman correlation of 0.58 (Figure 6). Fitting a linear model, we find that a 0.20 point increase in calibrated scores on average leads to a 10% increase in the antibody hit rate ($p < 0.0001$, Supplementary Figure S5).

We note that the residues at positions ~ 30-40 (Figure 6) score highly in DiscoTope but lacked observed epitopes. Upon further investigation into the IEDB database, we found this region to be part of a discontinuous epitope patch (including K31, R32, G34, D36, G37, G40 …) bound by a camelid antibody deposited under the PDB id 4I0C.

2.8 DiscoTope-3.0 web server

Finally, we deployed a DiscoTope-3.0 web server, which enables rapidly predicting epitopes on either solved or predicted structures (Figure 7). The web server currently accepts batches of up to 50 PDB files at a time, with any number of chains. Users may upload structures directly as PDB files, or automatically fetch existing structures submitted as a set of RCSB or AlphaFoldDB IDs. Output predictions are easily visualized through an interactive 3D

10.3389/fimmu.2024.1322712

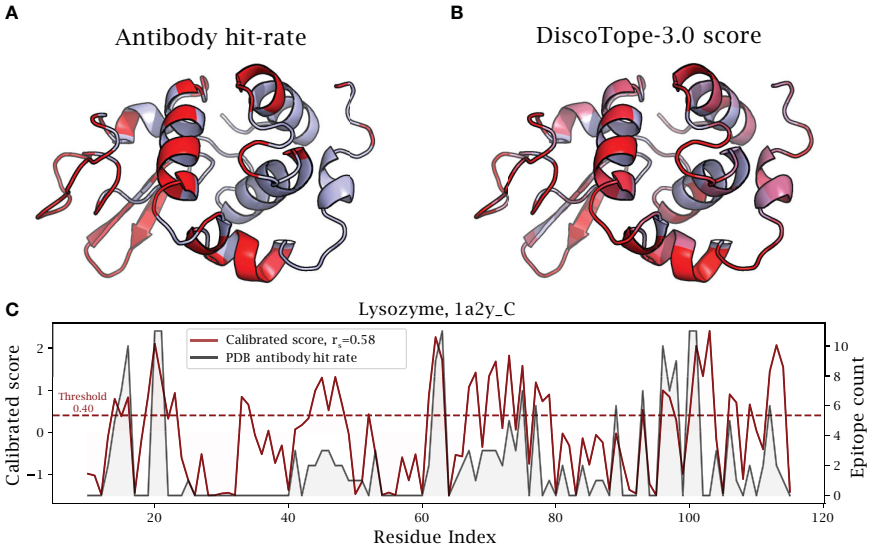


FIGURE 6
DiscoTope-3.0 score significantly correlates with antibody hit rate. Lysozyme epitope propensity as predicted by DiscoTope-3.0, excluding all lysozyme antigens from training. **(A)** PDB antibody hit rate mapped AlphaFold predicted structure (chain 1a2y_C), with increasing epitope propensity shown in red. **(B)** DiscoTope-3.0 score. **(C)** Epitope propensity visualized across the antigen sequence. Calibrated DiscoTope-3.0 score and antibody hit rate (epitope counts) shown, as measured from aligning the 12 epitope mapped lysozyme sequences (Spearman R = 0.58). Additional performance metrics available in Table 2.

view directly on the web server using Molstar (32), and predictions may be downloaded in both a CSV and PDB format.

3 Discussion

In this work, we present DiscoTope-3.0, a tool for improved B-cell epitope prediction. Our method exploits structure representations extracted from the ESM-IF1 inverse folding model. Extensive benchmarking of the tool demonstrated state of the art performance on both solved and predicted structures. Importantly the performance, in contrast to earlier proposed structure-based models, was found to be maintained when shifting to predicted and relaxed structures. This observation is of critical importance since it alleviates the need for experimentally solved structures imposed in current structure-based models, and allows for predicted structures to be applied for accurate B-cell epitope predictions. This extends the applicability of the tool by 3 orders of magnitude, from ~ 200K solved structures in the PDB (24), to ~ 200M predicted structures available in the AlphaFoldDB (20).

We note that other structure-based tools perform worse than the sequence-based BepiPred-3.0 in cases where only predicted structures and their sequences are available. This may arise from

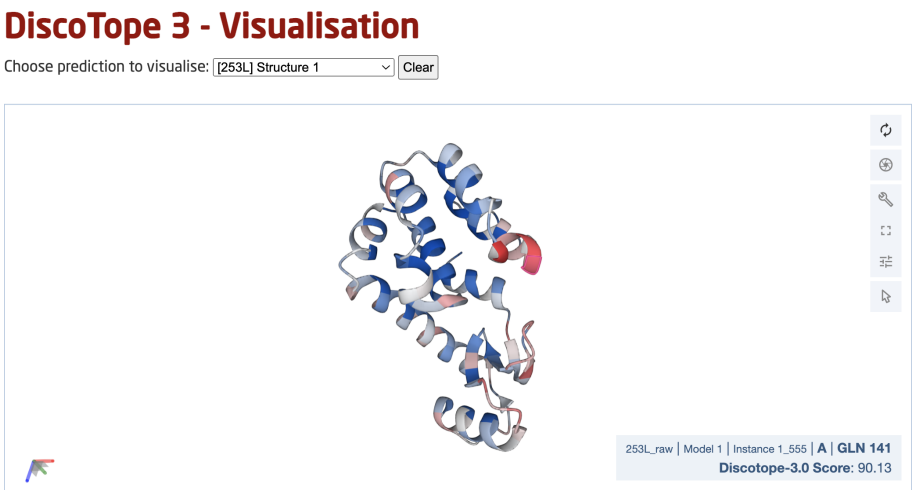


FIGURE 7
DiscoTope-3.0 web server interface. The web server provides an interactive 3D view for each predicted protein structure. DiscoTope-3.0 score on an example PDB, with increasing epitope propensity from blue to red. DiscoTope-3.0 is accessible at: https://services.healthtech.dtu.dk/service.php?DiscoTope-3.0, https://services.healthtech.dtu.dk/service.php?DiscoTope-3.0.

sensitivity to the quality of predicted structures, or relying on signals only present in solved or unrelaxed structures. DiscoTope-3.0's use of structure representations based on the protein backbone makes it robust to the predicted structural quality, and remarkably, able to perform similarly across solved, predicted and relaxed structures. It is, to the best of our knowledge, the first tool that presents highly accurate results on protein structural models.

We also find that other DiscoTope-like B-cell epitope prediction tools demonstrate a bias towards lower scores for longer antigens. After calibrating DiscoTope-3.0 scores for antigen length and surface residue scores, we provide calibrated score thresholds which provides the user with consistent expected epitope recall rates across shorter and longer antigens.

Finally, DiscoTope-3.0 interfaces with AlphaFoldDB and RCSB, enabling rapid batch processing across all currently cataloged proteins in UniProt and deposited solved structures. The web server is made freely available for academic use, accepting up to 50 input structures at a time, with any number of chains.

Our tool has been trained and evaluated on individual antigen chains. One could envision that, for multimeric antigen structures, it would be possible to further increase the tool performance by training and testing on the antigen complex. At this time, AlphaFold2 modeling accuracy for complexes is not yet on par with its accuracy on individual chains, and predicted complexes are not yet available in the AlphaFoldDB. As the science and technology behind the structural modeling progresses, it will be likely possible to further improve B-cell epitope predictions.

On the other hand, the positive-unlabelled learning strategy based on ensemble models and dataset bagging we use displays a remarkable boost in performance. We can imagine that, given the large dimension of the potential antibody space, the large gap between potential and observed epitopes will not be easily filled. An alternative strategy, that could circumvent this problem and provide valuable information to users, would be to perform antibody-specific epitope predictions. This approach has been tested by us and others in the past (33, 34), but the results are yet to provide a significant improvement in accuracy.

In summary, DiscoTope-3.0 is the first structure-based B-cell epitope prediction model that accepts and maintains state-of-the-art predictive power across solved, relaxed and predicted antigen structures. We believe this advance will serve as an important aid for the community in the quest for novel rational methods for the design of novel immunotherapeutics.

4 Methods

4.1 Training and evaluation of DiscoTope-3.0

The antigen training dataset as presented in BepiPred-3.0 was used as the starting point for our work. The dataset consists of 582 AbAg crystal structures from the PDB, filtered for a minimum resolution of 3.0 Å and R-factor 0.3. Epitopes are defined as any antigen residue containing at least 1 heavy atom within 4 Å of an antibody heavy atom. From this dataset, using the tool MMseqs2, we first remove any sequences with more than 20% sequence identity to the BepiPred-3.0 test set, resulting in 1406 chains. Next, the antigen sequences are clustered at 50% sequence identity. Each cluster has then been selected to be part of the validation (281 chains) or the training set (1125 chains).

In the ablation study, single XGBoost models (25) with default parameters were trained using representations from either the predicted structure or antigen sequence respectively. When testing feature combinations, ensemble size and effect of training on solved and predicted structures, error bars were estimated from re-training 20 times.

We manually adjusted three XGBoost hyperparameters from their defaults, guided by suggestions in the XGBoost documentation (??, xgb) and after observing improved performance on the validation set. Specifically, decision-tree *max_depth* was adjusted from 6 to 4, and the training data subsampling ratio *subsample* from 1.00 to 0.50 to reduce overfitting. *n_estimators* was adjusted from 100 to 200 trees after observing a plateauing improvement in the validation set AUC-ROC. The *gpu_hist* method was used to enable faster training on a GPU.

4.2 Dataset bagging and ensemble training

When sampling residues for each model in the ensemble, we randomly select 70% of available observed epitopes (positives) across the training dataset, then sample unlabelled residues (negatives) with a ratio of 5:2. When using both predicted and solved structures, these were sampled at a 1:1 ratio.

Ensembles were constructed by iteratively training independently trained XGBoost models on the randomly sampled datasets. When training an ensemble, we set a different random seed each time.

4.3 ESM-IF1 and ESM-2 representations

To generate per-residue ESM-IF1 structure representations, antigen structures were first split into single chains, and these inputted into ESM-IF1 following the instructions as listed on the official repository (Research, 35).

```
1. import esm.inverse_folding.
2. structure = esm.inverse_folding.util.load_structure(fpath, chain_id).
3. coords, seq = esm.inverse_folding.util.extract_coords_from_structure(structure).
4. rep = esm.inverse_folding.util.get_encoder_output(model, alphabet, coords).
```

For per-residue ESM-2 sequence representations, sequences were first extracted from all antigen chains and stored in a FASTA format. Next, the FASTA file was provided as input to

the official extract.py script (Research, 35) using the pre-trained ESM-2 650M parameter model.

```
./extract.py --model_location esm2_t33_650M_UR50D \
--fasta_file sequences.fasta \

--include per_tok \
--output_dir output/
```

4.4 Feature calculation and data filtering

Each isolated chain was processed as a single PDB file with ESM-IF1, extracting for each residue its latent representation from the ESM-IF1 encoder output. pLDDT values were either extracted from the PDB files in the case of AlphaFold structures, or set to 100 for solved structures. In the case of training on both solved and predicted structures, we include a binary input feature set to 1 if the input is an AlphaFold2 model, and 0 for solved structures.

Residue solvent accessible surface area was calculated using the Shrake-Rupley algorithm using Biotite (36), with default settings, and converted to relative surface accessibility using the Sander and Rost 1994 (37) scale as available in Biopython (38).

When training DiscoTope-3.0, we removed any antigen with less than 5 or more than 75 epitope residues, as well as PDBs with a mean pLDDT score below 85 or residues with a pLDDT below 70. No data filtering was performed during evaluation on the validation and external test datasets.

4.5 Calibration of DiscoTope-3.0 scores

When using calibrated scores, each antigen's DiscoTope-3.0 scores are normalized using the following formula:

$$Calibrated\ \ score = \frac{score - \mu}{std}$$

The values for μ and std are calculated for each antigen, using two separate linear generative additive models (GAMs) (39) fitted on the validation set. The length to μ model is fitted on antigen length versus mean score of antigen surface residues (RSA $>$ 20%), while the surface mean to std model is fitted on antigen mean surface residue score versus standard deviation of the same scores (Supplementary Figure S8).

4.6 External test set generation and evaluation

The external test set, used for comparing our tool to ScanNet and BepiPred-3.0, consists of solved antibody-antigen complexes deposited in either SAbDab and the PDB after April 2021 (collection date June 2022). Any antigen with more than 20% sequence identity to the training datasets used in this work, in ScanNet, or in BepiPred-3.0 were removed using MMseqs2. We annotated epitopes using the same approach as in BepiPred-3.0, which is common to all the tools.

We submitted either solved or AlphaFold2 predicted structures to the ScanNet web server (40), using the antibody-antigen binding mode and otherwise default parameters. BepiPred-3.0 predictions were generated from its online web server (5) using the antigen sequence and default parameters.

When evaluating DiscoTope-3.0 on the external test set, we retrained the final model with an ensemble size of 100, on the full training and validation set.

4.7 AlphaFold2 modeling and structural relaxation

Sequences for each antigen chain containing at least 1 epitope were extracted and modeled with the ColabFold implementation of AlphaFold2 at default settings. We picked the top ranking PDB after 5 independent iterations of 3 recycles, as ranked by AlphaFold2's internal quality measure.

For relaxation of the solved structures we used the foldx_20221231 version of FoldX, with the RepairPDB command for relaxing residues with bad torsion angles, van der Waals clashes or high total energy.

4.8 Data analysis

To calculate the mean epitope rank score, the predicted residue scores for an antigen were first ranked in ascending order. Next, we calculated the average of the rank scores for all epitope residues.

Exposed epitopes were defined as all epitopes with a relative surface accessibility exceeding 20%, while the remaining epitopes were defined as buried.

When reported, significance testing was performed with a one-sided paired t-test using scipy.stats.ttest_rel (41). The linear model on the mean antigen pLDDT vs mean epitope rank scores was fitted using a linear least-squares regression model (scipy.stats.linregress) with two-sided alternative hypothesis testing.

For confidence estimation with bootstrapping, the dataset was sampled fully with replacement 1000 times, with the bootstrapped datasets used to calculate means, epitope rank scores and standard deviation values.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding author.

Author contributions

MH: Conceptualization, Data curation, Software, Visualization, Writing – original draft, Writing – review & editing, Methodology. FG: Conceptualization, Data curation, Methodology, Software,

Writing – review & editing. JJ: Data curation, Methodology, Software, Writing – review & editing. CW: Data curation, Methodology, Software, Writing – review & editing. OW: Supervision, Writing – review & editing. MN: Conceptualization, Funding acquisition, Supervision, Writing – original draft. PM: Conceptualization, Funding acquisition, Supervision, Writing – original draft.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was in part funded by National Institute of Allergy and Infectious Diseases (NIAID), under award number 75N93019C00001. MH acknowledges the Sino-Danish Center [2021]. Funding for open access charge: Internal Funding from the University.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2024.1322712/full#supplementary-material

SUPPLEMENTARY FIGURE 1
Validation set performance up to ensemble size 20. Validation set gain in AUC-ROC from ensembling the full-feature model. Performance graphs are shown for training on either experimentally solved, AlphaFold predicted or both structures, and then evaluated on either the solved or predicted structure validation set.

SUPPLEMENTARY FIGURE 2
External test set AUC-ROC. Test set AUC-ROC, as evaluated on 24 antigens modeled with AlphaFold. For AUC-PR see.

SUPPLEMENTARY FIGURE 3
External test set PDB performances. Evaluation on 24 antigens modeled with AlphaFold (left) or experimentally solved structures (right). BepiPred-3.0 performances on antigen sequences only.

SUPPLEMENTARY FIGURE 4
DiscoTope-3.0 is robust towards modeling and relaxation. External test set change in mean epitope rank scores across PDBs, when **(A)** swapping predicted structures with their original solved structure or **(B)** solved structures with the same structure after FoldX relaxation (see Methods). Mean performance loss shown in percent.

SUPPLEMENTARY FIGURE 5
DiscoTope-3.0 score significantly correlates with antibody hit rate. Lysozyme case study on 1a2y_C, showing PDB antibody hit rate (ratio of times an epitope residue is observed across all 12 lysozyme chains) versus calibrated DiscoTope-3.0 scores. Model is trained excluding all lysozyme structures from training (see Methods).

SUPPLEMENTARY FIGURE 6
DiscoTope-3.0 calibrates for antigen length and surface area. Uncalibrated DiscoTope-3.0 surface scores are biased towards the antigen length, which may cause all residues to be assigned as positives/negatives for some antigens, if using a fixed, binary threshold. **(A)** Validations set DiscoTope-3.0 score distributionsbefore normalization and **(B)** after correcting for antigen length and surface scores (see Methods). **(C)** Calibrated score distributions in the validation set, for buried residues, exposed residues (relative surface accessibility > 20%) and exposed epitopes. The top 70th, 50th and 30th percentile scores for exposed epitopes are shown in red dashed lines **(A, B)**, as suggestive thresholds for binary epitope prediction.

SUPPLEMENTARY FIGURE 7
Benchmarking calibrated score thresholds on Lysozyme. Binary epitope prediction performance on the collapsed lyzosyme dataset, for different calibrated score thresholds. Recall of total observed epitopes shown in blue, with precision for any observed epitopes above the threshold. Green line shows the median epitope count per residue for residues above the given threshold (maximum 12). Red lines shown for the previously mentioned top 70th, 50th and 30th exposed epitope percentile scores from the validation set (Figure S6).

SUPPLEMENTARY FIGURE 8
Fitted GAM models for calibrating scores. Length to μ and surface to std fitted GAM models on the validation set, used for calibrating DiscoTope-3.0 scores (see Methods).

References

1. Galanis KA, Nastou KC, Papandreou NC, Petichakis GN, Pigis DG, Iconomidou VA. Linear b-cell epitope prediction for in silico vaccine design: A performance review of methods available via commandline interface. *Int J Mol Sci* (2019) 22. doi: 10.1101/833418

2. Sun P, Guo S, Sun J, Tan L, Lu C, Ma Z. Advances in in-silico b-cell epitope prediction. *Curr Topics Medicinal Chem* (2019) 19(2):105–15. doi: 10.2174/1568026619666181130111827

3. Jespersen MC, Peters B, Nielsen M, Marcatili P. Bepipred-2.0: improving sequence-based b-cell epitope prediction using conformational epitopes. *Nucleic Acids Res* (2017) 45(W1):W24–9. doi: 10.1093/nar/gkx346

4. Klausen MS, Jespersen MC, Nielsen H, Jensen KK, Jurtz VI, Sønderby CK, et al. Netsurfp-2.0: Improved prediction of protein structural features by integrated deep learning. *Proteins: Structure Function Bioinf* (2019) 87(6). doi: 10.1002/prot.25674

5. Clifford JN, Hoie MH, Deleuran S, Peters B, Nielsen M, Marcatili P. Bepipred-3.0: Improved b-cell epitope prediction using protein language models. *Protein Sci* (2022) 31(12):e4497. doi: 10.1002/pro.449

6. Lin Z, Akin H, Rao R, Hie B, Zhu Z, Lu W, et al. Evolutionary-scale prediction of atomic level protein structure with a language model. *bioRxiv* (2022). doi: 10.1101/2022.07.20.500902

7. Zhou C, Chen Z, Zhang L, Yan D, Mao T, Tang K, et al. 05. SEPPA 3.0—enhanced spatial epitope prediction enabling glycoprotein antigens. *Nucleic Acids Res* (2019) 47 (W1):W388–94. doi: 10.1093/nar/gkz413

8. Ponomarenko J, Bui HH, Li W, Fusseder N, Bourne PE, Sette A, et al. Ellipro: a new structure-based tool for the prediction of antibody epitopes. *BMC Bioinf* (2008) 9 (1). doi: 10.1186/1471-2105-9-514

9. Zhao L, Wong L, Lu L, Hoi SC, Li J. B-cell epitope prediction through a graph model. *BMC Bioinf* (2012) 13(S17). doi: 10.1186/1471-2105-13-s17-s20

10. Liang S, Zheng D, Standley DM, Yao B, Zacharias M, Zhang C. Epsvr and epmeta: prediction of antigenic epitopes using support vector regression and multiple server results. *BMC Bioinf* (2010) 11(1). doi: 10.1186/1471-2105-11-381

11. Kringelum JV, Claus Lundegaard OL, Nielsen M. Reliable b cell epitope predictions: Impacts of method development and improved benchmarking. *PloS Comput Biol* (2012) 8(12). doi: 10.1371/journal.pcbi.1002829

12. da Silva BM, Myung Y, Ascher DB, Pires DEV. epitope3d: a machine learning method for conformational b-cell epitope prediction. *Briefings Bioinf* (2021) 23(1). doi: 10.1093/bib/bbab423

13. Shashkova TI, Umerenkov D, Salnikov M, Strashnov PV, Konstantinova AV, Lebed I, et al. Sema: Antigen b-cell conformational epitope prediction using deep transfer learning. *Front Immunol* (2022) 13:960985. doi: 10.3389/fimmu.2022.960985

14. Tubiana J, Schneidman-Duhovny D, Wolfson HJ. Scannet: an interpretable geometric deep learning model for structure-based protein binding site prediction. *Nat Methods* (2022) 19(6). doi: 10.1038/s41592-022-01490-7

15. Dunbar J, Krawczyk K, Leem J, Baker T, Fuchs A, Georges G, et al. Sabdab: The structural antibody database. *Nucleic Acids Res* (2013) 42(D1). doi: 10.1093/nar/gkt1043

16. Ren J, Liu Q, Ellis J, Li J. Positive-unlabeled learning for the prediction of conformational b-cell epitopes. *BMC Bioinf* (2015) 16:(S18). doi: 10.1186/1471-2105-16-s18-s12

17. Li F, Dong S, Leier A, Han M, Guo X, Xu J, et al. 11. Positive-unlabeled learning in bioinformatics and computational biology: a brief review. *Briefings Bioinf* (2021) 23 (1). doi: 10.1093/bib/bbab461

18. Mordelet F, Vert JP. A bagging svm to learn from positive and unlabeled examples. *Pattern Recognition Lett* (2014) 37:201–9. doi: 10.1016/j.patrec.2013.06.010

19. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with alphafold. *Nature* (2021) 596(7873). doi: 10.1038/s41586-021-03819-2

20. Varadi M, Anyango S, Deshpande M, Nair S, Natassia C, Yordanova G, et al. 11. AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res* (2021) 50(D1): D439–44. doi: 10.1093/nar/gkab1061

21. Consortium, T.U. 11. UniProt: the universal protein knowledgebase in 2023. *Nucleic Acids Res* (2022) 51(D1):D523–31. doi: 10.1093/nar/gkac1052

22. Hsu C, Verkuil R, Liu J, Lin Z, Hie B, Sercu T, et al. Learning inverse folding from millions of predicted structures. *bioRxiv* (2022). doi: 10.1101/2022.04.10.487779

23. Vita R, Mahajan S, Overton JA, Dhanda SK, Martini S, Cantrell JR, et al. The immune epitope database (iedb): 2018 update. *Nucleic Acids Res* (2018) 47(D1). doi: 10.1093/nar/gky1006

24. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. 01. The protein data bank. *Nucleic Acids Res* (2000) 28(1):235–42. doi: 10.1093/nar/28.1.235

25. Chen T, Guestrin C. (2016). XGBoost: A scalable tree boosting system. In*Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (New York, NY,USA: ACM), KDD '16. pp. 785–94, ACM.

26. Claesen M, De Smet F, Suykens JA, De Moor B. A robust ensemble approach to learn from positive and unlabeled data using svm base models. *Neurocomputing* (2015) 160:73–84. doi: 10.1016/j.neucom.2014.10.081

27. Zhao Y, Zhang M, Zhang C, Chen W, Ye N, Xu M. A boosting algorithm for positive-unlabeled learning. (2022).

28. Dietterich TG. An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting, and randomization. *Mach Learn* (2000) 40(2):139–57. doi: 10.1023/a:1007607513941

29. Elkan C, Noto K. (2008). Learning classifiers from only positive and unlabelled data, in: *Proceeding of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD 08*, . doi: 10.1145/1401890.1401920

30. Huang F, Xie G, Xiao R. (2009). Research on ensemble learning, in: *2009 International Conference on Artificial Intelligence and Computational Intelligence*, , Vol. 3. pp. 249–52. doi: 10.1109/aici.2009.235

31. Schymkowitz J, Borg J, Stricher F, Nys R, Rousseau F, Serrano L. 07. The FoldX web server: an online force field. *Nucleic Acids Res* (2005) 33(suppl 2):W382–8. doi: 10.1093/nar/gki387

32. Sehnal D, Bittrich S, Deshpande M, Svobodová R, Berka K, Bazgier V, et al. 05. Mol* Viewer: modern web app for 3D visualization and analysis of large biomolecular structures. *Nucleic Acids Res* (2021) 49(W1):W431–7. doi: 10.1093/nar/gkab314

33. Krawczyk K, Liu X, Baker T, Shi J, Deane CM. Improving b-cell epitope prediction and its application to global antibody-antigen docking. *Bioinformatics* (2014) 30(16):2288–94. doi: 10.1093/bioinformatics/btu190

34. Jespersen MC, Mahajan S, Peters B, Nielsen M, Marcatili P. Antibody specific b-cell epitope predictions: Leveraging information from antibody-antigen protein complexes. *Front Immunol* (2019) 10:298. doi: 10.3389/fimmu.2019.00298

35. Meta Research. *Esm github repository*. (2023) Available at: https://github.com/facebookresearch/esm/ commit 2b369911bb5b4b0dda914521b9475cad1656.

36. Kunzmann P, Hamacher K. Biotite: A unifying open source computational biology framework in python. *BMC Bioinf* (2018) 19(1). doi: 10.1186/s12859-018-2367-z

37. Rost B, Sander C. Conservation and prediction of solvent accessibility in protein families. *Proteins: Structure Function Genet* (1994) 20(3). doi: 10.1002/prot.340200303

38. Cock PJ, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, et al. Biopython: Freely available python tools for computational molecular biology and bioinformatics. *Bioinformatics* (2009) 25(11):1422–3. doi: 10.1093/bioinformatics/btp163

39. Servén D, B. C. pygam: Generalized additive models in python. *J Mol Biol* (2018). doi: 10.5281/zenodo.1208723

40. Tubiana J, Schneidman-Duhovny D, Wolfson HJ. Scannet: A web server for structure-based prediction of protein binding sites with geometric deep learning. *J Mol Biol* (2022) 434(19):167758. doi: 10.1016/j.jmb.2022.167758

41. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, et al. Scipy 1.0: Fundamental algorithms for scientific computing in python. *Nat Methods* (2020) 17(3):261–72. doi: 10.1038/s41592-019-0686-2

frontiers | Frontiers in Immunology

TYPE Original Research
PUBLISHED 16 February 2024
DOI 10.3389/fimmu.2024.1285785

OPEN ACCESS

EDITED BY
Joe Hou,
Fred Hutchinson Cancer Center, United States

REVIEWED BY
Juan Carlo Santos Silva,
University of São Paulo, Brazil
Rui-Si Hu,
Guizhou Medical University, China

*CORRESPONDENCE
Sajitha Lulu S.
ssajithalulu@vit.ac.in

RECEIVED 30 August 2023
ACCEPTED 05 January 2024
PUBLISHED 16 February 2024

CITATION
Naidu A and Lulu S. S (2024) Systems and computational analysis of gene expression datasets reveals GRB-2 suppression as an acute immunomodulatory response against enteric infections in endemic settings. *Front. Immunol.* 15:1285785. doi: 10.3389/fimmu.2024.1285785

Systems and computational analysis of gene expression datasets reveals GRB-2 suppression as an acute immunomodulatory response against enteric infections in endemic settings

Akshayata Naidu and Sajitha Lulu S.*

Integrative Multi-omics Lab, Department of Biotechnology, Vellore Institute of Technology, Vellore, Tamil Nadu, India

Introduction: Enteric infections are a major cause of under-5 (age) mortality in low/middle-income countries. Although vaccines against these infections have already been licensed, unwavering efforts are required to boost suboptimalefficacy and effectiveness in regions that are highly endemic to enteric pathogens. The role of baseline immunological profiles in influencing vaccine-induced immune responses is increasingly becoming clearer for several vaccines. Hence, for the development of advanced and region-specific enteric vaccines, insights into differences in immune responses to perturbations in endemic and non-endemic settings become crucial.

Materials and methods: For this reason, we employed a two-tiered system and computational pipeline (i) to study the variations in differentially expressed genes (DEGs) associated with immune responses to enteric infections in endemic and non-endemic study groups, and (ii) to derive features (genes) of importance that keenly distinguish between these two groups using unsupervised machine learning algorithms on an aggregated gene expression dataset. The derived genes were further curated using topological analysis of the constructed STRING networks. The findings from these two tiers are validated using multilayer perceptron classifier and were further explored using correlation and regression analysis for the retrieval of associated gene regulatory modules.

Results: Our analysis reveals aggressive suppression of GRB-2, an adaptor molecule integral for TCR signaling, as a primary immunomodulatory response against *S. typhi* infection in endemic settings. Moreover, using retrieved correlation modules and multivariant regression models, we found a positive association between regulators of activated T cells and mediators of Hedgehog signaling in the endemic population, which indicates the initiation of an effector (involving differentiation and homing) rather than an inductive response upon infection. On further exploration, we found STAT3 to be instrumental in designating T-cell functions upon early responses to enteric infections in endemic settings.

Conclusion: Overall, through a systems and computational biology approach, we characterized distinct molecular players involved in immune responses to enteric infections in endemic settings in the process, contributing to the mounting evidence of endemicity being a major determiner of pathogen/vaccine-induced immune responses. The gained insights will have important implications in the design and development of region/endemicity-specific vaccines.

KEYWORDS

immune response, enteric infection, gene expression data analysis, network biology, machine learning methods, gene regulatory networks

1 Introduction

Enteric infections pose major challenges to global health as diarrheal diseases remain one of the major causes of under-5 (years) mortality in Sub-Saharan Africa and South Asia (1–3). In areas of high endemicity, the suboptimal vaccine efficacy/effectiveness of oral vaccines against enteric pathogens has been quite puzzling and concerning (4–6). Several second- and third-generation enteric vaccines are under development and evaluation and can greatly benefit from the establishment of reliable correlates of protection (CoP) and/or correlates of risk (CoR) (7, 8) during the phase of clinical testing. Since the advent of high-throughput technologies, many studies have aimed at establishing gene/molecular-level signatures to induced protective immune responses against multiple vaccines (9–11) and infections instead of solely relying on antibody titers as a protective biomarker. In the course of advancements in the field quite recently, the focus has shifted towards developing and assigning gene modules (functionally associated group of genes) to vaccine-induced immunological protection against several infections (12, 13).

Particularly for enteric infections, given that endemicity plays an important role in defining vaccine-induced immune responses (14), understanding the molecular mechanisms that are underplay in endemic settings after perturbation becomes absolutely essential (15). Hence, the objective of the study was to delineate these molecular mechanisms to distinguish between immune responses in endemic and non-endemic settings (against enteric pathogens). For this purpose, we employed a robust computational and network biology pipeline for the analysis of post-infection gene expression datasets (of the host) singularly and comprehensively. Through the analysis, we expect to exhibit meaningful insight and credible molecular signatures/regulatory modules that can distinguish immune responses in these two different settings with varied pathogen prevalence. In the process, we also put forward the used pipeline as an exploratory tool for future studies that involve meta-analysis of gene expression datasets and that particularly focus on studying immune responses to pathogens.

2 Materials and methods

2.1 Data collection and conceptual framework

Microarray and RNASeq datasets linked to host responses to prevalent enteric pathogens—*S. typhi*, ETEC, *Vibrio cholera*, and rotavirus infections—were collected from NCBI (GEO) and EMBL-EBI (ArrayExpress) databases using the following keywords: ["Salmonella" AND "Homo Sapiens"], ["Typhoid" AND "Homo Sapiens"], ["E. coli" AND "Homo Sapiens"], and ["Rotavirus" AND "Homo Sapiens"]. A total of 125 gene expression studies were retrieved. These studies were further filtered by excluding *in vitro* studies and only clinical studies were included with infected/challenged and control groups. Supplementary Figure S1 illustrates the detailed exclusion and inclusion criterion used for data screening and identification for the study for both endemic and non-endemic settings. The obtained gene expression datasets were segregated based on the study location and were labeled as "endemic" or "non-endemic" based on the pathogen prevalence as described in the literature. The two-tiered computational pipeline followed for the study is illustrated in Figure 1.

2.2 Data integration

For meta-dataset construction, gene expression datasets corresponding to acute stages of infection were derived from each of the studies and were integrated, and batch effect was corrected using the "sva" package's ComBat function in R (16).

2.3 Differential expression analysis

Differentially expressed genes (DEGs) for each of the dataset were obtained using the "GEOquery" (17) and "limma" package (18). Briefly, gene expression datasets were retrieved for each of the studies using the "fData" function, and rows with missing values

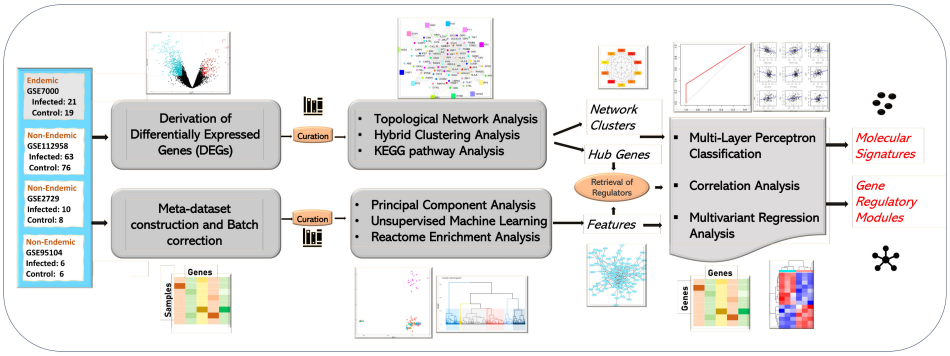


FIGURE 1
Study workflow of the analysis. The study was performed in two tiers. The first tier focused on the retrieval and topological network analysis of differentially expressed genes (at the acute stage) while the second tier focused on integration of all the infected samples in the form of a meta-dataset. The meta-dataset was used for feature selection using PCA and Random Forest Algorithm. The features/hub genes derived from the two tiers were further analyzed using correlation and multivariant regression analysis, and molecular signatures designating immune responses in endemic and non-endemic settings was derived using multilayer perceptron-based classification.

were omitted. Samples corresponding to acute responses to infections and controls were only considered for further analysis (Supplementary Figure 1). The four datasets were normalized using log2 transformation prior to the calculation of DEGs, which were corrected for false positives using the Benjamini & Hochberg method. The retrieved DEGs for the four tables were further filtered using logFC value (>1 and <−1) and *p*-values (0.05) and were visualized using volcano plots developed using the "ggplot2" package (19), and common and distinct DEGs were visualized using the "Venn diagram". Missing gene symbols from these datasets were obtained using the "biomaRt" package for further analysis (20). Supplementary File 1 provides the list of DEGs obtained for each of the cohorts in tabular format.

2.4 Functional enrichment analysis

The Gene Ontology database (Gene Ontology Resource) was used to prepare a master list of "biological processes" that are involved in immune responses against pathogens (Supplementary File 2) using the QuickGO interface (https://www.ebi.ac.uk/QuickGO/). A total of 248 biological processes were identified and used as a reference list. DEGs derived from the four datasets were individually fed to the DAVID database (https://david.ncifcrf.gov/) to derive enriched biological processes. The acquired lists (4) were manually curated to select "only" immune response-associated gene ontology terms using the drafted master list and were taken further for the analysis. Pathway enrichment analysis for all the four sets of DEGs was performed using the KEGG [KEGG PATHWAY Database (genome.jp)] (release 106.0) and Reactome (Home - Reactome Pathway Database) database (V86). Individual gene functions and associated pathways were derived from the GeneCards database (GeneCards - Human Genes | Gene Database | Gene Search).

2.5 Network analysis

Protein–protein interaction (PPI) networks were constructed using the STRING database [STRING: functional protein association networks (string-db.org)] and visualized and analyzed using Cytoscape (Cytoscape: An Open Source Platform for Complex Network Analysis and Visualization) plugins. The nodes of the network represent proteins and the edges represent the functional or physical associations the nodes have with each other as determined through text mining or experimental evidence and are represented and curated based on confidence scores. PPI networks were extended for up to 30 interacting partners per node (with 90% confidence score) to get a comprehensive functional understanding of the DEGs.

2.5.1 Topological network analysis

Hub nodes/genes in a network can be defined as the most influential nodes in terms of connectivity and influence and were calculated using the cytohubba plugin (21). For the four constructed network, hub genes were identified using three different algorithms. While the Maximum Clique Centrality (MCC) and Density of Maximum Neighborhood Compartment (DMNC) algorithms revealed nodes with maximum connectivity that were relevant in understanding influential proteins for each of the networks, the Bottleneck algorithm was especially important in extracting nodes that connected different subnetworks. The employed algorithms are detailed as follows:

- MCC is a local-based method for topological analysis where the MCC score for a node or $MCC(v)$ is defined as $MCC(v) = \sum C \in S(v)(|C|-1)!$, where $S(v)$ is the collection of maximal cliques that contain v, and $(|C|-1)!$ is the product of all positive integers less than $|C|$.
- DMNC is also a local-based method for topological analysis where the DMNC score or $DMNC(v)$ of a particular node is defined as $DMNC(v) = |E(MC(v))|/|V(MC(v))|^{\varepsilon}$, where $\varepsilon = 1.7$, $MC(v)$ is a maximum connected component of the $G[N(v)]$, and $G[N(v)]$ is the induced subgraph of G by $N(v)$ (total set of nodes). V is a collection of nodes and E is a collection of edges.
- The Bottleneck algorithm, on the other hand, is a global-based method for topological analysis where the Bottleneck

score $BN(v)$ is defined as $BN(v)=\sum s\in Vps(v)$, where $ps\ (v) = 1$ if more than $|V(Ts)|/4$ paths from node s to other nodes in Ts meet at the vertex v; otherwise, $ps(v) = 0$.

The PPI network clusters were detected using the MCODE algorithm available in the ClusterViz plugin in Cytoscape (22). The algorithm maps highly interconnected subnetworks of a network. In this algorithm, seed vertices are expanded based on the local neighborhood density and the density of the prospective cluster.

2.6 Feature selection through unsupervised machine learning algorithm

Firstly, principal component analysis (PCA) was performed on the constructed meta-dataset (section 2.2) to characterize the variance of gene expression profile in endemic and non-endemic settings. PCA is a dimension reduction technique used to derive key insights into big datasets based on the covariance of the variables involved based on the derived eigenvectors and values. Mathematically, covariance between two variables is defined as:

$$\mathrm{Cor}(x,y) \ = \ \mathrm{Sum}\ ((x_i - x*)\ (y_i - y*))/N$$

where x and y represent two variables, x^* and y^* represent their respective means, and N represents the total sample size of the study. PCA is generally used as a preliminary step to observe the underlining patterns of the large datasets and how these patterns are correlated with the phenotype/outcomes under consideration. The analysis was performed using the "prcomp" function in R.

Secondly, feature selection was performed using the Random Forest algorithm-based wrapper method that distinguished between gene expression profiles (with common 6,543 genes) from endemic and non-endemic settings using the "Boruta" package (23). Random Forest belongs to the family of decision trees where, based on numerical estimates, independent decision trees are constructed and evaluated for optimal classification performance. The importance of a variable is calculated based on the loss in accuracy in classification when the variable is dropped in a series of random permutations. The importance of each variable is determined using the Z score in the Boruta package. Mathematically, the Z score in the Boruta package can be defined as the average of the difference in real and predicted values of a variable (or the loss of accuracy) divided by the standard deviation. The higher the loss of accuracy computed for a variable, the poorer it seemed to have performed, and *vice versa*. The parameters used in the algorithms are optimized based on trial and error and are hence auto-optimized or auto-tuned.

Thirdly, hybrid clustering (using components of both k-means and hierarchical clustering algorithms) was performed on the logFC values of common genes between the four cohorts using the "FactoMineR" package (24). In hybrid clustering, small clusters are initially formed using the k-means algorithm (centroid-based clustering), which are later clustered on a larger scale based on the maximal distance between the formed clusters and come under hierarchical or connectivity-based clustering. Mathematically, k-means clustering relies on the calculation of Euclidean distance between two variables in order to assign variables to specific centroids. The Euclidean distance between two variables is computed as:

$$d^2(x,y) \ = \ (x_1 - y_1)^2 + (x_2 - y_2)^2 \ + \ (x_n - y_n)^2$$

where x and y represent the two variables (their values) in a plane and n represents the number of samples. On the other hand, maximal distance between two clusters in hierarchical clustering is computed as:

$$d(p,q) \ = \ T_{pq}/N_p + N_q$$

where p and q represent the two clusters, T represents the sum of the pairwise distances between the two clusters, and N represents the number of variables in the respective clusters.

The features/attributes/genes derived from the two algorithms (clustering and Random Forest) were used for the construction of the PPI network using the STRING database, and hub genes were retrieved through topological network analysis performed using the cytohubba plugin (Figure 1).

2.7 Machine learning based classification

Hub genes derived through the methods described in sections 2.5 and 2.6 specifically were used for the construction of classification models using the meta-dataset to distinguish between the endemic and non-endemic (infected) groups using the multilayer perceptron (MLP) algorithm on the WEKA platform with threefold cross-validation. Neural networks, specifically MLP, are well documented in the literature as good classifiers when gene expression datasets are used as input (25, 26). MLP is a deep machine learning algorithm that consists of an input layer, an output layer, and a hidden layer, and the neural network is trained using a feed-forward pathway. The activation function used for training was a sigmoid logistic function represented as:

$$\mathbf{F(x)} \ = \ 1/(1 \ + \ e^{-x})$$

which is a nonlinear function and represents an input variable in the range of 0 to 1. Activation functions are used to gauge and legitimize specific neurons or nodes of the neural network during training based on the weight and bias they hold for the classification. Thereafter, confusion matrices representing the performance of the classification were computed and visualized. The confusion matrix summarizes true positive (TP), false positive (FP), false negative (FN), and true negative (TN) values predicted by the model. The confusion matrix is used to compute Accuracy and Recall of the built classifier, where

$$\mathrm{Accuracy} \ = \ \mathrm{TP} \ + \ \mathrm{TN}/(\mathrm{TP} \ + \ \mathrm{TN} \ + \ \mathrm{FP} \ + \ \mathrm{FN})$$

$$\mathrm{Recall} \ = \ \mathrm{TP}/\mathrm{TP} \ + \ \mathrm{FN}$$

Accuracy represents the instances (out of total) where the classification predictions were correct, while Recall represents instances where the predictions were correct as compared to total positives (TP + FN). Genes were ranked based on the accuracy score of their respective models.

2.8 Correlation analysis

Correlation modules were retrieved using the "azolling/EBmodules" package (https://github.com/azolling/EBmodules) from the constructed meta-dataset, and modules with high-performing genes from the section above were identified (27). The algorithm behind the package combines gene–gene correlation matrices derived from different sets of microarray datasets with the sample-gene architecture using the Fischer transformation. From this constructed common correlation matrix, highly correlated genes or modules are derived using hierarchical clustering algorithm. The optimal number of modules to be derived from the correlation matrix is decided using the Gap statistical method that is discussed in detail elsewhere (https://joey711.github.io/phyloseq/gap-statistic.html), and for each cluster, Gap(k) is computed using:

$$\mathrm{Gap}(k) = (1/B)\,\mathrm{sum}(\log(W*) - log\,(Wk)$$

2.9 Multiple regression analysis

Genes correlated to high-performing genes (based on MLP classification) (or part of shared network clusters from section 2.5) and retrieved transcriptional factors for each of these genes were used for the construction of multivariant regression (MVR) models in R. MVR involves the prediction of a dependent variable based on a set of independent variables (instead of a single variable that is used in the single-variant regression analysis). Mathematically, regression models can be defined as:

$$Y = \beta 0 + \beta 1 x_i + \epsilon_i$$

where Y represents the dependent variable under investigation and x represents independent variables, while $\beta 0$ and $\beta 1$ represent the intercept and parameter of the model, respectively, and ϵ represents standard error. "i" indicates the number of independent variables being tested for the prediction of Y. For the highly influential genes derived from the steps above, MVR models were retrieved using a combinatorial approach where genes found to be correlated or associated with these genes of interest (throughout the analysis) were treated as independent variables to derive the best-performing model that could predict the pattern of expression of these influential genes. The aim of the analysis was to gain a deeper understanding of the underlying molecular mechanisms for the construction of robust gene regulatory modules associated with identified molecular signatures. MVR has been recently suggested as a robust method for deriving gene regulatory networks from gene expression datasets (28). The analysis was performed using the "lm" function in R.

2.10 Regulatory network inference

MVR models constructed in the above step with R^2 value > 0.50 were used for the inference of gene regulatory modules.

3 Results

Based on the criteria discussed in section 2.1, four gene expression studies—GSE7000, GSE112959, GSE2729, and GSE95104—were selected for the analysis. Here, GSE7000 study datasets were retrieved from subjects in Vietnam (a country endemic to *S. typhi* infection), whereas the latter three were from non-endemic settings. GSE112958 study datasets were derived from *S. typhi*-challenged adults in a controlled study conducted in Oxford (UK). GSE2729 datasets were retrieved from rotavirus-infected children from the USA and GSE95104 datasets were derived from ETEC-infected subjects from the USA (Table 1). Datasets from the earliest time points (post-symptom onset) for each of the four studies were used for the retrieval of DEGs and for the construction of the meta-dataset (Supplementary Figure 1). An integrated dataset (meta-dataset) with 6,543 common genes was constructed, and the batch effect was corrected for a total of 208 samples (all infected samples from the four datasets) (Supplementary Files 5 and 6) for meta-analysis of gene expression datasets. An online accessible processed dataset with 20 samples from GSE69529 (RNASeq) was reserved for validation (Supplementary File 7 and Supplementary Figure S1).

3.1 Retrieved differentially expressed genes, enriched pathways, and modules

At the early stage of infection, in the *S. typhi* cohort, there were 887 upregulated genes while there were 1,249 downregulated genes. For the *S. typhi* (Oxford) cohort, there were 258 upregulated genes and 34 downregulated genes. For the Rotavirus cohort, there were 139 upregulated genes and 207 downregulated genes. For the ETEC cohort, there were 80 upregulated genes and no genes were downregulated based on the set criterion (Supplementary Table S1). The retrieved DEGs from the four cohorts were illustrated as Volcano plots (Figure 2A). Briefly, for the *S. typhi* (Vietnam) cohort, there was upregulation of markers of activated lymphocytes and mediators of the NOTCH signaling pathways, and downregulation of mediators involved in acute inflammatory responses. For the *S. typhi* (Oxford) cohort, highly upregulated genes were inferred to an interferon-mediated inflammatory response along with the mediation of T-cell chemotaxis. For the Rotavirus cohort, we found upregulation of inflammatory cytokines, and for the ETEC cohort, we found upregulation of mediators involved in early stages of inflammation.

10.3389/fimmu.2024.1285785

TABLE 1 GEO Accession ID with description of the four microarray datasets used in the study along with a rnaseq dataset used for validation.

GEO Accession ID	Microarray platforms	Pathogen	No. of samples	Study population	Location	Reference
GSE2729	Affymetrix Human Genome U95 Version 2 Array	Rotavirus	23	Children, infected	USA	(29)
GSE95104	Affymetrix Human Genome U133A 2.0 Array	ETEC	72	Adults, challenged with unattenuated ETEC strain	USA	(30)
GSE7000 **(GLP4858)**	Stanford Human cDNA Microarray	*S. typhi*	183	Adults, INFECTED	Vietnam	(31)
GSE112958	Illumina HumanHT-12 V4.0 expression bead chip	*S. typhi*	178	Adults, challenged with *S. typhi* Quailes strain	UK	(*Diagnostic Host Gene Signature for Distinguishing Enteric Fever from Other Febrile Diseases—EMBO Molecular Medicine*, 2019)
GSE69529	Illumina HiSeq 2500	Multiple	204	Children, infected with multiple pathogens	Mexico	(32)

*RNA was extracted from PBMC samples in the first three studies and from the whole blood samples in the fourth study.

In terms of numbers, we found the least number of DEGs in the ETEC cohort and the highest number of DEGs in the *S. typhi* (Vietnam cohort). While the *S. typhi* (Vietnam) cohort had 59 DEGs in common with the Rotavirus cohort, there were only 34 DEGs common with the *S. typhi* (Oxford) cohort (Figure 2B).

Functional enrichment analysis was performed to gain biological insight into acute responses to pathogen in endemic and non-endemic cohorts. Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis on the DEGs acquired for the *S. typhi* (Vietnam) cohort revealed significant enrichment of multiple intracellular signaling pathways, top among which were the cGMP-PKG signaling pathway and the Calcium signaling pathway (Supplementary Table S2). Interestingly, pathway enriched analysis of "both" up-and downregulated genes separately for this cohort revealed enrichment of T-cell receptor signaling (at the acute state of infection). While CD40L, PI3K, SOS, HRAS, and PLC genes were upregulated, LCK and GRB2 were downregulated (Supplementary Figure S2) along with the downregulation of major signaling pathways conventionally associated with acute inflammatory responses (toll-like receptor signaling and cytokine/chemokine signaling pathway) (Supplementary Figure S3). For the *S. typhi* (Oxford) cohort, sensory signaling pathways—NOD-like receptor signaling pathways and the Cytosolic DNA-sensing signaling pathway—along with intracellular pathways involved in antigen processing and presentation were significantly enriched. On the other hand, in the Rotavirus cohort, enrichment of major inflammatory signaling pathways was observed upon KEGG pathway enrichment analysis. Importantly, pathways associated with PRR signaling and TCR/BCR signaling were also significantly enriched for this cohort. For the ETEC cohort, given the low number of DEGs derived for this cohort, no enriched KEGG signaling pathways were detected (Supplementary Table S2).

Enrichment and curation of GO biological processes based on the master list (section 2.4) yielded a total of 91 immune response-associated modules for the *S. typhi* (Vietnam) cohort, 117 modules for the *S. typhi* (Oxford) cohort, 118 modules for the Rotavirus cohort, and 6 modules for the ETEC cohort. The top curated enriched terms for the *S. typhi* (Vietnam) cohort were "inflammatory response", "positive regulation of cell migration", "cell surface receptor signaling pathways", "response to xenobiotic stimulus", and "neutrophil chemotaxis". Curated terms for *S. typhi* (Oxford) were "defense response to virus", "innate immune response", "response to virus", "negative regulation of viral genome replication", and "positive regulation of interferon beta production". For the Rotavirus cohort, the top enriched biological processes (after curation) were "chemokine-mediated signaling pathway", "cellular response to lipopolysaccharide", "negative regulation of MAPK cascade", "cytokine mediated signaling pathway", and "negative regulation of type 2 immune response". For the ETEC cohort, the top enriched (curated) terms were "regulation of phosphatidylinositol 3-kinase signaling", "positive regulation of innate immune response", "immune response", "acute-phase response", "regulation of immune system process", and "T-cell activation". Genes associated with curated GO terms were taken ahead for PPI network construction and analysis (Figure 3).

Overall, through the KEGG enrichment analysis, we found peculiar dysregulation of the TCR receptor signaling pathway in the endemic cohort as compared to the non-endemic cohort (Supplementary Figures S2 and S3). Furthermore, although all the four cohorts showed enrichment of biological processes involved in host responses to the pathogen and acute inflammatory responses, we observed specific enrichment of modules associated with cell migration in the endemic cohort.

3.2 Hub genes and network clusters

The list of genes derived for each of the cohorts after module screening and identification (Supplementary File 3) was used as input for the construction of PPI networks (as described in section 3.1) to retrieve genes of high influence or connectivity (hub genes)

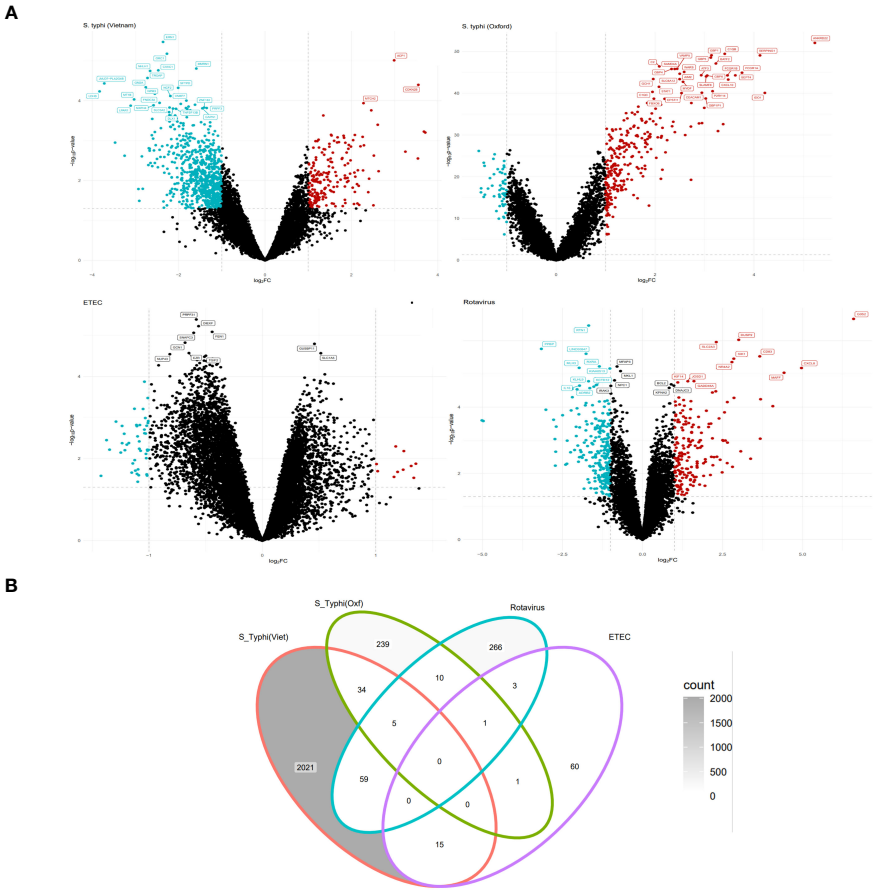


FIGURE 2
EXTRACTION OF DIFFERENTIALLY expressed genes (DEGs) (Tier 1). **(A)** Volcano plots depicting upregulated and downregulated genes derived from GSE7000, GSE112958, GSE95104, and GSE2729 (clockwise). **(B)** Venn diagram representing common and specific genes between the cohorts.

in immunologically relevant gene ontologies (for the four cohorts). Although PPI networks were constructed using a curated set of genes with high immunological relevance, for the *S. typhi* (Vietnam) cohort, topological analysis of the network did not derive any hub genes conventionally associated with immune responses. In fact, majority of the hub genes derived from the three topological algorithms were associated with cell cycle signaling (SOS1, HRAS, and KRAS), EGFR receptor-associated (EGFR and SRC), and MAPK/Erk (MAPK6/14) signaling pathways (Supplementary Table S3 and Figure 3A). For immune responses in the *S. typhi* (Oxford) cohort, hub genes using the MCC and DMNC algorithm were IRF1, IFIT1/3/4, and IFI35, and IRF1/4, IFIT5, and IFITM1/3, respectively. Both of these sets of genes are essential components of interferon-mediated signaling pathways (Supplementary Table S3 and Figure 3B). For the Rotavirus cohort, major inflammatory mediators—RELA, JUN, STAT3, CREBBP, IL6R, CXCL3/8, TNF, and STAT1—were revealed as hub genes of the constructed network (Supplementary Table S3 and Figure 3C). In the ETEC cohort, degree-based topological algorithms (MCC and DMNC) revealed adaptors and receptors involved in TCR (CD28, CD2, CD28, and CD247) and BCR (CD79A/B) signaling pathways as essential hub genes in the elicited immune response (Supplementary Table S3 and Figure 3D).

Network clusters derived from the four pathogen-specific PPI networks were filtered based on their clustering scores (>5 score); three clusters were retrieved from the *S. typhi* (Vietnam) cohort and one cluster (with a score of 40.55) was retrieved from the *S. typhi* (Oxford) cohort. From the Rotavirus cohort, three clusters were retrieved and two clusters were retrieved from the ETEC cohort. Fully annotated clusters are illustrated and described in Supplementary Figure S2 and Supplementary Table S4, respectively. Briefly, the highest-performing network cluster from the Vietnam cohort was enriched with genes belonging to the growth receptor signaling pathway (EGF, EGFR, MAPK, RHOA, KRAS, HRAS, GRB2, SHC, and PTPN11) and T-cell receptor signaling pathway (GRB2, LCK, SRC, MAPK, and HRAS). The highest-performing cluster in the *S. typhi* (Oxford) cohort was enriched with genes belonging to interferon-induced mediators, that in the Rotavirus cohort was enriched with cytokines and chemokines, and that with the ETEC cohort was enriched in surface mediators of lymphocyte signaling. Considering that the functional enrichment analysis pointed towards a dysregulated TCR signaling specifically in the *S. typhi* cohort, the highest performing cluster from the *S. typhi* (Vietnam) cohort (which was enriched with genes from tcr and growth factor receptor signalling) was considered as a

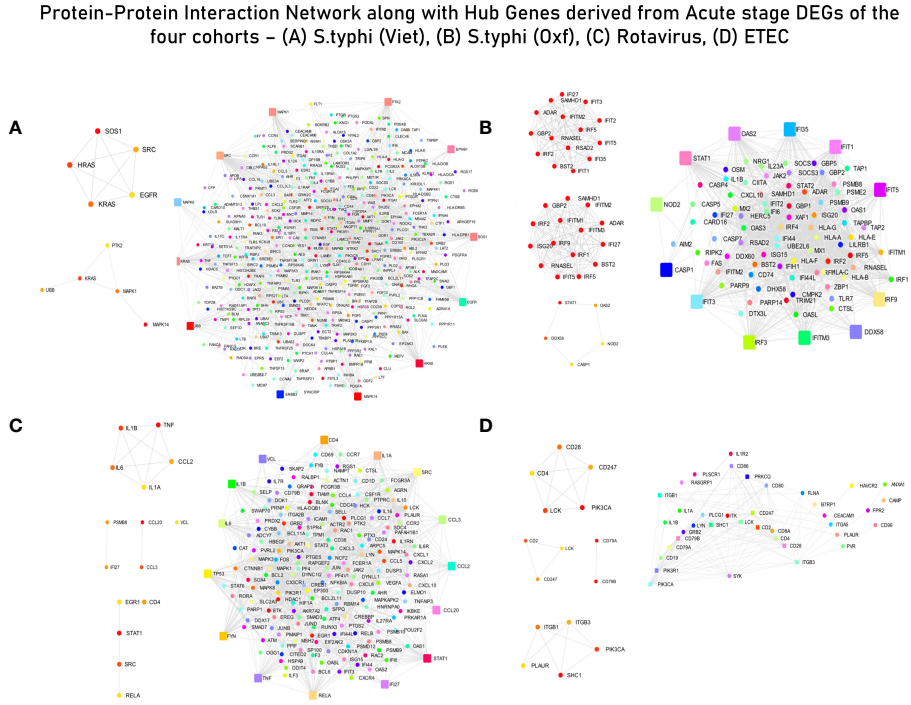


FIGURE 3
Protein–protein interaction (PPI) network with interacting partners (IPs) (Tier 1). **(A)** *S. typhi* (Vietnam) cohort—MCC hub genes and Bottleneck hub genes. **(B)** *S. typhi* (Oxford) cohort—MCC hub genes, DMNC hub genes, and Bottleneck hub genes. **(C)** Rotavirus cohort—MCC hub genes, DMNC hub genes, and Bottleneck hub genes. **(D)** ETEC cohort—MCC hub genes, DMNC hub genes, and Bottleneck hub genes (confidence score: 0.90).

distinguishing and peculiar feature of acute immune responses in the endemic cohort.

3.3 Features distinguishing immune responses in endemic and non-endemic settings

PCA of the integrated gene expression dataset (meta-dataset) revealed a high degree of variance in PC1 and PC2 and was performed to gauge covariances/eigenvectors corresponding to the four cohorts. While variance in component 1 was attributable to the differences in the gene expression profile between an adult cohort and a child cohort, variance in component 2 can be attributed to gene expression profiles triggered upon pathogen exposure in endemic versus non-endemic settings (Figure 4). Further analyses (by the employment of unsupervised ML algorithms) were performed to delineate gene expression profiles based on endemicity. For hybrid clustering (check method section) based on logFC values, the optimal number of clusters was pinned down to be six (based on the calculations of the "total with sum of square" values) (Supplementary Figure S5). Among the derived six clusters, cluster 2 was negatively associated with the *S. typhi* (Vietnam) cohort while cluster 4 was positively associated with this cohort. Figure 5A illustrates distinct gene expression patterns as observed in clusters 2 and 4, which distinguishes the *S. typhi* (Vietnam) cohort from the other three cohorts. Network construction and topological analysis of cluster 4 revealed ribosomal proteins (RPL22, RPS9, and RPS15) and genes associated with Hedgehog (JAG1, WNT2B, and ADAM17) signaling to be high-ranking hub genes as per the MCC and DMNC algorithm (Figure 5B). For cluster 2 (downregulated in the endemic cohort), the derived hub genes were mainly involved with growth factor receptor signaling (PTPN1, PTPN11, ERBB2, GRB2, FGF12, and PDGFRA), cell cycle signaling (WT1), and regulation of interferon signaling (SOCS1 and SOCS3). The findings of clustering analysis indicated upregulation of the Hedgehog signaling pathway and downregulation of growth factor receptor signaling to be specific attributes of the endemic cohort that distinguishes it from the other cohorts.

For deriving more reliable features, Random Forest-based feature selection was used on the meta-dataset to derive highly influential determiners (features/genes) in characterizing host responses to enteric pathogens in endemic and non-endemic settings. Network construction and analysis of the derived features revealed hub genes associated with growth factor receptor and PI3K/Akt signaling (ERBB2, ERBB3, FGFR2, PIK3CB, PIK3R1, PIK3CD, and PTPN11) and genes associated with the cell cycle (CCND1 and RET) (Figure 6). All the three groups of features were characterized via functional enrichment analysis using the reactome database, and their key regulators were then retrieved from the TRRUST database (Table 2). Interestingly, the Random Forest-based feature selection again pointed out towards growth factor receptor signaling as an integral distinguishing feature of the endemic cohort compared to the non-endemic cohort, further validating the findings of the clustering analysis.

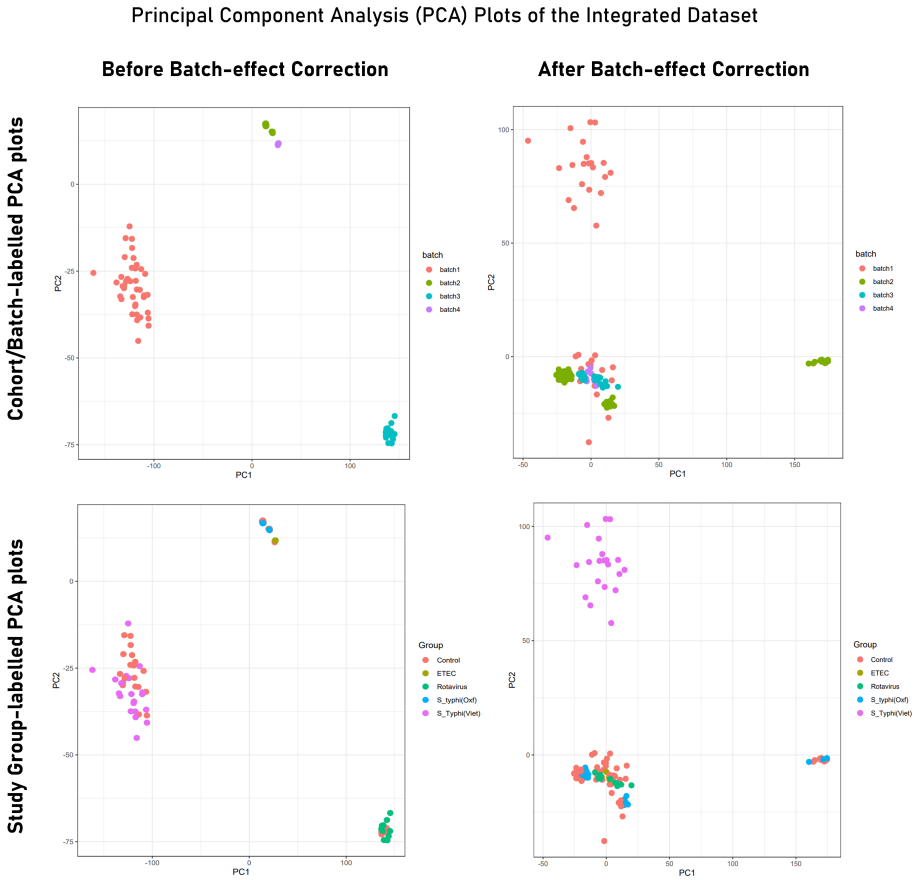


FIGURE 4
PCA plot illustrating variance in gene expression profiles (Tier 2) before and after batch correction. While PCA plots in the upper panel are labeled to indicate samples from different experiments/cohorts/batches, PCA plots from the lower panel are labeled with different study groups (infected and control). Here, Batch 1 = *S. typhi* (Viet) cohort (endemic); batch 2 = Rotavirus cohort; batch 3 = *S. typhi* (Oxf) cohort; batch 4 = ETEC cohort.

Based on the findings of the two unsupervised machine learning algorithms, the negative regulation of components of the growth factor receptor signaling pathways and the positive regulation of the Hedgehog/WNT signaling pathway were determined to be associated with immune responses in endemic settings. To investigate further if these mediators can act as primary determiners of differences in immune responses between endemic and non-endemic settings, we used neural network-based classification (MLP classifier).

3.4 Identification of highly influential genes using ML-based classification

Machine learning-based classification was performed on hub genes derived in sections 3.2 and 3.3, which were categorized as being "responsive" or "housekeeping" genes using the HRT Atlas (https://housekeeping.unicamp.br/) (Table 3). The "responsive" genes were then evaluated for their potential to act as a classifier of immune responses for the endemic cohort compared to the non-endemic cohort using multiple supervised machine learning algorithms. Neural network-based classification algorithms were used for the analysis because of their documented compatibility to accommodate, analyze, and evaluate gene expression data (26).

The performance of the classifiers was evaluated after the derivation of confusion matrices (based on the performed threefold classification). Based on accuracy and ROC, the genes were ranked based on their significance in differentiating immune responses in endemic and non-endemic settings grb2, an adaptor of tcr signalling was found to have the best performing score in classifying infected cohort from endemic and non-endemic setting (Table 4).

3.4.1 Validation of GRB2 as a classifier

To validate GRB2 as a high-performing classifier, two other machine learning algorithms were built to construct the classification model, where, again, GRB2 was classified with high accuracy (Supplementary Figure S7). To validate GRB2 suppression at the acute stage upon vaccination, the ImmuneSpace database was screened for trials that have reported GRB2 downregulation in the first 7 days after immunization. The findings of the survey are tabulated in Supplementary Table S7 where we found four clinical trials with indications of GRB2 suppresion at the acute stage post immunisation.

3.5 Correlation between TCR and Hedgehog/NOTCH signaling pathways

Based on the hypothesis generated in sections 3.2, 3.3 and 3.4, to derive the relationship between the two signaling pathways (TCR

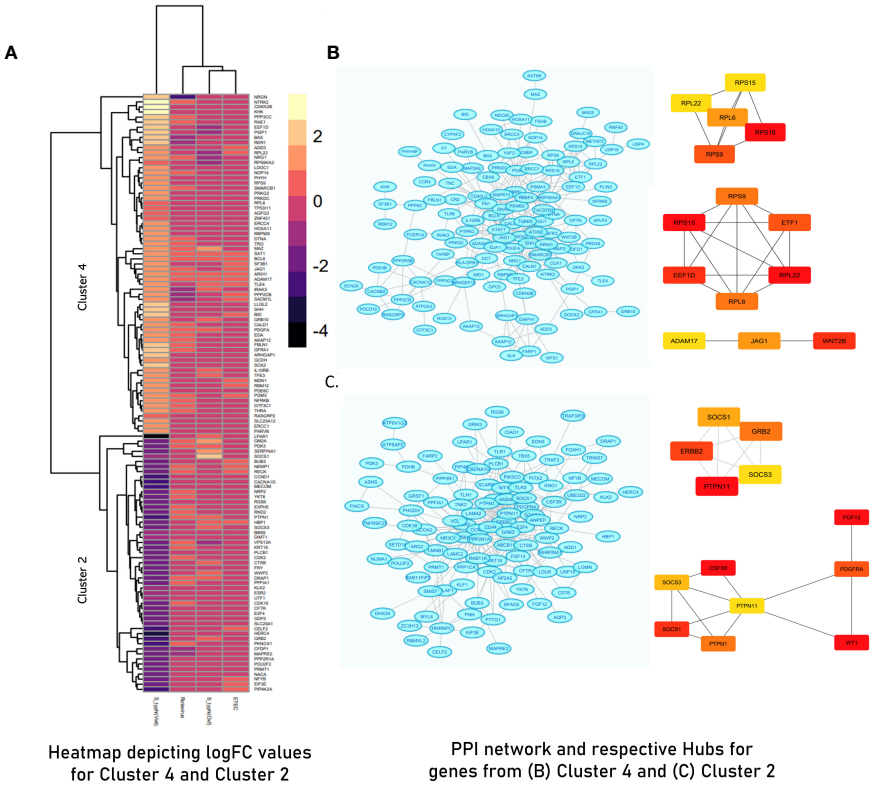


FIGURE 5
(A) Heatmap illustrating Cluster 2 and Cluster 4 derived from hybrid clustering (Tier 1) where yellow depicts logFC >2 and violet depicts logFC< −4. **(B)** STRING network and derived hub genes for Cluster 4. **(C)** STRING network and derived hub genes for Cluster 2 where light blue color nodes depict members of Cluster 4 and Cluster 2, respectively. For both clusters, hub genes were identified using MCC (up) and DMNC (down) algorithms where red-orange-yellow-colored nodes depict hub genes with high scores as calculated by respective algorithms with red- colored nodes depicting the highest scoring genes.

and Hedgehog), correlation studies were performed. A total of 20 correlation modules (group of genes) were identified in the integrated datasets. These modules were characterized using functional enrichment analysis and were filtered using the master list (Supplementary File 2) to derive immunologically relevant submodules (Supplementary Table S7). We found the curated submodule retrieved from module 3 to contain components of both TCR signaling (NFATC4 and NFATC1) and Hedgehog signaling (WNT2B, TLE4, MAFF, and ROR2) and to be highly correlated. NFATC1/4 are transcription factors associated with activated T cells, and their positive correlation with the components of the Hedgehog signaling pathway indicates activation of the latter in activated T cells. We also found CCL17, a known chemotactic agent of T cells, to be correlated with NFATC1/4 transcription factors (Figure 7).

3.6 Multivariant regression models to determine predictors of highly influential genes

For MVR analysis, housekeeping genes identified as highly influential genes in sections 3.2, 3.3 and 3.5 were taken as predictor variables and genes associated with effector functions (or are "responsive" to external stimuli) were taken ahead for the analysis as the response variables—GRB2, LCK, GLI (TF for WNT2B receptor) (Table 3). Potential predictor variables for these four genes were also retrieved from correlation modules in section 3.5. The MVR model for GRB2 yielded a high R^2 value of 0.7616 and its components/predictors were retrieved from network cluster 1 (Supplementary Figure S2). While other predictors showed a positive association with the target gene GRB2, LCK, MYB (TF of LCK), and HRAS showed a strong negative relation and were upregulated in the endemic cohort while the GRB2 was downregulated. The multiple regression model against GLI2 (a transcription factor for WNT2B) involving TLE4, BCL10, FOS, NRAS, PIK3R1, LCK, TNFRSF11A, ROR2, and CCL17 yielded an R^2 value of 0.708, and these predictors were retrieved from correlation module 3 (Figure 8). To investigate if there are common transcription factors that regulate both TCR signaling and the Hedgehog signaling pathway, univariate regression studies were performed for the mediators of the two signaling pathways. Although we did not find any single transcription factor as a common regulator of GRB2 and other mediators of Hedgehog signaling, we did find STAT3 to be negatively associated with LCK (another prominent adaptor in TCR signaling) and to be positively associated with GLI2 expression. Based on these findings, we inferred STAT3 to be a balancing transcription factor that, on

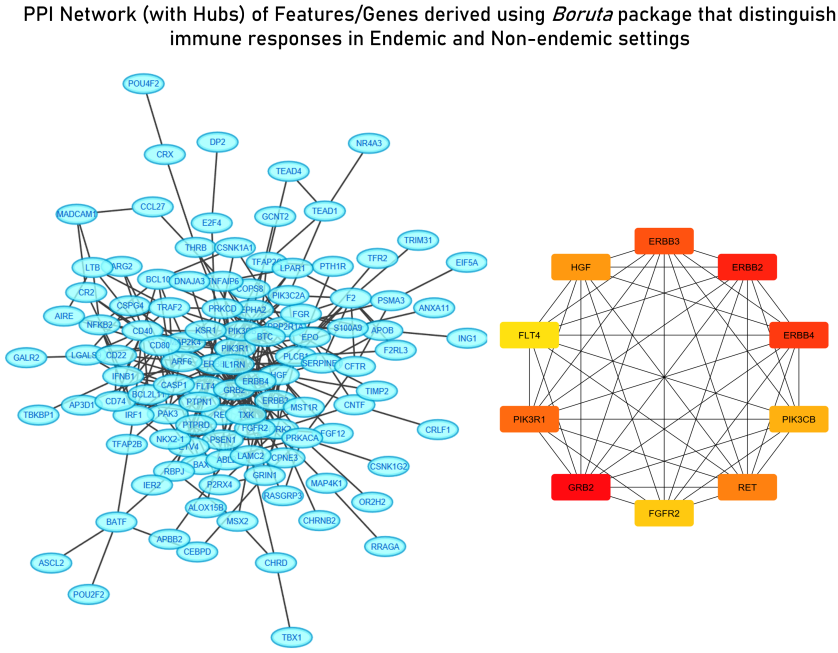


FIGURE 6
STRING network with derived hub genes of feature derived from the Random Forest algorithm distinguishing immune responses in endemic and non-endemic settings derived using the Boruta package.

one hand, regulates TCR signaling while promoting the induction of Hedgehog signaling on the other hand (Figure 8C).

3.7 Retrieved gene regulatory modules

Gene regulatory module 1 (GRM1) was inferred from the GRB2 multivariant model wherein, based on literature, the central role of GRB2 in TCR signaling was identified and key regulatory elements found in this study were integrated (Supplementary Figure S2). Several relevant findings from the obtained results were considered for module construction: (i) Genes involved in TCR signaling were both up- and downregulated upon KEGG pathway enrichment analysis (GRB2 being downregulated) (Supplementary Figures S3 and S4), (ii) downregulation of a cluster of genes (with GRB2 being a hub gene) involved in growth factor receptor signaling (Figure 5 and Table 2), (iii) GRB2 being one of the hub genes in the network obtained through Random Forest-based feature selection (Figure 6), and (iv) GRB2 performing perfectly as a classifier of immune responses in endemic and non-endemic settings (Table 4 and Supplementary Figure S7). Based on these findings, we hypothesize that GRB2 might play an integral role in downregulating growth factor receptor signaling and in negatively regulating downstream TCR signaling in the endemic cohort. Moreover, the MVR model derived for GRB2 (through a combinatorial approach) suggests that while PIK3R1, TP53, FYN, and RELA (from the model in Figure 8), which act downstream of TCR signaling (Supplementary Figures S3 and S4), would be affected by GRB2 suppression, other downstream mediators might actually act as negative regulators (HRAS, MYB, and LCK).

The second gene regulatory module (GRM2) was inferred using the MVR model for GLI2. Interactions of GLI2 with transcription factors and other mediators of TCR signaling and extracellular mediators involved in chemotaxis of lymphocytes were closely studied (Table 3). Through GRM2, we propose Hedgehog signaling pathways as primary differentiators of matured lymphocytes as compared to lymphocytes being freshly induced. Based on the results obtained from hybrid clustering (Figure 5), we propose them to be closely involved in T-cell function in endemic settings upon infection. The third gene regulatory module (GRM3) was specially retrieved based on the regulatory dynamics observed for STAT3 in two different regression models (Figure 8C). Based on our observations, we propose STAT3 as a primary determinant responsible for state switching of T cells upon infection by, on one hand, directly/indirectly negatively regulating TCR induction and, on the other hand, nudging towards Hedgehog signaling. Regulatory modules of GRB2 suppression and the negative association between STAT3 and LCK as derived from the meta-analysis were validated via the RNASeq dataset using a regression model (with an R^2 value of 0.5441) (Figure 9). The culmination of the key findings (which distinguish acute immune responses in endemic and non-endemic settings) from the study is illustrated in the form of a model in Figure 9. For the development of this model, established molecular interactions in TCR signaling were retrieved from literature (33).

Supplementary File 7 provides a more detailed rationale used for the construction of gene regulatory modules while Supplementary Figure S6 provides an illustrative summary of the entire study.

10.3389/fimmu.2024.1285785

TABLE 2 Enriched reactome pathways derived using different methodologies specific for endemic settings along with their key regulators (FDR< 0.05, strength > 0.90, top 10).

Methodology	Enriched reactome pathways	Regulators
Network Topological Analysis of DEGs	•Signaling by FGFR3 fusions in cancer (**HSA-8853334**) •Signaling by PDGFRA transmembrane, juxta-membrane, and kinase domain mutants (**HSA-9673767**) •Activated NTRK2 signals through RAS (**HSA-9026519**) Signaling by FGFR4 in disease (**HSA-5655291**) •Constitutive signaling by overexpressed ERBB2 (**HSA-9634285**) Constitutive signaling by EGFRvIII (**HSA-5637810**) •MET activates PI3K/AKT signaling (**HSA- 8851907**)	MYB, SP1
Hybrid Clustering based on LogFC values (Cluster 2)	•Regulation of IFNG signaling (**HSA-877312**) •Signaling by CSF3 (G-CSF) (**HSA-9674555**) •Spry regulation of FGF signaling (**HSA-1295596**) •Regulation of KIT signaling (**HSA-1433559**) •Inactivation of CSF3 (G-CSF) signaling (**HSA-9705462**) •Regulation of IFNA/IFNB signaling (**HSA-912694**) •CTLA4 inhibitory signaling (**HSA-389513**) •Growth hormone receptor signaling (**HSA-982772**) •Signaling by PTK6 (**HSA-8848021**) •Signaling by SCF-KIT (**HSA-1433557**)	MYB, SP1, SP3, SMARCA4, HIF1A, ETS1, GLI1, CTTNB1, PAX2, STAT5B, ETS2, RELA, NFKB1, NR2C1, SP4, STAT1, YY1, AR, HOXA10, ATF3, DDIT3, GLI2, EP300, ELK1, KLF6, NR1H4, E2F4, ATF1, HDAC3, PGR, TCF4, HDAC1, TFAP2A, CTCF, **STAT3**, JUND, RUNX1, TP53, VDR, USF2, CEBPA, IRF1, BRCA1, GATA1, CEBPB, EGR1, CREB1, MYC
Hybrid Clustering based on LogFC values (Cluster 4)	•Hedgehog ligand biogenesis (**HSA-5358346**) •TP53 regulates transcription of cell death genes (**HSA-5633008**) •Release of Hh-Np from the secreting cell (**HSA-5362798**) •Activation, translocation, and oligomerization of BAX (**HSA-114294**) •Nonsense mediated decay (NMD) independent of the Exon	SP1, SMAD4, RELA, CTCF, ABL1, SNAI1, JUND, NR3C1, CREB5, E2F3, STAT5A, ZEB1, HIF1A, SNAI1, STAT1, FOSL2, BCL6, FOXO3, FOS, WT1, SOX9, SP3, FOXO1, NFKB1, PARP1, LEF1, CIITA, REST, ETS1, ATF, **STAT3**, JUN, EZH2, VDR, MYCN, BRCA1, SPI1, PPARG, HDAC1, ESR1, CREB1, AR, E2F1, TP53
	Junction Complex (EJC) (**HSA-975956**)	
Features from Wrapper Algorithm with Random Forest	•SHC1 events in ERBB2 signaling (**HAS-1250196**) •PI3K events in ERBB2 signaling (**HAS-1963642**) •ERBB2 activates PTK6 signaling (**HAS-8847993**) •MET activates PI3K/AKT signaling (**HAS-8851907**) •Activated NTRK2 signals through PI3K (**HAS- 9028335**) •GRB7 events in ERBB2 signaling (**HSA-1306955**) •GRB2 events in ERBB2 signaling (**HSA-1963640**) •ERBB2 regulates cell motility (**HSA-6785631**) •CD28-dependent Vav1 pathway (**HSA-389359**)	RELA, NFKB1, SP1, FOXA1, STAT1, TFAP2A, AR, NCOS, TRERF1, CUX1, SP3, BTF2, TFAP2C, IRF7, HIF1A, CREB1, NR4A1, FOXA2, NFKBIA, PML, ELK1, CEBPB, ETV4, ATF1, SRF, SAMD4, YBX1, SMAD3, YY1, PPARA, TP53, USF2, IRF1, EP300, SPI1, USF1, PPARG1, STAT3, JUN. ESR1, ETS1, E2F1

4 Discussion

Enteric vaccines have been reported to show low efficacy in regions that are highly endemic to pathogens (4–6). Apart from enteric infections, vaccines against other infectious diseases have also shown similar tendencies. For example, in a study, the YF-17D, the yellow fever vaccine, showed low vaccine efficacy in an African cohort, which the author attributed to an "activated" microenvironment in the study population—including "differentiated T and B cells and pro-inflammatory cytokine secreting monocytes" (34). On similar lines, recently, it has been observed that infection with SARS-CoV-2 with its different variants generates cross-reactive T cells, which are not necessarily protective, but had a direct impact on vaccine effectiveness (35, 36). These findings imply that pre-existing immunity against specific pathogens can have a direct impact on immune responses to subsequent immunization attempts. With SARS-CoV-2 becoming endemic worldwide, the design and development of the next generation of COVID-19 vaccines and advanced vaccines against other endemic infections would require keen consideration to pre-existing protective/semi-protective/non-protective immunity against these pathogens in the target population.

Hence, understanding the immunological dynamics of re-infection in general and the possible impact of immunization in a chronically exposed population becomes absolutely essential for the development of future vaccines that are region- and population-specific (15, 37). In this regard, several studies have investigated immune responses against malaria and other helminth infection in a previously exposed population. One of these studies reported acute upregulation of co-stimulatory molecules (like CD40, CD80, and CD86) upon stimulation of dendritic cells in experienced (38).

10.3389/fimmu.2024.1285785

TABLE 3 List of hub genes specific for the endemic cohort derived using different methodologies along with their corresponding functional roles and regulators (as identified from TRRUST database).

Source	Hub genes*	Biological process	Role	Key regulators
Network Topological Analysis of DEGs	**HRAS**	GO:0000165: MAPK cascade	Housekeeping	N/A
	SOS1	GO:0002260: Lymphocyte homeostasis	Housekeeping	N/A
	KRAS	GO:0000165: MAPK cascade	Housekeeping	N/A
	SRC	GO:0002376: Immune system processes	**Responsive**	SP1, TAF1
	EGFR	GO:0038134: ERBB2- *EGFR* signaling	**Responsive**	AR, BCL3 BRAC1, CREBBP, EGR1, ESR1, HDAC1/3, HOXB7, JUN, JUNB, KLF10, LRRFIP1, MTA1, NFKB1, NR3C2, PGR, PML, PPARG, RELA, SP1
	MAPK1	GO:0000165: MAPK cascade	Housekeeping	**N/A**
	MAPK14	GO:0000165: MAPK cascade	Housekeeping	N/A
	PTK2	GO:0001932: Regulation of protein phosphorylation	**Responsive**	N/A
	UBB	GO:0016567: Protein ubiquitination	Housekeeping	N/A
Hybrid Clustering based on LogFC values (Cluster 2)	**GRB2**	GO:0007173: EGFR signaling	**Responsive**	**N/A**
	ERBB2	GO:0004714: Transmembrane receptor protein tyrosine kinase activity	**Responsive**	AR, ATF, CREB1, DENND4A, ELF1, EP300, ETV4, FOXP3, GATA4, JUND, MYB, NCOA3, PAX2, PGR, PURA, SP1, TFAP2A, VDR, XRCC5, YBX1, YY1
	PTPN11	GO:0000077: DNA damage checkpoint signaling	Housekeeping	N/A
	SOCS1	GO:001817: Regulation of cytokine production	**Responsive**	GL1/2, HIF1A,IRF1, SP1, STAT3/6
	PIK3CD	GO:0002250: Adaptive immune response	**Responsive**	RUNX1
	SOCS3	GO:001817: Regulation of cytokine production GO:0000082:	**Responsive**	CEBPA, NFKB1, RELA, SP3, STAT1/3/4
	CCND1	G1/S transition mitotic cell cycle	Housekeeping	N/A
	PDGFRA	GO:0001775: Cell activation	Housekeeping	N/A
	CSF3R	HSA:9674555: Signaling by CSF3	**Responsive**	CEBPA, ETS1, MYB, SPI1
	LCK	HAS:389356: CD28 co-stimulation	**Responsive**	MYB
	FGF13	GO:0000165: MAPK cascade	Housekeeping	N/A
	WT1	HAS:9675108: Nervous system development	**Responsive**	CTCF, EP300, ETS1, GATA1/2, HDAC4/5, HOXA10, IFI6, MYB, NFKB1, PAX2/8, RELA, SP1, TFCP2
	PHGDH	GO0006541: Glutamine metabolic process GO:0033209:	**Responsive**	HOXA10, SP1
	KRT18	Tumor necrosis Factor-mediated signaling pathway	**Responsive**	BRCA1, CTBP1, SP1
	PTPN1	HAS:163615: PKA activation	Housekeeping	N/A

(Continued)

10.3389/fimmu.2024.1285785

TABLE 3 Continued

Source	Hub genes*	Biological process	Role	Key regulators
Hybrid Clustering based on LogFC values (Cluster 4)	**RPS16**	GO:0006364: rRNA processing	Housekeeping	N/A
	RPL6	Same as above	Housekeeping	N/A
	RPS9	Same as above	Housekeeping	N/A
	RPL22	Same as above	Housekeeping	N/A
	RPS15	Same as above	Housekeeping	N/A
	ETF1	GO:0006415: Translational Termination	Housekeeping	N/A
	SOX2	HAS-452271: Signaling by WNT	**Responsive**	ID4, KDM2A, POU5F1
	FN1	GO:0006953: Acute-phase response	**Responsive**	AR, ATF2, CEBPA, EGR1, KLF8, NFKB1, PARP1, RELA, SNAI1, SOX17, TWIST1/2
	HSP90AA1	GO:0002218: Activation of innate immune response	Housekeeping	N/A
	EEF1D	GO:0009299: Translational elongation	Housekeeping	N/A
	WNT2B	HAS:3238698: WNT ligand biogenesis and trafficking	**Responsive**	GLI2
	JAG1	HAS:2979096:NOTCH2 activation and transcriptional signal to the Nucleus	**Responsive**	KDM4C, PPARG, RUNX3, SNAI2
	TLR6		**Responsive**	HIF1A
Features from Wrapper Algorithm with Random Forest	**GRB2**	GO:0007173: EGFR signaling	**Responsive**	
	ERBB2	See above	**Responsive**	See above
	ERBB4	GO: 0006916: Apoptotic process	**Responsive**	WWP1
	ERBB3	GO:0007162: Negative regulation of cell adhesion	**Responsive**	AR, TWIST1/2, YBX1
	PIK3R1	GO:0002687: Positive regulation of leukocyte migration	**Responsive**	N/A
	RET	GO:0000165: MAPK cascade	**Responsive**	ESR1, FOXA1, SOX10, NKX2-1, TFAP2C
	TXK	GO0001819: Positive regulation of cytokine production	Housekeeping	N/A
	MST1R	GO:0002376: Immune system processes	Responsive	N/A

Functional roles identified from: https://housekeeping.unicamp.br/.
N/A, Not Available.

Another study indicated the important role of γδ T cells in secondary immune responses to malaria in endemic settings (39). Moreover, an immunomodulatory effect of chronic exposure to parasitic infections has also been reported against parasitic infections (40). Such studies are still lagging behind for enteric infections in endemic settings. Using an intensive systems and computational pipeline, we have designated molecular signatures and transcriptional regulatory networks that delineate acute immune responses in endemic settings in comparison to those induced in non-endemic settings, taking enteric infections as a case study. Importantly, we show that (i) there is a negative feedback regulation of downstream signaling pathway associated with T-cell activation through GRB2 downregulation (GRM1), (ii) WNT receptor expression in activated T cells is under the influence of CCL17 (GRM2), and (iii) STAT3 mediated the state change of activated T cells through the upregulation of WNT receptor (GRM3).

To elaborate on the first regulatory module (GRM1), GRB2 is an adaptor molecule assembled and recruited near the intracellular chains of growth factor receptors involved in the activation of RAS, which unleashes the downstream signaling pathways. GRB2 also plays an essential role in TCR signaling by propagating activation/

TABLE 4 MLP classification evaluation of the identified hub genes based on threefold classification.

Gene_LIST	Accuracy	Precision	Recall	*F*-measure	ROC area
GRB2	100%	1	1	1	1
PIK3R1	98.86%	1	0.952	0.976	0.973
ERBB3	97.72%	0.952	0.952	0.952	0.971
ERBB4	97.72%	0.952	0.952	0.952	0.999
RET	95.45%	0.947	0.857	0.9	0.925
ERBB2	94.31%	0.86	0.905	0.884	0.99
TLR6	92.04%	0.889	0.763	0.821	0.902
SOX2	90.90%	0.741	0.952	0.833	0.942
EGFR	89.77%	0.8	0.762	0.78	0.979
PTK2	88.63%	1	0.524	0.688	0.728
SOCS1	88.63%	0.824	0.667	0.737	0.841
PIK3CD	87.50%	0.917	0.524	0.667	0.781
PHGDH	85.22%	0.682	0.714	0.698	0.84
CSF3R	78.40%	0.583	0.333	0.424	0.768
KRT18	77.27%	0.667	0.095	0.167	0.569
FN1	77.27%	0.52	0.619	0.565	0.741
WNT2B	76.13%	NA	NA	NA	0.599
JAG1	76.13%	NA	NA	NA	0.482
SOCS3	76.13%	0.5	0.238	0.323	0.841
LCK	75%	0	0	0	0.385
WT1	72.72%	0.385	0.238	0.294	0.731

Model construction and evaluation were performed using the WEKA software.
N/A, Not Available.

proliferation signals intracellularly after synapse formation of the TCR complex with the peptide–MHC complex through the activation of MAPK signaling pathway. Upon TCR/co-receptor stimulation of LCK, an SRC family tyrosine kinase,* gets activated and, through a short series of phosphorylation, recruits ZAP-70, which, in turn, facilitates the assembly of downstream scaffolds that includes the Linker Activator of T-cells (LAT). LAT provides a platform for GRB2 (and for other adaptor molecules) assembly where GRB2 relays the received signals through RAS activation (41). Because of its early involvement in signaling events, GRB2 has been designated as a rate-limiting and essential component of the TCR-induced MAPK/ERK signaling pathway, which is essential for lymphocyte selection, proliferation, and differentiation (42–44).

Owing to the constitutive and ubiquitous nature of the MAPK pathway and risk associated with its overexpression, several negative regulatory circuits have evolved throughout the signaling pathway downstream of TCR activation (45). Broadly, there are two channels of negative regulation that involve the phosphorylation-based functional inactivation of upstream mediators by activated ERK and, secondly, the transcriptional regulation of upstream mediators. In terms of GRB2 suppression, phosphorylation of LAT, which leads to its disassociation with GRB2, has been previously reported, which is an example of the former, and induction of SPRY protein (through ERK pathway activation) that binds and disables GRB2 action can be considered as an example of the latter (41, 46). Although post-translational regulation of GRB2 is well documented (46, 47), transcriptional regulation of GRB2 expression remains quite elusive in the literature.

Our study, particularly MVR analysis focusing on GRB2 expression using the gene expression dataset, indicates that high expression levels of HRAS, MYB (downstream mediators of growth factor receptor signaling), and LCK (adaptor for the TCR receptor) negatively affect GRB2 expression upon perturbation (antigenic exposure), which might negatively impact T-cell activation and proliferation. This observation is further validated by the fact that GRB2 was peculiarly downregulated at the acute stage of infection in an endemic setting and the fact that the TCR signaling pathway was also seen to be downregulated in this endemic cohort (Supplementary Figure 8). The molecular and transcriptional mechanism for suppression of GRB2 expression needs further investigation. Although MIR200a and microRNA have been reported to suppress the expression of GRB2, consequently negatively regulating the MAPK signaling pathway (48), its relevance in this particular setting is not known.

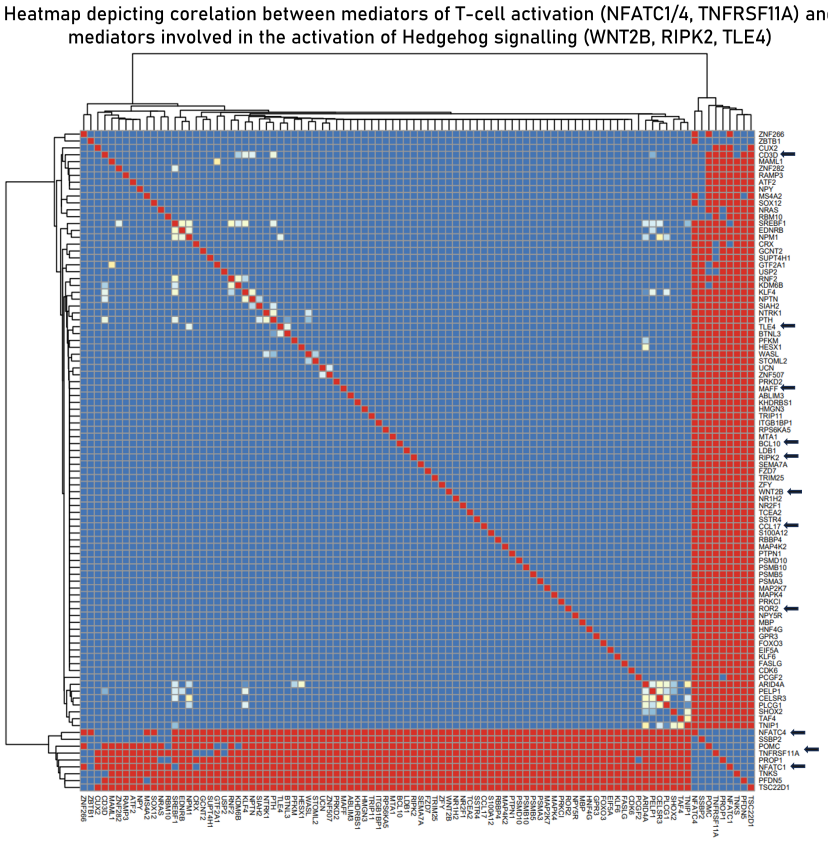


FIGURE 7
Curated submodule derived from module 3 correlation module derived from the *EBModules* package that shows the positive associations of positive regulators of T-cell activation with mediators of the Hedgehog signaling pathways. Here, red bricks indicate a high correlation coefficient of 1, blue bricks indicate a correlation coefficient of 0, and yellow bricks indicate intermediate correlation coefficient.

To further explore if the described phenomenon occurs upon vaccination as well, investigation of GRB2 expression levels in other vaccine clinical trials in the ImmuneSpace database was conducted. We found that clinical trials with ImmunPort accession IDs SDY299, SDY1328, SDY1276, and SDY180 (out of 47 studies reporting GRB2 expression levels) also report GRB2 downregulation at early time points of vaccination (Supplementary Table S7), validating GRB2 suppression as an acute immunomodulatory response in certain conditions. Gene expression datasets (post vaccination) from endemic settings were not available in the ImmuneSpace database (Supplementary Figure S8).

The outcomes of our analysis specifically might have profound implications in the vaccine design and development of endemicity/region-specific vaccines as it would provide explanation to previously ambiguous vaccine trial outcomes where unexpectedly suboptimal T-cell responses were observed (as discussed above). Importantly, as baseline-heightened immunological profile in the endemic cohorts is very well documented, we hypothesize that further perturbation/exposure/attack of pathogen might push TCR signaling into an auto-regulatory loop. This would imply that suboptimal vaccine efficacy observed in these regions would be the inherent characteristic of the vaccinees, and hence, increasing the dosage of a vaccine or using high adjuvanticity might not have the expected result and might actually disrupt the biological "seesaw" or balance put in place to check for immune hyperactivity or even autoimmunity. This is worth considering particularly because several autoinflammatory and autoimmune diseases have been attributed to GRB2-linked molecular assemblies (41, 49). In the same line, in mice, it has been demonstrated that GRB2-induced MAPK/Erk signaling pathway might switch to hyperactivity if not negatively regulated by LCK (50) (negative association of LCK and GRB2 was demonstrated through our analysis) (Figures 8 and 9).

While GRB2 suppression solely would have indicated a regulatory immune response to infection in these settings, the observed GRM2 indicates a more multidimensional effector function of T cells. Overall, these findings suggest a biphasic transformative nature of T cells, which is dependent on the pathogenic load of the environment. In this regard, we propose STAT3 to be a key determiner of biphasic T-cell function in endemic settings based on its negative association with LCK expression and positive association with GLI2 (transcription factor for WNT2B receptor expression). Our findings are validated by the fact that STAT3 has been reported to dampen

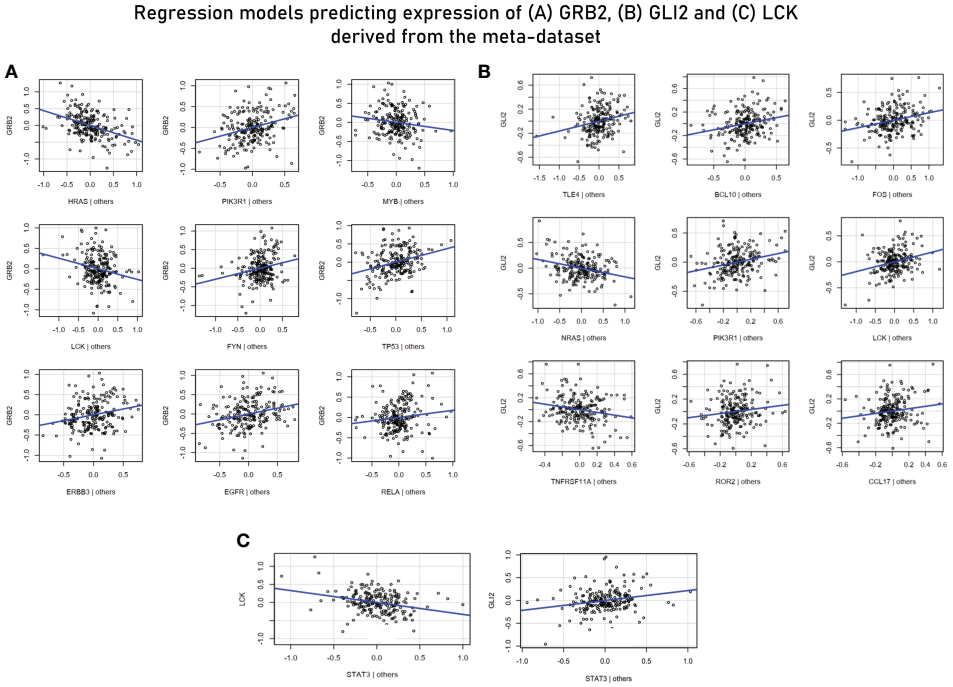


FIGURE 8
(A) Multivariant regression model for GRB2 (R^2 = 0.76). While the rest of the predictors showed a positive association with GRB2, HRAS, MYB (TF for LCK), and LCK demonstrated a negative association. **(B)** Multivariant regression model for GLI2 (TF for WNT2B) (R^2 = 0.70). The model demonstrated positive associations of GL2 with key mediators of TCR signaling: BCL10, PIK3R1, LCK, and TNFRSF11A. **(C)** Regression model predicting the association of LCK and GLI2 with STAT3.

immune responses, which, in this case, can be a result of frequent exposure to enteric pathogens in pathogen-prevalent regions. STAT3 has also been reported to promote the activation of regulatory T-cell responses (51). Besides this, a strong indication of the WNT signaling pathway being involved in immune responses in endemic settings is an intriguing finding. Recently, WNT signaling has been reported to be activated in the local mucosa in subjects affected by environmental enteropathy, which is prominent in regions with endemicity of enteric infections (52). WNT signaling pathways have been reported to play an integral role in the differentiation and functioning of mature T cells particularly in the context of cell-to-cell communication and in cell migration/homing (53, 54). Given this, activation of these signaling pathways could mediate the induction of regulatory T cells (differentiation) as an immunomodulatory response to re-infection. These signaling pathways, especially the WNT signaling pathway, can also be involved in T-cell trafficking towards infected mucosa under the influence of activated leukocytes and, resultantly, cytokine secretion. Through our work, we also established positive associations between the induction of these pathways and the chemokine ligand CCL17, which is an established lymphocyte chemoattractant (GRM2) (55, 56).

Although the robust computational pipeline provides novel insights into the key molecular mechanisms that might be peculiar to endemic settings, the study is restricted by the sample size secured for the endemic population due to the unavailability/inaccessibility of immune response-linked gene expression datasets from these settings even after the systemic screening of public repositories. Another major limitation of the study is the loss of genes to a mere 6,543 genes in the meta-dataset, which could be considered as a "cost-of-merger" of heterogeneous gene expression datasets. We suspect that, like GRB2, we might come across other key molecular mediators that play an essential role in distinguishing immune responses in endemic and non-endemic populations that can only be uncovered by multicohort studies (from endemic and non-endemic settings) where pre- and post-infection/vaccination RNASeq data are retrieved for all the study groups.

Despite the mentioned limitations, in conclusion, through a novel methodical analytical pipeline, we demonstrate that gene expression datasets provide an unprecedented opportunity to understand variations in gene regulatory modules involved in immune responses to pathogens in different environmental settings (with a different pathogenic load). We used an amalgamation of systems (in the form of STRING networks) and advanced computational approaches (hybrid clustering, wrapper method for feature selection, MLP classification, correlation, and MVR analysis) to delineate immune responses specific to the endemic cohort of the study. Based on the findings of the study, we propose that perhaps the basal immune system and subsequent post-infection/vaccination immune responses diverge upon varying

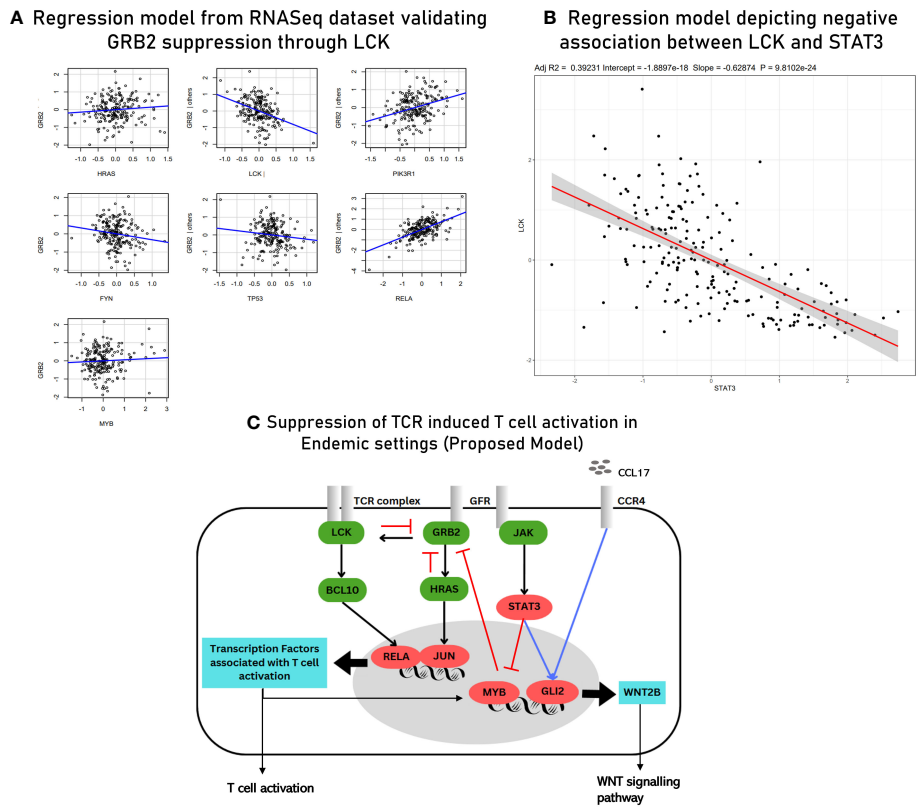


FIGURE 9
(A) Multivariant regression model for GRB2 suppression with R^2 = 0.54 derived from RNASeq data (validation). **(B)** Scatter plot depicting a negative association between LCK and STAT3 derived using RNASeq data (validation). **(C)** Proposed model of TCR signaling upon acute infection in endemic settings. The known/established regulatory associations in TCR signaling are depicted with black arrows. The negative regulation of GRB2 (depicted with red inhibitory arrows) is inferred from gene regulatory module 1. The induction of GLI2 by the transcription factors associated with activated T cells and through CCL17-based signaling (blue arrow) is inferred from gene regulatory module 2. STAT3-mediated inhibition of MYB (transcription factor for LCK) (red arrow) and the positive regulation of GLI2 (blue arrow) are inferred from the findings of gene regulatory module 3. The three gene regulatory modules are described in detail in the Supplementary File 8.

levels of previous exposures. Consequently, detailed insight into the reasons and principles behind these divergences should form the basis for the design and development of the "next-gen" precise vaccines. We put forward acute GRB2 suppression as a divergent (immunomodulatory) path the immune system evolves to take in endemic settings as one of the divergent paths the immune system evolves to take. While these observations are specific for *S. typhi* (intracellular bacterial) infection that attacks the enteric mucosa, further studies that look into the induction of the discussed regulatory molecules in other mucosal infections (possibly other enteric infections) can be an exciting start towards the development of endemicity-specific vaccines. From a global health standpoint, these studies should also include infections induced in the lung mucosa because of seasonal or perennial prevalence by pathogens like the influenza virus and quite recently by SARS CoV-2.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: GSE7000, GSE112958, GSE95104, GSE2729, GSE69529.

Ethics statement

Ethical approval was not required for the study involving humans in accordance with the local legislation and institutional requirements. Written informed consent to participate in this study was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and the institutional requirements.

Author contributions

AN: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Visualization, Writing – original draft. SL: Conceptualization, Methodology, Supervision, Validation, Writing – review & editing.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2024.1285785/full#supplementary-material

References

1. *Child mortality and causes of death*. Available at: https://www.who.int/data/gho/data/themes/topics/topic-details/GHO/child-mortality-and-causes-of-death.

2. *Diarrhoea - UNICEF DATA*. Available at: https://data.unicef.org/topic/child-health/diarrhoeal-disease/.

3. Alam MM, Aktar A, Afrin S, Rahman MA, Aktar S, Uddin T, et al. Antigen-specific memory B-cell responses to enterotoxigenic escherichia coli infection in Bangladeshi adults. *PloS Negl Trop Dis* (2014) 8(4). doi: 10.1371/journal.pntd.0002822

4. Lopman BA, Pitzer VE, Sarkar R, Gladstone B, Patel M, Glasser J, et al. Understanding reduced rotavirus vaccine efficacy in low socio-economic settings. *PloS One* (2012) 7(8). doi: 10.1371/journal.pone.0041720

5. Naylor C, Lu M, Haque R, Mondal D, Buonomo E, Nayak U, et al. Environmental enteropathy, oral vaccine failure and growth faltering in infants in Bangladesh. *EBioMedicine* (2015) 2(11):1759–66. doi: 10.1016/J.EBIOM.2015.09.036

6. *Weekly epidemiological record Relevé épidémiologique hebdomadaire*. (2017). Available at: http://www.who.

7. Holmgren J, Parashar UD, Plotkin S, Louis J, Ng SP, Desauziers E, et al. Correlates of protection for enteric vaccines. *Vaccine* (2017) 35(26):3355–63. doi: 10.1016/J.VACCINE.2017.05.005

8. Riddle MS, Chen WH, Kirkwood CD, MacLennan CA. Update on vaccines for enteric pathogens. *Clin Microbiol Infection* (2018) 24(10):1039–45. doi: 10.1016/J.CMI.2018.06.023

9. Kazmin D, Nakaya HI, Lee EK, Johnson MJ, van der Most R, Van Den Berg RA, et al. Systems analysis of protective immune responses to RTS,S malaria vaccination in humans. *Proc Natl Acad Sci USA* (2017) 114(9):2425–30. doi: 10.1073/pnas.1621489114

10. Mottram L, Lundgren A, Svennerholm A-M, Leach S. A systems biology approach identifies B cell maturation antigen (BCMA) as a biomarker reflecting oral vaccine induced IgA antibody responses in humans. *Front Immunol* (2021) 12:647873. doi: 10.3389/FIMMU.2021.647873

11. Zhu H, Chelysheva I, Cross DL, Blackwell L, Jin C, Gibani MM, et al. Molecular correlates of vaccine-induced protection against typhoid fever. *J Clin Invest* (2023) 133 (16):e169676. doi: 10.1172/JCI169676

12. Li S, Rouphael N, Duraisingham S, Romero-Steiner S, Presnell S, Davis C, et al. Molecular signatures of antibody responses derived from a systems biology study of five human vaccines. *Nat Immunol* (2014) 15(2):195–204. doi: 10.1038/ni.2789

13. Liu YE, Darrah PA, Zeppa JJ, Kamath M, Laboune F, Douek DC, et al. Blood transcriptional correlates of BCG-induced protection against tuberculosis in rhesus macaques. *Cell Rep Med* (2023) 4(7):101096. doi: 10.1016/j.xcrm.2023.101096

14. Naidu A, Lulu S S. Mucosal and systemic immune responses to Vibrio cholerae infection and oral cholera vaccines (OCVs) in humans: a systematic review. *Expert Rev Clin Immunol* (2022) 18(12):1307–18. doi: 10.1080/1744666X.2022.2136650

15. Ragonnet R, Trauer JM, Denholm JT, Geard NL, Hellard M, McBryde ES. Vaccination programs for endemic infections: modelling real versus apparent impacts of vaccine and infection characteristics. *Sci Rep 2015 5:1* (2015) 5(1):1–11. doi: 10.1038/srep15468

16. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* (2012) 28(6):882–3. doi: 10.1093/BIOINFORMATICS/BTS034

17. Sean D, Meltzer PS. GEOquery: a bridge between the gene expression omnibus (GEO) and bioConductor. *Bioinformatics* (2007) 23(14):1846–7. doi: 10.1093/BIOINFORMATICS/BTM254

18. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* (2015) 43(7):e47–7. doi: 10.1093/NAR/GKV007

19. Wickham H. Getting started with qplot. *Ggplot2* (2009), 9–26. doi: 10.1007/978-0-387-98141-3_2

20. Durinck S, Spellman PT, Birney E, Huber W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat Protoc* (2009) 4(8):1184–91. doi: 10.1038/nprot.2009.97

21. Chin CH, Chen SH, Wu HH, Ho CW, Ko MT, Lin CY. cytoHubba: identifying hub objects and sub-networks from complex interactome. *BMC Syst Biol* (2014) 8 Suppl 4(Suppl 4):S11. doi: 10.1186/1752-0509-8-S4-S11

22. Wang J, Zhong J, Chen G, Li M, Wu F-X, Pan Y. ClusterViz: a cytoscape APP for cluster analysis of biological network. *IEEE/ACM Trans Comput Biol Bioinf* (2015) 12:815–22. doi: 10.1109/TCBB.2014.2361348

23. Kursa MB, Rudnicki WR. Feature selection with the boruta package. *J Stat Software* (2010) 36(11):1–13. doi: 10.18637/jss.v036.i11

24. Lê S, Josse J, Husson F. FactoMineR: An r package for multivariate analysis. *J Stat Software* (2008) 25(1):1–18. doi: 10.18637/jss.v025.i01

25. Alharbi F, Vakanski A. Machine learning methods for cancer classification using gene expression data: A review. *Bioengineering (Basel).* (2023) 10(2):173. doi: 10.3390/bioengineering10020173

26. Carreras J, Hamoudi R. Artificial neural network analysis of gene expression data predicted non-hodgkin lymphoma subtypes with high accuracy. *Mach Learn Knowledge Extraction* (2021) 3:720–39. doi: 10.3390/make3030036

27. Zollinger A, Davison AC, Goldstein DR. Automatic module selection from several microarray gene expression studies. *Biostatistics* 19:153–68. doi: 10.1093/biostatistics/kxx032

28. Mbebi AJ, Nikoloski Z. Gene regulatory network inference using mixed-norms regularized multivariate model with covariance selection. *PloS Comput Biol* (2023) 19 (7):e1010832. doi: 10.1371/journal.pcbi.1010832

29. Wang Y, Dennehy PH, Keyserling HL, Tang K, Gentsch JR, Glass RI, et al. Rotavirus infection alters peripheral t-cell homeostasis in children with acute diarrhea. *J Virol* (2007) 81(8):3904–12. doi: 10.1128/JVI.01887-06

30. Yang WE, Suchindran S, Nicholson BP, McClain MT, Burke T, Ginsburg GS, et al. Transcriptomic analysis of the host response and innate resilience to enterotoxigenic escherichia coli infection in humans. *J Infect Dis* (2016) 213 (9):1495–504. doi: 10.1093/infdis/jiv593

31. Thompson LJ, Dunstan SJ, Dolecek C, Perkins T, et al. Transcriptional response in the peripheral blood of patients infected with salmonella enterica serovar typhi. *Proc Natl Acad Sci U.S.A.* (2009) 106(52):22433–8.

32. Hanafusa H., Torii S., Yasunaga T., et al. Sprouty1 and Sprouty2 provide a control mechanism for the Ras/MAPK signalling pathway. *Nat Cell Biol* (2002) 4::580–8. doi: 10.1038/ncb867

33. Courtney AH, Lo WL, Weiss A. TCR signaling: mechanisms of initiation and propagation. *Trends Biochem Sci* (2018) 43(2):108–23. doi: 10.1016/J.TIBS.2017.11.00

34. Muyanja E, Ssemaganda A, Ngauv P, Cubas R, Perrin H, Srinivasan D, et al. Immune activation alters cellular and humoral responses to yellow fever 17D vaccine. *J Clin Invest* (2014) 124(10):1–1. doi: 10.1172/JCI77956

35. Kundu R, Narean JS, Wang L, Fenn J, Pillay T, Fernandez ND, et al. Cross-reactive memory T cells associate with protection against SARS-CoV-2 infection in COVID-19 contacts. *Nat Commun* (2022) 13(1). doi: 10.1038/s41467-021-27674-x

36. Murray SM, Ansari AM, Frater J, Klenerman P, Dunachie S, Barnes E, et al. The impact of pre-existing cross-reactive immunity on SARS-CoV-2 infection and vaccine responses. *Nat Rev Immunol* (2023) 23(5):304–16. doi: 10.1038/s41577-022-00809-x

37. Driciru E, Koopman JPR, Cose S, Siddiqui AA, Yazdanbakhsh M, Elliott AM, et al. Immunological considerations for schistosoma vaccine development:

transitioning to endemic settings. *Front Immunol* (2021) 12:635985/BIBTEX. doi: 10.3389/FIMMU.2021.635985/BIBTEX

38. Turner TC, Arama C, Ongoiba A, Doumbo S, Doumtabé D, Kayentao K, et al. Dendritic cell responses to Plasmodium falciparum in a malaria-endemic setting. *Malaria J* (2021) 20(1):1–13. doi: 10.1186/S12936-020-03533-W/FIGURES/6

39. Kurup SP, Harty JT. γδ T cells and immunity to human malaria in endemic regions. *Ann Trans Med* (2015) 3(Suppl 1):S22–2. doi: 10.3978/J.ISSN.2305-5839.2015.02.22

40. Loke P, Lee SC, Oyesola OO. Effects of helminths on the human immune response and the microbiome. *Mucosal Immunol 2022 15:6* (2022) 15(6):1224–33. doi: 10.1038/s41385-022-00532-9

41. Bilal MY, Houtman JCD. Transmission of T cell receptor-mediated signaling via the GRB2 family of adaptor proteins. In: *Signaling mechanisms regulating T cell diversity and function*. Boca Raton (FL): CRC Press/Taylor & Francis (2018), p. 147–75. doi: 10.1201/9781315371689-9

42. Jang IK, Zhang J, Chiang YJ, Kole HK, Cronshaw DG, Zou Y, et al. Grb2 functions at the top of the T-cell antigen receptor-induced tyrosine kinase cascade to control thymic selection. *Proc Natl Acad Sci USA* (2010) 107(23):10620–5. doi: 10.1073/pnas.0905039107

43. Rozengurt E, Soares HP, Sinnet-Smith J. Suppression of feedback loops mediated by pi3k/mtor induces multiple overactivation of compensatory pathways: An unintended consequence leading to drug resistance. *Mol Cancer Ther* (2014) 13 (11):2477–2488). doi: 10.1158/1535-7163.MCT-14-0330

44. Radtke D, Lacher SM, Szumilas N, Sandrock L, Ackermann J, Nitschke L, et al. Grb2 is important for T cell development, th cell differentiation, and induction of experimental autoimmune encephalomyelitis. *J Immunol* (2016) 196(7):2995–3005. doi: 10.4049/jimmunol.1501764

45. Reth M, Brummer T. Feedback regulation of lymphocyte signalling. *Nat Rev Immunol* (2004) 4(4):269–77. doi: 10.1038/nri1335

46. Shin SY, Rath O, Choo SM, Fee F, McFerran B, Kolch W, et al. Positive- and negative-feedback regulations coordinate the dynamic behavior of the Ras-Raf-MEK-ERK signal transduction pathway. *J Cell Sci* (2009) 122(3):425–35. doi: 10.1242/jcs.036319

47. Zhou J, Tu D, Peng R, Tang Y, Deng Q, Su B, et al. RNF173 suppresses RAF/MEK/ERK signaling to regulate invasion and metastasis via GRB2 ubiquitination in Hepatocellular Carcinoma. *Cell Communication Signaling* (2023) 21(1). doi: 10.1186/s12964-023-01241-x

48. Liu Y, Liu Q, Jia W, Chen J, Wang J, Ye D, et al. MicroRNA-200a regulates grb2 and suppresses differentiation of mouse embryonic stem cells into endoderm and mesoderm. *PloS One* (2013) 8(7). doi: 10.1371/journal.pone.0068990

49. Shan X, Miao Y, Fan R, Song C, Wu G, Wan Z, et al. Suppression of Grb2 expression improved hepatic steatosis, oxidative stress, and apoptosis induced by palmitic acid *in vitro* partly through insulin signaling alteration. *In Vitro Cell Dev Biol - Anim* (2013) 49(8):576–82. doi: 10.1007/s11626-013-9646-9

50. Methi T, Ngai J, Vang T, Torgersen KM, Taskén K. Hypophosphorylated TCR/CD3ζ signals through a Grb2-SOS1-Ras pathway in Lck knockdown cells. *Eur J Immunol* (2007) 37(9):2539–48. doi: 10.1002/eji.200636973

51. Oweida AJ, Darragh L, Phan A, Binder D, Bhatia S, Mueller A, et al. STAT3 modulation of regulatory T cells in response to radiation therapy in head and neck cancer. *JNCI: J Natl Cancer Institute* (2019) 111(12):1339–49. doi: 10.1093/JNCI/DJZ036

52. Kummerlowe C, Mwakamui S, Hughes TK, Mulugeta N, Mudenda V, Besa E, et al. Single-cell profiling of environmental enteropathy reveals signatures of epithelial remodeling and immune activation. *Sci Trans Med* (2022) 14(660). doi: 10.1126/SCITRANSLMED.ABI8633/SUPPL_FILE/SCITRANSLMED.ABI8633_DATA_FILE_S1.ZIP

53. van Loosdregt J, Coffer PJ. The role of WNT signaling in mature T cells: T cell factor is coming home. *J Immunol* (2018) 201(8):2193–200. doi: 10.4049/JIMMUNOL.1800633

54. Vanderbeck A, Maillard I. Notch signaling at the crossroads of innate and adaptive immunity. *J Leukocyte Biol* (2021) 109(3):535–48. doi: 10.1002/JLB.1RI0520-138R

55. Mendez-Enriquez E, García-Zepeda EA. The multiple faces of CCL13 in immunity and inflammation. *Inflammopharmacology* (2013) 21(6):397–406.doi: 10.1007/S10787-013-0177-5/TABLES/2

56. Kohli K, Pillarisetty VG, Kim TS. Key chemokines direct migration of immune cells in solid tumors. *Cancer Gene Ther 2021 29:1* (2021) 29(1):10–21. doi: 10.1038/s41417-021-00303-x

frontiers | Frontiers in Immunology

TYPE Original Research
PUBLISHED 08 March 2024
DOI 10.3389/fimmu.2024.1357342

OPEN ACCESS

EDITED BY
Joe Hou,
Fred Hutchinson Cancer Center, United States

REVIEWED BY
Vikram Dalal,
Washington University in St. Louis, United States
Khaled Mohamed Darwish,
Suez Canal University, Egypt

*CORRESPONDENCE
Abdelali Agouni
aagouni@qu.edu.qa

RECEIVED 17 December 2023
ACCEPTED 08 February 2024
PUBLISHED 08 March 2024

CITATION
Khan A, Zahid MA, Mohammad A and Agouni A (2024) Structure-guided engineering and molecular simulations to design a potent monoclonal antibody to target aP2 antigen for adaptive immune response instigation against type 2 diabetes. *Front. Immunol.* 15:1357342. doi: 10.3389/fimmu.2024.1357342

Structure-guided engineering and molecular simulations to design a potent monoclonal antibody to target aP2 antigen for adaptive immune response instigation against type 2 diabetes

Abbas Khan[1], Muhammad Ammar Zahid[1], Anwar Mohammad[2] and Abdelali Agouni[1]*

[1]Department of Pharmaceutical Sciences, College of Pharmacy, QU Health, Qatar University, Doha, Qatar, [2]Department of Biochemistry and Molecular Biology, Dasman Diabetes Institute, Kuwait City, Kuwait

Introduction: Diabetes mellitus (DM) is recognized as one of the oldest chronic diseases and has become a significant public health issue, necessitating innovative therapeutic strategies to enhance patient outcomes. Traditional treatments have provided limited success, highlighting the need for novel approaches in managing this complex disease.

Methods: In our study, we employed graph signature-based methodologies in conjunction with molecular simulation and free energy calculations. The objective was to engineer the CA33 monoclonal antibody for effective targeting of the aP2 antigen, aiming to elicit a potent immune response. This approach involved screening a mutational landscape comprising 57 mutants to identify modifications that yield significant enhancements in binding efficacy and stability.

Results: Analysis of the mutational landscape revealed that only five substitutions resulted in noteworthy improvements. Among these, mutations T94M, A96E, A96Q, and T94W were identified through molecular docking experiments to exhibit higher docking scores compared to the wild-type. Further validation was provided by calculating the dissociation constant (K_D), which showed a similar trend in favor of these mutations. Molecular simulation analyses highlighted T94M as the most stable complex, with reduced internal fluctuations upon binding. Principal components analysis (PCA) indicated that both the wild-type and T94M mutant displayed similar patterns of constrained and restricted motion across principal components. The free energy landscape analysis underscored a single metastable state for all complexes, indicating limited structural variability and potential for high therapeutic efficacy against aP2. Total binding free energy (TBE) calculations further supported the superior performance of the T94M mutation, with TBE values demonstrating the enhanced binding affinity of selected mutants over the wild-type.

Discussion: Our findings suggest that the T94M substitution, along with other identified mutations, significantly enhances the therapeutic potential of the CA33 antibody against DM by improving its binding affinity and stability. These results not only contribute to a deeper understanding of antibody-antigen interactions in the context of DM but also provide a valuable framework for the rational design of antibodies aimed at targeting this disease more effectively.

KEYWORDS

AP2, CA33, antibody, structural engineering, docking, simulation, free energy calculation

1 Introduction

Diabetes mellitus (DM) is considered the oldest chronic disease that is characterized by high glucose levels in the blood. It mainly occurs due to the scarcity of insulin production and can be classified into two types: type 1 (T1DM) and type 2 (T2DM) DM (1, 2). The condition arises from the destruction of pancreatic beta-cells which consequently cannot produce insulin. In T2DM, insulin production is decreased but not completely abolished. The delay in diagnosis or management of diabetes may lead to serious complications such as diabetic neuropathy, retinopathy, diabetic foot ulcer, and cardiovascular diseases. DM is also considered as a socioeconomic burden and recent data revealed that by 2049, there will be 629 million people suffering from DM worldwide (3). The major contributing risk factors in the development of this condition include genetic predisposition, obesity, and a sedentary lifestyle. The distorted metabolic functioning and regulation of the adipose tissue are also considered another important aspect contributing to the pathophysiology of DM (4, 5).

Adipose tissue is an endocrine organ that maintains the homeostasis of various other tissues such as the brain, pancreas, and liver (6). Adipocytes respond to metabolic and immune cues by mobilizing their fat stores through lipolysis and by secreting a variety of hormones known as adipokines (7). Such signals interact with the target tissues to regulate several important processes such as glucose or insulin production. Integration of systematic metabolic regulation with adipocytes is primarily controlled by a (FAPB4) fatty acid binding protein 4 or aP2 (8). Since its discovery, the role of aP2 has been depicted in lipid metabolism and the pathogenesis of several metabolic diseases such as atherosclerosis, fatty liver, and diabetes (9–11). Improved liver function, increased sensitivity to insulin, and reduced fatty liver have been reported in mice deficient with aP2 protein thus showing the essential role of this protein in chronic metabolic disorders. The connection between aP2 and T2DM is further corroborated by genetic investigation studies conducted in diverse populations (12). These studies have shown that individuals with a rare haplo-sufficiency mutation in the *aP2* gene experience metabolic and cardiovascular advantages (13). This finding further confirms the involvement of aP2 in the pathogenesis of metabolic diseases. Being an intracellular protein, aP2 also acts as an active adipokine, a peptide that is secreted by adipose tissue that regulates hepatic glucose production and systematic glucose homeostasis. It has also been reported that aP2 contributes to insulin resistance as its serum levels are significantly elevated in obese mice and T2DM (14). In human-based investigations, the role of aP2 was observed in metabolic and cardiovascular disorders. Nonetheless, in a population-based study, reduced expression of aP2 was found to protect against cardiovascular disease and diabetes. Taken together, these findings underline that the biological and hormonal roles of aP2 are evolutionarily conserved and hold relevance in the context of human pathophysiology. Furthermore, the presence of secreted aP2 indicates a robust and promising therapeutic target for the development of therapeutics for diabetes (10, 15). Additionally, this paradigm-shifting evidence about aP2 biology underscores the potential for designing novel therapeutics based on anti-aP2 monoclonal antibodies (mAb) and offers potential solutions to the existing challenges in diabetes treatment (16).

Targeting aP2 therapeutically is a formidable task; however, Burak et al. identified a mAb, CA33, specifically targeting aP2 that was reported to improve glucose metabolism, increase insulin sensitivity, reduce fat mass, and ameliorate liver steatosis in obese mouse models (16). They reported that the novel mAb, CA33, binds to the aP2 through a direct interaction with the light chain and an indirect interaction with the heavy chain. Improving the specificity and binding of CA33 may yield better therapeutic outcomes and elicit stronger immune response. Therefore, using state-of-the-art computational methods is a promising approach to engineer therapeutic proteins for improved bindings. *In silico* saturation, mutagenesis offers a faster and more accurate way to improve the binding by inducing specific mutations. For instance, such methods have been used to engineer different proteins in different diseases such as stomach ulcers, cancer, and SARS-CoV-2 (17–20).

Computational methods have greatly accelerated the identification and development of therapeutic agents against various diseases (21, 22). As proof of the principle of this therapeutic direction, the current study uses in silico mutagenesis approaches by employing the graph signature-based algorithm to

determine the impact of novel substitutions on the binding of CA33 with aP2. We resolved the mutated structures by using Chimera software and the interaction of the mutated CA33 with aP2 was predicted through the HADDOCK algorithm. A mutational landscape of 57 mutants was constructed which revealed that only 4 substitutions were able to improve the binding. The mutations designed to enhance affinity were subsequently examined through the utilization of dissociation constant calculations and molecular simulations. These analyses have confirmed the efficacy of the four most prominent mutants, namely T94M, T94W, A96Q, and A96GE, in their ability to enhance the binding affinity of CA33 with aP2. These mutant variants may be deemed suitable for experimental verification in the context of therapeutic applications.

2 Materials and methods

2.1 Structure retrieval, preparation, and interface analysis

The crystallographic coordinates of the aP2-CA33 complex were retrieved from the Protein Databank (RCSB) using the accession number 5C0N. the native structure contains three chains including the aP2 which comes in direct contact with the light (L) chain of the antibody and a heavy (H) chain of the antibody which interacts indirectly with the aP2 (23). The structures were assessed before further processing and the L chain has some missing residues so Modeler was used to model the missing loops. The structure was minimized and prepared in Chimera using the Conjugate gradients and steepest descent algorithms to relax the contacts and address deformity (24). The final prepared structure was submitted to PDBsum and analyzed for the contacts using PyMOL visualization software. The interface residues were retrieved using the PDBsum and PyMOL consensually (25, 26).

2.2 Graph-based signature algorithm for antibodies modeling

For the flexible and robust recognition and binding of the CA33 antibody by aP2, we employed a computational algorithm, graph-based signatures, available as mCSM-Ab2 (http://structure.bioc.cam.ac.uk/mcsm_ab) which uses experimental data to predict the impact of a particular mutation on the binding of antigen and antibody (27). The interface residues were scanned for predicting the essential contacts which revealed three residues important for recognition while the other three contacts are supplementary. We generated a mutational landscape of 57 mutants by replacing the Glu27, Thr94, and Ala96 with the remaining 19 amino acids to understand the impact on stability and binding affinity. The two contacts Tyr92 and Asp28 were kept the same as they are required for the recognition of the antigen. Among the 57 mutants only top mutations that affect the overall binding (increase) were selected for subsequent analysis. The top-scoring residues that increase the binding of the antibody were modeled in Chimera using the Dunbrack rotamers library based on the proper sidechain torsion (chi) and probability value (24). For optimization purposes, rotamer sampling and side-chain flexibility were applied.

2.3 Antigen-Ab docking using HADDOCK

To model the biomolecular complexes of the antigen (aP2) and antibodies, we used a high ambiguity-driven protein-protein docking (HADDOCK) algorithm. This approach utilized the biophysical and biochemical data to model the interactions and gives the results based on chemical shift perturbation data obtained from NMR titration experiments of mutagenesis data. The obtained information is then incorporated into the docking process such as Ambiguous interaction Restraints (AIRs). An AIR is specifically characterized as an uncertain distance constraint involving all residues that have been identified as participants in the interaction. For docking the protonation states were set as default (*"authohis = true"*). The Z-positioning restraints were also set to default as experimental restraints. The surface contact restraint was set as *"surfrest = true"* while the dihedral angles were also set as default. The top-scoring complexes based on the HADDOCK docking score and Z-scores were retrieved analyzed and subjected to interactions and subsequent analysis (28). The residues Glu27, Asp28, Tyr92, Thr94, and Ala96 were selected as the interface residues for the heavy and light chain of CA33 while the residues Lys9, Leu10, Val11, Lys37, and Glu129 were selected as the active residues for aP2 interaction.

2.4 Determination of the binding strength through dissociation constant prediction

The dissociation constant is an essential aspect of determining the pharmacological potential of antigen-antibodies complexes modeling and the results provide essential insights into the impact of a particular mutation on the recognition and binding. We used PRODIGY, a contact-based predictor, for modeling the binding strength of the native and mutated CA33 antibody with aP2 (29). The Prodigy server is the most widely and highly accurate server used for predicting the dissociation constant of a macromolecular complex. The server uses the interatomic contacts with 5.5Å and combines them with the non-interacting surface (NIS) to derive essential knowledge regarding the binding strength of K_D.

2.5 All-atoms molecular simulation and analysis

We assessed the dynamic characteristics of the wild-type, T94M, T94W, A96Q, and A96E complexes in conjunction with E4R using the AMBER 21 software. To prepare the system, we employed the "tleap" module from AmberTools to generate topology and coordinate files. Missing atoms and hydrogens were added via the LEaP builder tool. To achieve charge

neutrality, we introduced counterions using the AddToBox module, and for solvation, we incorporated an optimal point charge (OPC) model of the water box using the SolvateBox module. Initially, we conducted an energy minimization of the system, employing both the steepest descent and conjugate gradient algorithms. This minimization process ran for 10,000 and 8,000 steps or until the energy change became less than 0.1 kcal/mol. Subsequently, we subjected the system to a 10 ns equilibration period. During the initial 100 ps of equilibration, we applied Langevin dynamics with a collision frequency of 1.0 ps^{-1} to raise the system's temperature from 0K to 300 K. Following this, we maintained a constant pressure of 1 atm using the Parrinello-Rahman barostat for 1 ns. This was succeeded by sustaining a constant temperature of 300K through Langevin dynamics for an additional 1 ns. Finally, a 7 ns equilibration simulation was performed utilizing an NPT ensemble with PME electrostatics and a non-bonded cutoff of 10 Å. After achieving equilibration, we conducted a 300 ns production simulation under the same parameters used during equilibration. To accelerate the simulation, we employed PMEMD.CUDA and saved the coordinates every 10 ps for subsequent analysis.

2.6 Essential dynamics

To understand the dynamics variation and atomic motion of the whole trajectories the similar conformations were clustered and presented as Principal components by using the principal component analysis approach (30). This approach clusters the simulation trajectories and has been widely used in large-scale data analysis. To further understand the stable and metastable states the two principal components i.e., PC1 and PC2 were used to determine the free energy landscape (FEL). It has been widely used to determine the lowest conformational state and variations as compared to the native conformation. For this purpose, CPPTRAJ was used and the *g_sham* module of Gromacs was used for the PC's construction.

2.7 Calculation of binding free energies

The strength of a protein interacting with its biologically significant ligand/protein, or a small inhibitor significantly impacts the drug discovery and understanding of protein coupling mechanisms (31, 32). For protein-protein and protein-ligand complexes, this property is frequently represented by the binding free energies (BFE). In this work, it is calculated as the difference between the free energies of the bound aP2-CA33 complex ($G_{complex,\ solvated}$) and the unbound states of aP2 ($G_{aP2,\ solvated}$) and CA33 ($G_{CA33,\ solvated}$), as shown in equation (i). For each complex, the hydrogen bonding and distances with energetic contribution were calculated from a relaxed structure. The following equation was used to calculate each term:

$$\Delta G_{bind} = G_{(complex,\ solvated)} - G_{(aP2,\ solvated)} - G_{(CA33,\ solvated)} \quad \text{(i)}$$

This equation can be used to determine the contribution of interaction in the complex and can be expressed as equation (ii):

$$G = E_{Molecular\ Mechanics} - G_{solvated} - TS \quad \text{(ii)}$$

This equation can be further restructured to calculate the specific energy term.

$$\Delta G_{bind} = \Delta E_{Molecular\ Mechanics} + \Delta G_{solvated} - \Delta TS$$
$$= \Delta G_{vaccum} + \Delta G_{solvated} \quad \text{(iii)}$$

$$\Delta E_{Molecular\ Mechanics} = \Delta E_{int} + \Delta E_{electrostatic} + \Delta E_{vdW} \quad \text{(iv)}$$

$$\Delta G_{solvated} = \Delta G_{Generalized\ born} + \Delta G_{surface\ area} \quad \text{(v)}$$

$$\Delta G_{surface\ area} = \gamma . SASA + b \quad \text{(vi)}$$

$$\Delta G_{vaccum} = \Delta E_{Molecular\ Mechanics} - T\Delta S \quad \text{(vii)}$$

Specifically, we represent the free energy associated with the total binding of proteins as ΔG_{bind} (iii, v, vii). This encompasses the cumulative gas phase energy, which consists of $\Delta E_{internal}$, $\Delta E_{electrostatic}$, and ΔE_{vdW}, and is denoted as ΔE_{MM} (iv). The combined contributions from the polar (ΔGPB/GB) and nonpolar (ΔG_{SA}) components of solvation are expressed as ΔG_{sol} (v). The conformational binding entropy, typically evaluated through normal-mode analysis, is denoted as -TΔS. The internal energy, resulting from various bonds, angles, and dihedrals in the molecular mechanics (MM) force field, is encapsulated in $\Delta E_{internal}$. Notably, in calculations involving MM/PBSA and MM/GBSA, this value remains consistently zero, as observed in the single trajectory of a complex calculation. $\Delta E_{electrostatic}$ and ΔE_{vdW} represent the electrostatic and van der Waals energies, respectively, computed using MM. Meanwhile, ΔGPB/GB signifies the polar contribution to the solvation-free energy, computed employing Poisson–Boltzmann (PB) or generalized Born (GB) methods. Lastly, ΔG_{SA} quantifies the nonpolar solvation-free energy, usually determined using a linear function based on solvent-accessible surface area (SASA) (vi). It's worth noting that the calculation of conformational entropy is often omitted due to its computational expense and susceptibility to inaccuracies.

3 Results and discussion

3.1 CA33 mutants prediction and docking with aP2

Structural engineering of a protein has always been a great tool to increase the binding affinity and specificity for therapeutic purposes. Using graph-based signatures we generated the structural mutant of the L chain of the CA33 antibody. The complex as depicted in Figures 1A, B (cartoon and surface presentation) shows the binding of aP2 with the L and H chains of CA33. It was observed that the L chain only interacts directly with the binding residues of aP2 while the H chain comes in indirect

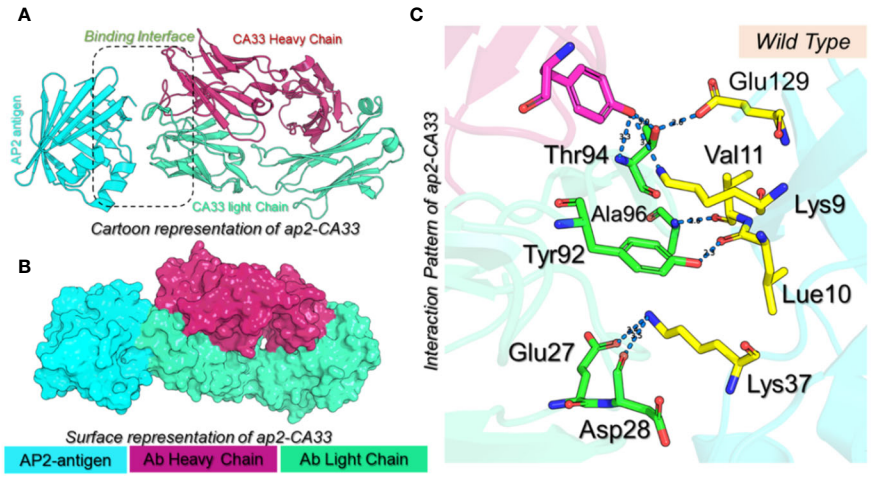


FIGURE 1
(A) Cartoon presentation of the aP2-CA33 complex. The aP2 antigen is shown in cyan color, the heavy chain of CA33 is shown in pomegranate color while the light chain is given in light green color. **(B)** shows the surface representation of the aP2-CA33 complex. **(C)** represents the interaction pattern for the aP2-CA33 complex, where the green color represents the L chain, magenta represents the H chain and the yellow represents aP2. The hydrogen bonding interactions are given in blue dashes with the bonding distances.

contact with aP2 through non-bonded contacts. Before modeling the novel mutants, we analyzed the binding interface which revealed that the residues Lys9, Leu10, Val11, Thr56, Glu129, and Lys37 are involved in interaction with the aP2. Among these six hydrogen bonds were formed by Lys9-Thr94 (2.72Å), Lys9-Thr94 (3.26Å), Leu10-Tyr92 (2.26Å), Val11-Asp28 (2.89Å) and Glu129-Thr94 (2.55Å). The only salt bridge was reported between Lys37-Glu27 with a bonding distance of 3.41 Å. Considering this interaction paradigm we mutated the selected residues in the L chain of the antibody. We observed that mutating Tyr92 and Asp28 abolish the interactions while the others Glu27, Thr94, and Ala96 are non-essential contacts and favorable for substitutions that could result in higher binding affinity than the native complex. Among the predicted mutants 30 mutants were predicted to increase the binding affinity while the rest were predicted to decrease the binding affinity. We set a threshold of Predicted ΔΔG>1 that will be considered while the others should be considered as non-essential substitutions. Using this criterion, Thr94Met was observed to increase the binding affinity with the predicted ΔΔG of 1.24 to be the highest among all. The Ala96Gln replacement reported an affinity change in the predicted ΔΔG of 1.09 while the Ala96Thr, and Ala96Ile, Ala96Glu reported ΔΔG of 1.035, 1.202 and 1.02 respectively. The Thr94Trp substitution reported ΔΔG of 1.022 respectively. These top-scoring mutants were generated by using Chimera software and subjected to aP2-antibody docking using HADDOCK. The interaction pattern for the wild-type CA33 and aP2 is illustrated in Figure 1C while the predicted affinity change for top residues with RSA (accessible surface area) is provided in Figures 2A, B. The predicted Ramachandran plot (Clash Score, Ramachandran Favored/Outliers, rotamer Outliers) for dihedral angle analysis, and MolProbity Scores are summarized in Supplementary Table S1.

Next, we generated the mutants (Figure 2) that increase the binding affinity and modeled by using Modeler software embedded Chimera tool. To obtain the docking scores for the wild-type we submitted the native complex to the HADDOCK server and used a

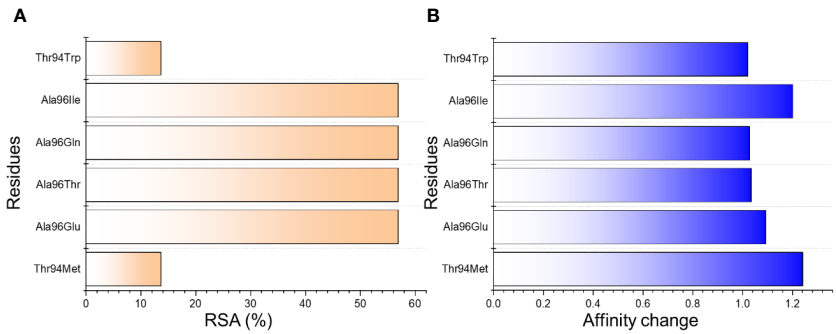


FIGURE 2
The predicted top mutants increase the binding affinity upon the substitution. **(A)** shows the relative surface area change in percent while **(B)** shows the affinity change due to each substitution.

10.3389/fimmu.2024.1357342

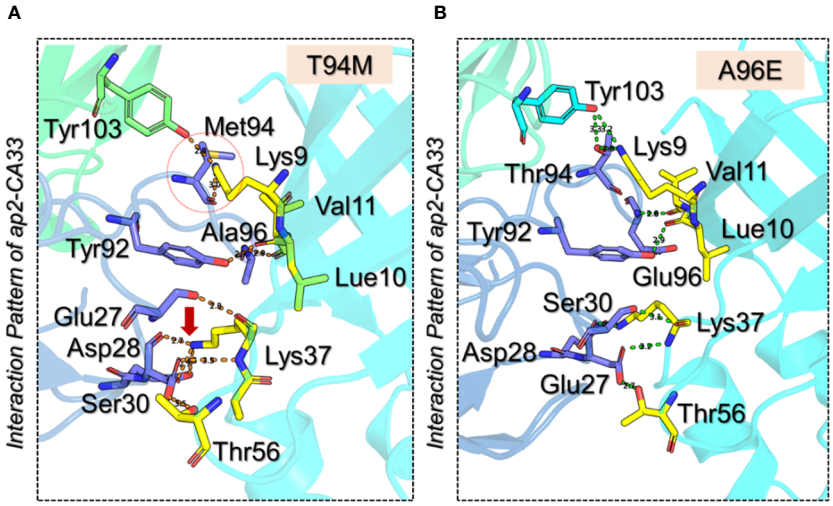


FIGURE 3
3D interaction paradigm for the T94M and A96E mutants in complex with aP2. **(A)** represent the interaction pattern of T94M with aP2. In this panel, the yellow sticks represent aP2, the blue sticks represent the L chain, and the green stick represents the H chain. **(B)** represents the interaction pattern of A96E with aP2. In this panel, the yellow sticks represent aP2, the blue sticks represent the L chain, and the cyan stick represents the H chain.

refinement option to get the results for the wild-type and use as a comparison for the further mutant's selection. The HADDOCK server predicted the docking score for the wild-type of -364.90 ± 3.0 kcal/mol with the vdW (Van Der Waals) score of -184.70 ± 4.0 kcal/mol and the electrostatic energy of -498.00 ± 28.2 kcal/mol. The other parameters are provided in Table 1. Considering the total docking score of the wild-type (-364.90 ± 3.0 kcal/mol) the top-scoring mutants were selected based on this threshold. Among the selected mutants the two i.e., Ala96Leu reported a docking score of -363.50 ± 2.0 kcal/mol and Ala96Thr reported a docking score of -361.70 ± 5.3 kcal/mol which is a higher than the control (wild-type) and were excluded from the further analysis. The mutant Thr94Met predicted the best docking score among all. The docking score for the Thr94Met was calculated to be -372.00 ± 3.7 kcal/mol with ten hydrogen bonds and 2 salt bridges. A total of 51 non-bonded contacts were reported in this complex. In this complex Tyr103 established a hydrogen bond with Lys9 (2.8 Å) from the H chain while the L chain established the remaining nine hydrogen bonding contacts. Among these Glu27-Lys37 (2.8 Å), Asp28-Lys37 (2.7 Å), Ser30-Lys37 (3.5 Å), Ser30-Thr56 (3.5 Å), Tyr92-Leu10 (2.7 Å), Met94-Lys9 (2.99 Å), Ala96-Leu10 (2.9 Å) and Ala96-Val11 (2.9 Å) respectively. The only salt bridge was established between Lys37-Glu27 with a bonding distance of 2.70 Å. Interestingly the mutated residues Met94 directly interact with the aP2 and additional contacts have been established such as Ser30 interaction with Lys37 and Thr56. The interaction paradigm for the Thr94Met is shown in Figure 3A. For this complex the vdW was estimated to be -194.40 ± 6.1 kcal/mol while the electrostatic energy was calculated to be -459.7 ± 18.1 kcal/mol. In contrast to the native complex, this mutant presented a better vdW energy that particularly contributed to the robust binding of this mutant than the wild-type. On the other hand, the Ala96Glu with a docking score of -371.4 ± 1.9 was ranked as the second-best mutant that has a lower docking score than the wild-type. The rationale behind the increase in the docking is that this complex involved the highest number of non-bonded contacts with additional hydrogen bonds and the conserved salt bridge. The hydrogen bonding paradigm reported eight hydrogen bonds Lys9-Thr94 (2.83 Å), Leu10-Tyr92 (2.93 Å), Val11-Glu96 (2.79 Å), Lys37-Asp28 (3.17 Å), Lys37-Glu27 (2.85 Å), Lys37-Asp28 (2.84 Å) and Thr56-Asp28 (2.75 Å) respectively. The only salt bridge was established between Lys37-Glu27 with a bonding distance of 2.80 Å. Additionally, a hydrogen bond was also reported between the heavy chain Tyr103 and Lys9 residue of aP2. These additional hydrogen bonding contacts consequently increase the binding and neutralization of aP2 antigen through the recognition of essential immune epitopes. The vdW and electrostatic energies for this complex were calculated to be -192.30 ± 4.7 and -471.10 ± 17.8 kcal/mol respectively which are lower than the control native aP2-CA33 complex thus inducing stronger binding and neutralization. The interaction paradigm for the Ala96Glu is shown in Figure 3B. The docking scores and other parameters for these mutants are provided in Table 1.

We further evaluated the binding patterns of T94W and A96Q mutants with aP2. The T94W with the docking score of -366.1 ± 2.0 kcal/mol reported eight hydrogen bonds involving Glu27-Lys37 (2.7 Å), Asp28-Lys37 (2.7 Å), Asp28-Thr56 (3.3 Å), Ser30-Lys37 (3.1 Å), Tyr92-Leu10 (2.8 Å), Tyr103-Lys9 (2.8 Å) and Tyr103-Glu129 (3.4 Å) respectively. Interestingly the heavy chain established two direct hydrogen bonds with the two residues of aP2 thus showing differential binding of this mutant. Moreover, the Ala96 interaction with Val11 was observed to be demolished while the extra contacts by the Ser30 can be seen in the complex. The Lys37-Glu27 (2.74 Å) salt bridge remained conserved here too. The vdW energy for this complex was observed to be -192.9 ± 4.4 kcal/mol while the electrostatic energy was -437.0 ± 26.8 kcal/mol respectively. The interaction pattern of T94W is shown in Figure 4A. On the other

10.3389/fimmu.2024.1357342

TABLE 1 The predicted docking score for each substitution using HADDOCK. The bonding residues and distances for each complex.

Parameters	Wild-type-aP2	T94M-aP2	A96E-aP2	A96Q-aP2	T94W-aP2	A96L-aP2	A96T-aP2
HADDOCK score	-364.9 ± 3.0	-372.0 ± 3.7	-371.4 ± 1.9	-369.2 ± 2.3	-366.1 ± 2.0	-363.5 ± 2.0	-361.7 ± 5.3
Cluster size	20	20	20	20	20	20	20
RMSd from the overall lowest-energy structure	0.5 ± 0.3	0.5 ± 0.3	0.5 ± 0.3	0.6 ± 0.3	0.6 ± 0.3	0.5 ± 0.3	0.5 ± 0.3
Van der Waals energy	-184.7 ± 4.0	-194.4 ± 6.1	-192.3 ± 4.7	-190.9 ± 1.5	-192.9 ± 4.4	-189.3 ± 4.5	-185.7 ± 5.0
Electrostatic energy	-498.0 ± 28.2	-459.7 ± 18.1	-471.1 ± 17.8	472.8 ± 16.5	-437.0 ± 26.8	-432.3 ± 22.9	-495.6 ± 25.5
Desolvation energy	-80.7 ± 3.6	-85.6 ± 1.1	-84.9 ± 2.8	-83.8 ± 1.5	-85.8 ± 3.6	-87.8 ± 1.4	-76.9 ± 4.4
Restraint's violation energy	0.2 ± 0.1	0.1 ± 0.1	0.2 ± 0.1	0.3 ± 0.3	0.2 ± 0.2	0.2 ± 0.1	0.3 ± 0.2
Buried Surface Area	4748.5 ± 44.2	4694.9 ± 49.8	4776.0 ± 48.2	4774.2 ± 67.2	4686.1 ± 32.8	4722.5 ± 62.5	4668.1 ± 76.5
Z-Score	0	0	0	0	0	0	0
Dissociation constant (K_D)	$1.2E^{-8}$	$0.9E^{-10}$	$1.1E^{-9}$	$1.1E^{-9}$	$1.2E^{-6}$	–	–
Hydrogen Bonds	Lys9-Thr94 (2.72Å), Lys9-Thr94 (3.26Å), Leu10-Tyr92 (2.26Å), Val11-Asp28 (2.89Å) and Glu129-Thr94 (2.55Å)	Glu27-Lys37 (2.8 Å), Asp28-Lys37 (2.7 Å), Ser30-Lys37 (3.5 Å), Ser30-Thr56 (3.5 Å), Tyr92-Leu10 (2.7 Å), Met94-Lys9 (2.99 Å), Ala96-Leu10 (2.9 Å) and Ala96-Val11 (2.9 Å)	Lys9-Thr94 (2.83 Å), Leu10-Tyr92 (2.93 Å), Val11-Glu96 (2.79 Å), Lys37-Asp28 (3.17 Å), Lys37-Glu27 (2.85 Å), Lys37-Asp28 (2.84 Å) and Thr56-Asp28 (2.75 Å)	Lys9-Thr94 (3.10 Å), Leu10-Tyr92 (3.23 Å), Val11-Gln96 (2.97 Å), Lys37-Asp28 (3.30 Å), Lys37-Asp28 (2.68 Å), Thr56-Asp28 (2.71 Å), Lys37-Asp28 (3.00 Å), Glu129-Thr94 (2.69 Å), and Glu129-Tyr103 (3.10 Å)	Glu27-Lys37 (2.7 Å), Asp28-Lys37 (2.7 Å), Asp28-Thr56 (3.3 Å), Ser30-Lys37 (3.1 Å), Tyr92-Leu10 (2.8 Å), Tyr103-Lys9 (2.8 Å) and Tyr103-Glu129 (3.4 Å)	–	–
Salt bridges	Lys37-Glu27 (3.41 Å)	Lys37-Glu27 (3.41 Å)	Lys37-Glu27 (2.80 Å)	Lys37-Glu27 (2.74 Å)	Lys37-Glu27 (2.74 Å)	–	–

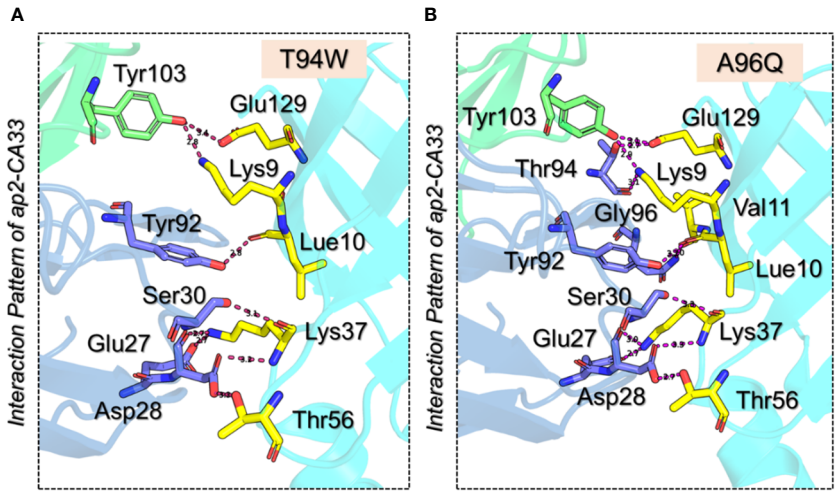


FIGURE 4
3D interaction paradigm for the T94W and A96Q mutants in complex with aP2. **(A)** represent the interaction pattern of T94W with aP2. In this panel, the yellow sticks represent aP2, the blue sticks represent the L chain, and the green sticks represent the H chain. **(B)** represents the interaction pattern of A96Q with aP2. In this panel, the yellow sticks represent aP2, the blue sticks represent the L chain, and the green sticks represent the H chain.

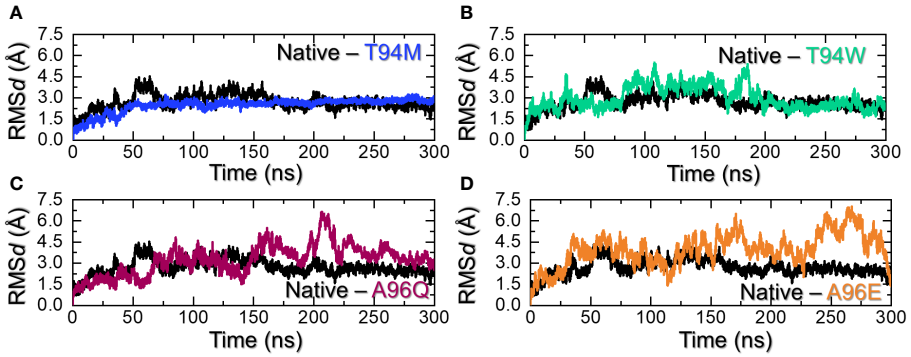


FIGURE 5
Dynamic stability assessment of the wild-type and mutants. **(A)** shows the RMS*d* graphs for the wild-type and T94M, **(B)** shows the RMS*d* graphs for the wild-type and T94W, **(C)** shows the RMS*d* graphs for the wild-type and A96Q while **(D)** shows the RMS*d* graphs for the wild-type and A96E.

hand, the Ala96Gln reported a docking score of -369.2 ± 2.3 kcal/mol, vdW of -190.9 ± 1.5, and electrostatic energy of -472.8 ± 16.5 kcal/mol respectively. Investigation of the binding pattern revealed ten hydrogen bonds among which 2 were established by the H chain and the remaining 8 by the L chain. The other differences include the direct interaction of the H chain with the aP2. Among the hydrogen bonds Lys9-Thr94 (3.10 Å), Leu10-Tyr92 (3.23 Å), Val11-Gln96 (2.97 Å), Lys37-Asp28 (3.30 Å), Lys37-Asp28 (2.68 Å), Thr56-Asp28 (2.71 Å), Lys37-Asp28 (3.00 Å), Glu129-Thr94 (2.69 Å), and Glu129-Tyr103 (3.10 Å) respectively. The Lys37-Glu27 (2.74 Å) salt bridge remained conserved here too. The interaction pattern of A96Q is depicted in Figure 4B. The docking scores and other parameters for these mutants are summarized in Table 1. Overall, the current findings show that both the vdW and electrostatic energy terms are increased which consequently causes the robust binding of CA33 to the aP2. The current findings highlight the importance of protein engineering in the design of novel and effective therapeutics for the development of specific antibodies against T2DM.

3.2 Calculation of binding strength through K_D

The binding strength was further validated by using the dissociation constant calculation based on the AI-powered algorithm trained with the experimental data. The results demonstrated that the K_D value for the wild-type was $1.2E^{-8}$ while for the T94M, the K_D was estimated to be $0.9E^{-10}$. For the T94W the K_D was estimated to be $1.1E^{-9}$, for the A96Q the K_D was computed to be $1.1E^{-9}$ and for the A96E the K_D was computed to be $1.2E^{-6}$. This shows the higher binding strength for the mutants except A96E and therefore demonstrates a robust immune response by interacting with aP2.

3.3 Dynamic stability assessment of the wild-type and mutant complexes

Determining complex stability during simulation is an essential step towards the understanding of the pharmacological efficiency of a therapeutic molecule. It is considered as important for stable binding and therefore is necessary to estimate the system's stability. Considering the importance of dynamic stability, we calculated root mean square deviation (RMS*d*) as a function of time using the simulation trajectory. As shown in Figure 5A, the wild-type antibody stabilized at 3.0 Å at 75ns. The complex initially demonstrated a higher RMS*d* with minor deviations, it stabilized and maintained the same level until the end of the simulation. An average RMS*d* for the wild-type was calculated to be 2.74 Å. On the other hand, the T94M stabilized at 2.25 Å at 37ns. The complex reported no significant perturbation and the average RMS*d* for this complex was calculated to be 2.40 Å. This indicates that the introduction of this mutant causes structural stabilization and thus the binding is further stabilized. Hence, this mutation is more favorable for enhancing the binding and instigation of a stronger immune response against aP2. Moreover, the T94W mutant reported a comparatively destabilized behavior than the T9M but was more stable than the wild-type at the end of the simulation. The trajectory started from 0 and reached 4.3 Å at 40ns. The complex then exhibited a stable behavior but after reaching 75ns the RMS*d* increased again and maintained the same level till 175ns. An abrupt rise in RMS*d* at 180ns was followed by a subsequent decline. After 190ns, the complex attained stability and maintained a uniform level until the end of the simulation. An average RMS*d* for this complex was calculated to be 2.95 Å. The RMS*d* results for the T94W are shown in Figure 5B. Interestingly, the A96Q and A96E substitutions were found to show dynamically unstable behavior with a reported RMS*d* higher than the wild-type and T94M/W mutants. For instance, the RMS*d* pattern for the A96Q reported significant structural perturbations with a higher RMS*d* level of 6.2 Å. The structure started with 1.5 Å until 50ns and then an abrupt increase/decrease was experienced. An average RMS*d* for the A96Q complex was estimated to be 3.24 Å. The A96E complex was observed to be the most destabilized complex with a reported RMS*d* of 6.5 Å. With significant structural perturbation, this complex maintained a higher RMS*d* level than all the complexes, with an average RMS*d* (4.58 Å) being observed. The RMS*d* graphs for the A96Q are shown in Figure 5C while the RMS*d* graph for the A96E is depicted in Figure 5D. It can be observed that the T94M is the most stable substitution which increases the binding stability throughout the

simulation while the T94W also exhibited comparatively a dynamically stable behavior. The superimposed structures of each complex retrieved at different time intervals were further compared with the native to understand the structural variations. As shown in Supplementary Figure S1, it can be noted that the interface in all the complexes remains intact while the tail of the L chain folds and unfolds inward and outward to cause deviation from the native structure. Moreover, the flipping of the beta sheets in the aP2 also causes deviation from the native structure. This shows that the aP2-CA33 remains bound during simulation however the movement of some secondary structural elements causes the drift in the RMS*d* pattern. In sum, these two substitutions are more favorable for the enhanced and stabilized binding of the CA33 antibody than the A96E and A96Q and therefore should be further investigated for clinical purposes.

3.4 Structural compactness assessment

Calculation of the structural compactness by using a radius of gyration (R*g*) over the simulation time is an important parameter that determines the binding and unbinding events during the simulation. It is an essential step to determine the pharmacological potential of a therapeutic molecule. Considering the application of R*g* in determining structural stability and compactness, we also calculated Rg using the simulation trajectory. Interestingly, the R*g* results for the wild-type aligned with the RMS*d* results. The R*g* started from 29.80 Å and steadily decreased over time. The highest Rg value was observed at 70ns and then a continuous decline in the R*g* value was observed. An average Rg for the wild-type was calculated to be 29.85 Å. On the other hand, the R*g* for T94M mutant started at 30.0 Å and continued to decrease till 26.8Å at 50ns. The complex then reported a uniform straight graph for Rg values and no deviation was observed. This indicates that the complex maintained a rigid and stabilized compact structure and therefore had minimal unbinding events during the simulation. The R*g* results strongly align with the RMS*d* results, with stability maintained throughout the simulation. An average R*g* for the T94M complex was estimated to be 27.0 Å as shown in Figure 6A. The T94W initially reported a lower Rg (30.0 Å) behavior by keeping the R*g* at 30.0 Å up to 75ns. The R*g* then gradually increased and continued to report a similar behavior until 225ns. Like the RMS*d* results, the Rg also maintained a stable and lower level during the last part of the simulation. The increase in the Rg pattern determines the unwinding of the tail of the CA3 which causes a significant increase in the protein size. The R*g* for the T94W is shown in Figure 6B. Interestingly, the A96Q comparatively reported a stabilized protein size during the first 75ns and then gradually increased up to 32.0 Å. This R*g* level was maintained for the remaining simulation time showing the unwinding of the CA33 tail and then rewind. An average R*g* for the A96Q was calculated to be 31.5 Å (Figure 6C). The Rg results for the A96E also reported a similar behavior to the findings of RMS*d*. The R*g* remained higher than all the complexes. This complex maintained an Rg level of ~34.50 Å throughout the simulation. An average R*g* for the A96E was calculated to be 34.45Å (Figure 6D). Overall, these findings strongly corroborate with the RMS*d* and show that T94M and T94W are the most favorable that not only increase the binding but also increase the stability. Interestingly, the higher binding mutant remained the most compact avoiding the unbinding events while the three other substitutions i.e., T94W, A96Q, and A96E caused structural instability. Thus, substitutions that increase the structural stability increase the binding significantly.

3.5 Hydrogen bonding analysis

Hydrogen bonding calculation is one of the key assessments that help in determining the pharmacological potential of a drug or inhibitor. It is an essential approach to reveal the potency and binding strength of the interacting molecules. This approach has been widely applied to understand the pharmacological mechanism of a particular drug, and the interaction mechanism of two or more proteins to reveal the mechanism of a disease or bio-catalytic process (33–37). Considering the essential role of this approach, we used a similar approach to calculate the total number of hydrogen bonds in each complex. The average number of hydrogen bonds in each complex was calculated to be 231 in the wild-type, 236 in the T94M, 229 in the T94W, 232 in the A96Q, and 231 in A96E. It can be observed that the hydrogen bonds in the predicted mutants are more than the wild-type thus implying that

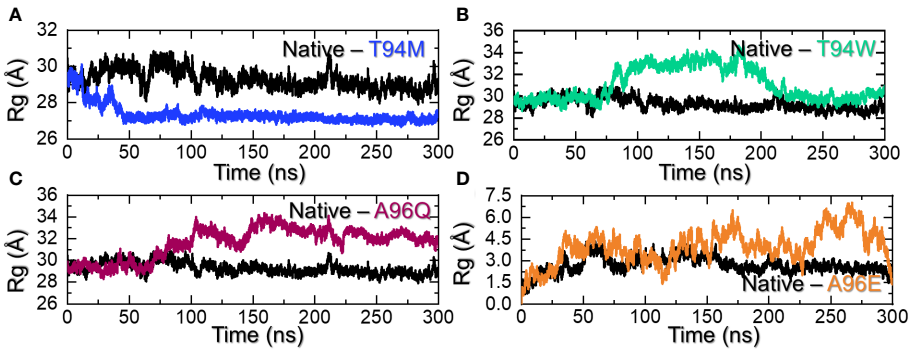


FIGURE 6
Structural compactness assessment of the wild-type and mutants. **(A)** shows the R*g* graphs for the wild-type and T94M, **(B)** shows the R*g* graphs for the wild-type and T94W, **(C)** shows the R*g* graphs for the wild-type and A96Q while **(D)** shows the R*g* graphs for the wild-type and A96E.

these mutants increase the binding. Although the number of bonds is increased in the three mutants T94M is the more favorable substitution that increases the binding stability with the number of hydrogen bonds. The hydrogen bonding results for all the complexes are presented in Figures 7A–D. Additional information about the hydrogen bonding, distances, and half-life information are summarized in Supplementary Table S2.

3.6 Root mean square fluctuation calculation

Residue fluctuation indexing is an essential factor in determining the role of particular residues in molecular recognition, protein inhibition, ligand recognition, and opening and closing switches. For instance, this approach has been widely used to determine the impact of different mutations on the binding and internal fluctuation of different receptors (38). Herein, we also calculated residual flexibility using the simulation trajectory. The RMSF results presented in Figure 8A demonstrate that the internal fluctuation of the aP2 has been stabilized and thus minimal fluctuations are produced by the wild-type and T94M complexes while the other complexes have produced higher fluctuations. The regions 35-225 and 230-335 determined major fluctuations in the T94W, A96Q, and A96E. We further dissected the RMSF profiles of each mutated residue in each complex. The results shown in Figure 8B indicate that the mutated residues demonstrated higher fluctuation than the wild-type and therefore result in better conformational optimization for enhanced binding. Interestingly, the RMSF results also corroborate with the binding results and indicate that wild-type and T94M are better immune response-provoking agents than the other mutants.

3.7 Principal component analysis for trajectories motions clustering

The analysis of data distribution within the component space yields valuable insights into the fundamental dynamics of the underlying system. Notably, both the wild-type and T94M had comparable patterns of constraint and restricted motion across each principal component. It further shows that these two systems are more stable and controlled in these dimensions. The conformational space is divided into two states i.e., the pink color which is separated by the purple color (transition state) from the blue color. On the other hand, the T94W, A96Q, and A96E determined differential trajectories clustering and therefore presented an unstable state for each complex. This indicates that mutant T94M behaves more like the wild-type but presents favorable variations that cause more robust binding of T94M than the control. These findings also corroborate the residues' flexibility and docking results. The PCA graphs are presented in Figures 9A–E.

3.8 Free energy landscape analysis

In the context of molecular mechanics and simulation, the free energy landscape is used to understand and visualize the energy landscape of each system. It provides a visual presentation of the relationship between the potential energy and its collective variables. It determines the possible lowest energy configuration state and determines the protein folding. All the complexes presented a single metastable (lowest energy state) during the simulation which indicates that the system does not readily transit through multiple conformations. This demonstrates limited structural variability and underscores the therapeutic antibody's efficacy against aP2. The FEL graphs are presented in Figures 10A–E.

3.9 Binding free energy analysis

We calculated the binding free energy for each complex which revealed that vdW values of -160.82 kcal/mol, -173.49 kcal/mol, -165.69 kcal/mol, -170.83 kcal/mol, -168.67 kcal/mol were calculated for wild-type, and T94M, A96Q, and A96E mutants, respectively. This indicates that the rise in the number of hydrogen bonds leads to a corresponding increase in the vdW energy within

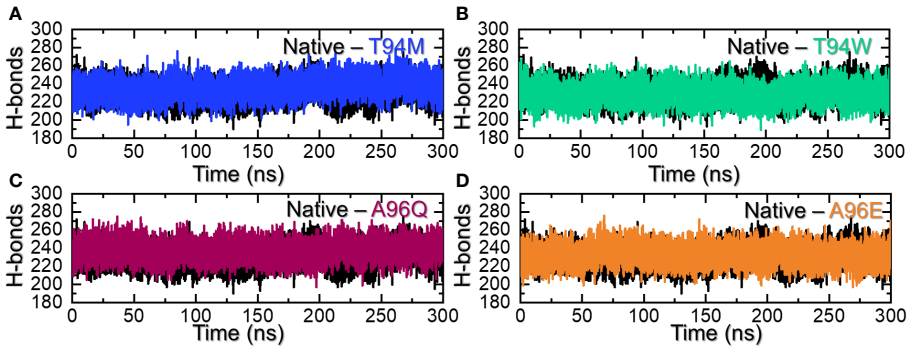


FIGURE 7
Hydrogen bonding analysis of the wild-type and mutants. **(A)** shows the H-bonds graphs for the wild-type and T94M, **(B)** shows the H-bonds graphs for the wild-type and T94W, **(C)** shows the H-bonds graphs for the wild-type and A96Q while **(D)** shows the H-bonds graphs for the wild-type and A96E.

10.3389/fimmu.2024.1357342

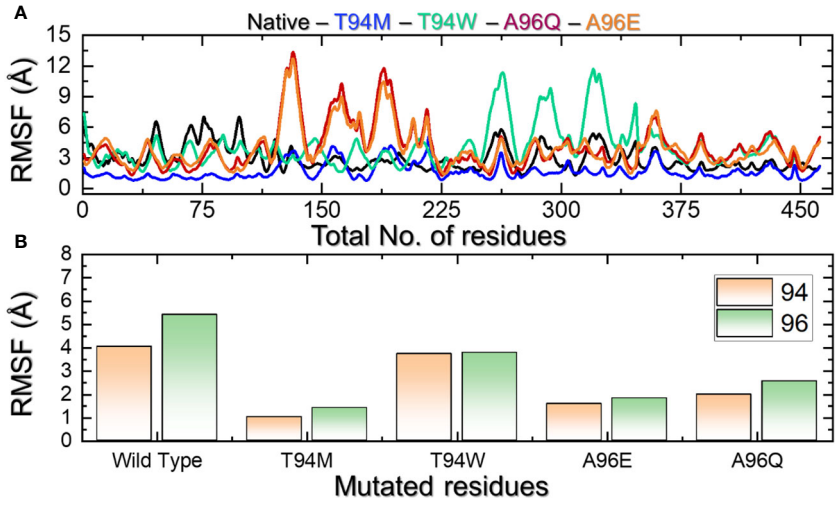


FIGURE 8
(A) Residue's flexibility analysis of the wild-type and mutants. All the complexes are differently colored. **(B)** shows the RMSF pattern for the mutated residues in each complex.

each complex, causing the binding affinity to strengthen. On the other hand, the electrostatic energy calculations showed Elec values of -20.36 kcal/mol, -19.27 kcal/mol, -18.39 kcal/mol, -19.35 kcal/mol, -18.48 kcal/mol for wild-type, T94M, T94W, A96Q, and A96E mutant, respectively. To provide conclusive evidence on the role of the introduced mutations and their impact on the binding, we calculated the total binding free energy for each complex to accurately evaluate the binding strength of each complex. The results strongly corroborate with the docking scores and dissociation constant (K_D) results. The TBE for the wild-type was computed to be -279.84 kcal/mol, for the T94M the highest binding free energy was estimated to be -295.22 kcal/mol. For the T94W, the binding free energy was computed to be -281.67 kcal/mol, and for the A96Q the TBE was -289.44 kcal/mol while for the A96E the TBE was estimated to be -277.29 kcal/mol. This shows that the predicted substitutions strongly corroborate with the hypothesis of affinity-increasing mutants that consequently cause an enhanced binding of the CA33-engineered antibody to the aP2 antigen. The binding free energy results for each complex are shown in Figure 11. The specific energy contribution is summarized in Supplementary Table S2.

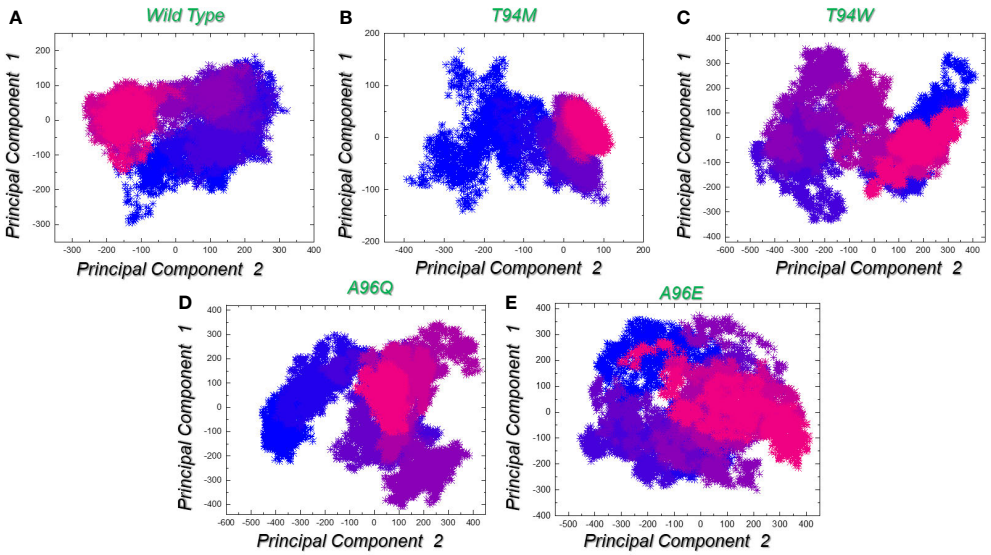


FIGURE 9
Trajectories clustering and motion using principal component analysis (PCA). **(A)** represents the trajectory distribution for the wild-type complex in X and Y dimensions given as PC1 and PC2. **(B)** represents the trajectory distribution for the T94M complex in X and Y dimensions given as PC1 and PC2. **(C)** represents the trajectory distribution for the T94W complex in X and Y dimensions given as PC1 and PC2. **(D)** represents the trajectory distribution for the A96Q complex in X and Y dimensions given as PC1 and PC2. **(E)** represents the trajectory distribution for the A96E complex in X and Y dimensions given as PC1 and PC2.

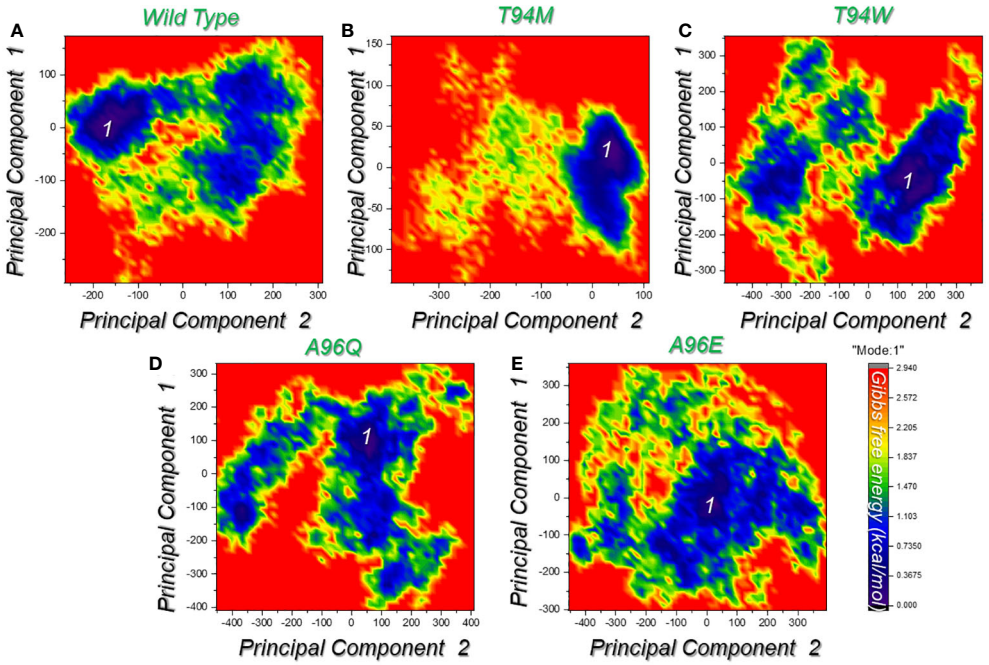


FIGURE 10
Free energy landscape (FEL) analysis of the wild-type and the designed mutated antibodies. **(A)** represents the FEL for the wild-type complex in X and Y dimensions given as PC1 and PC2. **(B)** represents the FEL for the T94M complex in X and Y dimensions given as PC1 and PC2. **(C)** represents the FEL for the T94W complex in X and Y dimensions given as PC1 and PC2. **(D)** represents the FEL for the A96Q complex in X and Y dimensions given as PC1 and PC2. **(E)** represents the FEL for the A96E complex in X and Y dimensions given as PC1 and PC2. Each graph represents the only conformational state attained by each complex.

4 Conclusions

The current study utilized structure-guided engineering strategies to enhance the CA33 antibody, leveraging graph-signature-based algorithms for rationale antibody design. The mutational landscape was subjected to a thorough examination, which revealed the presence of only four substitutions that were found to be significant. These alterations include T94M, T94W, A96Q, and A96E. Additional validation was conducted using post-prediction molecular simulations, which confirmed that the T94M substitution was the most favorable. Significantly, this change not only enhanced the docking score but also demonstrated exceptional stability throughout the simulation. To bolster the robustness of our results, we employed K_D estimates to quantify the binding affinity, introducing an additional level of validation to our investigation.

Future directions for this research involve investigating similar antibodies and exploring diverse diabetes-related biotargets. Analyzing additional antibodies using similar structurally guided engineering approaches promises a more thorough understanding of potential improvements. Expanding the study with a broader range of mutations and rigorous experimental validation can address the limitations and enhance the robustness of the findings. A comprehensive exploration of various diabetes-related biotargets will contribute to a holistic approach to antibody design. Although the findings of this study have the potential to offer significant insights into the strategic design of diabetes-targeting antibodies, collaborative efforts with experimentalists for *in vitro* and *in vivo* validations are anticipated, paving the way for the translation of these insights into clinical trials and practical applications.

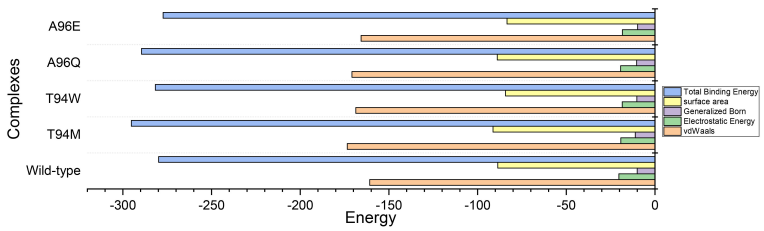


FIGURE 11
Total binding free energy results for each complex using the MM-GBSA approach. All the energies are given in kcal/mol.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding author.

Author contributions

AK: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. MZ: Conceptualization, Data curation, Investigation, Methodology, Visualization, Writing – original draft. AM: Conceptualization, Data curation, Formal Analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. AA: Conceptualization, Data curation, Formal Analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by Qatar University grant No. QUPD-CPH-23/24-592. M.A.Z is supported by a graduate assistantship from the Office of Graduate Studies of Qatar University. The statements made herein are solely the responsibility of the authors. Open Access funding is provided by the Qatar National Library.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2024.1357342/full#supplementary-material

References

1. Bastaki S. Diabetes mellitus and its treatment. *Dubai Diabetes And Endocrinol J.* (2005) 13:111–34. doi: 10.1159/000497580

2. Kumar R, Saha P, Kumar Y, Sahana S, Dubey A, Prakash O. A review on diabetes mellitus: type1 & Type2. *World J Pharm Pharm Sci.* (2020) 9:838–50.

3. Zimmet P, Alberti KG, Magliano DJ, Bennett PH. Diabetes mellitus statistics on prevalence and mortality: facts and fallacies. *Nat Rev Endocrinol.* (2016) 12:616–22. doi: 10.1038/nrendo.2016.105

4. Lovic D, Piperidou A, Zografou I, Grassos H, Pittaras A, Manolis A. The growing epidemic of diabetes mellitus. *Curr Vasc Pharmacol.* (2020) 18:104–9. doi: 10.2174/1570161117666190405165911

5. Harding JL, Pavkov ME, Magliano DJ, Shaw JE, Gregg EW. Global trends in diabetes complications: a review of current evidence. *Diabetologia.* (2019) 62:3–16. doi: 10.1007/s00125-018-4711-2

6. Chadt A, Al-Hasani H. Glucose transporters in adipose tissue, liver, and skeletal muscle in metabolic health and disease. *Pflügers Archiv-European J Physiol.* (2020) 472:1273–98. doi: 10.1007/s00424-020-02417-x

7. Ertunc ME, Sikkeland J, Fenaroli F, Griffiths G, Daniels MP, Cao H, et al. Secretion of fatty acid binding protein aP2 from adipocytes through a nonclassical pathway in response to adipocyte lipase activity. *J Lipid Res.* (2015) 56:423–34. doi: 10.1194/jlr.M055798

8. Dahlström EH, Saksi J, Forsblom C, Uglebjerg N, Mars N, Thorn LM, et al. The low-expression variant of FABP4 is associated with cardiovascular disease in type 1 diabetes. *Diabetes.* (2021) 70:2391–401. doi: 10.2337/db21-0056

9. Wu LE, Samocha-Bonet D, Whitworth PT, Fazakerley DJ, Turner N, Biden TJ, et al. Identification of fatty acid binding protein 4 as an adipokine that regulates insulin secretion during obesity. *Mol Metab.* (2014) 3:465–73. doi: 10.1016/j.molmet.2014.02.005

10. Furuhashi M, Tuncman G, Görgün CZ, Makowski L, Atsumi G, Vaillancourt E, et al. Treatment of diabetes and atherosclerosis by inhibiting fatty-acid-binding protein aP2. *Nature.* (2007) 447:959–65. doi: 10.1038/nature05844

11. Furuhashi M, Hotamisligil GS. Fatty acid-binding proteins: role in metabolic diseases and potential as drug targets. *Nat Rev Drug Discovery.* (2008) 7:489–503. doi: 10.1038/nrd2589

12. Crunkhorn S. Targeting aP2 reverses diabetes. *Nat Rev Drug Discovery.* (2016) 15:86–6. doi: 10.1038/nrd.2016.4

13. Tuncman G, Erbay E, Hom X, De Vivo I, Campos H, Rimm EB, et al. A genetic variant at the fatty acid-binding protein aP2 locus reduces the risk for hypertriglyceridemia, type 2 diabetes, and cardiovascular disease. *Proc Natl Acad Sci.* (2006) 103:6970–5. doi: 10.1073/pnas.0602178103

14. Yang Q, Graham TE, Mody N, Preitner F, Peroni OD, Zabolotny JM, et al. Serum retinol binding protein 4 contributes to insulin resistance in obesity and type 2 diabetes. *Nature.* (2005) 436:356–62. doi: 10.1038/nature03711

15. Cao H, Sekiya M, Ertunc ME, Burak MF, Mayers JR, White A, et al. Adipocyte lipid chaperone AP2 is a secreted adipokine regulating hepatic glucose production. *Cell Metab.* (2013) 17:768–78. doi: 10.1016/j.cmet.2013.04.012

16. Burak MF, Inouye KE, White A, Lee A, Tuncman G, Calay ES, et al. Development of a therapeutic monoclonal antibody that targets secreted fatty acid–binding protein aP2 to treat type 2 diabetes. *Sci Trans Med.* (2015) 7:319ra205–319ra205. doi: 10.1126/scitranslmed.aac6336

17. Khan A, Randhawa AW, Balouch AR, Mukhtar N, Sayaf AM, Suleman M, et al. Blocking key mutated hotspot residues in the RBD of the omicron variant (B. 1.1. 529) with medicinal compounds to disrupt the RBD-hACE2 complex using molecular screening and simulation approaches. *RSC Adv.* (2022) 12:7318–27. doi: 10.1039/D2RA00277A

18. Humayun F, Kumar V, Dhankhar P, Dalal V. Abrogation of SARS-CoV-2 interaction with host (NRP1) Neuropilin-1 receptor through high-affinity marine natural compounds to curtail the infectivity: A structural-dynamics data. *Comput Biol Med.* (2022) 141:104714. doi: 10.1016/j.compbiomed.2021.104714

19. Dalal V, Golemi-Kotra D, Kumar P. Quantum mechanics/molecular mechanics studies on the catalytic mechanism of a novel esterase (FmtA) of Staphylococcus aureus. *J Chem Inf Modeling.* (2022) 62:2409–20. doi: 10.1021/acs.jcim.2c00057

20. Kumari R, Kumar V, Dhankhar P, Dalal V. Promising antivirals for PLpro of SARS-CoV-2 using virtual screening, molecular docking, dynamics, and MMPBSA. *J Biomolecular Structure Dynamics*. (2023) 41:4650–66. doi: 10.1080/07391102.2022.2071340

21. Celik I, Khan A, Dwivany FM, Fatimawali, Wei D-Q, Tallei TE. Computational prediction of the effect of mutations in the receptor-binding domain on the interaction between SARS-CoV-2 and human ACE2. *Mol Diversity*. (2022) 26:3309–24. doi: 10.1007/s11030-022-10392-x

22. Khan A, Rehman Z, Hashmi HF, Khan AA, Junaid M, Sayaf AM, et al. An integrated systems biology and network-based approaches to identify novel biomarkers in breast cancer cell lines using gene expression data. *Interdiscip Sciences: Comput Life Sci*. (2020) 12:155–68. doi: 10.1007/s12539-020-00360-0

23. Burley SK, Bhikadiya C, Bi C, Bittrich S, Chen L, Crichlow GV, et al. RCSB Protein Data Bank: powerful new tools for exploring 3D structures of biological macromolecules for basic and applied research and education in fundamental biology, biomedicine, biotechnology, bioengineering and energy sciences. *Nucleic Acids Res*. (2021) 49:D437–51. doi: 10.1093/nar/gkaa1038

24. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J Comput Chem*. (2004) 25:1605–12. doi: 10.1002/jcc.20084

25. Yuan S, Chan HS, Hu Z. Using PyMOL as a platform for computational drug design. *Wiley Interdiscip Reviews: Comput Mol Sci*. (2017) 7:e1298. doi: 10.1002/wcms.1298

26. Laskowski RA. PDBsum: summaries and analyses of PDB structures. *Nucleic Acids Res*. (2001) 29:221–2. doi: 10.1093/nar/29.1.221

27. Myung Y, Rodrigues CH, Ascher DB, Pires DE. mCSM-AB2: guiding rational antibody design using graph-based signatures. *Bioinformatics*. (2020) 36:1453–9. doi: 10.1093/bioinformatics/btz779

28. De Vries SJ, Van Dijk M, Bonvin AM. The HADDOCK web server for data-driven biomolecular docking. *Nat Protoc*. (2010) 5:883–97. doi: 10.1038/nprot.2010.32

29. Xue LC, Rodrigues JP, Kastritis PL, Bonvin AM, Vangone A. PRODIGY: a web server for predicting the binding affinity of protein–protein complexes. *Bioinformatics*. (2016) 32:3676–8. doi: 10.1093/bioinformatics/btw514

30. Altis A, Nguyen PH, Hegger R, Stock G. Dihedral angle principal component analysis of molecular dynamics simulations. *J Chem Phys*. (2007) 126:244111. doi: 10.1063/1.2746330

31. Sun H, Li Y, Tian S, Xu L, Hou T. Assessing the performance of MM/PBSA and MM/GBSA methods. 4. Accuracies of MM/PBSA and MM/GBSA methodologies evaluated by various simulation protocols using PDBbind data set. *Phys Chem Chem Phys*. (2014) 16:16719–29. doi: 10.1039/C4CP01388C

32. Hou T, Wang J, Li Y, Wang W. Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. The accuracy of binding free energy calculations based on molecular dynamics simulations. *J Chem Inf Modeling*. (2011) 51:69–82. doi: 10.1021/ci100275a

33. Khan A, Mao Y, Tahreem S, Wei DQ, Wang Y. Structural and molecular insights into the mechanism of resistance to enzalutamide by the clinical mutants in androgen receptor (AR) in castration-resistant prostate cancer (CRPC) patients. *Int J Biol Macromol*. (2022) 218:856–65. doi: 10.1016/j.ijbiomac.2022.07.058

34. Prekovic S, van Royen ME, Voet AR, Geverts B, Houtman R, Melchers D, et al. The effect of F877L and T878A mutations on androgen receptor response to enzalutamide Molecular analysis of androgen receptor mutants. *Mol Cancer Ther*. (2016) 15:1702–12. doi: 10.1158/1535-7163.MCT-15-0892

35. Selvaraj D, Muthu S, Kotha S, Siddamsetty RS, Andavar S, Jayaraman S. Syringaresinol as a novel androgen receptor antagonist against wild and mutant androgen receptors for the treatment of castration-resistant prostate cancer: molecular docking, in-vitro and molecular dynamics study. *J Biomolecular Structure Dynamics*. (2021) 39:621–34. doi: 10.1080/07391102.2020.1715261

36. Hu X, Chai X, Wang X, Duan M, Pang J, Fu W, et al. Advances in the computational development of androgen receptor antagonists. *Drug Discovery Today*. (2020) 25:1453–61. doi: 10.1016/j.drudis.2020.04.004

37. Gim HJ, Park J, Jung ME, Houk K. Conformational dynamics of androgen receptors bound to agonists and antagonists. *Sci Rep*. (2021) 11:1–15. doi: 10.1038/s41598-021-94707-2

38. Khan A, Waris H, Rafique M, Suleman M, Mohammad A, Ali SS, et al. The Omicron (B.1.1.529) variant of SARS-CoV-2 binds to the hACE2 receptor more strongly and escapes the antibody response: Insights from structural and simulation data. *Int J Biol Macromol*. (2022) 200:438–48. doi: 10.1016/j.ijbiomac.2022.01.059

frontiers | Frontiers in Immunology

TYPE Original Research
PUBLISHED 03 April 2024
DOI 10.3389/fimmu.2024.1358477

OPEN ACCESS

EDITED BY
Helder Nakaya,
University of São Paulo, Brazil

REVIEWED BY
Eduardo L. V. Silveira,
University of São Paulo, Brazil
Andre Aquime Goncalves,
University of Oxford, United Kingdom
Patrícia Gonzalez-Dias,
University of Oxford, United Kingdom

*CORRESPONDENCE
Richard B. Kennedy
kennedy.rick@mayo.edu

RECEIVED 19 December 2023
ACCEPTED 19 March 2024
PUBLISHED 03 April 2024

CITATION
Haralambieva IH, Chen J, Quach HQ, Ratishvili T, Warner ND, Ovsyannikova IG, Poland GA and Kennedy RB (2024) Early B cell transcriptomic markers of measles-specific humoral immunity following a 3rd dose of MMR vaccine.
Front. Immunol. 15:1358477.
doi: 10.3389/fimmu.2024.1358477

Early B cell transcriptomic markers of measles-specific humoral immunity following a 3rd dose of MMR vaccine

Iana H. Haralambieva[1], Jun Chen[2], Huy Quang Quach[1], Tamar Ratishvili[1], Nathaniel D. Warner[2], Inna G. Ovsyannikova[1], Gregory A. Poland[1] and Richard B. Kennedy[1]*

[1]Mayo Clinic Vaccine Research Group, Department of Internal Medicine, Mayo Clinic, Rochester, MN, United States, [2]Department of Quantitative Health Sciences, Mayo Clinic, Rochester, MN, United States

B cell transcriptomic signatures hold promise for the early prediction of vaccine-induced humoral immunity and vaccine protective efficacy. We performed a longitudinal study in 232 healthy adult participants before/after a 3rd dose of MMR (MMR3) vaccine. We assessed baseline and early transcriptional patterns in purified B cells and their association with measles-specific humoral immunity after MMR vaccination using two analytical methods ("per gene" linear models and joint analysis). Our study identified distinct early transcriptional signatures/genes following MMR3 that were associated with measles-specific neutralizing antibody titer and/or binding antibody titer. The most significant genes included: the interleukin 20 receptor subunit beta/*IL20RB* gene (a subunit receptor for IL-24, a cytokine involved in the germinal center B cell maturation/response); the phorbol-12-myristate-13-acetate-induced protein 1/*PMAIP1*, the brain expressed X-linked 2/*BEX2* gene and the B cell Fas apoptotic inhibitory molecule/*FAIM*, involved in the selection of high-affinity B cell clones and apoptosis/regulation of apoptosis; as well as *IL16* (encoding the B lymphocyte-derived IL-16 ligand of CD4), involved in the crosstalk between B cells, dendritic cells and helper T cells. Significantly enriched pathways included B cell signaling, apoptosis/regulation of apoptosis, metabolic pathways, cell cycle-related pathways, and pathways associated with viral infections, among others. In conclusion, our study identified genes/pathways linked to antigen-induced B cell proliferation, differentiation, apoptosis, and clonal selection, that are associated with, and impact measles virus-specific humoral immunity after MMR vaccination.

KEYWORDS

MMR vaccine, measles vaccine, measles virus, humoral immunity, gene expression, B cells

1 Introduction

Omics and systems biology studies in vaccinology investigate how immune parameters are perturbed after vaccination at the whole systems level, and endeavor to identify transcriptomic/omics markers and models that can serve as immune response "signatures" correlated with or predictive of outcomes such as vaccine immunogenicity and/or protective efficacy (1–5). Most of these studies focus on humoral immune responses, as they are crucial for protection against many viral pathogens. Humoral immunity is conferred by antibodies (Ab) and the B lymphocytes/plasma cells that produce them, with the important contribution of $CD4^+$ T cell help (6). Both the initial plasmablast response and the generated pools of long-lived plasma cells and memory B cells have significant role in protection, in maintaining Ab responses, and in carrying out the anamnestic response upon subsequent viral exposure.

Measles virus (MV) is part of the live attenuated MMR vaccine containing measles, mumps, and rubella, which has been effective in reducing the morbidity and mortality associated with these three pathogens, although with differing degrees of success (1, 2). A third dose of MMR vaccine (MMR3) is administered in outbreak settings to control mumps, and more rarely during measles outbreaks (7). Here in this study, we used MMR vaccine as a probe and a model system to study transcriptomic signatures of the recall B cell response in individuals known to be high and low antibody responders to the measles component of the vaccine. We comprehensively investigated early transcriptional events in purified B cells of 232 study participants and their impact/association with humoral immunity after MMR vaccination. Our results demonstrate distinct transcriptional patterns after receipt of MMR3, which are correlated with, and may explain the observed inter-individual differences in, measles vaccine-induced humoral immunity.

2 Materials and methods

The described methods are similar or identical to the ones in our previously published studies (8–13). Our study design/workflow and analysis methodology are outlined in Figure 1.

2.1 Study participants

The study cohort has been previously described in detail (11, 13). It is comprised of 232 healthy subjects from Olmsted County (MN, USA) with two prior documented doses of MMR vaccine. Study subjects provided blood samples prior to the receipt of MMR3 vaccine (Day 0, baseline) and at Day 8 and Day 28 following vaccination. Demographic and clinical variables were collected, including age, sex, race, ethnicity, and MMR vaccination history, as described in our previous study (13). The study was approved by The Mayo Clinic Institutional Review Board. All enrolled participants for the study provided written informed consent.

2.2 Measles virus-specific binding antibody and avidity

MV-specific IgG antibody titer was measured using the Zeus ELISA Measles IgG Test System (Zeus Scientific, Inc., Branchburg, NJ), and results are presented as sample index (SI), as previously described (13). Per the kit's instructions, a sample index greater than 1.1 indicates a seropositive sample. The assay had an intra-assay coefficient of variability (CV) of 6.7% and inter-assay CV of 7.2% in our laboratory.

MV-specific IgG avidity was measured using the Zeus ELISA Measles IgG Test System as previously described (13). Avidity was calculated as the percentage of the absorbance value with and without diethylamine (DEA) in washing buffer. Low avidity (below 30%) and moderate/high avidity (above 30%) were defined arbitrarily using a previously established avidity threshold (11).

2.3 Measles virus-specific neutralizing antibody

Neutralizing antibodies were measured using an optimized MV Edmonston-specific fluorescence-based plaque reduction microneutralization assay, as previously described (8, 10). The 50% neutralizing dose (ND_{50}) was calculated using Karber's formula, and the ND_{50} titer was converted to mIU/mL using the 3^{rd} anti-measles serum international standard (NIBSC code No. 97/648) (8, 10). The assay had a CV of 5.7% and a limit of detection of 15 mIU/mL in our laboratory.

2.4 mRNA sequencing

Next-generation mRNA sequencing was performed in purified B cells as previously described (12). B cells were first isolated from PBMCs via negative selection using the Miltenyi Biotec's B cell isolation kit and MidiMACS™ Separator. This process yielded B cells with an average cell viability (measured by Trypan blue exclusion test) of 98% and average B cell purity (assessed by flow cytometry) of 93%. Total RNA was extracted from the isolated bulk B cells using the RNeasy Plus Mini Kit (Qiagen, Valencia, CA), and evaluated for quality/concentration on an Agilent 2010 Bioanalyzer (Agilent, Palo Alto, CA).

cDNA libraries were generated at the Mayo Clinic's Gene Sequencing Core according to the manufacturer's protocol using the TruSeq® Stranded mRNA Library Prep v2 kit (Illumina, San Diego, CA). Illumina's NovaSeq 6000 S2 Reagent Kit (100 cycles) was used to perform paired-end read sequencing on the Illumina NovaSeq 6000 Instrument. The MAP-RSeq version 3.0 pipeline was applied to align reads using STAR to the hg38 human reference genome, and gene expression counts were obtained using featureCounts utilizing the gene definition files from Ensembl v78 (14). Conditional Quantile Regression was used for normalization (15).

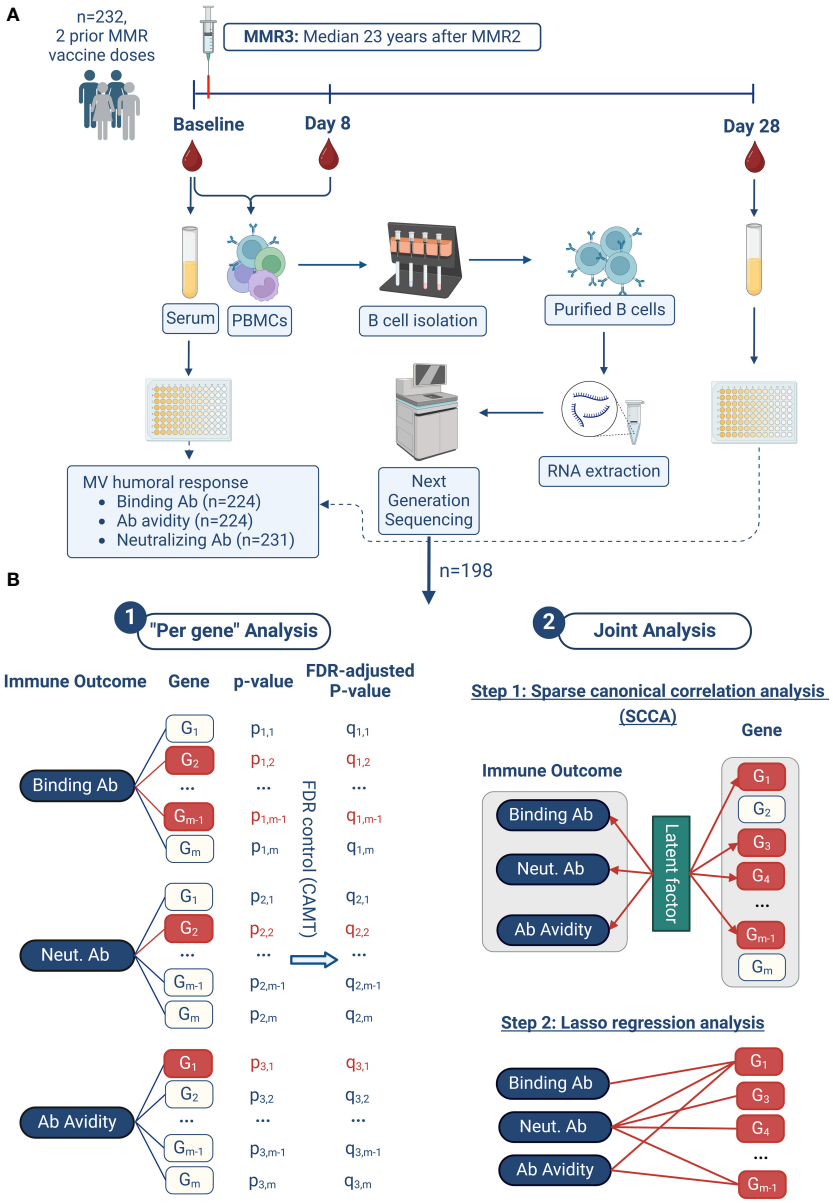


FIGURE 1
Study design and analysis approach. The workflow of our study is illustrated in **(A)**. Our two-pronged analysis approach in summarized in **(B)** and consists of two major steps: 1. "per gene" analysis, and 2. joint analysis. 1. In the "Per gene" analysis model, a linear regression model is fitted for each immune outcome and each gene, with the immune outcome as the dependent variable and the gene expression as the independent variable, adjusting for other covariates. Group-adaptive false discovery rate control (FDR) using the CAMT procedure is then performed based on these individual association P-values. Genes with FDR-adjusted P-values less than 0.1 are considered significant (highlighted in red). 2. In the joint analysis model, the three immune outcomes and all gene expressions are analyzed together, promoting the selection of genes associated with multiple humoral immune outcomes. This analysis proceeds in two steps. Firstly, sparse canonical correlation analysis (SCCA) is applied to select a subset of genes (highlighted in red) whose expressions are most correlated with the three immune outcomes through latent factors. Secondly, lasso sparse regression is applied for each immune outcome based on the SCCA-selected genes from the previous step. The result of this second step is a detailed association network between the three immune outcomes and the associated genes. This figure was created with BioRender.com.

2.5 Statistical analysis

Genes with low abundance or less variability were filtered out (median count <16 at each timepoint or <20th percentile of CV), and a total of 10,174 genes were included in the analyses. The analysis was performed separately for the Baseline and Day 8 gene expression data. The immune outcome was defined as the difference between Day 28 immune outcome and baseline (i.e., Day 28 – Day 0 difference/change on the linear scale). The immune outcomes assessed included: change in MV-specific binding Ab (anti-MV IgG) presented as sample index (SI), change in MV-specific IgG avidity calculated as the percentage/ratio of the ELISA absorbance value with and without the chaotropic agent/DEA (avidity index/AI) and change in anti-MV nAb in mIU/mL (Neut. Ab mIU/ml), as previously described (13).

Our analysis approach is summarized in Figure 1B. It consisted of two major steps. In the first step, we focused on the "per gene"

model since this statistical approach is standard and commonly applied. The advantage of this approach is the maturity of the statistical method (linear regression), the explicit error control (false discovery rate control) and the ability to retrieve correlated genes (thus facilitating enrichment analysis), while the disadvantage is reduction of statistical power (discussed below and in Results). In the second step of our analysis, we applied the joint analysis approach, which addresses some of the limitations of the "per gene" analysis and applies a selection of a sparse subset of genes, which jointly have the highest correlation with the overall vaccine-induced immunity (the three humoral immune outcomes).

"Per gene" model was fitted using multiple linear regression with each immune outcome as the response, and the gene expression as the predictor, controlling for batch, age, and gender effects. Linear model-based t-test was used to calculate "per gene" p-values, followed by multiple testing correction using group-adaptive false discovery rate (FDR) control based on the Covariate Adaptive Multiple Testing/CAMT procedure (16, 17). Here, the group structure is specified by the immune outcome the p-values come from. FDR-corrected p-value or q-value less than 0.1 was used as the significance cutoff. Enriched gene pathways were identified using the Gene Set Enrichment Analysis method (GSEA, (18)), as implemented in the "gseKEGG" (GSEA of KEGG) function of the R Bioconductor package "clusterProfiler" v4.6.2 (19). In comparison to the over-representation test based on the significant genes only, GSEA examines the ranks of the effect sizes (e.g., log2 fold change) for all genes in a specific pathway, and if the rank is overall higher or lower than what would be expected from a random distribution, it indicates that the pathway is activated or suppressed. In our gene expression dataset, Entrez Gene IDs were available for 9,479 of the 10,174 analyzed genes, which were then used in the GSEA. Gene coefficient estimates from the "per gene" models were used as the effect size. FDR control (Benjamini-Hochberg procedure) was performed based on the enrichment p-values (20) to correct for multiple testing.

To complement the "per gene" modeling results, we performed joint analysis of all genes and the three immune outcomes together with the goal to reveal additional biological insights into the influence of gene expression on vaccine-induced immune response outcomes (21). Since the same gene could simultaneously be associated with multiple immune outcomes, joint analysis of all the three immune outcomes could increase the statistical power to identify such co-associated genes. To do this, we first used sparse canonical correlation analysis (SCCA) (22), which selects a sparse subset of genes that explains the most correlation between the gene expression data from a specific timepoint (baseline or Day 8) and the three humoral immune outcomes (using the R "PMA" package v1.2-2). Permutation test was used to select the sparsity tuning parameter as implemented in the "cca.permute" function of the R "PMA" package. Since SCCA does not associate the genes to a specific humoral immune outcome, we further proceeded to identify the genes associated with each of the specific humoral immune outcomes. We applied the least absolute shrinkage and selection operator/lasso regression model to the SCCA-selected genes (all selected genes or the top 500 genes based on the largest SCCA coefficients, if more than 500 genes were selected) for each humoral immune outcome (R "glmnet" package 4.1-8) (23). To account for covariates, linear regression was used to control for confounding variables (effects of batch, gender and age), and the residuals were used in the SCCA. Cross-validation was used to select the sparsity tuning parameter as implemented in the "cv.glmnet" function of the R "glmnet" package. All the statistical analyses were performed in R 4.1.2.

3 Results

3.1 Characteristics of the study cohort and humoral immune response outcomes after MMR3

The study cohort has been previously described in detail and is comprised of two subcohorts as previously described (13). The demographic characteristics of our study cohort was reflective of the demographics of the Olmsted County, MN population (U.S.). According to their racial characteristics the study participants were mostly White (96.5%), and their ethnicity was mostly non-Hispanics or Latino (95.3%). The study cohort included 62.9% females, the median age at enrollment was 35.95 years (IQR 31.95, 40.9) and the study participants' median body mass index (BMI) was 27.9. Median ages at the first dose and second dose of MMR were 15.59 months (IQR 15, 17.71), and 12.5 years (IQR 11.43, 17.15), respectively. In 1998, the American Academy of Pediatrics recommended the current MMR vaccine schedule (2nd dose at 4-6 years of age). A significant portion of our cohort was older than 4-6 years of age at the time of these recommendations and therefore received the 'catch-up' dose (second dose of MMR vaccine) upon entering their next school (middle school/junior high or high school). In the course of this study participants received a third MMR vaccine dose approximately 23 years (median 23.45 years) after their second MMR vaccine dose (Figure 1). The immune outcomes for the study cohort are summarized in Supplementary Figure 1. All humoral immune outcomes increased significantly from baseline to Day 28 following MMR3 vaccination ($p < 2.3E^{-08}$ for all immune outcomes), indicating a significant boost of measles-specific humoral immunity, as previously described (13). At baseline the median nAb titer for the study cohort was 535 mIU/mL (IQR: 260, 1250), and at the peak (Day 28) of antibody response after MMR3, the median of nAb titer was 845 mIU/mL (IQR: 421, 1694). The median Day 28 sample index was 3.47 (IQR: 2.55, 4.21) and the median Ab avidity was 42.8% (IQR: 33.74, 55.43), as previously described (13). Importantly, considerable variation in each humoral immune outcome was observed in our study cohort (13) – providing an ideal scenario for evaluating the potential role of MMR3-induced transcriptional changes in B cells in association with such immune response variability.

Of the study cohort, 198 participants had gene expression data on Day 0 and Day 8 (see Figure 1), as well as neutralizing antibody measure (Day 0 and Day 28) and were used in the transcriptomic association analysis (with neutralizing Ab). Of the subjects with gene expression data, MV-specific binding Ab (SI) and Avidity measures were available for 191 subjects at Day 0 and for 194 subjects at Day 28, and therefore the transcriptomic association analysis with these immune measures was performed in 191 subjects. The Day 0 (baseline) and Day 8 gene expression patterns (heatmaps) across covariates (sex, age, subcohort) and

MV-specific immune response outcomes (neutralizing Ab, binding Ab and avidity) are displayed in Supplementary Figures 2, 3.

3.2 Baseline B cell transcriptomic markers associated with MV-specific humoral immune response following MMR vaccination ("per gene" linear models)

For our analyses, the humoral immune response to vaccination (MV-specific binding IgG Ab, IgG avidity, and neutralizing Ab) was defined as the difference/change of Day 28 immune outcome with respect to baseline (i.e., Day 28 – Day 0 defined as a difference).

3.2.1 Results of "per gene" linear model analysis reveal the impact of B cell Day 0/baseline gene expression on MV-specific humoral immunity after MMR vaccination

First, we assessed the "per gene" associations between baseline/Day 0 gene expression and the Day 28 – Day 0 humoral immune response outcomes. The "per gene" linear model was fitted for each humoral immune response outcome separately. We identified 1,152 B-cell genes displaying significant associations (q-value < 0.1) with measures of MV-specific vaccine-induced humoral immunity, although their individual effect (see Coefficient, Table 1) on the immune response was relatively small (Table 1; Figures 2A, B). Of the most statistically significant genes, several (e.g., B cell linker/BLNK, interferon regulatory factor 5/IRF5, phosphatidylinositol-5-phosphate 4-kinase type 2 alpha/PIP4K2A, all with q-value = 0.0185) are known to impact various B cell activities and functions.

3.2.2 Pathway enrichment analysis on Day 0 gene expression

From a systems biology point of view, pathways and even seemingly unrelated "pools" of different genes may be collectively important. To untangle the biological processes behind our "per gene" models we performed pathway enrichment analysis on the Day 0 genes associated with anti-MV binding or neutralizing antibody titer after MMR vaccination, as described in Statistical analysis. Our assessment confirmed the enrichment of genes involved in metabolic pathways and basic cellular/organelle functions (lysosome, phagosome), as well as signal transduction pathway genes linked to inflammation/autoimmunity (e.g., NOD-like receptor signaling pathway) and/or host innate and adaptive immune response, including the B cell receptor signaling pathway (Figures 2C, D; Supplementary Table 1).

3.3 Early/Day 8 B-cell transcriptomic markers associated with MV-specific humoral immune response following MMR vaccination ("per gene" linear models)

3.3.1 Results of "per gene" linear model analysis reveal the impact of early B cell gene expression on MV-specific humoral immunity after MMR vaccination

The early transcriptional events in B cells upon antigenic stimulation are of critical importance for the generation and maintenance of humoral immunity. Since other studies have reported associations between plasmablast transcriptional response (peaking at Day 7-8) and antibody titers following vaccination, our study sought to identify early (Day 8, plasmablast) transcriptional signatures in B cells that are highly correlated with vaccine-induced humoral immune outcomes (24–27). To achieve this, we fit linear models for each gene with the Day 8 gene expression as the covariate. While this "per gene" analysis yielded a smaller number of significantly associated genes (n=318, Day 8 genes with statistically significant associations at q-value < 0.1), compared to baseline genes, their individual gene effects/weights (reported as an estimated effect of each gene/Coefficient, Table 2; Figures 3A, B) on the immune response outcome were relatively large, which is consistent with the substantial contribution of specific Day 8 B-cell genes to the measured immune outcome. Of note, among the top 30 most significant findings we identified interleukin 20 receptor subunit beta/*IL20RB*, phorbol-12-myristate-13-acetate-induced protein 1/*PMAIP1* and brain expressed X-linked 2/*BEX2* gene involved in apoptosis, proteasome 26S subunit, non-ATPase 12/*PSMD12*, involved in ubiquitination and replication of influenza virus, and other genes linked to antigen-induced proliferation, differentiation, apoptosis, commitment to different B cell lineages and clonal selection (Table 2; Figure 3).

3.3.2 Pathway enrichment analysis on Day 8 gene expression

This assessment identified 29 significantly enriched pathways (q<0.05) among the genes associated with MV-specific binding antibody and 9 significantly enriched pathways among the genes associated with MV-specific nAb. We observed a moderate overlap with the enriched Day 0 gene expression pathways/Figure 3C; Supplementary Table 2) consisting of basic metabolic and cellular function-related pathways. Among the identified enriched pathways, there were also five pathways associated with different viral infections (measles virus, herpes simplex virus, Kaposi sarcoma-associated herpesvirus, human T cell leukemia virus 1 and Epstein-Barr virus) and multiple pathways related to metabolism, basic cellular functions, signaling pathways and lymphocyte immune activity (Figure 3C; Supplementary Table 2).

3.4 Results from joint analysis of B-cell transcriptomic markers associated with MV-specific humoral immune response following MMR vaccination

"Per gene" model tests one gene at a time and requires multiple testing correction that may result in reduced statistical power. In addition, the "per gene" model aggregates the effects of other relevant genes into the error term, thus increasing the variance of the error term and reducing the statistical power to identify genes with moderate effects. If a gene is associated with multiple immune outcomes, the "per gene" model is not able to use such information, leading to further loss of statistical power. "Per gene" model also does not account for correlations among genes, and highly correlated genes tend to be selected together. Thus, it has limited ability to identify genes, whose associations are independent of other genes. Joint analysis of gene expression, on the other hand, could address some of the listed

TABLE 1 "Per gene" linear model analysis for Day 0 gene expression in B cells and Day 28 – Day 0 humoral immune response outcomes.

Gene Symbol	Description	Outcome	p-value	q-value	Coefficient*
DIRC2	solute carrier family 49 member 4	SI- ELISA binding Ab	3.25E-07	0.0185	0.386
INTS8	integrator complex subunit 8	SI- ELISA binding Ab	2.75E-06	0.0185	-0.361
MORC3	MORC family CW-type zinc finger 3	SI- ELISA binding Ab	3.52E-06	0.0185	-0.357
RPS6KA1	ribosomal protein S6 kinase A1	SI- ELISA binding Ab	3.60E-06	0.0185	-0.372
WDR48	WD repeat domain 48	SI- ELISA binding Ab	4.92E-06	0.0185	0.364
SNX9	sorting nexin 9	SI- ELISA binding Ab	8.64E-06	0.0185	0.334
CRTAP	cartilage associated protein	SI- ELISA binding Ab	8.94E-06	0.0185	-0.331
MPC1	mitochondrial pyruvate carrier 1	SI- ELISA binding Ab	1.04E-05	0.0185	0.324
SYNRG	synergin gamma	SI- ELISA binding Ab	1.14E-05	0.0185	-0.352
TCOF1	treacle ribosome biogenesis factor 1	SI- ELISA binding Ab	1.90E-05	0.0185	-0.322
CASZ1	castor zinc finger 1	SI- ELISA binding Ab	2.34E-05	0.0185	0.326
BLNK	B cell linker	SI- ELISA binding Ab	2.40E-05	0.0185	-0.334
GLT8D1	glycosyltransferase 8 domain containing 1	SI- ELISA binding Ab	3.07E-05	0.0185	-0.316
ACADM	acyl-CoA dehydrogenase medium chain	SI- ELISA binding Ab	3.15E-05	0.0185	-0.318
PIP4K2A	phosphatidylinositol-5-phosphate 4-kinase type 2 alpha	SI- ELISA binding Ab	3.55E-05	0.0185	-0.333
C4orf46	chromosome 4 open reading frame 46	SI- ELISA binding Ab	3.76E-05	0.0185	0.315
METTL7A	methyltransferase like 7A	SI- ELISA binding Ab	3.76E-05	0.0185	-0.336
MAPRE1	microtubule associated protein RP/EB family member 1	SI- ELISA binding Ab	3.92E-05	0.0185	0.304
CYSLTR1	cysteinyl leukotriene receptor 1	SI- ELISA binding Ab	3.94E-05	0.0185	-0.333
ZNF106	zinc finger protein 106	SI- ELISA binding Ab	4.21E-05	0.0185	-0.339
IPPK	inositol-pentakisphosphate 2-kinase	SI- ELISA binding Ab	5.16E-05	0.0185	0.305
SLC15A3	solute carrier family 15 member 3	SI- ELISA binding Ab	6.03E-05	0.0185	-0.301
APEH	acylaminoacyl-peptide hydrolase	SI- ELISA binding Ab	6.33E-05	0.0185	-0.313
C11orf73	heat shock protein nuclear import factor hikeshi	SI- ELISA binding Ab	6.99E-05	0.0185	-0.310
MTMR9	myotubularin related protein 9	SI- ELISA binding Ab	7.16E-05	0.0185	0.309
PLD4	phospholipase D family member 4	SI- ELISA binding Ab	7.51E-05	0.0185	-0.341
CHRAC1	chromatin accessibility complex subunit 1	SI- ELISA binding Ab	7.73E-05	0.0185	0.309
IRF5	interferon regulatory factor 5	SI- ELISA binding Ab	7.92E-05	0.0185	-0.288
CAPG	capping actin protein, gelsolin like	SI- ELISA binding Ab	8.09E-05	0.0185	-0.298
C1orf198	chromosome 1 open reading frame 198	SI- ELISA binding Ab	8.48E-05	0.0185	0.319

The top 30 displayed genes/findings with significant associations are genes associated with SI/anti-MV IgG as an immune outcome (see Statistical analysis).
*Coefficient can be interpreted as the change of the immune outcome measurement in response to one standard deviation change of the gene expression.

limitations and reveal additional biological insights. To achieve this, we performed a sparse canonical correlation analysis (SCCA) to first select genes jointly impacting the three MV-specific humoral immune outcomes (neutralizing Ab, binding Ab/SI and avidity/AI). Since the same gene could simultaneously be associated with multiple immune outcomes, joint analysis of all the three immune outcomes could increase the statistical power to identify these co-associated genes. We then performed lasso regression analysis on the SCCA-selected genes to identify genes associated with a specific humoral immune outcome (see Statistical analysis).

3.4.1 Results from joint gene expression analysis of the impact of baseline B-cell gene expression on MV-specific humoral immunity after vaccination

The SCCA analysis on the baseline B cell gene expression resulted in the selection of 172 genes simultaneously associated with all three measures of MV-specific humoral immunity (nAb, binding Ab, and antibody avidity Figure 4A). The lasso regression analysis of these genes resulted in the identification of 40 genes associated with MV-specific neutralizing antibody, 31 genes

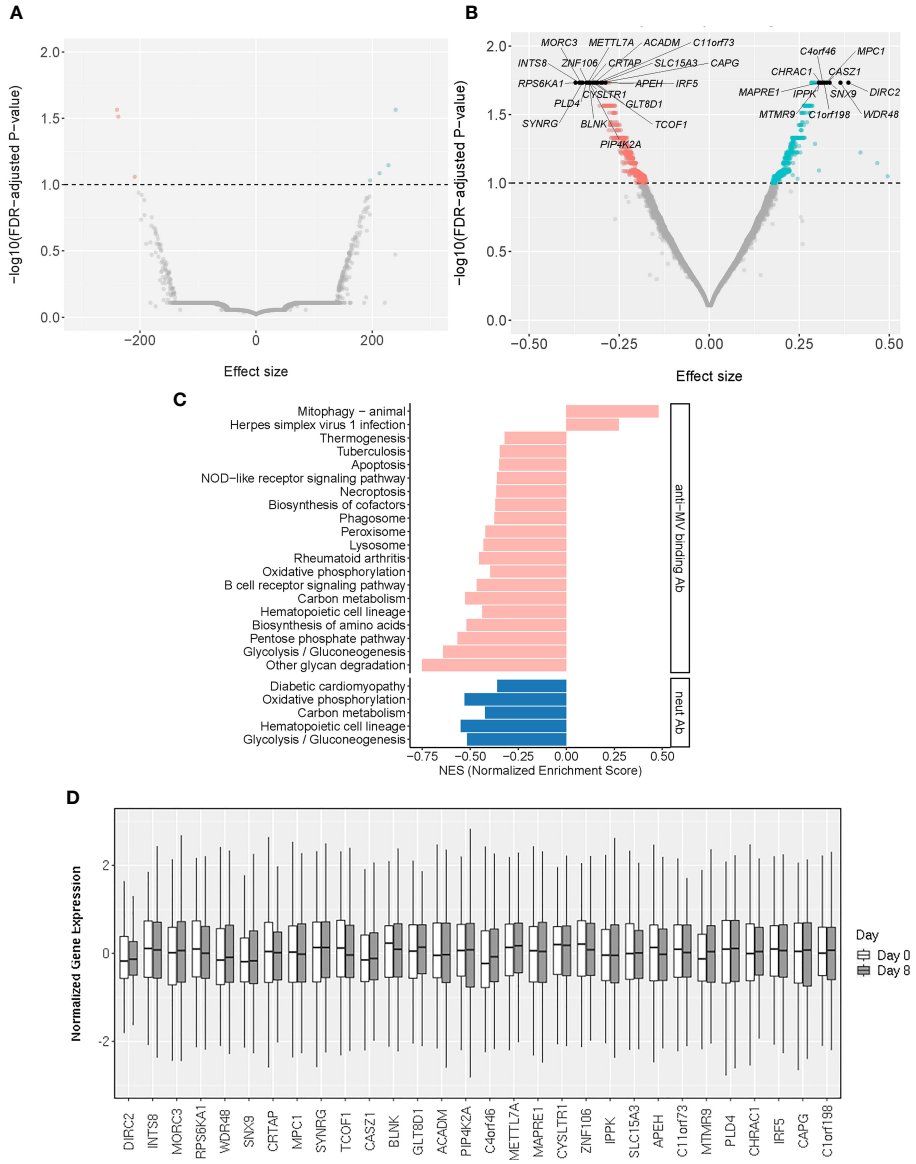


FIGURE 2
Baseline transcriptomic markers associated with MV-specific humoral immunity. **(A, B)** The volcano plots illustrate the association of baseline (Day 0) gene expression with neutralizing Ab **(A)** and MV-specific binding Ab **(B)**. The effect size represents the coefficient from the "per gene" linear regression analysis. The top 30 significant genes are designated with their gene symbols. **(C)** Pathway enrichment analysis/GSEA plots of hallmark pathways for Day 0 gene associations with MV-specific humoral immunity (binding Ab/pink or neutralizing Ab/blue). NES was calculated based on the coefficients from "per gene" analysis. **(D)** Normalized gene expression box plots of the top (most significant) 30 Day 0 (baseline) genes associated with MV-specific humoral immunity across two timepoints (Day 0 and Day 8).

associated with MV-specific binding antibody and 22 genes associated with antibody avidity, including predominantly metabolic genes and genes involved in different cell signaling cascades (Supplementary Table 3).

3.4.2 Results from joint gene expression analysis of the impact of early/Day 8 B-cell gene expression on MV-specific humoral immunity after vaccination

The SCCA assessment on the Day 8 B-cell gene expression led to the selection of many genes associated with the three measures of MV-specific humoral immunity (n=7,716). Although it is possible that a large number of genes are associated with the activation of B cells, each with weak effects, we recognize that the large number of genes selected could also be due to the limitation of the lasso sparsity penalty used in SCCA, where it tends to produce a denser model in order to retain those truly associated genes. Thus, we focused our further analysis on the top 500 genes with largest SCCA coefficients (Figure 4B). Focusing on the top 500 selected genes, the lasso regression analysis identified 94 genes associated with MV-specific neutralizing antibody and 66 genes associated with MV-specific binding antibody (Supplementary Table 4). Of note,

TABLE 2 "Per gene" linear model results for Day 8 B-cell gene expression and Day 28 – Day 0 humoral immune outcomes.

Gene Symbol	Description	Outcome	p-value	q-value	*Coefficient
ATAD2	ATPase family AAA domain containing 2	Neut. Ab miu/ml	1.51E-06	0.053	263.35
C1orf162	chromosome 1 open reading frame 162	Neut. Ab miu/ml	3.23E-06	0.053	-255.78
CA2	carbonic anhydrase 2	SI ELISA binding Ab	5.71E-06	0.053	0.34
WHAMM	WASP homolog associated with actin, golgi membranes and microtubules	SI ELISA binding Ab	8.44E-06	0.053	0.35
BEX2	brain expressed X-linked 2	Neut. Ab miu/ml	1.26E-05	0.053	228.33
PMAIP1	phorbol-12-myristate-13-acetate-induced protein 1	SI ELISA binding Ab	1.44E-05	0.053	0.33
SMIM15	small integral membrane protein 15	Neut. Ab miu/ml	2.37E-05	0.053	229.99
TDG	thymine DNA glycosylase	Neut. Ab miu/ml	2.56E-05	0.053	219.91
DDIT3	DNA damage inducible transcript 3	SI ELISA binding Ab	2.91E-05	0.053	0.31
C4orf46	chromosome 4 open reading frame 46	SI ELISA binding Ab	3.26E-05	0.053	0.32
INO80D	INO80 complex subunit D	SI ELISA binding Ab	3.48E-05	0.053	0.31
CRTAP	cartilage associated protein	SI ELISA binding Ab	3.69E-05	0.053	-0.30
PUS7	pseudouridine synthase 7	Neut. Ab miu/ml	4.47E-05	0.053	-213.79
TMEM14A	transmembrane protein 14A	SI ELISA binding Ab	5.57E-05	0.053	-0.34
GORAB	golgin, RAB6 interacting	Neut. Ab miu/ml	6.10E-05	0.053	218.02
PTBP2	polypyrimidine tract binding protein 2	SI ELISA binding Ab	6.11E-05	0.053	0.32
AGPAT1	1-acylglycerol-3-phosphate O-acyltransferase 1	Neut. Ab miu/ml	6.58E-05	0.053	-216.29
PSMD12	proteasome 26S subunit, non-ATPase 12	Neut. Ab miu/ml	7.07E-05	0.053	227.86
TMEM39A	transmembrane protein 39A	Neut. Ab miu/ml	7.23E-05	0.053	218.20
RHEBL1	RHEB like 1	SI ELISA binding Ab	7.51E-05	0.053	0.33
LACTB2	lactamase beta 2	SI ELISA binding Ab	7.88E-05	0.053	-0.33
CGRRF1	cell growth regulator with ring finger domain 1	Neut. Ab miu/ml	8.26E-05	0.053	211.89
MCPH1	microcephalin 1	SI ELISA binding Ab	8.57E-05	0.053	0.31
CHMP2B	charged multivesicular body protein 2B	Neut. Ab miu/ml	8.71E-05	0.053	208.85
VPS26A	VPS26, retromer complex component A	Neut. Ab miu/ml	8.78E-05	0.053	213.23
IL20RB	interleukin 20 receptor subunit beta	Neut. Ab miu/ml	9.84E-05	0.053	201.08
KLF7	Kruppel like factor 7	SI ELISA binding Ab	9.99E-05	0.053	0.29
C1GALT1C1	C1GALT1 specific chaperone 1	SI ELISA binding Ab	0.0001	0.053	-0.30
TMEM123	transmembrane protein 123	SI ELISA binding Ab	0.0001	0.053	0.28
RND1	Rho family GTPase 1	SI ELISA binding Ab	0.0001	0.053	0.29

Top 30 genes/findings included associations with SI/anti-MV IgG and neutralizing Ab (see Statistical analysis).
*Coefficient can be interpreted as the change of the immune outcome measurement in response to one standard deviation change of the gene expression.

although the analytical approaches we used were different, we observed a reasonable number of overlapping results with the identified genes via "per gene" linear models. For example, of the 94 genes associated with MV-specific nAb (identified via the joint analysis approach), 6 genes (*PUS7, TDG, PTBP2, BEX2, CRTAP, INO80D*) were among the 30 top genes (representing 20% of the top genes associated with MV-specific humoral immunity) that were identified via "per gene" linear models (see Table 2). Thirty one of the 94 genes (approximately 33%) overlapped with the list of the 318 significantly associated genes (q <0.1) with humoral immunity via "per gene" linear models. Among these overlapping genes, most had an unknown link to B cells and/or the generation/maintenance of humoral immunity, however a few were known apoptotic genes (e.g., BEX2, involved in the regulation of mitochondrial apoptosis, as well as the Fas apoptotic inhibitory molecule/FAIM, Supplementary Table 4) that may have implications on the

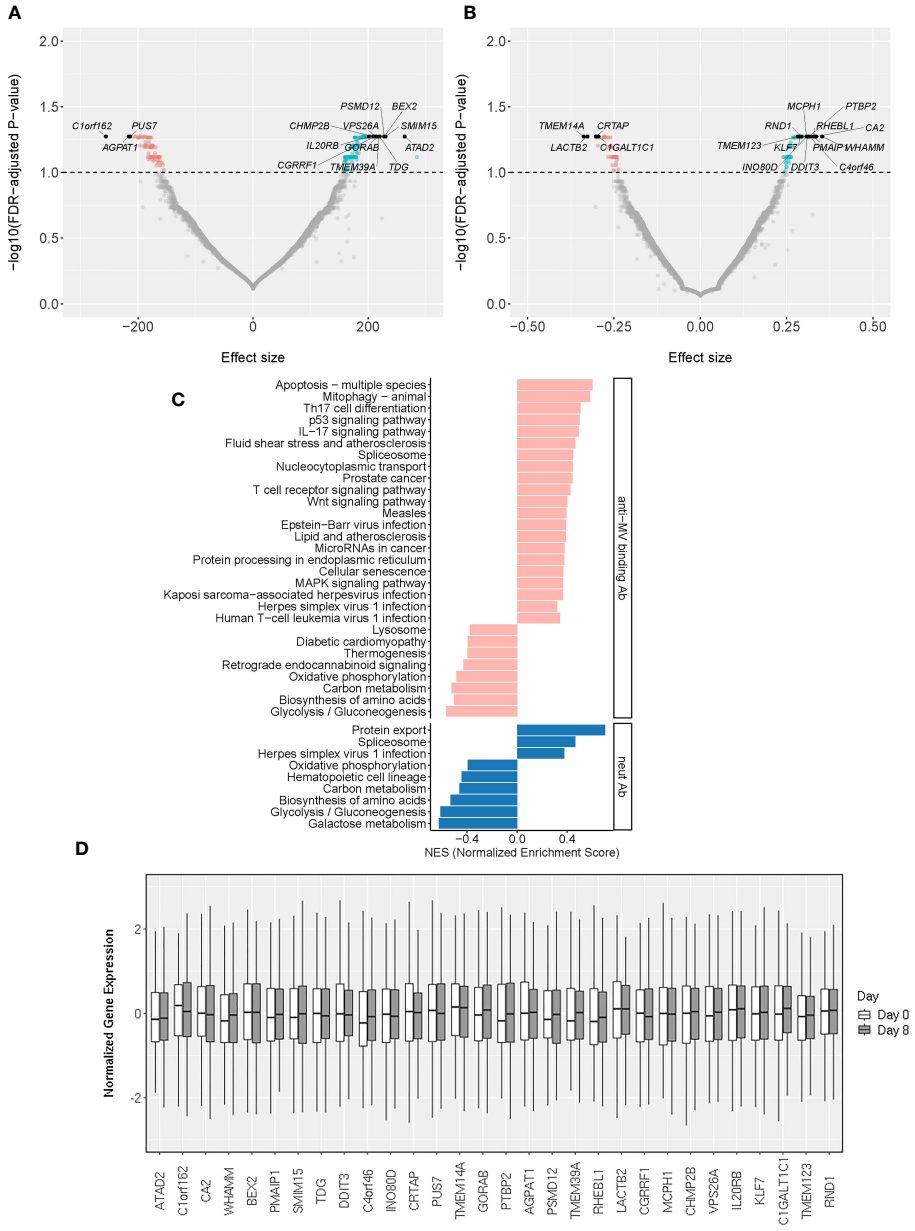


FIGURE 3
Early/Day 8 transcriptomic markers associated with MV-specific humoral immunity. **(A, B)** The volcano plots illustrate the association of Day 8 gene expression with neutralizing Ab **(A)** and MV-specific binding Ab **(B)**. The effect size represents the coefficient from the "per gene" linear regression analysis. The top 30 significant genes are designated with their gene symbols. **(C)** Pathway enrichment analysis/GSEA plots of hallmark pathways for Day 8 gene associations with MV-specific humoral immunity (binding Ab/pink or neutralizing Ab/blue). NES was calculated based on the coefficients from "per gene" analysis. **(D)** Normalized gene expression box plots of the top (most significant) 30 Day 8 genes associated with MV-specific humoral immunity across two timepoints (Day 0 and Day 8).

process of apoptosis in the B cell lineage. Finally, the joint analysis approach identified also many non-shared (with the results from the linear models) genes, among those the interleukin 16/*IL16* gene associated with the nAb titer (Supplementary Table 4).

4 Discussion

The discovery of genes/genetic signatures or other "omics" measurements associated with and/or predictive of immune response after vaccination has been the goal and the subject of cutting-edge systems-level vaccine research for over a decade (3–5, 28).

The current study identified multiple key biomarkers/factors and pathways that contribute to and shape inherent B cell activity and functions necessary for generating and/or maintaining optimal vaccine-induced humoral immunity. We focused our study design on the B cell compartment in order to identify intrinsic B cell factors driving the recall immune response to vaccination and highly associated with MV-specific humoral immunity. We acknowledge

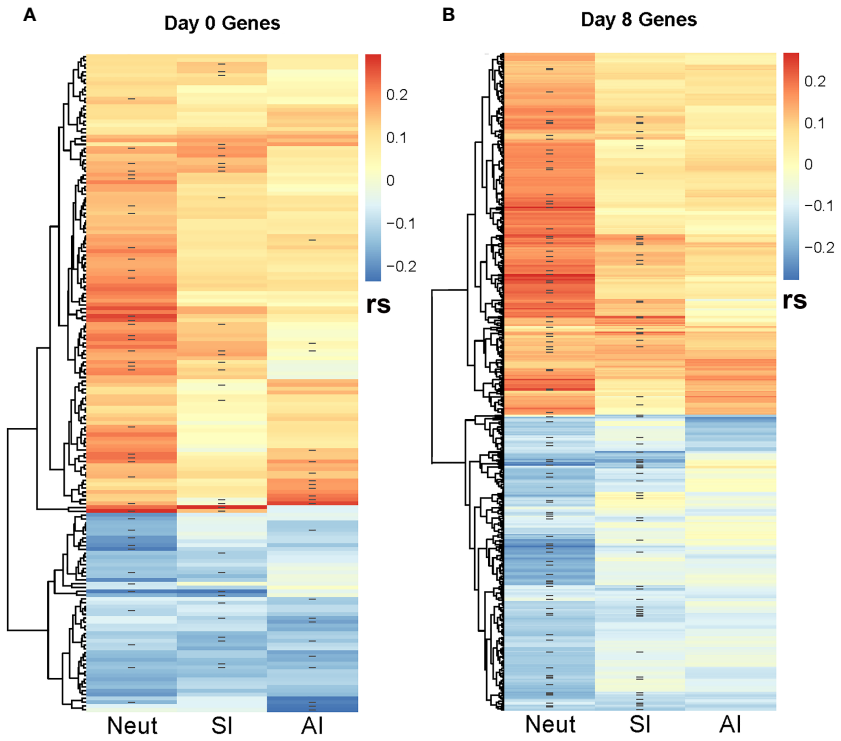


FIGURE 4
Correlation heatmap between SCCA-selected genes and MV-specific humoral immune response outcomes. The heatmaps illustrate the Spearman correlations (rs) between the SCCA-selected genes and the three immune response outcomes (neutralizing Ab/NeutAb, binding antibody sample index/SI and avidity index/AI). '-' denotes the genes selected by the Lasso regression analysis as associated with each immune outcome. **(A)** illustrates the Spearman correlations with immune outcomes for the Day 0 (baseline) genes. **(B)** illustrates the Spearman correlations with immune outcomes for the Day 8 genes. As shown, many genes are co-associated with multiple immune outcomes.

that our study design (i.e., measuring gene expression in purified B cells) supported the identification of B cell-specific genes, even so this study identified a range of distinct early transcriptional activities (transcriptional factors) and specific molecular/cellular processes, that influence recall measles-specific humoral immunity. One of our most important findings is the discovery of *IL20RB* gene (encoding a cytokine receptor subunit of the heterodimeric complex required for IL-19, IL-20 and IL-24 binding and activity) as an early transcriptional biomarker in B cells that was highly associated with MV-specific nAb titer. The interleukin/IL-20 subfamily consists of IL-19, IL-20, IL-22, IL-24 and IL-26, and its members are involved in inflammatory and innate immune (including antiviral) activity, tissue repair/homeostasis, cell communication, proliferation and differentiation, and oncogenesis (29). Of the known cytokines using this receptor/subunit, IL-24 has been described as a pivotal B cell immunoregulatory cytokine, directly involved in the processes of germinal center B cell maturation (30). This multifunctional cytokine signals through two heterodimeric receptors IL-20RA/IL-20RB and IL-20RB/IL22RA1 (both include the subunit encoded by *IL20RB*) and is known to mediate inflammatory and autoimmune responses, as well as to regulate a variety of immune cell functions (including in B cells, T cells, NK cells, and macrophages) (29). Although not specifically linked to vaccine-induced immunity, IL20RB and the associated signaling pathway have been identified as critical in the protection and host defense against mucosal pathogens (31). It has been postulated that BCR activation/CD40 engagement (CD40-CD40L ligation) in follicular B cells (in particular $CD27^+$ memory B cells and $CD5^+$ B cells) is associated with high expression of IL-24, which plays an important role in supporting germinal center T-dependent antigen B cell proliferation (30). The ligand IL-24 has been shown to hinder plasma cell/terminal B cell differentiation and antibody production by favoring the maturation of memory B cells (30). Interestingly, we previously identified IL-24 (*IL24*) along with CD93 as markers of differential MV-specific transcriptional response (in PBMCs) in 15 high *vs.* 15 low antibody responders to measles vaccination (32). Hence, it is likely that the differential expression of IL-24 receptor and/or IL-24 by specific B cell subsets during B cell activation (early post-measles vaccine), physiologically "fine-tunes" the balance between plasma cell and MBC commitment, thus affecting antigen-specific plasmablast/plasma cell response and antibody production. We speculate that further investigation in this direction can potentially lead to the development of improved vaccine candidates by modulating the production of IL-24 via: incorporating an adjuvant that stimulates IL-24 production; incorporating a recombinant IL-24 lacking apoptosis-inducing properties (33); or generating a recombinant virus, expressing IL-24 or a factor silencing IL-24 for testing in future studies. Another interesting early B cell transcriptional marker associated with antibody response is the phorbol-12-myristate-13-acetate-induced protein 1/*PMAIP1* (Noxa), encoding a pro-apoptotic member of the BCL-2 protein family with significant involvement in the selection

of high-affinity B cell clones upon antigenic stimulation (34, 35). The ablation of the encoded protein leads to increased survival of low-affinity clones at the expense of high-affinity clones *in vivo*, in a mouse model following influenza vaccination (34–36).

The two analytical approaches (linear models and joint analysis) used in our study have discovered that approximately 20-30% of the identified genes overlap and are associated with immune outcome/nAb titer, which builds confidence in our findings. Both approaches identified genes involved in the apoptosis/regulation of apoptosis. For example, the *BEX2* gene is a known regulator of mitochondrial apoptosis and G1 cell cycle, while *FAIM* (cloned as an inhibitor/regulator of Fas-mediated apoptosis in B cells) has a significant role in the regulation of germinal center B cell response and the plasma cell compartment response (37–39). Another important gene, *IL16* (encoding the B lymphocyte-derived IL-16 ligand of CD4), identified via the joint analysis approach, has been demonstrated to play a significant role in the crosstalk and attraction/recruitment of dendritic cells and helper T cells to initiate and achieve an optimal humoral immune response (40, 41). A vaccine study in solid organ transplant patients, found that IL-16 levels (among other cytokines) were significantly lower in subjects with very low antibody response to mRNA-based COVID-19 vaccine compared to subjects with normal immune response, suggesting that this cytokine is associated with the optimal development of humoral immunity after COVID-19 vaccination (42).

Another highlighted finding in our study has been identified as a novel virus-specific host factor. The proteasome 26S subunit, non-ATPase 12/*PSMD12*, has been previously implicated in regulation of the replication/budding of influenza virus through K63-specific ubiquitination of the matrix/M1 viral structural protein (43). It is plausible that it may impact the budding/replication of other enveloped RNA viruses, and thus affect antigenic abundance/host response. Factors associated with anti-vital immunity, such as IRF5 (identified in our study) were found to be part of a molecular signature induced by LAIV influenza vaccination (44, 45). In agreement with the identified genes/cellular functions, our pathway enrichment analysis of Day 0 and Day 8 genes/gene expression pointed to enriched pathways associated with different viral infections, as well as to multiple cytokines and immune/B cell signaling pathways, apoptosis/regulation of apoptosis, metabolic pathways and cell cycle-related pathways, among others. As expected, we found pathways and gene expression patterns that have been previously identified with other viral vaccines and immune response studies. A member of the B cell signaling pathway triggered upon B cell activation (TNFRSF17, a receptor for BLyS-BAFF) was identified as a key predictive factor for neutralizing antibody response to yellow fever vaccination (24). B cell signaling modules were also identified as important for the optimal response to influenza vaccination (46). Other identified pathways in our study (apoptosis/regulation of apoptosis) have also been found to impact immune response to vaccination by others. Furman et al., identified the regulation of apoptosis as an essential pathway prognostic of responsiveness to influenza vaccine (47). Vaccine adjuvants and vaccine components, conversely, were demonstrated to induce damage-associated molecular patterns (DAMPs) and cell death-associated signaling pathways, that were found to be important for augmenting immunogenicity after vaccination (48, 49). As separate studies found associations between genes regulating apoptosis and immune response to vaccination, it is likely that apoptosis and its regulation may play a role (perhaps through increased survival of antibody producing cells) in vaccine-induced immunity. This warrants further investigation.

The strengths of our study include the relatively large (for transcriptomic studies) sample size, the acquisition of high-quality transcriptomic information/data from purified B cells before/after MMR vaccination and the use of two different analytical approaches to identify biologically relevant gene signatures. An important limitation is the possibility of false positive findings, which is alleviated with the reporting of FDR-adjusted p-values or q-values and the implementation of the joint analysis. Using the FDR control, the percentage of false positives is controlled in the "per gene" analysis. While the joint analysis method does not offer explicit FDR control, by jointly analyzing the immune outcomes and the genes together, the method promotes the selection of genes associated with multiple outcomes, thus pooling association evidence across immune outcomes. The two steps of our analysis ("per gene" model and joint analysis) are complementary rather than competitive. Together, the results they produce provide a better understanding of the important transcriptional factors underlying measles vaccine-induced humoral immunity. Another important point to mention is the confounding effect of simultaneous immune stimulation during MMR vaccination (measles, mumps, and rubella). In this regard, it will be important to study the effect of the identified genes on rubella virus and mumps virus-specific immune outcomes. It is also important to note, that although our goal was to study Day 8 (plasmablast) transcriptional response in terms of association with humoral immunity, the assessment of earlier transcriptional programs in B cells (collected at earlier timepoints) could provide additional valuable insights into the generation of recall immune response after vaccination. Validation of our major findings through functional studies is necessary to determine the contribution of specific gene/genes (e.g., *IL20RB*) to MV-specific humoral immunity. Another avenue to explore is the assessment of transcriptional patterns (including the identified genes of high interest and other genes) in different B cell subsets at different timepoints following vaccination, which will help to better understand the gene expression dynamics in the B cell compartment and its contribution to humoral immunity.

In summary, our study identified important early B lymphocyte-derived transcriptomic signatures (*IL20RB*, *PMAIP1*, *BEX2*, *FAIM*, and *IL16*) associated with functional immunity/MV-specific neutralizing antibody response and other measures of humoral immunity following MMR vaccination. We suggest that such molecular signatures can serve as early biomarkers of optimal vaccine immunogenicity and hold promise for (potentially) improving vaccine-induced immunity through providing useful information for the development of next-generation vaccine candidates (3, 46, 50, 51).

Data availability statement

The data presented in this study are deposited in Synapse.org (Sage Bionetworks) under Project SynID syn54153421, at https://www.synapse.org/#!Synapse:syn54153421/wiki/626686. The publicly available dataset consists of subjects who consented to data sharing.

Ethics statement

The studies involving humans were approved by The Mayo Clinic Institutional Review Board. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

IH: Data curation, Investigation, Methodology, Writing – original draft. JC: Formal analysis, Software, Writing – review & editing. HQ: Data curation, Investigation, Methodology, Writing – review & editing. TR: Investigation, Methodology, Writing – review & editing. NW: Formal analysis, Software, Writing – review & editing. IO: Investigation, Project administration, Supervision, Writing – review & editing. GP: Conceptualization, Funding acquisition, Writing – review & editing. RK: Conceptualization, Funding acquisition, Supervision, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. Research reported in this publication was supported by the National Institute of Allergy and Infectious Diseases of the National Institutes of Health under awards R01 AI033144, R01 AI138965, R01 AI121054 and R01 AI127365.

Acknowledgments

We thank all participants enrolled for this study. Research reported in this publication was supported by the National Institute of Allergy and Infectious Diseases of the National Institutes of Health under awards R01 AI033144, R01 AI138965, R01 AI121054 and R01 AI127365.

Conflict of interest

Dr. GP offers consultative advice to Johnson & Johnson/Janssen Global Services LLC, and is the chair of a Safety Evaluation Committee for novel investigational vaccine trials being conducted by Merck Research Laboratories. Dr. GP also offers consultative advice on vaccine development to Merck & Co., Medicago, GlaxoSmithKline, Sanofi Pasteur, Emergent Biosolutions, Dynavax, Genentech, Eli Lilly and Company, Kentucky Bioprocessing Inc, Bavarian Nordic, AstraZeneca, Exelixis, Regeneron, Janssen, Vyriad, Moderna, and Genevant Sciences, Inc. Drs. GP and IO hold patents related to vaccinia and measles peptide vaccines. Drs. RK, GP, and IO hold a patent related to vaccinia peptide vaccines. Drs. GP, RK, IO and IH hold a patent related to the impact of single nucleotide polymorphisms on measles vaccine immunity. Drs. GP, RK, and IO have received grant funding from ICW Ventures for preclinical studies on a peptide-based COVID-19 vaccine. Dr. RK has received funding from Merck Research Laboratories to study waning immunity to mumps vaccine. Dr. RK also offers consultative advice on vaccine development to Merck & Co. and Sanofi Pasteur. These activities have been reviewed by the Mayo Clinic Conflict of Interest Review Board and are conducted in compliance with Mayo Clinic Conflict of Interest policies. This research has been reviewed by the Mayo Clinic Conflict of Interest Review Board and was conducted in compliance with Mayo Clinic Conflict of Interest policies.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2024.1358477/full#supplementary-material

SUPPLEMENTARY FIGURE 1
Immune response summary of the study subjects. Box plots summarizing: **(A)** Day 0 and Day 28 binding antibody sample index/SI; **(B)** Day 0 and Day 28 antibody avidity index; and **(C)** Day 0 and Day 28 Neutralizing Ab. The line indicates the median of the immune response measure in our cohort, while the whiskers indicate 25% and 75% IQR. The p-values (Wilcoxon signed rank test) demostrate the significant upregulation of Day 28 immune outcomes (post MMR3) compared to baseline (Day 0) immune outcomes.

SUPPLEMENTARY FIGURE 2
Heatmap of Day 0 gene expression patterns. Heatmap of Day 0 (baseline) gene expression patterns of the significant genes (FDR < 0.1) from "per gene" analysis across covariates (sex, age, subcohort) and MV-specific immune response outcomes (Day 28 – Day 0 difference): Neutralizing Ab (Naut.Ab),SI (Sample Index/Binding Ab) and AI (Avidity Index).

SUPPLEMENTARY FIGURE 3
Heatmap of Day 8 gene expression patterns. Heatmap of Day 8 gene expression patterns of the significant genes (FDR < 0.1) from "per gene" analysis across covariates (sex, age, subcohort) and MV-specific immune response outcomes (Day 28 – Day 0 difference): Neutralizing Ab (Naut.Ab),SI (Sample Index/Binding Ab) and AI (Avidity Index).

References

1. Haralambieva IH, Kennedy RB, Ovsyannikova IG, Whitaker JA, Poland GA. Variability in humoral immunity to measles vaccine: new developments. *Trends Mol Med*. (2015) 21:789–801. doi: 10.1016/j.molmed.2015.10.005

2. Haralambieva IH, Kennedy RB, Ovsyannikova IG, Schaid DJ, Poland GA. Current perspectives in assessing humoral immunity after measles vaccination. *Expert Rev Vaccines*. (2019) 18:75–87. doi: 10.1080/14760584.2019.1559063

3. Kennedy RB, Ovsyannikova IG, Palese P, Poland GA. Current challenges in vaccinology. *Front Immunol*. (2020) 11:1181. doi: 10.3389/fimmu.2020.01181

4. Pulendran B, Davis MM. The science and medicine of human immunology. *Science*. (2020) 369. doi: 10.1126/science.aay4014

5. Arunachalam PS, Scott MKD, Hagan T, Li C, Feng Y, Wimmers F, et al. Systems vaccinology of the BNT162b2 mRNA vaccine in humans. *Nature*. (2021) 596:410–6. doi: 10.1038/s41586-021-03791-x

6. Pulendran B, Ahmed R. Immunological mechanisms of vaccination. *Nat Immunol*. (2011) 12:509–17. doi: 10.1038/ni.2039

7. Marlow MA, Marin M, Moore K, Patel M. CDC guidance for use of a third dose of MMR vaccine during mumps outbreaks. *J Public Health Manag Pract*. (2020) 26:109–15. doi: 10.1097/PHH.0000000000000962

8. Haralambieva IH, Ovsyannikova IG, O'byrne M, Pankratz VS, Jacobson RM, Poland GA. A large observational study to concurrently assess persistence of measles specific B-cell and T-cell immunity in individuals following two doses of MMR vaccine. *Vaccine*. (2011) 29:4485–91. doi: 10.1016/j.vaccine.2011.04.037

9. Voigt EA, Ovsyannikova IG, Haralambieva IH, Kennedy RB, Larrabee BR, Schaid DJ, et al. Genetically defined race, but not sex, is associated with higher humoral and cellular immune responses to measles vaccination. *Vaccine*. (2016) 34:4913–9. doi: 10.1016/j.vaccine.2016.08.060

10. Haralambieva IH, Ovsyannikova IG, Kennedy RB, Larrabee BR, Zimmermann MT, Grill DE, et al. Genome-wide associations of CD46 and IFI44L genetic variants with neutralizing antibody response to measles vaccine. *Hum Genet*. (2017) 136:421–35. doi: 10.1007/s00439-017-1768-9

11. Haralambieva IH, Ovsyannikova IG, Kennedy RB, Goergen KM, Grill DE, Chen MH, et al. Rubella virus-specific humoral immune responses and their interrelationships before and after a third dose of measles-mumps-rubella vaccine in women of childbearing age. *Vaccine*. (2020) 38:1249–57. doi: 10.1016/j.vaccine.2019.11.004

12. Haralambieva IH, Quach HQ, Ovsyannikova IG, Goergen KM, Grill DE, Poland GA, et al. T cell transcriptional signatures of influenza A/H3N2 antibody response to high dose influenza and adjuvanted influenza vaccine in older adults. *Viruses*. (2022) 14 (12):2763. doi: 10.3390/v14122763

13. Quach HQ, Chen J, Monroe JM, Ratishvili T, Warner ND, Grill DE, et al. The influence of sex, body mass index, and age on cellular and humoral immune responses against measles after a third dose of measles-mumps-rubella vaccine. *J Infect Dis*. (2022) 227:141–50. doi: 10.1093/infdis/jiac351

14. Kalari KR, Nair AA, Bhavsar JD, O'brien DR, Davila JI, Bockol MA, et al. MAP-RSeq: mayo analysis pipeline for RNA sequencing. *BMC Bioinf*. (2014) 15:224. doi: 10.1186/1471-2105-15-224

15. Hansen KD, Irizarry RA, Wu Z. Removing technical variability in RNA-seq data using conditional quantile normalization. *Biostatistics*. (2012) 13:204–16. doi: 10.1093/biostatistics/kxr054

16. Huang J, Bai L, Cui B, Wu L, Wang L, An Z, et al. Leveraging biological and statistical covariates improves the detection power in epigenome-wide association testing. *Genome Biol*. (2020) 21:88. doi: 10.1186/s13059-020-02001-7

17. Zhang X, Chen J. Covariate adaptive false discovery rate control with applications to omics-wide multiple testing. *J Am Stat Assoc*. (2022) 117:411–27. doi: 10.1080/01621459.2020.1783273

18. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U.S.A*. (2005) 102:15545–50. doi: 10.1073/pnas.0506580102

19. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics*. (2012) 16:284–7. doi: 10.1089/omi.2011.0118

20. Benjamini Y, Hochberg Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J R Stat Society Ser B (Methodological)*. (1995) 57:289–300. doi: 10.1111/j.2517-6161.1995.tb02031.x

21. Priya S, Burns MB, Ward T, Mars R, Adamowicz B, Lock EF, et al. Identification of shared and disease-specific host gene-microbiome associations across human diseases using multi-omic integration. *Nat Microbiol*. (2022) 7:780–95. doi: 10.1038/s41564-022-01121-z

22. Witten DM, Tibshirani R, Hastie T. A penalized matrix decomposition, with applications to sparse principal components and canonical correlation analysis. *Biostatistics*. (2009) 10:515–34. doi: 10.1093/biostatistics/kxp008

23. Tibshirani R. Regression shrinkage and selection via the lasso. *J R Stat Society Ser B (Methodological)*. (1996) 58:267–88. doi: 10.1111/j.2517-6161.1996.tb02080.x

24. Querec TD, Akondy RS, Lee EK, Cao W, Nakaya HI, Teuwen D, et al. Systems biology approach predicts immunogenicity of the yellow fever vaccine in humans. *Nat Immunol*. (2009) 10:116–25. doi: 10.1038/ni.1688

25. Nakaya HI, Hagan T, Duraisingham SS, Lee EK, Kwissa M, Rouphael N, et al. Systems analysis of immunity to influenza vaccination across multiple years and in diverse populations reveals shared molecular signatures. *Immunity*. (2015) 43:1186–98. doi: 10.1016/j.immuni.2015.11.012

26. Popper SJ, Strouts FR, Lindow JC, Cheng HK, Montoya M, Balmaseda A, et al. Early transcriptional responses after dengue vaccination mirror the response to natural infection and predict neutralizing antibody titers. *J Infect Dis*. (2018) 218:1911–21. doi: 10.1093/infdis/jiy434

27. Hagan T, Gerritsen B, Tomalin LE, Fourati S, Mulè MP, Chawla DG, et al. Transcriptional atlas of the human immune response to 13 vaccines reveals a common predictor of vaccine-induced antibody responses. *Nat Immunol*. (2022) 23:1788–98. doi: 10.1101/2022.04.20.488939

28. Pulendran B, Li S, Nakaya HI. Systems vaccinology. *Immunity*. (2010) 33:516–29. doi: 10.1016/j.immuni.2010.10.006

29. Zhong Y, Zhang X, Chong W. Interleukin-24 immunobiology and its roles in inflammatory diseases. *Int J Mol Sci*. (2022) 23(2):627. doi: 10.3390/ijms23020627

30. Maarof G, Bouchet-Delbos L, Gary-Gouy H, Durand-Gasselin I, Krzysiek R, Dalloul A. Interleukin-24 inhibits the plasma cell differentiation program in human germinal center B cells. *Blood*. (2010) 115:1718–26. doi: 10.1182/blood-2009-05-220251

31. Beute JE, Kim AY, Park JJ, Yang A, Torres-Shafer K, Mullins DW, et al. The IL-20RB receptor and the IL-20 signaling pathway in regulating host defense in oral mucosal candidiasis. *Front Cell Infect Microbiol*. (2022) 12:979701. doi: 10.3389/fcimb.2022.979701

32. Haralambieva IH, Zimmermann MT, Ovsyannikova IG, Grill DE, Oberg AL, Kennedy RB, et al. Whole transcriptome profiling identifies CD93 and other plasma cell survival factor genes associated with measles-specific antibody response after vaccination. *PloS One*. (2016) 11:e0160970. doi: 10.1371/journal.pone.0160970

33. Kreis S, Philippidou D, Margue C, Rolvering C, Haan C, Dumoutier L, et al. Recombinant interleukin-24 lacks apoptosis-inducing properties in melanoma cells. *PloS One*. (2007) 2:e1300. doi: 10.1371/journal.pone.0001300

34. Wensveen FM, Derks IA, Van Gisbergen KP, De Bruin AM, Meijers JC, Yigittop H, et al. BH3-only protein Noxa regulates apoptosis in activated B cells and controls high-affinity antibody formation. *Blood*. (2012) 119:1440–9. doi: 10.1182/blood-2011-09-378877

35. Wensveen FM, Slinger E, Van Attekum MH, Brink R, Eldering E. Antigen-affinity controls pre-germinal center B cell selection by promoting Mcl-1 induction through BAFF receptor signaling. *Sci Rep*. (2016) 6:35673. doi: 10.1038/srep35673

36. Nakagawa R, Calado DP. Positive selection in the light zone of germinal centers. *Front Immunol*. (2021) 12:661678. doi: 10.3389/fimmu.2021.661678

37. Kaku H, Rothstein TL. Fas apoptosis inhibitory molecule expression in B cells is regulated through IRF4 in a feed-forward mechanism. *J Immunol*. (2009) 183:5575–81. doi: 10.4049/jimmunol.0901988

38. Naderi A, Liu J, Bennett IC. BEX2 regulates mitochondrial apoptosis and G1 cell cycle in breast cancer. *Int J Cancer*. (2010) 126:1596–610. doi: 10.1002/ijc.24866

39. Huo J, Xu S, Lam KP. FAIM: an antagonist of fas-killing and beyond. *Cells*. (2019) 8(6):541. doi: 10.3390/cells8060541

40. Kaser A, Dunzendorfer S, Offner FA, Ludwiczek O, Enrich B, Koch RO, et al. B lymphocyte-derived IL-16 attracts dendritic cells and Th cells. *J Immunol*. (2000) 165:2474–80. doi: 10.4049/jimmunol.165.5.2474

41. Zou X, Sun G, Huo F, Chang L, Yang W. The role of dendritic cells in the differentiation of T follicular helper cells. *J Immunol Res*. (2018) 2018:7281453. doi: 10.1155/2018/7281453

42. Karaba AH, Zhu X, Benner SE, Akinde O, Eby Y, Wang KH, et al. Higher proinflammatory cytokines are associated with increased antibody titer after a third dose of SARS-coV-2 vaccine in solid organ transplant recipients. *Transplantation*. (2022) 106:835–41. doi: 10.1097/TP.0000000000004057

43. Hui X, Cao L, Xu T, Zhao L, Huang K, Zou Z, et al. PSMD12-mediated M1 ubiquitination of influenza A virus at K102 regulates viral replication. *J Virol*. (2022) 96: e0078622. doi: 10.1128/jvi.00786-22

44. Nakaya HI, Wrammert J, Lee EK, Racioppi L, Marie-Kunze S, Haining WN, et al. Systems biology of vaccination for seasonal influenza in humans. *Nat Immunol*. (2011) 12:786–95. doi: 10.1038/ni.2067

45. Pulendran B, Oh JZ, Nakaya HI, Ravindran R, Kazmin DA. Immunity to viruses: learning from successful human vaccines. *Immunol Rev*. (2013) 255:243–55. doi: 10.1111/imr.12099

46. HIPC-CHI Signatures Project Team and HIPC-I Consortium. Multicohort analysis reveals baseline transcriptional predictors of influenza vaccination responses. *Sci Immunol*. (2017) 2(14). doi: 10.1126/sciimmunol.aal4656

47. Furman D, Jojic V, Kidd B, Shen-Orr S, Price J, Jarrell J, et al. Apoptosis and other immune biomarkers predict influenza vaccine responsiveness. *Mol Syst Biol*. (2013) 9:659. doi: 10.1038/msb.2013.15

48. Jounai N, Kobiyama K, Takeshita F, Ishii KJ. Recognition of damage-associated molecular patterns related to nucleic acids during inflammation and vaccination. *Front Cell Infect Microbiol*. (2012) 2:168. doi: 10.3389/fcimb.2012.00168

49. Iurescia S, Fioretti D, Rinaldi M. Targeting cytosolic nucleic acid-sensing pathways for cancer immunotherapies. *Front Immunol*. (2018) 9:711. doi: 10.3389/fimmu.2018.00711

50. Poland GA, Ovsyannikova IG, Kennedy RB. Personalized vaccinology: A review. *Vaccine*. (2018) 36:5350–7. doi: 10.1016/j.vaccine.2017.07.062

51. Pezeshki A, Ovsyannikova IG, Mckinney BA, Poland GA, Kennedy RB. The role of systems biology approaches in determining molecular signatures for the development of more effective vaccines. *Expert Rev Vaccines*. (2019) 18:253–67. doi: 10.1080/14760584.2019.1575208

frontiers | Frontiers in Immunology

TYPE Original Research
PUBLISHED 08 April 2024
DOI 10.3389/fimmu.2024.1331474

OPEN ACCESS

EDITED BY
José Roberto Mineo,
Federal University of Uberlandia, Brazil

REVIEWED BY
Giorgia Moschetti,
University of Milan, Italy
Kelly Rausch,
National Institute of Allergy and Infectious Diseases (NIH), United States
Alexandra Jane Spencer,
The University of Newcastle, Australia

*CORRESPONDENCE
Eduardo L. V. Silveira
eduardosilveira@usp.br
Irene S. Soares
isoares@usp.br

†PRESENT ADDRESS
Katia S. Françoso,
IQVIA, São Paulo, Brazil
Rodolfo F. Marques,
Department of Parasitology, Institute of Biomedical Sciences, University of São Paulo, São Paulo, Brazil

‡These authors share last authorship

RECEIVED 01 November 2023
ACCEPTED 18 March 2024
PUBLISHED 08 April 2024

CITATION
Costa-Gouvea TBL, Françoso KS, Marques RF, Gimenez AM, Faria ACM, Cariste LM, Dominguez MR, Vasconcelos JRC, Nakaya HI, Silveira ELV and Soares IS (2024) Poly I:C elicits broader and stronger humoral and cellular responses to a *Plasmodium vivax* circumsporozoite protein malaria vaccine than Alhydrogel in mice.
Front. Immunol. 15:1331474.
doi: 10.3389/fimmu.2024.1331474

Poly I:C elicits broader and stronger humoral and cellular responses to a *Plasmodium vivax* circumsporozoite protein malaria vaccine than Alhydrogel in mice

Tiffany B. L. Costa-Gouvea[1], Katia S. Françoso[1†], Rodolfo F. Marques[1†], Alba Marina Gimenez[1], Ana C. M. Faria[1], Leonardo M. Cariste[2], Mariana R. Dominguez[1], José Ronnie C. Vasconcelos[2], Helder I. Nakaya[1,3,4], Eduardo L. V. Silveira[1*‡] and Irene S. Soares[1*‡]

[1]Department of Clinical and Toxicological Analyses, School of Pharmaceutical Sciences, University of São Paulo, São Paulo, Brazil, [2]Laboratório de Vacinas Recombinantes, Departamento de Biociências, Universidade Federal de São Paulo, Santos, Brazil, [3]Institut Pasteur São Paulo, São Paulo, Brazil, [4]Hospital Israelita Albert Einstein, São Paulo, Brazil

Malaria remains a global health challenge, necessitating the development of effective vaccines. The RTS,S vaccination prevents *Plasmodium falciparum* (Pf) malaria but is ineffective against *Plasmodium vivax* (Pv) disease. Herein, we evaluated the murine immunogenicity of a recombinant PvCSP incorporating prevalent polymorphisms, adjuvanted with Alhydrogel or Poly I:C. Both formulations induced prolonged IgG responses, with IgG1 dominance by the Alhydrogel group and high titers of all IgG isotypes by the Poly I:C counterpart. Poly I:C-adjuvanted vaccination increased splenic plasma cells, terminally-differentiated memory cells (MBCs), and precursors relative to the Alhydrogel-combined immunization. Splenic B-cells from Poly I:C-vaccinated mice revealed an antibody-secreting cell- and MBC-differentiating gene expression profile. Biological processes such as antibody folding and secretion were highlighted by the Poly I:C-adjuvanted vaccination. These findings underscore the potential of Poly I:C to strengthen immune responses against Pv malaria.

KEYWORDS

adjuvant, TLR3-ligand, malaria, antibody-secreting cells, memory B cells

1 Introduction

Malaria continues to exert a substantial global health burden in tropical and subtropical regions worldwide. According to the World Health Organization (WHO), this disease affected an alarming 3.2 billion individuals in 84 countries in 2021, highlighting that 40% of the world population live in areas at risk of infection. Nearly, 620,000 individuals were killed, especially children, by this illness in the African sub-Saharan region. Among the *Plasmodium* parasites capable of transmitting malaria to humans, five species stand out: *Plasmodium falciparum* (Pf), *Plasmodium vivax* (Pv), *Plasmodium ovale*, *Plasmodium malariae*, and *Plasmodium knowlesi*. Pf, the deadliest of these species, commands attention, but Pv, with its wide distribution and status as the second most prevalent species, presents unique challenges. Contrary to historical perceptions of Pv malaria as benign, recent observations reveal severe symptoms, including cerebral damage, acute kidney injury, anemia, and respiratory complications in afflicted individuals (1). Notably, data from the WHO indicate 4.9 million Pv infections diagnosed annually in Asia, the Western Pacific, the Mediterranean, Central, and South America (2). Adding complexity to the Pv malaria landscape, the parasite can establish dormant hypnozoites in the liver, which may reactivate and lead to recurrent malaria episodes (3).

Malaria elimination and, ultimately, eradication require a multifaceted approach. While vector management and timely diagnostics and treatment remain pivotal, the development of a protective and universally effective malaria vaccine stands as a critical objective long pursued by the scientific community. The circumsporozoite protein (CSP), expressed abundantly on *Plasmodium* sporozoites during the pre-erythrocytic stage of infection, has emerged as a leading vaccine candidate (4). Its central-repeat portion, the most immunogenic region, has demonstrated the ability to generate antibodies capable of neutralizing sporozoites, thereby inhibiting hepatocyte invasion and preventing subsequent morbidity and mortality. Due to the antigen density in the blood-stage of infection and ability to evade infection, residents of malaria-endemic regions tend to develop an increased frequency of antibody-secreting cells (ASCs) and memory B cells (MBCs) specific to non-CSP targets over CSP (reviewed by 5). To overcome this issue, the RTS,S vaccine was conceived. This AS01 adjuvanted-vaccine comprises virus-like particles (VLP), encoded by the hepatitis B virus antigen, expressing different portions of the Pf circumsporozoite protein (CSP): the central-repeat domain and the C-terminal region containing T-cell epitopes. While the full RTS,S vaccination displayed variable efficacy depending on the local parasitic transmission levels, its protection proved to be of limited duration (6). Importantly, high antibody titers specific to the central-repeat region of CSP have been considered RTS,S-derived correlates of protection against Pf malaria (7). Notably, children aged 5-17 months exhibited higher anti-PfCSP IgG titers and protection following a full RTS,S vaccination regimen compared to their 6-12 week-old counterparts (8). Hence, the WHO has approved the implementation of RTS,S vaccination in malaria endemic areas of the African sub-Saharan region (9). However, the central-repeat region of PvCSP has a particularity relative to its Pf counterpart. While Pf sporozoites display a conserved central-repeat region of CSP, polymorphisms have been associated with the Pv sporozoite origin (10–12). Despite this diversity, neutralizing antibody-specific epitopes have been identified within the PvCSP central-repeat region (13, 14), further emphasizing the need for a universal vaccine against Pv malaria.

In the pursuit of a malaria vivax vaccine, two distinct approaches have been explored: PvCSP-derived peptides and virus-like particles (VLPs). The former has demonstrated safety and immunogenicity, stimulating both humoral and cellular responses in a naive population (15). Additionally, the Qβ-peptide platform has induced robust humoral responses and protection against minimal PvCSP peptides (16). On the other hand, VLPs consist of a vector system to display foreign antigens as viral to the host immune system. This strategy has been extensively evaluated, being remarkably effective in generating protection in numerous animal models of infections, including malarial Pv sporozoites. In the latter, immunization with VLP-expressing Rv21 provided a high degree of protection against virulent Pv sporozoite challenges in mice, with Rv21-specific IgG2a antibodies associated with protection, even in the absence of PvCSP-specific T cell responses (17). Moreover, our group revealed that the Poly I:C-adjuvanted immunization with a recombinant PvCSP, encoding its central-repeat region composed by sequences of the 3 major alleles (VK210, VK247, and *P. vivax*-like) and the C-terminal region, elicited high and long-lasting IgG responses against all alleles in mice (18). Overall, this immunization conferred partial protection against parasitic challenges with transgenic *P. berghei* (Pb) sporozoites expressing VK210 or VK247 or *P. vivax*-like PvCSP alleles in their central-repeat region (18–20). Also, the fusion of these 3 PvCSP alleles with the mumps viral nucleocapsid protein formed stable nucleocapsid-like particles (NLP) and protected mice against a malarial challenge with transgenic Pb sporozoites expressing VK210 when combined with Poly I:C (21). However, the precise mechanisms of protection associated with these vaccines remain elusive.

The adjuvant selection is a critical step in vaccine development, with multiple adjuvants described, some advancing to clinical trials, and a few approved for human use. Among them, aluminum salts are widely used adjuvants, comprising amorphous aluminum hydroxyphosphate sulfate, aluminum phosphate, potassium aluminum sulfate, and aluminum hydroxide (including Alhydrogel). Regarding their adjuvant properties, aluminum salts were initially thought to present a slow and continuous antigen release (depot effect) to recruit antigen-presenting cells (22) and eosinophils to the inoculum site (23). Nowadays, it is accepted that their mechanism of action is linked to the activation of NLRP3 inflammasome (24). More specifically, aluminum salts are phagocyted by dendritic cells (DCs) at the injection site, leading to their lysosome blockade and necrosis. Monosodium urate derived from a damage-associated molecular pattern, such as uric acid, can also inhibit DC lysosomes, facilitating the release of antigens and cathepsin B in those necrotic cells. Finally, cathepsin B stimulates the potassium flux that triggers the NLRP3 inflammasome (25–27). Another promising adjuvant is the Poly

I:C, a synthetic double-stranded RNA molecule recognized by Toll-like receptor 3 (28, 29) and the cytoplasmic melanoma differentiation-associated protein-5 (MDA-5) (30). This adjuvant stimulates the production of IL-12 and type I IFN, intensifying the innate immunity (31) and vaccine-derived immune responses (32, 33). After interaction, TLR3 dimers cluster along Poly I:C, enabling TRIF recruitment (34, 35) and assembly for the proper downstream signaling through TRAF (36). Furthermore, adjuvants based on the Poly I:C structure have reached clinical trials in humans (37).

In this context, we embark on a comparative analysis, examining the humoral and cellular immune responses elicited by immunizations with yPvCSP-All$_{CT}$ epitopes combined with Poly I:C or Alhydrogel. In addition, we conduct transcriptomic analysis on splenocytes from mice vaccinated with yPvCSP-All$_{CT}$ epitopes or yNLP-PvCSP$_{CT}$ adjuvanted with Poly I:C, or Poly I:C alone, shedding light on the mechanisms underlying these B-cell responses. These findings hold the potential to enhance the development of efficient malaria vivax vaccine formulations and bring us closer to the ultimate goal of malaria eradication.

2 Materials and methods

2.1 Animals

Six to eight-week-old female C57Bl/6 mice were purchased from the mouse facility at the School of Medicine at the University of São Paulo (USP). The animals were housed under specific pathogen-free conditions at the animal facility of the School of Pharmaceutical Sciences and Biochemistry Institute, USP, with unrestricted access to water and food. All experiments and procedures were performed in accordance with guidelines approved by the local ethics committee (CEUA/FCF 055.2019-P594 and CEUA/FCF 74.2016-P531).

2.2 Production of the vaccine antigen

The yPvCSP-All$_{CT}$ epitopes recombinant protein was expressed and purified from *Pichia pastoris* yeast (y) as previously described (18), following good laboratory practices by The Biological Process Development Facility, The College of Engineering at the University of Nebraska (USA).

2.3 Immunizations and sampling

To evaluate both humoral and cellular responses, C57Bl/6 mice underwent three intramuscular (i.m.) immunizations with a 2-week interval between each dose. Each vaccine dose consisted of 10 micrograms of yPvCSP-All$_{CT}$ epitopes adjuvanted with 50 micrograms of Poly I:C HMW (Invivogen) or a 1:1 volume of Alhydrogel (Invivogen), totaling 100 microliters. Half of this volume was administered into each thigh muscle. Plasma samples were collected from immunized animals one day before each vaccination dose through submandibular vein puncture. To investigate the Poly I:C effect on the splenic transcriptome of vaccinees, C57Bl/6 mice were immunized three times, two-weeks apart, with 10 micrograms of recombinant protein (yPvCSP-All$_{CT}$ epitopes or yNLP-PvCSP$_{CT}$) adjuvanted with 50 micrograms of Poly I:C HMW (Invivogen) in both cases via the subcutaneous (s.c.) route (38). Spleens were excised after different time points after the 2nd or 3rd vaccine doses for the analysis of cellular responses or transcriptome.

2.4 ELISA

Enzyme-linked immunosorbent assays (ELISAs) were conducted to determine titers of plasma IgG antibodies and their isotypes (IgG1, IgG2b, IgG2c, and IgG3) specific to the vaccine antigen (yPvCSP-All$_{CT}$ epitopes). These assays followed a standard operating procedure (SOP) developed by the Clinic Parasitology Laboratory staff (led by Dr. Irene Soares, School of Pharmaceutical Sciences at USP) with modifications. Briefly, ELISA plate wells (Costar high-binding - REF 3590) were coated with 1μg/mL of the recombinant protein used in immunization (yPvCSP-All$_{CT}$ epitopes). Following overnight incubation at 4°C, plate wells were washed four times with PBS and four times with PBS containing 0.5% Tween 20 (0.5% PBS-T20). Subsequently, they were blocked with a 2-hour incubation in blocking solution (PBS supplemented with 10% FBS) at room temperature. Plasma serial dilutions from immunized mice, ranging from 1:100 in blocking solution, were individually added to each plate well and incubated for 90 minutes at room temperature. Plate wells were washed four times with 0.5% PBS-T20, followed by a 90-minute incubation with anti-mouse IgG, IgG1, IgG2b, IgG2c, or IgG3 antibodies conjugated with peroxidase (Southern Technologies, Chattanooga, TN, USA) diluted 1:3,000 in blocking solution at room temperature and in the dark. The final washing steps included four washes with 0.5% PBS-T20 and four washes with PBS. Revelation was carried out using 1mg/mL of O-phenylenediamine (OPD) diluted in phosphate-citrate buffer (pH 5.0) containing 0.03% hydrogen peroxide. The addition of 4N sulfuric acid to each plate well halted the reaction. Plates were immediately read in an ELISA reader (Awareness Technology, model Stat Fax 3200, USA) at an optical density of 492 nm. We considered the end-point titer of a tested sample when its respective dilution presented an optical density (OD) value equal or higher than three-times the blank counterpart.

To estimate the avidity of vaccine-derived antibodies, we conducted an ELISA as described above with the following modifications. After the 90-minute incubation with selected dilutions of day 90-derived plasma samples that generated optical density ratios (450nm/630nm) nearly 1.0, plate wells were washed twice with 0.5% PBS-T20, followed by two washing steps with PBS. Different urea concentrations (6 M, 2 M, and 0.66 M), diluted in PBS, were individually added to each plate well and incubated for 30 minutes at room temperature. Plate wells were washed twice with PBS, followed by the incubation with peroxidase-conjugated anti-mouse IgG antibodies as described above. Values corresponding to the plate wells incubated with no urea represented maximum antibody avidity.

2.5 Measuring spleen areas

To estimate the size of the spleen areas, we used the ImageJ software and performed the following steps: 1) A picture of a murine spleen was always taken with a ruler on its side; 2) Image was duplicated, gray-scale transformed (8-bit images), and had its scale adjusted to cm^2 with the aid of a line of known length; 3) Image was cropped, had its defective region segmented through a manual-adjusting threshold, and the respective remaining area was measured.

2.6 ELISPOT

To enumerate antibody-secreting cells specific to the yPvCSP-All_{CT} epitopes recombinant protein used in immunization, the enzyme-linked immunosorbent spot (ELISPOT) assay was employed, following a previously described protocol (39) with modifications. Briefly, 10 μg/mL of the vaccine antigen (yPvCSP-All_{CT} epitopes) were diluted in PBS to coat individual wells of ELISPOT plates (Millipore - cat. MSHAN4B50). After overnight incubation at 4°C, plate wells were washed four times with PBS containing 0.05% Tween 20 (0.05% PBS-T20), followed by four washes with PBS. Plate wells were blocked for 2 hours with RPMI 1640 cell culture medium supplemented with 10% FBS (blocking solution) in a 5% CO_2 incubator at 37°C. After blocking, the solution was removed, and 10^6 splenocytes from each immunized mouse were diluted in blocking solution and added to the first-row wells of the ELISPOT plates. Serial cell dilutions, with a 3-fold factor, were performed across the remaining rows, and the plates were incubated overnight at 37°C in a 5% CO_2 incubator. Subsequently, cells were removed from the ELISPOT plates, and wells were washed four times with 0.05% PBS-T20. An anti-mouse IgG secondary antibody conjugated with biotin (Thermo Fisher Scientific - Cat. B2763), diluted 1:1,000 in PBS containing 0.05% Tween 20 and 2% FBS, was added to the plate wells and incubated for 90 minutes at room temperature. Plate wells were washed four times with 0.05% PBS-T20 and incubated with Avidin-D-HRP (Vector labs), diluted 1:1,000 in 1X PBS containing 0.05% Tween 20 and 2% FBS, for 3 hours in the dark at room temperature. Following incubation, plate wells were washed four times with 0.05% PBS-T20 and four times with PBS. Revelation was carried out by adding the 3-amino-9-ethyl carbazole (AEC) substrate (BD Cat. # 551015) to the plate wells as recommended by the manufacturer. Plate wells were washed with running water and dried before images were obtained using an AID ELISPOT plate reader (KS ELISPOT, Zeiss, Oberkochen, Germany).

2.7 Cell staining

After anesthesia and euthanasia, the spleens were removed and macerated in PBS. Red blood cells (RBCs) were eliminated after a 5-minute incubation with the Ack lysis buffer (Lonza) at room temperature. The remaining splenocytes were washed twice with PBS supplemented with 2% FBS (PBS-2% FBS) before staining for distinct B cell subsets. An antibody cocktail was added to the samples for 30 minutes in the dark at 4°C, including anti-B220-APC-Cy7 (clone RA3-6B2 - BD), anti-CD3-FITC (clone 17A2 - Biolegend), anti-F4/80-FITC (clone BM8 - Biolegend), anti-CD138-PE (clone 281.2 - Biolegend), and anti-CD38-APC (clone 90 - eBioscience). Stained cells were washed twice with PBS-2% FBS and fixed with a 4% paraformaldehyde solution. Event acquisition was performed using a FACSCelesta (BD), and data analysis was conducted with FlowJo software.

2.8 RNA extraction, cDNA library preparation, and sequencing

Mice immunized with yPvCSP-All_{CT} epitopes or yNLP-$PvCSP_{CT}$ adjuvanted with Poly I:C, or Poly I:C alone had their spleens excised two weeks after the immunization regimen as well as naive mice. Splenic B-cells were purified using MagniSort ™Mouse B cell Enrichment (ThermoFisher Scientific), resuspended in RNAlater solution (ThermoFisher Scientific), and stored at -80°C until use. Total RNA was extracted using the Quick - RNA Miniprep kit (Zymo Research, USA) following the manufacturer's instructions. RNA integrity was verified for each sample using the Agilent 2100 BioAnalyzer and Agilent RNA 6000 Nano Chips (Agilent). mRNA preparation was performed using the rRNA depletion technique with the Agilent DNA 1000 kit and Agilent 2100 BioAnalyzer equipment. cDNA library preparation and sequencing were conducted by Quick Biology Inc (Pasadena, CA, USA) using the HiSeq 4000 equipment, generating approximately 24 million reads.

2.9 Systems biology analysis

Differentially expressed genes (DEGs) were identified using the edgeR program (40). A gene was considered differentially expressed when the p-value was < 0.05 and the fold change (FC) was > 1.5 times compared to naive mice. Functional enrichment analysis utilized the Reactome database (41) through the EnrichR tool (http://amp.pharm.mssm.edu/Enrichr/), with an adjusted p-value < 0.05 indicating statistically significant enrichment. Protein-protein interaction networks were constructed using the NetworkAnalyst 3.0 platform (42) with IMEX interactome curated from the InnateDB database (43), considering only experimental evidence and a 900 confidence-score cutoff. Transcription factor-DEG interaction networks were also defined using the NetworkAnalyst 3.0 platform with the ENCODE ChIP-seq data package following set-up: peak intensity signal <500 and predicted regulatory potential score <1 (through the BETA Minus algorithm). Based on particular parameters, such as degree and betweenness centrality, the resulting networks were visualized with Cytoscape

version 3.7.2 (44), and the subnetworks illustrated only immunity-related pathways. DEGs exclusively detected in mice immunized with yPvCSP-All$_{CT}$ epitopes + Poly I:C were highlighted in red, while DEG-associated transcription factors or DEG-relative protein-associated proteins were represented in purple or yellow, respectively.

2.10 Statistical analysis

Statistical analyses were performed using GraphPad Prism for Windows, version 6.0 (GraphPad Software, Inc., La Jolla, CA, USA) using a Two-way ANOVA with multiple comparisons through Sidak's test, computing confidence interval and significance. A p-value ($p<0.05$) indicated a significant difference between the two groups evaluated.

3 Results

3.1 Poly I:C-adjuvanted vaccination induced a balanced and durable IgG response compared to Alhydrogel

To assess the immunogenicity of a malaria vaccine targeting PvCSP (yPvCSP-All$_{CT}$ epitopes) with different adjuvants, we conducted a comprehensive study involving 12 C57Bl/6 mice immunized via intramuscular injection with three doses administered at 14-day intervals. The vaccine formulations were combined with either Poly I:C or Alhydrogel as adjuvants. We closely monitored the immune responses of these mice for nearly 500 consecutive days. Plasma samples were collected at various time points before, during, and after vaccination to measure total IgG titers specific to the vaccine antigen using ELISA (Figure 1A).

As expected, we observed a significant increase in IgG titers after each vaccination, regardless of the adjuvant used. Interestingly, the presence of Poly I:C as an adjuvant resulted in a slightly faster onset of the vaccine-induced humoral response compared to Alhydrogel. Specifically, Poly I:C-adjuvanted vaccination led to the peak of IgG titers at day 42, maintaining this elevated level until day 120. In contrast, the group that received Alhydrogel had a delayed peak in antibody titers, occurring at day 90 (Figure 1B). Importantly, both adjuvants induced IgG antibodies with similar avidity against the vaccine antigen (yPvCSP-All$_{CT}$ epitopes) (Supplementary Figure 1A). Antigen-specific IgG titers declined significantly by day 150 but were maintained at a certain level until day 495 for both adjuvants. This suggests that vaccination with either adjuvant can induce durable humoral responses in mice (Figure 1B).

3.2 IgG isotype profile highlights differential immune responses

The vaccine formulations elicited distinct IgG isotype profiles, shedding light on the nature of the immune response induced by

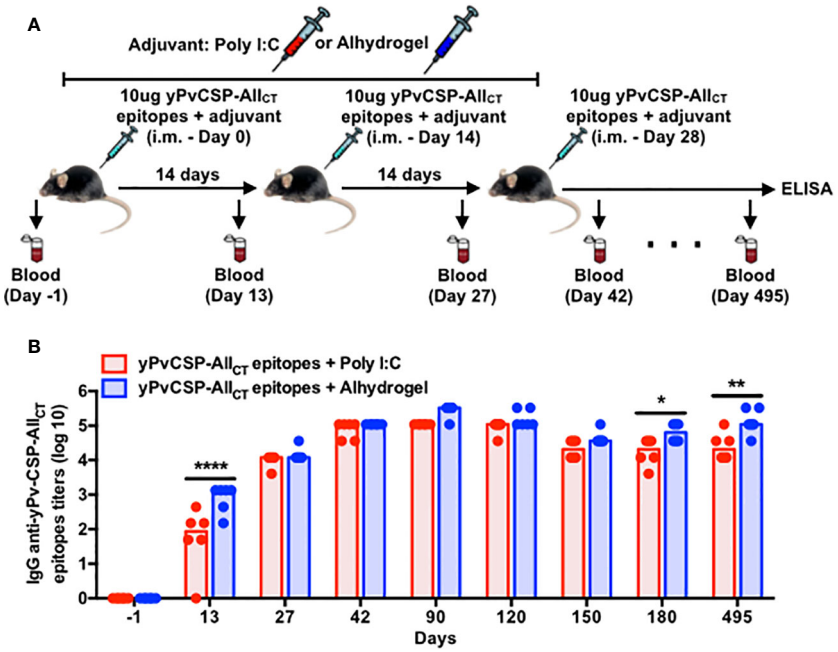


FIGURE 1
Adjuvanted-malaria vaccine specific to *P. vivax* circumsporozoite protein elicits long-lasting IgG responses in mice. **(A)** Outline of the blood draws and intramuscular vaccination with the recombinant yPvCSP-All$_{CT}$ epitopes protein combined with Poly I:C (n=6) or Alhydrogel (n=6). **(B)** IgG titers specific to the vaccine antigen were measured before, during and after vaccination in plasma samples through ELISA. Dots and columns represent individual values detected for each mouse and median, respectively. Red and blue colors indicate animals immunized with yPvCSP-All$_{CT}$ epitopes + Poly I:C or Alhydrogel, respectively. *p<0.05; **p<0.01; ****p<0.0005.

each adjuvant. Notably, IgG1 dominated the humoral response in Alhydrogel-adjuvanted vaccinees, with significantly higher titers observed at day 42 compared to those in the Poly I:C-adjuvanted group (Figure 2A). In contrast, while IgG1 displayed the highest titer among the IgG isotypes in Poly I:C-adjuvanted vaccinees, IgG2c, IgG2b, and IgG3 titers followed a hierarchical pattern, peaking also at day 42, with higher magnitudes and a more balanced IgG1/IgG2c ratio (Th1/Th2 profile) compared to Alhydrogel counterparts (Supplementary Figure 1B). These antibody titers significantly declined by day 150 (IgG1) or day 180 (IgG2b, IgG2c, and IgG3), becoming undetectable at day 495 (IgG3) in the Poly I:C-adjuvanted group. In contrast, Alhydrogel-adjuvanted vaccinees initiated the decline a bit earlier (day 90) for IgG1, but their remaining IgG isotypes maintained low titers, as observed at day 42, except for IgG3, which was undetected by day 495 (Figures 2B–D).

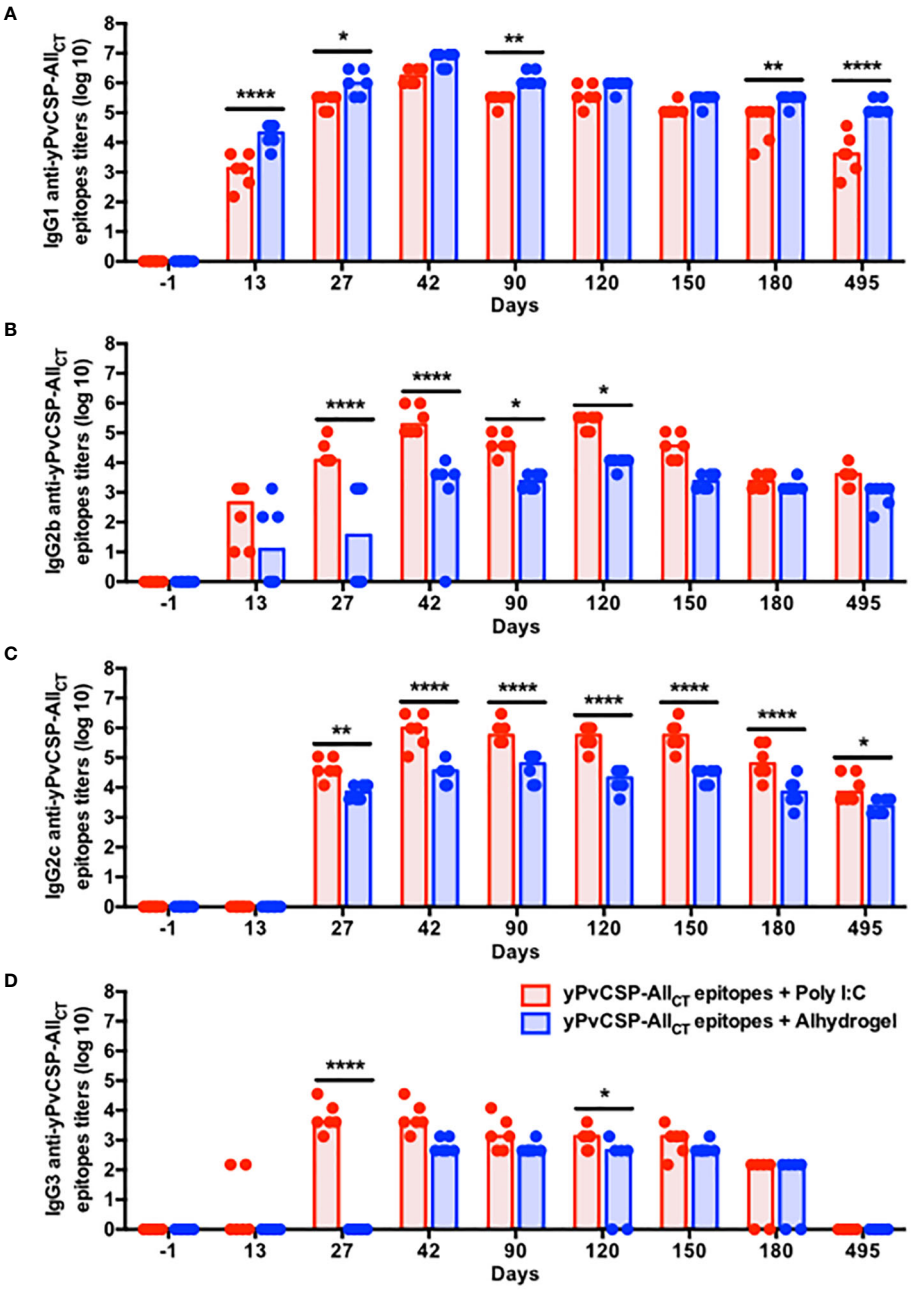


FIGURE 2
Poly I:C-adjuvanted malaria vaccine triggers a broader isotypic diversification of IgG responses via the intramuscular route in comparison to the Alhydrogel counterpart. Mice were immunized via intramuscular with the recombinant yPvCSP-All$_{CT}$ epitopes protein combined with Poly I:C (n=6) or Alhydrogel (n=6). **(A)** IgG1; **(B)** IgG2b; **(C)** IgG2c; **(D)** IgG3) titers specific to the vaccine antigen were measured before, during and after vaccination in plasma samples through ELISA. Red and blue colors indicate animals immunized with yPvCSP-All$_{CT}$ epitopes + Poly I:C or Alhydrogel, respectively. Dots and columns represent individual values detected for each mouse and median, respectively. *p<0.05; **p<0.01; ****p<0.0005.

3.3 Poly I:C-adjuvanted vaccination enhances the frequency of antibody-secreting cells and memory B cells

To investigate the impact of adjuvants on the spleen cellularity and on the frequency of B cell subsets, 18 animals were immunized with half receiving each adjuvant (Figure 3A). Disregarding the adjuvant used, spleen areas tended to increase from 2nd to final vaccination (day 5). Five days later, those organs returned to initial measures (Supplementary Figures 2A, B). Since B cells, involved with antibody responses, are the most abundant immune cells in murine spleens (45), we quantified the frequency and absolute number of several B-cell subsets at various time points following immunization with Poly I:C or Alhydrogel using flow cytometry (Supplementary Figure 3). Short-lived plasmablasts (PBs) typically follow specific kinetics upon immunization in different mammals (46–49). In this

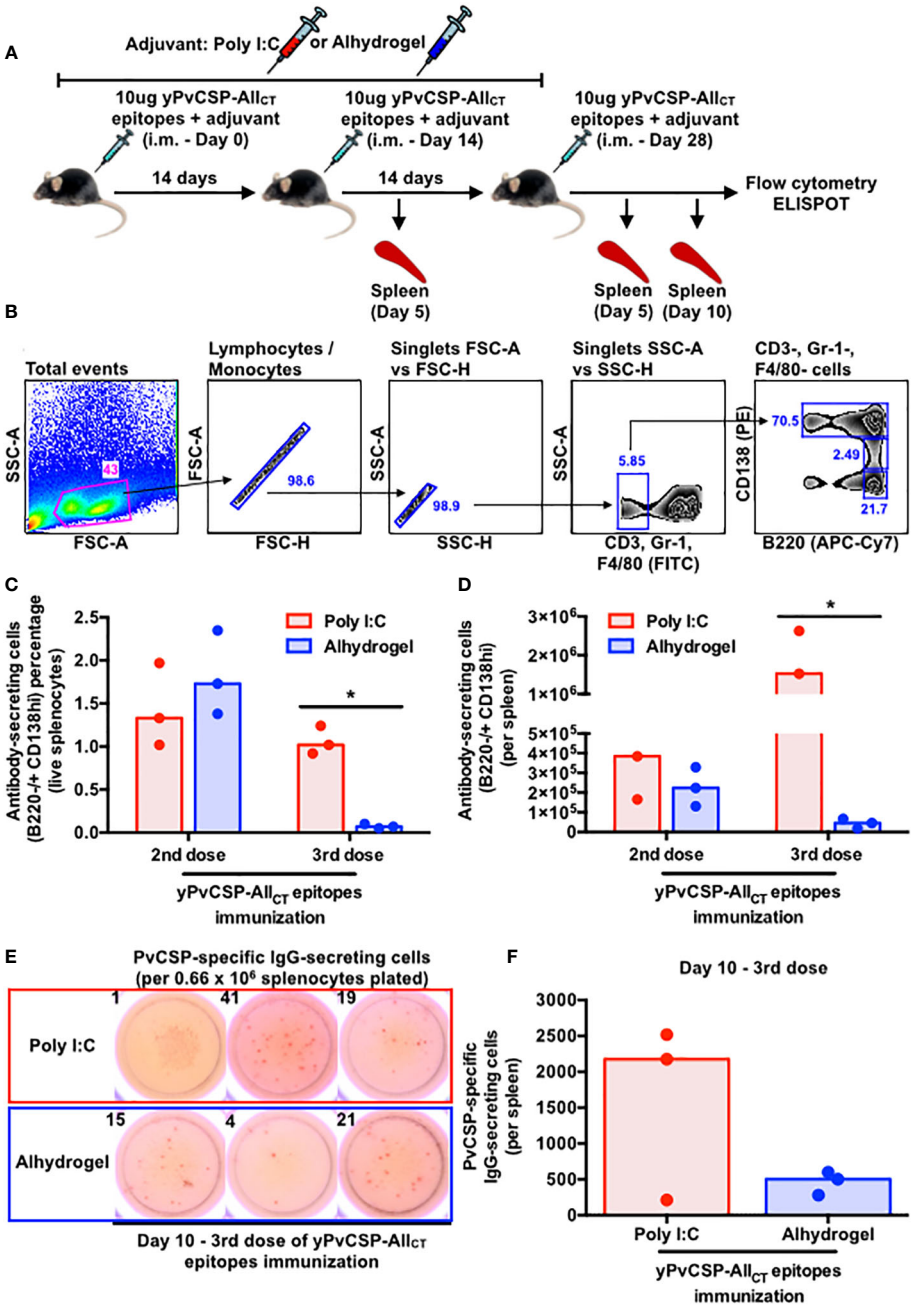


FIGURE 3
Poly I:C-adjuvanted malaria vaccine induces a more potent antibody-secreting cell response in the mouse spleen via the intramuscular route than the Alhydrogel counterpart. **(A)** Outline of intramuscular vaccination with the recombinant yPvCSP-All$_{CT}$ epitopes protein combined with Poly I:C (red - n=9) or Alhydrogel (blue - n=9) and tissue sampling (n=3 per group per time point). **(B)** Sequential gating strategy to enumerate plasma cells (PCs) through flow cytometry. **(C)** Percentage and **(D)** absolute number of splenic PCs at different time points upon vaccination through flow cytometry. **(E)** Representative images of the ELISPOT results for PvCSP-specific IgG-secreting cells at day 10 of the third vaccine dose (left panel). The numbers on top of each image indicate the quantity of spot-forming cells enumerated per well plated with 0.66 x 106 mouse splenocytes. **(F)** Magnitude of PvCSP-specific IgG-secreting cells per spleen of immunized mice (right panel). Dots and bars represent the totality of splenic PvCSP-specific IgG-secreting cells individually detected for each mouse and median, respectively. *p<0.05

study, both B cells (B220+) and PBs (B220+ CD138int CD38+) tended to increase in the spleens of mice vaccinated with Poly I:C compared to those receiving Alhydrogel, particularly after boosters (Supplementary Figures 4A–D). In contrast, Poly I:C-adjuvanted vaccinees maintained a similar percentage and count of long-lived plasma cells (PCs) at the same period, while Alhydrogel-adjuvanted vaccinees exhibited a significant decrease in both parameters at day 5 after the third immunization (Figures 3B, C). To address the specificity of these splenic antibody-secreting cells (ASCs), we enumerated IgG-secreting cells specific to the vaccine antigen at day 10 after the third vaccination using ELISPOT. The Poly I:C-adjuvanted vaccine induced a higher, though not statistically significant, number of IgG-secreting cells specific to the vaccine antigen compared to the Alhydrogel group (Figures 3D, E).

The secretion of IgG antibodies relies on the activation and differentiation of follicular B cells, which participate in germinal center (GC) reactions with follicular T cells (TFh) (reviewed by 50), into ASCs. We measured the frequency of important GC players and observed increasing trends in the percentage and absolute numbers of FoBs (B220+ CD23+), GC-Bs (B220+ CD138- CD38- GL7+) and TFh (GC-TFh (CD3+ CD4+ GL7+ CD40L+ CXCR5+) and non-GC TFh (CD3+ CD4+ GL7- CD40L+ CXCR5+)) with the last booster for both adjuvants (Supplementary Figures 3, 4E–H, 5), although without statistical significance.

Critical for the durability of vaccine-derived responses and protection, we also evaluated the frequency of memory B cell (MBC) precursors (B220+ CD138- CD38+ GL7+) and terminally-differentiated MBCs (B220+ CD138- CD38+ GL7-) in the spleens of vaccinees. A significantly lower percentage and absolute number of MBC precursors and MBCs were observed in Alhydrogel-adjuvanted vaccinees after the third vaccine dose compared to their Poly I:C-adjuvanted counterparts (Figures 4A–D).

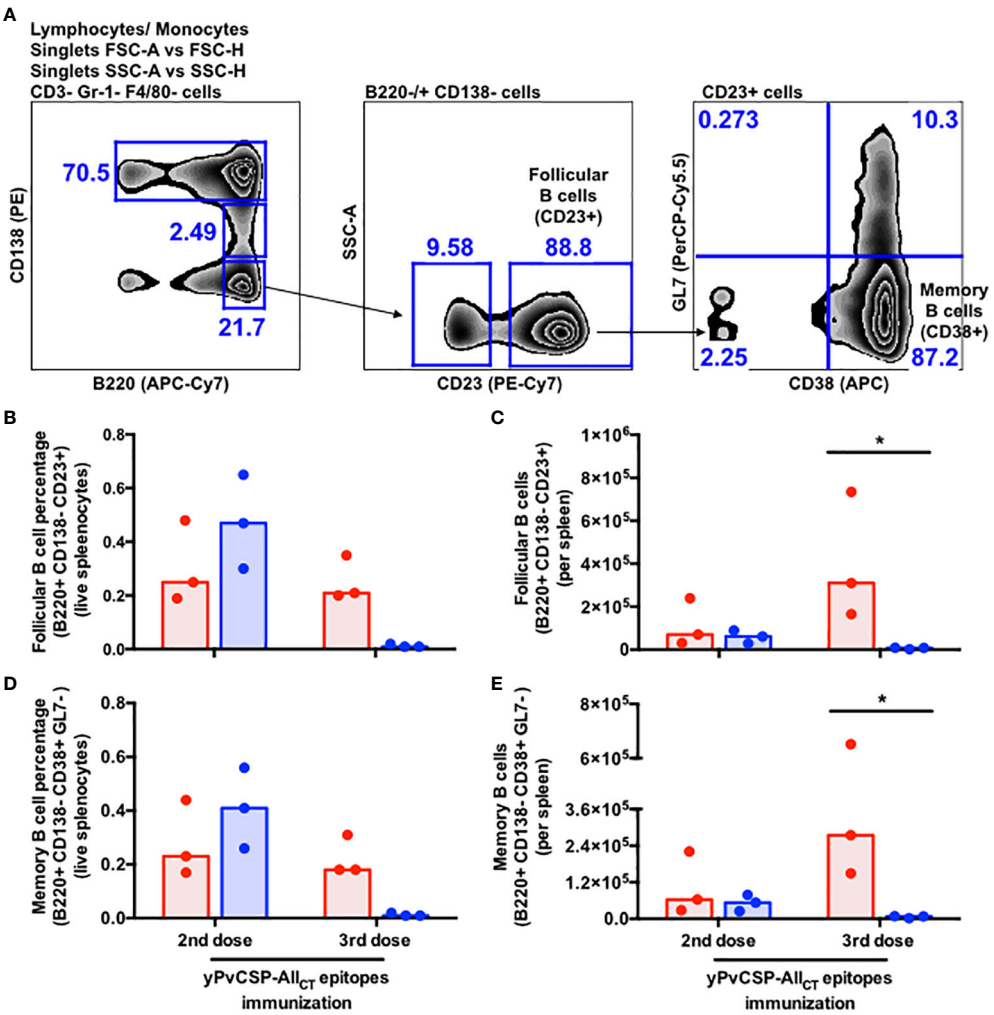


FIGURE 4
Poly I:C-adjuvanted malaria vaccine induces a stronger memory B-cell response via the intramuscular route relative to the Alhydrogel counterpart. **(A)** Sequential gating strategy to enumerate follicular B cells and memory B cells through flow cytometry. Percentage **(B, D)** absolute number of cells **(C, E)** detected in the spleen of mice at different time points upon vaccination through flow cytometry. Red and blue colors indicate animals immunized with yPvCSP-All$_{CT}$ epitopes + Poly I:C or Alhydrogel, respectively. Dots and columns represent individual values detected for each mouse and median, respectively. * p<0.05

3.4 Poly I:C-adjuvanted vaccination modulates the expression of genes associated with antibody-secreting cells and memory B cells

Subcutaneous vaccination with yPvCSP-All$_{CT}$ epitopes combined with Poly I:C demonstrated similar immunogenicity to that obtained via the intramuscular route (data not shown) and protection against transgenic *P. berghei* sporozoites expressing PvCSP alleles (VK210, VK247, or *P. vivax*-like) (18–20). Additionally, a vaccine formulation based on the fusion of the mumps viral nucleocapsid and yPvCSP-All$_{CT}$ epitopes (yNLP-PvCSPCT$_{CT}$) protected mice against parasitic challenges (21). To gain insights into the molecular mechanisms underlying these responses, we compared the splenic B-cell transcriptome of mice vaccinated with yPvCSP-All$_{CT}$ epitopes + Poly I:C, yNLP-PvCSPCT$_{CT}$ + Poly I:C, or Poly I:C alone (Figure 5A). This analysis identified nearly 120 differentially expressed genes (DEGs) that were either exclusive to each group or shared among groups (Figure 5B; Supplementary Tables 1–3). Among the 33 exclusive DEGs derived from animals immunized with yPvCSP-All$_{CT}$ epitopes + Poly I:C, 16 were upregulated, and 17 were downregulated (Figure 5C). Of these exclusive DEGs, 6 were associated with facilitating B-cell differentiation into ASCs (Col18a1, Hspa2, Pstk, S100a8, Zfp457, and Tubb4a), while others were linked to MBC generation (Gpr3, Hmgb1-rs17, and Igsf23) (Figure 5D). Gene ontology analysis indicated that these 33 exclusive DEGs were involved in processes related to cell localization, protein secretion, wound response, and cation homeostasis (Figure 5E). At the molecular level, the activities of protein dimerization and transmembrane transport were associated with these DEGs (Figure 5F). Gene networks revealed interactions between some of these DEGs, transcription factors (IRF4 and S100a8), or proteins (CamK2a and Cdk1) respectively critical for B-cell differentiation into ASCs or MBCs (Figures 5G, H).

4 Discussion

The durability of vaccine-induced immune responses is a critical factor in assessing the long-term protective efficacy of vaccination and the potential need for booster doses. Our study demonstrates that both Poly I:C and Alhydrogel adjuvants can elicit robust and long-lasting humoral responses following immunization with the yPvCSP-All$_{CT}$ epitopes formulation. Notably, anti-PvCSP IgG titers persisted for extended periods, declining only after 120 days post-vaccination and remaining stable for almost 350 days thereafter for both adjuvants (Figures 1, 2). This suggests that the number of antibody-secreting cells (ASCs), particularly plasma cells (PCs), generated by yPvCSP-All$_{CT}$ epitopes vaccination with Poly I:C or Alhydrogel does not significantly decrease in the bone marrow of vaccinated individuals within the first year, a phenomenon observed in humans vaccinated against influenza (51). Furthermore, it is plausible that PCs originating from both Poly I:C- and Alhydrogel-adjuvanted vaccinations maintain similar levels of the ZBT720 transcription factor, which is known to sustain humoral responses (52).

Different than the concern raised by the excessive amount of serum anti-CSP antibodies induced by RTS,S vaccination before completing the entire regimen, which hinder the increase of the humoral response (53), all booster doses of yPvCSP-All$_{CT}$ epitopes + Poly I:C or Alhydrogel triggered an enhancement of anti-PvCSP IgG titers (Figure 1B). Interestingly, each adjuvant induced a distinct IgG profile specific to PvCSP. The Poly I:C-adjuvanted vaccine triggered a balanced production of PvCSP-specific IgG1 and IgG2c, along with a notable IgG2b response, whereas the Alhydrogel-adjuvanted vaccine was dominated by IgG1 (Figure 2). This suggests a potential Th1/Th2 immune profile, which may be advantageous for protection against PvCSP. In comparison to the PfCSP-specific response, the RTS,S vaccination stimulates higher secretion of IgG1, and some IgG3 and IgG2 in humans, being protective when specific to the central-repeat or C-terminal region of the PfCSP. However, these antibody titers significantly wane in less than 8 months and continue to gradually decline in subsequent years. IgG2 and IgG4 have been associated with increased Pf malaria risk and are detected at lower magnitudes than IgG1 and IgG3 (54–56). Regarding the IgG subclasses induced by another malaria vaccine formulation to be implemented (R21 + Matrix-M), they remain elusive in humans. In mice, this latter vaccine elicited higher humoral and cellular responses, culminating with higher protection against transgenic sporozoites compared to R21 + Alhydrogel (57) or R21 alone (58). In this case, the non-protective R21 alone triggered an IgG1-dominated profile (Th2 type) (58) as well as our immunization with yPvCSP-All$_{CT}$ epitopes + Alhydrogel. When other adjuvants, such as SQ or LMQ, were combined with R21, they protected Balb/c mice against a malaria challenge. While the humoral response induced by R21 + SQ was dominated by IgG1 (Th2 profile), the R21+LMQ immunization resulted in comparable titers of IgG2a, IgG1, and IgG3 (balanced Th1/Th2 profile) (58). Notably, our immunization with yPvCSP-All$_{CT}$ epitopes + Poly I:C elicited a similar humoral response, Th profile, and ability to protect against a malaria challenge (18, 19) as R21+LMQ does. Considering that human IgG1 and IgG3 and murine IgG2 are cytophilic, fix complement (59) and interact with Fcγ-receptors on phagocytes, adjuvants capable of triggering distinct Th profiles can eventually facilitate protection against Pv malaria. Moreover, these functional properties of anti-PvCSP antibodies have not been explored yet.

Serum anti-CSP antibodies derived from individuals living in malaria-endemic regions or those immunized with different formulations have been shown to possess neutralizing capabilities against Pf sporozoites (reviewed by 5, 60), reduce the hypnozoite burden, and delay the onset of blood-stage Pv infection (61). Recent molecular dynamics simulations and crystallography analyses suggest that anti-PvCSP neutralizing antibodies efficiently interact with their epitopes, despite the structural disorder of the central-repeat portion of PvCSP (62). However, a non-neutralizing anti-PfCSP monoclonal antibody, isolated from immunized mice, was

10.3389/fimmu.2024.1331474

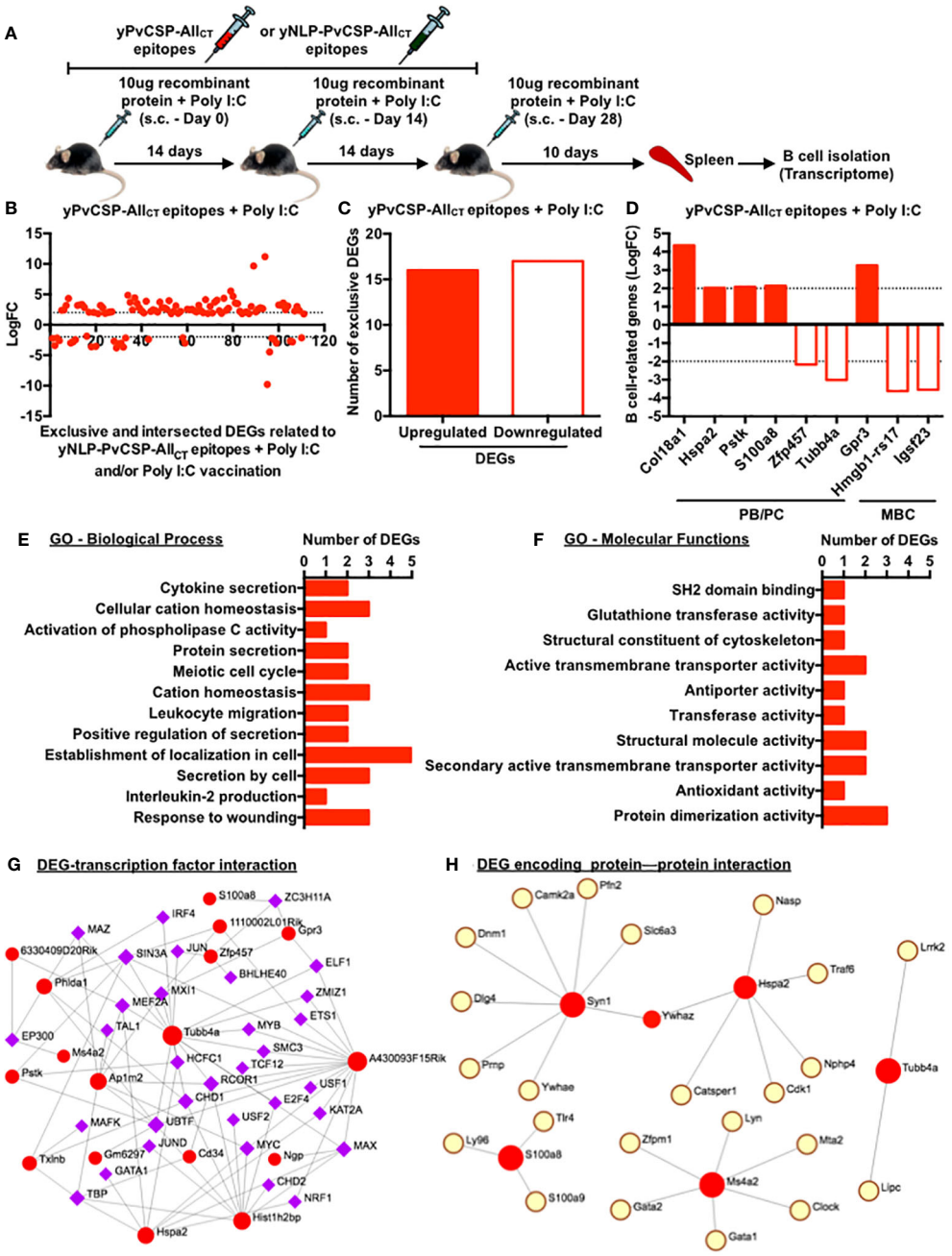


FIGURE 5
Poly I:C-adjuvanted malaria vaccine elicits modifications in the transcriptome of splenic B cells, enhancing their differentiation into antibody-secreting or/and memory B cells. **(A)** Subcutaneous immunization with yPvCSP-All$_{CT}$ epitopes + Poly I:C, yNLP-PvCSP$_{CT}$ + Poly I:C, or Poly I:C alone, number of doses and their intervals, and euthanasia time for spleen excision, B-cell isolation and freezing for further RNA extraction. **(B)** Log fold-change (FC) of differential expressed genes (DEGs) exclusively induced by the yPvCSP-All$_{CT}$ epitopes + Poly I:C vaccination or mutually induced by yPvCSP-All$_{CT}$ epitopes + Poly I:C and one of the remaining immunizations. **(C)** Number of DEGs exclusively detected in splenic B cells of mice vaccinated with yPvCSP-All$_{CT}$ epitopes + Poly I:C. **(D)** LogFC of DEGs associated with B-cell differentiation into antibody-secreting cells (PB/PC) or memory B cells (MBC) detected upon yPvCSP-All$_{CT}$ epitopes + Poly I:C vaccination. **(E)** Major biological processes and **(F)** molecular functions of splenic B-cell DEGs derived from mice vaccinated with yPvCSP-All$_{CT}$ epitopes + Poly I:C through Gene Ontology analyses. Interaction networks between B-cell-derived DEGs (red dots) elicited upon yPvCSP-All$_{CT}$ epitopes + Poly I:C vaccination with transcription factors **(G)** and DEG-encoding protein with proteins **(H)**. Dotted lines represent LogFC values ≥ -2 and ≤ 2. .

recently demonstrated to abrogate protection against Pf sporozoites, even in the presence of neutralizing counterparts (63). Given that previous subcutaneous immunizations with yPvCSP-All$_{CT}$ epitopes combined with Poly I:C provided only partial protection in mice exposed to transgenic PvCSP-expressing sporozoites (18–21), it remains unclear whether the vaccine-induced humoral response specific to yPvCSP-All$_{CT}$ epitopes includes non-neutralizing anti-PvCSP antibodies.

Another crucial mechanism of malaria immunity is the opsonization of sporozoites mediated by anti-CSP antibodies. Human anti-PfCSP IgG1 and IgG3 have been shown to interact with neutrophils via FcRIIa and FcRIII, as well as to a lesser extent with monocytes and NK cells, facilitating parasite clearance (64). Antibody-dependent complement activation and fixation are also vital components of effective immunity. Human IgG1 and IgG3 specific to the N-terminal, central-repeat, and C-terminal regions of PfCSP have been demonstrated to fix complement (65). However, it remains unexplored whether anti-PvCSP antibodies induced by yPvCSP-All$_{CT}$ epitopes immunizations can execute these functions, regardless of the adjuvant employed.

Despite the observed discrepancies in IgG responses with the two tested adjuvants, PvCSP vaccination did not result in differences in splenic sizes (Supplementary Figure 2). B cells are the most prevalent immune cells within this organ in mice, and various B cell subsets may have their frequencies altered following infection or vaccination. To elicit protective immunity against malaria, a combination of multiple B cell subsets is required. For instance, immunization with irradiated sporozoites (IrSpz), which have CSP as the immunodominant antigen (4), provides protection to several murine models of disease and humans. In mice, the IrSpz-derived response triggers an increased number of CSP-specific plasmablasts and long-lasting germinal center (GC) B cells. The functionality of that cellular response seems to be dependent on T cells, as CD28 KO mice displayed reduced numbers of GC B cells and plasmablasts, and an ensuing higher susceptibility to wild-type (WT) Spz infection (66). The blood stage of malaria is another parameter known to alter the composition of B cell subsets, increasing susceptibility to infection. Mice infected with WT Spz present reduced anti-CSP antibody titers upon the establishment of the blood stage due to an inhibition of the CSP-specific GC B cell response (67). Straight infection with infected red-blood cells also elicits a detrimental GC B cell response (68). Consequently, plasmablasts show a faster decline and only a reduced number of memory B cells (MBCs) are maintained. If mice are treated with atovaquone during the blood-stage of the infection, parasitemia is cleared and animals present a subsequent enhancement in the number of splenic B cells, GC B cells, plasmablasts and anti-CSP antibody titers as observed with IrSPz-immunized mice (67). Notably, a fine tuning for metabolites between plasmablasts and GC B cells seems to occur for prompting protection against malaria. During the blood stage of infection in mice, plasmablasts rapidly proliferate, diminishing levels of blood L-glutamine. Somehow, this scenario delays the proliferation of GC B cells, resulting in reduced numbers of MBCs and plasma cells, and higher-peak parasitemia. On the other hand, if plasmablast depletion or an L-glutamine treatment is done during the beginning of the blood stage of infection, it triggers an effective proliferation of GC B cells and follicular helper T cells, culminating with increased numbers of MBCs and plasma cells, and lower parasitemia peak (69). In this study, Poly I:C-adjuvanted vaccinees displayed significantly higher absolute numbers of PCs, follicular B cells, and terminally-differentiated MBCs compared to Alhydrogel counterparts (Figures 3, 4). Regarding PCs, both qualitative (flow cytometry) and quantitative assays (ELISPOT) exhibited similar kinetics (Figure 3), parallel to what has been observed in vaccinated macaques (47) and humans (49). However, the specificities of follicular B cells and terminally-differentiated MBCs induced by our vaccination require further investigation. Moreover, the study sheds light on the cellular aspects of immunity, indicating that Poly I:C may enhance the generation of higher-affinity memory and long-lasting protection against PvCSP relative to Alhydrogel.

The differences in the gene expression profiles of B cells between the two adjuvant groups provide valuable insights into the mechanisms underlying the observed immune responses. Beyond the DEGs identified as B-cell markers in the EMBL-EBI public data repository (Figure 5D; https://www.ebi.ac.uk/), several others were exclusively found in splenic B-cells derived from mice of the Poly I:C group, reflecting the robust B-cell response elicited by this adjuvant when compared to Alhydrogel. The enhanced and sustained humoral responses in Poly I:C-adjuvanted vaccinees may be associated with the downregulation of Syn1, which reduces its interaction with CamK2a (Figure 5H). This may hinder the transmission of calcium ions within B cells, impacting the regulation of B-cell activation and differentiation (reviewed by 70, 71). Additionally, the downregulation of Hmgb1-rs17 may contribute to the accumulation of splenic PCs and MBCs (Figures 3, 4) by inhibiting B-cell egress from lymphoid tissues, such as Peyer's patches (72). The regulation of vaccine-derived responses by regulatory T cells (Tregs) could also be affected, as indicated by the downregulation of Gm10408 and Gm14391 (Supplementary Table 1), potentially limiting their frequency or functionality in the spleens of Poly I:C-adjuvanted vaccinees (Supplementary Table 1). Other downregulated DEGs in Poly I:C-adjuvanted vaccinees represent long non-coding RNAs (Gm6297, 1110002L01Rik, 5830416I19Rik, 6330409D20Rik, and A430093F15Rik), which are more highly expressed in T cells than in B lymphocytes (https://www.ebi.ac.uk/). About the upregulated DEGs, Lilrb4 has been associated with attenuated PRDM1 expression and antibody production. It is possible that the recognition of Poly I:C by TLR3 or MDA5 may maintain Lilrb4 expression at a dysfunctional level. Additionally, Hspa2, which interacts with Cdk1 (Figure 5H), is essential for the transcriptional regulation of PC function (73). The positive expression of Col18a1 suggests signaling toward PB formation, particularly when compared to MBCs and naive B cells. Notably, this DEG also interacts with DENV proteins based on disease severity, a condition that leads to a massive PB expansion (74). Phlda1 is a transcription factor with hierarchical expression in naive B cells, followed by MBCs and PBs, and complexes with the IRF4 transcription factor (Figure 5G), a fundamental marker for ASC differentiation. S100a8 is highly expressed on the surface of B cells in patients with systemic lupus erythematosus, with its expression decreasing upon disease treatment (75). However, S100a8 displays lower expression in splenic ASCs than in bone marrow counterparts (76). Therefore, the downregulation of genes associated with B-cell

activation, calcium ion transmission, and B-cell egress, as well as the upregulation of genes involved in PC formation and ASC differentiation in the Poly I:C group, contribute to our understanding of the enhanced humoral and cellular responses elicited by this adjuvant.

The administration of several vaccine adjuvants, such as products from aluminum hydroxide, has demonstrated to be safe in humans. However, their immunogenicity is long-away off the levels displayed by other adjuvants. For instance, Poly I:C activates immune responses through TLR3 signaling that result in the IFN-α and MDA-5 production (30). In our model, this adjuvant clearly enhances humoral and cellular responses against PvCSP in such levels that immunized mice are protected from malaria challenges (18, 19). Toxicological studies have also supported our vaccination regimen as a safe immunogen (data not shown). However, analogs of Poly I:C have been preferred in clinical trials, such as Hiltonol (also called Poly I:C/L:C), due to its higher stability against serum nucleases present in the plasma of primates, and higher immunogenicity than Poly I:C (77). Thus, the establishment of a clinical trial in which individuals from *P. vivax*-endemic or non-endemic areas be vaccinated with yPvCSP-All$_{CT}$ epitopes + Hiltonol seems to be a critical and subsequent step. An important question to answer is whether the vaccinees would develop high titers of IgG against all repeat domains contained within the yPvCSP-All$_{CT}$ epitopes as observed in mice (18, 19), characteristics that attribute the universality aspect and protection to our vaccine formulation.

In conclusion, our murine model of PvCSP vaccination presents compelling evidence that Poly I:C surpasses Alhydrogel as an adjuvant, eliciting a more balanced and long-lasting humoral response, as well as a more robust cellular memory and an effective response. This provides a strong rationale for further investigation and optimization of adjuvant formulations in the pursuit of a potent and effective vaccine against *P. vivax* malaria. We believe that the insights gained from this comprehensive and longitudinal study will contribute to the accelerated development of a much-needed protective vaccine, ultimately reducing the burden of *P. vivax* malaria in endemic regions and improving global health outcomes.

Data availability statement

The transcriptomic data presented in the study are deposited in the GEO repository (https://www.ncbi.nlm.nih.gov/bioproject/), accession numbers BioProject PRJNA1092424 (ID 1092424) and BioProject PRJNA839078 (GSE GSE203218).

Ethics statement

The animal study was approved by the Animal Ethics Committee (CEUA) from the School of Pharmaceutical Sciences, University of São Paulo (CEUA/FCF 055.2019-P594 and CEUA/FCF 74.2016-P531). The study was conducted in accordance with the local legislation and institutional requirements.

Author contributions

TC-G: Project administration, Writing – review & editing, Methodology, Investigation, Formal analysis, Data curation. KF: Writing – review & editing, Visualization, Software, Project administration, Methodology, Investigation, Formal analysis, Data curation. RM: Writing – review & editing, Methodology, Investigation. AG: Writing – review & editing, Methodology, Investigation. AF: Writing – review & editing, Methodology. LC: Writing – review & editing, Methodology. MD: Writing – review & editing, Methodology. JV: Writing – review & editing, Supervision, Resources, Methodology, Funding acquisition, Formal analysis. HN: Supervision, Software, Resources, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Writing – original draft, Visualization, Writing – review & editing. ES: Writing – review & editing, Writing – original draft, Visualization, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. IS: Writing – review & editing, Writing – original draft, Visualization, Supervision, Software, Resources, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This study was supported by São Paulo Research Foundation (FAPESP) grants awarded to AG (Process 2014/18102-7), HN (Process 2018/14933-2), ES (Process 2017/11931-6), and IS (Process 2012/13032-5) and the National Council for Scientific and Technological Development (CNPq) awarded to ES (Process 304377/2021-0). ES was also supported by PIPAE (2021.1.10424.1.9) and Programa de Apoio aos Novos Docentes grants from the University of São Paulo. IS was also funded by the National Institute of Science and Technology of Vaccines (CNPq 465293/2014-0) and Finep (1208/21). TC-G was a Fapesp (Process 2019/06494-1) and CNPq fellow (Process 162652/2021-6). AF was a Fapesp fellow (Process 2019/25373-0)

Acknowledgments

We thank Dr. Esper Kallas (University of São Paulo and Butantan Institute, Brazil) and his staff for the usage and expert assistance with the ELISPOT plate reader (KS ELISPOT, Zeiss, Oberkochen, Germany). We also thank Eric Liao and the team at Quick Biology Inc. (via Science Exchange) for providing the Illumina Sequencing of mouse B-cell transcriptome and Dr. André N. A. Gonçalves (University of São Paulo and Institute Pasteur São Paulo) for expert assistance with RNA-seq preprocessing data.

Conflict of interest

RM, AG, and IS are co-inventors of the potential P. vivax vaccine evaluated in this study. The patent is under evaluation process, application number BR102022005915-2.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2024.1331474/full#supplementary-material

SUPPLEMENTARY FIGURE 1
Similar IgG avidity against the *P. vivax* circumsporozoite protein and differential T-helper (Th) cytokine response patterns triggered by Poly I:C- or Alhydrogel-adjuvanted vaccination. Red and blue colors indicate animals immunized with yPvCSP-All$_{CT}$ epitopes + Poly I:C or Alhydrogel, respectively. Plasma samples (Day 42 **(A)** and Day 90 **(B)**) from mice vaccinated with Poly I:C- or Alhydrogel-adjuvanted malaria vaccine were evaluated for **(A)** IgG and **(B)** IgG1 and IgG2c binding to the vaccine antigen in the presence of different concentrations of urea through ELISA. **(A)** Dots and error bars represent average ± SEM, respectively. **(B)** Dots and columns represent individual values detected for each mouse and their median, respectively. ** p<0.01.

SUPPLEMENTARY FIGURE 2
Similar spleen area in mice immunized with a *P. vivax* circumsporozoite protein-specific malaria vaccine adjuvanted with Poly I:C or Alhydrogel. Red and blue colors indicate animals immunized with yPvCSP-All$_{CT}$ epitopes + Poly I:C or Alhydrogel, respectively. **(A)** Representative images of murine spleens collected upon first or second boosters. **(B)** Dots and columns represent individual values detected for each mouse and their median, respectively.

SUPPLEMENTARY FIGURE 3
Sequential gating strategy to enrich distinct splenic murine B-cell subsets: (1) lymphocytes and monocytes; (2 and 3) singlets; (4) B220+ and B220- cells; (5a) Plasma cells (PCs - B220- CD138hi); (5b) B220+ MZBs (CD23-) and FoBs (CD23+); (5c) B220+ cells (CD138- and CD138int); (6a) B220+ CD138- lymphocytes (GCs (GL7+ CD38-), MBC precursors (GL7+ CD38+), and MBCs (GL7- CD38+); and (6b) Plasmablasts (PBs - CD138int CD38+).

SUPPLEMENTARY FIGURE 4
Similar increasing trend for B cells **(A)**, plasmablasts **(B)**, follicular B cells **(C)**, and germinal center B cells **(D)** in mice immunized with a *P. vivax* circumsporozoite protein-specific malaria vaccine adjuvanted with Poly I:C or Alhydrogel. Red and blue colors indicate the frequency **(A, C, E, G)** and absolute number **(B, D, F, H)** of cells derived from animals immunized with yPvCSP-All$_{CT}$ epitopes + Poly I:C or Alhydrogel, respectively. Dots and columns represent individual values detected for each mouse and their median, respectively. * p<0.05; ** p<0.01.

SUPPLEMENTARY FIGURE 5
Similar increasing trend for follicular helper T cells in mice immunized with a *P. vivax* circumsporozoite protein-specific malaria vaccine adjuvanted with Poly I:C or Alhydrogel. **(A)** Sequential gating strategy to enrich distinct splenic murine B-cell subsets: (1) lymphocytes and monocytes; (2 and 3) singlets; (4) T lymphocytes (CD3+ CD4+); 5) Activated CD4+ T cells (CD40L+ GL7- and CD40L+ GL7+); 6a) Non-germinal center follicular helper T cells (CXCR5+ GL7-); and 6B) germinal center follicular helper T cells (CXCR5+ GL7+). Red and blue colors indicate the frequency **(B, D)** and absolute number **(C, E)** of cells derived from animals immunized with yPvCSP-All$_{CT}$ epitopes + Poly I:C or Alhydrogel, respectively. Dots and columns represent individual values detected for each mouse and their median, respectively.

SUPPLEMENTARY TABLE 1
Exclusive differential expressed genes and their respective log fold-change (FC) values detected in splenic B cells upon distinct immunizations.

SUPPLEMENTARY TABLE 2
Similar differential expressed genes and their respective log fold-change (FC) values mutually detected in splenic B cells upon distinct immunizations. **(A)** yPvCSP-All$_{CT}$ epitopes + Poly I:C and yNLP-PvCSP$_{CT}$ + Poly I:C. **(B)** yPvCSP-All$_{CT}$ epitopes + Poly I:C and Poly I:C alone. **(C)** yNLP-PvCSP$_{CT}$ + Poly I:C and Poly I:C alone.

SUPPLEMENTARY TABLE 3
Similar differential expressed genes and their respective log fold-change (FC) values detected in splenic B cells upon distinct immunizations.

References

1. Phyo AP, Dahal P, Mayxay M, Ashley EA. Clinical impact of vivax malaria: A collection review. *PloS Med*. (2022) 19:e1003890. doi: 10.1371/journal.pmed.1003890

2. World malaria report 2022. *World Health Organization World malaria report*. Geneva, Switzerland: The World Health Organization (2022).

3. Krotoski WA, Collins WE, Bray RS, Garnham PC, Cogswell FB, Gwadz RW, et al. Demonstration of hypnozoites in sporozoite-transmitted Plasmodium vivax infection. *Am J Trop Med Hyg*. (1982) 31:1291–3. doi: 10.4269/ajtmh.1982.31.1291

4. Kumar KA, Sano G-I, Boscardin S, Nussenzweig RS, Nussenzweig MC, Zavala F, et al. The circumsporozoite protein is an immunodominant protective antigen in irradiated sporozoites. *Nature*. (2006) 444:937–40. doi: 10.1038/nature05361

5. Silveira ELV, Dominguez MR, Soares IS. To B or not to B: Understanding B cell responses in the development of malaria infection. *Front Immunol*. (2018) 9:2961. doi: 10.3389/fimmu.2018.02961

6. Bejon P, White MT, Olotu A, Bojang K, Lusingu JPA, Salim N, et al. Efficacy of RTS,S malaria vaccines: individual-participant pooled analysis of phase 2 data. *Lancet Infect Dis*. (2013) 13:319–27. doi: 10.1016/S1473-3099(13)70005-7

7. Kazmin D, Nakaya HI, Lee EK, Johnson MJ, van der Most R, van den Berg RA, et al. Systems analysis of protective immune responses to RTS,S malaria vaccination in humans. *Proc Natl Acad Sci USA*. (2017) 114:2425–30. doi: 10.1073/pnas.1621489114

8. White MT, Verity R, Griffin JT, Asante KP, Owusu-Agyei S, Greenwood B, et al. Immunogenicity of the RTS,S/AS01 malaria vaccine and implications for duration of vaccine efficacy: secondary analysis of data from a phase 3 randomised controlled trial. *Lancet Infect Dis*. (2015) 15:1450–8. doi: 10.1016/S1473-3099(15)00239-X

9. Adepoju P. RTS,S malaria vaccine pilots in three African countries. *Lancet*. (2019) 393:1685. doi: 10.1016/S0140-6736(19)30937-7

10. Arnot DE, Barnwell JW, Tam JP, Nussenzweig V, Nussenzweig RS, Enea V. Circumsporozoite protein of Plasmodium vivax: gene cloning and characterization of the immunodominant epitope. *Science*. (1985) 230:815–8. doi: 10.1126/science.2414847

11. Rosenberg R, Wirtz RA, Lanar DE, Sattabongkot J, Hall T, Waters AP, et al. Circumsporozoite protein heterogeneity in the human malaria parasite Plasmodium vivax. *Science*. (1989) 245:973–6. doi: 10.1126/science.2672336

12. Qari SH, Shi YP, Povoa MM, Alpers MP, Deloron P, Murphy GS, et al. Global occurrence of Plasmodium vivax-like human malaria parasite. *J Infect Dis*. (1993) 168:1485–9. doi: 10.1093/infdis/168.6.1485

13. Yadava A, Hall CE, Sullivan JS, Nace D, Williams T, Collins WE, et al. Protective efficacy of a Plasmodium vivax circumsporozoite protein-based vaccine in Aotus nancymaae is associated with antibodies to the repeat region. *PloS Negl Trop Dis*. (2014) 8:e3268. doi: 10.1371/journal.pntd.0003268

14. Nardin EH, Nussenzweig V, Nussenzweig RS, Collins WE, Harinasuta KT, Tapchaisri P, et al. Circumsporozoite proteins of human malaria parasites Plasmodium falciparum and Plasmodium vivax. *J Exp Med*. (1982) 156:20–30. doi: 10.1084/jem.156.1.20

15. Herrera S, Bonelo A, Perlaza BL, Fernández OL, Victoria L, Lenis AM, et al. Safety and elicitation of humoral and cellular responses in Colombian malaria-naive volunteers by a Plasmodium vivax circumsporozoite protein-derived synthetic vaccine. *Am J Trop Med Hyg*. (2005) 73:3–9. doi: 10.4269/ajtmh.2005.73.3

16. Atcheson E, Reyes-Sandoval A. Protective efficacy of peptides from Plasmodium vivax circumsporozoite protein. *Vaccine*. (2020) 38:4346–54. doi: 10.1016/j.vaccine.2020.03.063

17. Salman AM, Montoya-Díaz E, West H, Lall A, Atcheson E, Lopez-Camacho C, et al. Rational development of a protective P. vivax vaccine evaluated with transgenic rodent parasite challenge models. *Sci Rep*. (2017) 7:46482. doi: 10.1038/srep46482

18. Gimenez AM, Lima LC, Françoso KS, Denapoli PMA, Panatieri R, Bargieri DY, et al. Vaccine containing the three Allelic variants of the Plasmodium vivax circumsporozoite antigen induces protection in mice after challenge with a transgenic rodent malaria parasite. *Front Immunol*. (2017) 8:1275. doi: 10.3389/fimmu.2017.01275

19. de Camargo TM, de Freitas EO, Gimenez AM, Lima LC, de Almeida Caramico K, Françoso KS, et al. Prime-boost vaccination with recombinant protein and adenovirus-vector expressing Plasmodium vivax circumsporozoite protein (CSP) partially protects mice against Pb/Pv sporozoite challenge. *Sci Rep*. (2018) 8:1–14. doi: 10.1038/s41598-017-19063-6

20. Gimenez AM, Salman AM, Marques RF, López-Camacho C, Harrison K, Kim YC, et al. A universal vaccine candidate against Plasmodium vivax malaria confers protective immunity against the three PvCSP alleles. *Sci Rep*. (2021) 11:17928. doi: 10.1038/s41598-021-96986-1

21. Marques RF, Gimenez AM, Aliprandini E, Novais JT, Cury DP, Watanabe I-S, et al. Protective malaria vaccine in mice based on the Plasmodium vivax circumsporozoite protein fused with the mumps nucleocapsid protein. *Vaccines (Basel)*. (2020) 8:1–19. doi: 10.3390/vaccines8020190

22. Kool M, Soullié T, van Nimwegen M, Willart MAM, Muskens F, Jung S, et al. Alum adjuvant boosts adaptive immunity by inducing uric acid and activating inflammatory dendritic cells. *J Exp Med*. (2008) 205:869–82. doi: 10.1084/jem.20071087

23. Wang H-B, Weller PF. Pivotal advance: eosinophils mediate early alum adjuvant-elicited B cell priming and IgM production. *J Leukoc Biol*. (2008) 83:817–21. doi: 10.1189/jlb.0607392

24. Li H, Willingham SB, Ting JP-Y, Re F. Cutting edge: inflammasome activation by alum and alum's adjuvant effect are mediated by NLRP3. *J Immunol*. (2008) 181:17–21. doi: 10.4049/jimmunol.181.1.17

25. Eisenbarth SC, Colegio OR, O'Connor W, Sutterwala FS, Flavell RA. Crucial role for the Nalp3 inflammasome in the immunostimulatory properties of aluminium adjuvants. *Nature*. (2008) 453:1122–6. doi: 10.1038/nature06939

26. Marrack P, McKee AS, Munks MW. Towards an understanding of the adjuvant action of aluminium. *Nat Rev Immunol*. (2009) 9:287–93. doi: 10.1038/nri2510

27. Flach TL, Ng G, Hari A, Desrosiers MD, Zhang P, Ward SM, et al. Alum interaction with dendritic cell membrane lipids is essential for its adjuvanticity. *Nat Med*. (2011) 17:479–87. doi: 10.1038/nm.2306

28. Alexopoulou L, Holt AC, Medzhitov R, Flavell RA. Recognition of double-stranded RNA and activation of NF-kappaB by Toll-like receptor 3. *Nature*. (2001) 413:732–8. doi: 10.1038/35099560

29. Matsumoto M, Kikkawa S, Kohase M, Miyake K, Seya T. Establishment of a monoclonal antibody against human Toll-like receptor 3 that blocks double-stranded RNA-mediated signaling. *Biochem Biophys Res Commun*. (2002) 293:1364–9. doi: 10.1016/S0006-291X(02)00380-7

30. Gitlin L, Barchet W, Gilfillan S, Cella M, Beutler B, Flavell RA, et al. Essential role of mda-5 in type I IFN responses to polyriboinosinic:polyribocytidylic acid and encephalomyocarditis picornavirus. *Proc Natl Acad Sci USA*. (2006) 103:8459–64. doi: 10.1073/pnas.0603082103

31. McCartney S, Vermi W, Gilfillan S, Cella M, Murphy TL, Schreiber RD, et al. Distinct and complementary functions of MDA5 and TLR3 in poly(I:C)-mediated activation of mouse NK cells. *J Exp Med*. (2009) 206:2967–76. doi: 10.1084/jem.20091181

32. Matsumoto M, Seya T. TLR3: interferon induction by double-stranded RNA including poly(I:C). *Adv Drug Deliv Rev*. (2008) 60:805–12. doi: 10.1016/j.addr.2007.11.005

33. Salem ML, EL-Naggar SA, Kadima A, Gillanders WE, Cole DJ. The adjuvant effects of the toll-like receptor 3 ligand polyinosinic-cytidylic acid poly (I:C) on antigen-specific CD8+ T cell responses are partially dependent on NK cells with the induction of a beneficial cytokine milieu. *Vaccine*. (2006) 24:5119–32. doi: 10.1016/j.vaccine.2006.04.010

34. Yamamoto M, Sato S, Mori K, Hoshino K, Takeuchi O, Takeda K, et al. Cutting edge: a novel Toll/IL-1 receptor domain-containing adapter that preferentially activates the IFN-beta promoter in the Toll-like receptor signaling. *J Immunol*. (2002) 169:6668–72. doi: 10.4049/jimmunol.169.12.6668

35. Oshiumi H, Matsumoto M, Funami K, Akazawa T, Seya T. TICAM-1, an adaptor molecule that participates in Toll-like receptor 3-mediated interferon-beta induction. *Nat Immunol*. (2003) 4:161–7. doi: 10.1038/ni886

36. Lim CS, Jang YH, Lee GY, Han GM, Jeong HJ, Kim JW, et al. TLR3 forms a highly organized cluster when bound to a poly(I:C) RNA ligand. *Nat Commun*. (2022) 13:6876. doi: 10.1038/s41467-022-34602-0

37. Zhao T, Cai Y, Jiang Y, He X, Wei Y, Yu Y, et al. Vaccine adjuvants: mechanisms and platforms. *Signal Transduct Target Ther*. (2023) 8:283. doi: 10.1038/s41392-023-01557-7

38. Marques RF, de Melo FM, Novais JT, Soares IS, Bargieri DY, Gimenez AM. Immune system modulation by the adjuvants Poly (I:C) and Montanide ISA 720. *Front Immunol*. (2022) 13:910022. doi: 10.3389/fimmu.2022.910022

39. Fabris AL, Nunes AV, Schuch V, de Paula-Silva M, Rocha G, Nakaya HI, et al. Hydroquinone exposure alters the morphology of lymphoid organs in vaccinated C57Bl/6 mice. *Environ pollut*. (2020) 257:113554. doi: 10.1016/j.envpol.2019.113554

40. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. (2010) 26:139–40. doi: 10.1093/bioinformatics/btp616

41. Jassal B, Matthews L, Viteri G, Gong C, Lorente P, Fabregat A, et al. The reactome pathway knowledgebase. *Nucleic Acids Res*. (2020) 48(D1):D498–503. doi: 10.1093/nar/gkz1031

42. Zhou G, Soufan O, Ewald J, Hancock REW, Basu N, Xia J. NetworkAnalyst 3.0: a visual analytics platform for comprehensive gene expression profiling and meta-analysis. *Nucleic Acids Res*. (2019) 47:W234–41. doi: 10.1093/nar/gkz240

43. Breuer K, Foroushani AK, Laird MR, Chen C, Sribnaia A, Lo R, et al. InnateDB: systems biology of innate immunity and beyond—recent updates and continuing curation. *Nucleic Acids Res*. (2013) 41:D1228–33. doi: 10.1093/nar/gks1147

44. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res*. (2003) 13:2498–504. doi: 10.1101/gr.1239303

45. Hensel JA, Khattar V, Ashton R, Ponnazhagan S. Characterization of immune cell subtypes in three commonly used mouse strains reveals gender and strain-specific variations. *Lab Invest*. (2019) 99:93–106. doi: 10.1038/s41374-018-0137-1

46. Kasturi SP, Skountzou I, Albrecht RA, Koutsonanos D, Hua T, Nakaya HI, et al. Programming the magnitude and persistence of antibody responses with innate immunity. *Nature*. (2011) 470:543–7. doi: 10.1038/nature09737

47. Silveira ELV, Kasturi SP, Kovalenkov Y, Rasheed AU, Yeiser P, Jinnah ZS, et al. Vaccine-induced plasmablast responses in rhesus macaques: Phenotypic characterization and a source for generating antigen-specific monoclonal antibodies. *J Immunol Methods*. (2015) 416:69–83. doi: 10.1016/j.jim.2014.11.003

48. Kasturi SP, Kozlowski PA, Nakaya HI, Burger MC, Russo P, Pham M, et al. Adjuvanting a simian immunodeficiency virus vaccine with Toll-like receptor ligands encapsulated in nanoparticles induces persistent antibody responses and enhanced protection in TRIM5α restrictive macaques. *J Virol*. (2017) 91:1–25. doi: 10.1128/jvi.01844-16

49. Wrammert J, Smith K, Miller J, Langley WA, Kokko K, Larsen C, et al. Rapid cloning of high-affinity human monoclonal antibodies against influenza virus. *Nature*. (2008) 453:667–71. doi: 10.1038/nature06890

50. Victora GD, Nussenzweig MC. Germinal centers. *Annu Rev Immunol*. (2022) 40:413–42. doi: 10.1146/annurev-immunol-120419-022408

51. Davis CW, Jackson KJL, McCausland MM, Darce J, Chang C, Linderman SL, et al. Influenza vaccine–induced human bone marrow plasma cells decline within a year after vaccination. *Science*. (2020) 370:237–41. doi: 10.1126/science.aaz8432

52. Wang Y, Bhattacharya D. Adjuvant-specific regulation of long-term antibody responses by ZBTB20. *J Exp Med*. (2014) 211:841–56. doi: 10.1084/jem.20131821

53. McNamara HA, Idris AH, Sutton HJ, Vistein R, Flynn BJ, Cai Y, et al. Antibody feedback limits the expansion of B cell responses to malaria vaccination but drives diversification of the humoral response. *Cell Host Microbe*. (2020) 28:572–585.e7. doi: 10.1016/j.chom.2020.07.001

54. Ubillos I, Ayestaran A, Nhabomba AJ, Dosoo D, Vidal M, Jiménez A, et al. Baseline exposure, antibody subclass, and hepatitis B response differentially affect malaria protective immunity following RTS,S/AS01E vaccination in African children. *BMC Med*. (2018) 16:1–18. doi: 10.1186/s12916-018-1186-4

55. Mugo RM, Mwai K, Mwacharo J, Shee FM, Musyoki JN, Wambua J, et al. Seven-year kinetics of RTS, S/AS01-induced anti-CSP antibodies in young Kenyan children. *Malar J*. (2021) 20:1–8. doi: 10.1186/s12936-021-03961-2

56. Kurtovic L, Agius PA, Feng G, Drew DR, Ubillos I, Sacarlal J, et al. Induction and decay of functional complement-fixing antibodies by the RTS,S malaria vaccine in children, and a negative impact of malaria exposure. *BMC Med*. (2019) 17:1–14. doi: 10.1186/s12916-019-1277-x

57. Collins KA, Snaith R, Cottingham MG, Gilbert SC, Hill AVS. Enhancing protective immunity to malaria with a highly immunogenic virus-like particle vaccine. *Sci Rep*. (2017) 7:1–15. doi: 10.1038/srep46621

58. Reinke S, Pantazi E, Chappell GR, Sanchez-Martinez A, Guyon R, Fergusson JR, et al. Emulsion and liposome-based adjuvanted R21 vaccine formulations mediate protection against malaria through distinct immune mechanisms. *Cell Rep Med*. (2023) 4:101245. doi: 10.1016/j.xcrm.2023.101245

59. Schwenk R, DeBot M, Porter M, Nikki J, Rein L, Spaccapelo R, et al. IgG2 antibodies against a clinical grade Plasmodium falciparum CSP vaccine antigen associate with protection against transgenic sporozoite challenge in mice. *PloS One*. (2014) 9:e111020. doi: 10.1371/journal.pone.0111020

60. Cohen S, McGREGOR IA, Carrington S. Gamma-globulin and acquired immunity to human malaria. *Nature*. (1961) 192:733–7. doi: 10.1038/192733a0

61. Schäfer C, Dambrauskas N, Reynolds LM, Trakhimets O, Raappana A, Flannery EL, et al. Partial protection against P. vivax infection diminishes hypnozoite burden and blood-stage relapses. *Cell Host Microbe*. (2021) 29:752–756.e4. doi: 10.1016/j.chom.2021.03.011

62. Kucharska I, Hossain L, Ivanochko D, Yang Q, Rubinstein JL, Pomès R, et al. Structural basis of Plasmodium vivax inhibition by antibodies binding to the circumsporozoite protein repeats. *Elife*. (2022) 11:1–29. doi: 10.7554/elife.72908

63. Vijayan K, Visweswaran GRR, Chandrasekaran R, Trakhimets O, Brown SL, Watson A, et al. Antibody interference by a non-neutralizing antibody abrogates humoral protection against Plasmodium yoelii liver stage. *Cell Rep*. (2021) 36:109489. doi: 10.1016/j.celrep.2021.109489

64. Feng G, Wines BD, Kurtovic L, Chan J-A, Boeuf P, Mollard V, et al. Mechanisms and targets of Fcγ-receptor mediated immunity to malaria sporozoites. *Nat Commun*. (2021) 12:1–16. doi: 10.1038/s41467-021-21998-4

65. Kurtovic L, Drew DR, Dent AE, Kazura JW, Beeson JG. Antibody targets and properties for complement-fixation against the circumsporozoite protein in malaria immunity. *Front Immunol*. (2021) 12:775659. doi: 10.3389/fimmu.2021.775659

66. Fisher CR, Sutton HJ, Kaczmarski JA, McNamara HA, Clifton B, Mitchell J, et al. T-dependent B cell responses to Plasmodium induce antibodies that form a high-avidity multivalent complex with the circumsporozoite protein. *PloS Pathog*. (2017) 13: e1006469. doi: 10.1371/journal.ppat.1006469

67. Keitany GJ, Kim KS, Krishnamurty AT, Hondowicz BD, Hahn WO, Dambrauskas N, et al. Blood stage malaria disrupts humoral immunity to the pre-erythrocytic stage circumsporozoite protein. *Cell Rep*. (2016) 17:3193–205. doi: 10.1016/j.celrep.2016.11.060

68. Fontana MF, Ollmann Saphire E, Pepper M. Plasmodium infection disrupts the T follicular helper cell response to heterologous immunization. *Elife*. (2023) 12:1–22. doi: 10.7554/elife.83330

69. Vijay R, Guthmiller JJ, Sturtz AJ, Surette FA, Rogers KJ, Sompallae RR, et al. Infection-induced plasmablasts are a nutrient sink that impairs humoral immunity to malaria. *Nat Immunol*. (2020) 21:790–801. doi: 10.1038/s41590-020-0678-5

70. Baba Y, Kurosaki T. Impact of Ca2+ signaling on B cell function. *Trends Immunol*. (2011) 32:589–94. doi: 10.1016/j.it.2011.09.004

71. Ulbricht C, Leben R, Rakhymzhan A, Kirchhoff F, Nitschke L, Radbruch H, et al. Intravital quantification reveals dynamic calcium concentration changes across B cell differentiation stages. *Elife*. (2021) 10:1–26. doi: 10.7554/elife.56020

72. Spagnuolo L, Puddinu V, Boss N, Spinetti T, Oberson A, Widmer J, et al. HMGB1 promotes CXCL12-dependent egress of murine B cells from Peyer's patches in homeostasis. *Eur J Immunol*. (2021) 51:1980–91. doi: 10.1002/eji.202049120

73. Covens K, Verbinnen B, Geukens N, Meyts I, Schuit F, Van Lommel L, et al. Characterization of proposed human B-1 cells reveals pre-plasmablast phenotype. *Blood*. (2013) 121:5176–83. doi: 10.1182/blood-2012-12-471953

74. Zhang Y, Guo J, Gao Y, Li S, Pan T, Xu G, et al. Dynamic transcriptome analyses reveal m6A regulated immune non-coding RNAs during dengue disease progression. *Heliyon*. (2023) 9:e12690. doi: 10.1016/j.heliyon.2022.e12690

75. Kitagori K, Oku T, Wakabayashi M, Nakajima T, Nakashima R, Murakami K, et al. Expression of S100A8 protein on B cells is associated with disease activity in patients with systemic lupus erythematosus. *Arthritis Res Ther*. (2023) 25:1–12. doi: 10.1186/s13075-023-03057-z

76. Shi W, Liao Y, Willis SN, Taubenheim N, Inouye M, Tarlinton DM, et al. Transcriptional profiling of mouse B cell terminal differentiation defines a signature for antibody-secreting plasma cells. *Nat Immunol*. (2015) 16:663–73. doi: 10.1038/ni.3154

77. Tewari K, Flynn BJ, Boscardin SB, Kastenmueller K, Salazar AM, Anderson CA, et al. Poly(I:C) is an effective adjuvant for antibody and multi-functional CD4+ T cell responses to Plasmodium falciparum circumsporozoite protein (CSP) and αDEC-CSP in non human primates. *Vaccine*. (2010) 28:7256–66. doi: 10.1016/j.vaccine.2010.08.098

frontiers | Frontiers in Immunology

TYPE Original Research
PUBLISHED 22 May 2024
DOI 10.3389/fimmu.2024.1388721

OPEN ACCESS

EDITED BY
Helder Nakaya,
University of São Paulo, Brazil

REVIEWED BY
Sayan Das,
University of Maryland, United States
Jutamas Shaughnessy,
University of Massachusetts Medical School, United States

*CORRESPONDENCE
Ekaterina A. Kurbatova
kurbatova6162@yandex.ru
Nikolay E. Nifantiev
nen@ioc.ac.ru

RECEIVED 20 February 2024
ACCEPTED 06 May 2024
PUBLISHED 22 May 2024

CITATION
Akhmatova NK, Kurbatova EA, Zaytsev AE, Akhmatova EA, Yastrebova NE, Sukhova EV, Yashunsky DV, Tsvetkov YE and Nifantiev NE (2024) Synthetic BSA-conjugated disaccharide related to the *Streptococcus pneumoniae* serotype 3 capsular polysaccharide increases IL-17A Levels, γδ T cells, and B1 cells in mice. *Front. Immunol.* 15:1388721. doi: 10.3389/fimmu.2024.1388721

Synthetic BSA-conjugated disaccharide related to the *Streptococcus pneumoniae* serotype 3 capsular polysaccharide increases IL-17A Levels, γδ T cells, and B1 cells in mice

Nelli K. Akhmatova[1], Ekaterina A. Kurbatova[1]*, Anton E. Zaytsev[1], Elina A. Akhmatova[2], Natalya E. Yastrebova[1], Elena V. Sukhova[2], Dmitriy V. Yashunsky[2], Yury E. Tsvetkov[2] and Nikolay E. Nifantiev[2]*

[1]Laboratory of Therapeutic Vaccines, Mechnikov Research Institute for Vaccines and Sera, Moscow, Russia, [2]Laboratory of Glycoconjugate Chemistry, N. D. Zelinsky Institute of Organic Chemistry, Russian Academy of Science, Moscow, Russia

The disaccharide (β-D-glucopyranosyluronic acid)-(1→4)-β-D-glucopyranoside represents a repeating unit of the capsular polysaccharide of *Streptococcus pneumoniae* serotype 3. A conjugate of the disaccharide with BSA (di-BSA conjugate) adjuvanted with aluminum hydroxide induced — in contrast to the non-adjuvanted conjugate — IgG1 antibody production and protected mice against *S. pneumoniae* serotype 3 infection after intraperitoneal prime-boost immunization. Adjuvanted and non-adjuvanted conjugates induced production of Th1 (IFNγ, TNFα); Th2 (IL-5, IL-13); Th17 (IL-17A), Th1/Th17 (IL-22), and Th2/Th17 cytokines (IL-21) after immunization. The concentration of cytokines in mice sera was higher in response to the adjuvanted conjugate, with the highest level of IL-17A production after the prime and boost immunizations. In contrast, the non-adjuvanted conjugate elicited only weak production of IL-17A, which gradually decreased after the second immunization. After boost immunization of mice with the adjuvanted di-BSA conjugate, there was a significant increase in the number of CD45+/CD19+ B cells, TCR+ γδ T cell, CD5+ B1 cells, and activated cells with MHC II+ expression in the spleens of the mice. IL-17A, TCR + γδ T cells, and CD5+ B1 cells play a crucial role in preventing pneumococcal infection, but can also contribute to autoimmune diseases. Immunization with the adjuvanted and non-adjuvanted di-BSA conjugate did not elicit autoantibodies against double-stranded DNA targeting cell nuclei in mice.

10.3389/fimmu.2024.1388721

Thus, the molecular and cellular markers associated with antibody production and protective activity in response to immunization with the di-BSA conjugate adjuvanted with aluminum hydroxide are IL-17A, TCR+ γδ T cells, and CD5+ B1 cells against the background of increasing MHC II+ expression.

KEYWORDS

Streptococcus pneumoniae serotype 3, synthetic disaccharide, cytokine, γδ T cells, B1 Cells, interleukin 17A, antibody, mice immunoprotection

1 Introduction

Streptococcus pneumoniae (pneumococcus) cause pneumonia, bacteremia, septic arthritis, meningitis, sinusitis, otitis media and some other diseases in humans (1, 2). The incidence of community-acquired pneumonia is one per one thousand adults. The mortality rate for pneumococcal pneumonia among hospitalized patients is 5–7% (3–7). Symptoms of pneumococcal infection depend on the localization of the infection. These may include fever, cough, chest pain, a stiff neck, chills, ear pain and others.

Pneumococcal polysaccharide and conjugate vaccines, which contain capsular polysaccharides (CPs) from clinically significant *S. pneumoniae* serotypes, are available (8). *S. pneumoniae* serotype 3 is predominant among other serotypes in various countries (9–12). Epidemiological data suggests a high incidence of disease caused by *S. pneumoniae* serotype 3 (13–15). However, the widespread use of pneumococcal vaccines should help to reduce the incidence of this disease (16–19). Improving the quality of *S. pneumoniae* type 3 in the composition of pneumococcal vaccines is essential.

Bacterial CPs contain a diverse mixture of oligosaccharides with varying chain lengths and frame shifts (20). Although their chemical preparation is practically possible (see, for example (21),), synthetic oligosaccharide derivatives represent more convenient antigenic components for the design of conjugate carbohydrate vaccines (22–25). Currently, a number of semisynthetic vaccines are under development, including those against *Staphylococcus*, *Clostridium*, *Klebsiella*, *Shigella*, and *Enterococcus* (25–33). The semi-synthetic glycoconjugate vaccine, Quimi-Hib, for the prevention of *H. influenzae* type b infection is licensed for use in Cuba (34). Optimization of the composition of pneumococcal vaccines using synthetic oligosaccharides conjugated with a protein carrier is a priority in contemporary vaccinology (25, 35–38).

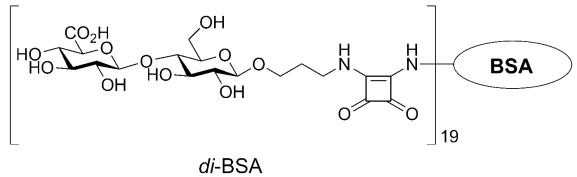
Moreover, synthetic oligosaccharides with precisely defined chemical structures enable the study of the effect of bacterial antigens (39, 40), yielding a better understanding of the innate and cellular immunity, the antibody (Ab) response, and protective activity of CPs.

Immunization with glycoconjugate vaccines partially mimics the development of natural infection without actually causing the disease. In a mouse model, γδ T cells and natural killer T cells (NKT) have been shown to play a crucial role in anti-pneumococcal immunity by producing Th1 and/or Th17-related cytokines (41). The ability of semisynthetic glycoconjugates to stimulate cytokine production *in vivo* and their influence on the activation of cellular immunity remain unknown. Here, we report on the effect of a conjugate of the synthetic disaccharide, which represents a repeating unit of S. pneumoniae serotype 3 (42), on production of Th1/Th2/Th17 cytokines in mice, changes in expression of surface molecules on splenocytes, antibody response, and protection against *S. pneumoniae* infection. We also investigated the production of autoantibodies against double-stranded (ds) DNA.

2 Materials and methods

2.1 The synthetic disaccharide and its conjugate

The synthetic disaccharide (35, 43) was coupled to BSA (Sigma-Aldrich, St. Louis, MO, USA), as previously described (35, 44). The structure of the conjugate is illustrated in Figure 1. BSA is often used



di-BSA

FIGURE 1
The structure of the BSA conjugate with the disaccharide that corresponds to a repeating unit of the CP from *S. pneumoniae* serotype 3.

Abbreviations: Ab, antibody; BSA, bovine serum albumin; CP, capsular polysaccharide; CTL, cytotoxic T cells; EU, endotoxin unite; IL, interleukin; IFN γ, interferon gamma; IL-2 R, IL-2 receptor; MALDI-TOF, matrix assisted laser desorption ionization time-of-flight; MHC II, major histocompatibility complex class II; NK, natural killer cells; NKT, natural killer T cells; TCR, T cell receptor; TNF α, tumor necrosis factor alfa; Treg, T regulatory cells.

as a protein carrier in engineered immunogenic glycoconjugates and other biological systems (45). Previous studies using MALDI-TOF mass spectrometry have shown that the di-BSA conjugate contains, on average, 19 oligosaccharide ligands per protein molecule, which corresponds to a 9% carbohydrate content by weight (43, 44). The lyophilized di-BSA conjugate remains stable at +4°C, with no decrease in activity, for at least three years (i.e., observation period).

2.2 Bacterial capsular polysaccharide

Bacterial CP was isolated from the *S. pneumoniae* type 3 laboratory strain, #10196, which was isolated on June 30, 2011, from the blood culture of a child suffering from bacteremia in the microbiology department of the "Scientific Center for Children's Health" in Moscow, Russia. The strain had been grown in a semi-synthetic growth medium. The isolation process for CP has been previously described elsewhere (46). The presence of CP in the preparation was confirmed by NMR spectrometry.

2.3 Animals

BALB/c male mice, aged 6–8 weeks (n=162), were purchased from the Scientific and Production Centre for Biomedical Technologies in Moscow, Russia, and kept in the vivarium at the Mechnikov Institute for Vaccines and Sera. Housing, breeding, blood collection, and euthanasia conditions followed European Union guidelines for laboratory animal care and use. Experimental designs were reviewed and approved (Protocol No. 2, dated February 12th, 2019) by the Ethics Committee at the Institute.

2.4 Conjugated disaccharide-induced cytokine production

Quantitative determination of cytokines was performed as previously described (46). Male BALB/c mice (n=6) were sacrificed, and serum was collected and stored at −20°C until further quantification of cytokine levels. Using the Flow Cytomix Mouse Th1/Th2 10-plex test system, cytokine levels were measured by adding beads coated with monoclonal antibodies to IL-1α, IL-1β, IL-2, IL-4, IL-5, IL-6, IL-10, IL-12p70, IL-13, IL-17A, IL-21 and IL-22, as well as IFNγ and TNFα, following the manufacturer's instructions (eBioscience, San Diego, USA) using a Cytomix FC-500 flow cytometer (Beckman Coulter, Brea, USA).

2.5 Immunization

Mice were intraperitoneally immunized with the di-BSA conjugate, either adjuvanted or not, with aluminum hydroxide (Sigma-Aldrich). The amount of carbohydrate in 0.5 mL of the experimental semisynthetic vaccine was 20 μg, BSA ~200 μg; aluminum hydroxide, as an adjuvant, standardized for aluminum, was added in an amount of 250 μg. The single immunizing dose per mice was 0.5 mL of the di-BSA conjugate. Animals were given the vaccine twice, on days 0 and 14 of the study.

Similar immunization schedules were used for the pneumococcal conjugate vaccine Prevnar 13 (Pfizer, New York, NY, USA), which contains aluminum phosphate as an adjuvant. A 0.5 mL dose contains 2.2 μg of polysaccharides from serotypes 1, 3, 4, 5, 6A, 7F, 9V, 14, 18C, 19A, 19F, and 23F, as well as 4.4 μg of the polysaccharide from serotype 6B. The vaccine also contains 32 μg of the carrier protein, CRM_{197}, and 125 μg of aluminum as aluminum phosphate. Mice were immunized twice with a single dose of 1.1 μg of CP from *S. pneumoniae* type 3 per inoculation (equivalent to half of the recommended human dose). Control mice were injected with saline.

2.6 Content of bacterial endotoxins in glycoconjugates

Detection of bacterial endotoxin impurities in the di-BSA conjugate was performed using the Limulus amebocyte lysate ENDOCHROME ^{TM} (Charles River Endosafe Div. of Charles River Laboratories, Inc., Charleston, US) test obtained from the Collective Usage Center of the Mechnikov Research Institute for Vaccine and Sera (Moscow, Russia), in accordance with the manufacturer's instructions. The di-BSA conjugate contained 0.08–0.11 EU/mL of endotoxin (LAL-Center, Moscow, Russia).

2.7 Measurement of antibody response to the disaccharide conjugate

Antibody titers for CP in post-immunization sera were measured using ELISA. Briefly, plates coated with *S. pneumoniae* type 3 CP were incubated with antisera from 6 immunized mice (42). Wells were washed and secondary antibody was added, followed by incubation and washing. The results were then analyzed. Enzyme substrate aliquots (100 μL) were added, followed by incubation for 20 minutes at 22°C. The reactions were quenched with 1 M H_2SO_4. Optical densities (ODs) were determined using an iMark microplate absorbance reader (Bio-Rad, Osaka, Japan) at a wavelength of 450 nm. Antibody titers are expressed as the dilution of serum in which the antibody was detected.

2.8 Expression of surface molecules on splenic mononuclear cells

Splenocytes were isolated from mice that had been vaccinated with the glyconjugate either in the absence of or in the presence of aluminum hydroxide, one and seven days after primary and booster immunizations. Single-cell suspensions of splenocytes were

prepared by manually mashing the spleens using the plunger from a disposable syringe. The ground spleen was then passed through a nylon mesh and the cells were suspended in PBS. Splenic single-cell suspensions were then stained with antibodies conjugated to phycoerythrin (PE) or fluorescein isothiocyanate (FITC) to detect specific proteins in the cells: CD3e-FITC (clone 145–2C11), CD4-FITC (clone GK1.5), CD8a-FITC (clone 53–6.7), CD19-FITC (eBio1D3), CD5-PE (clone 53–7.3), NK.1.1 (clone PK136), CD3/CD16/CD32 (NKT), CD25-PE (PC61.5), CD4/CD25/Foxp3 (Treg), γδT (clone γδ TCR-PE, eBioGL3), and MHCII-PE (I-EK) (clone 14–4-45). Treg cells were stained with CD4-FITC (clone GK1.5), together with CD25-PE (PC61.5), and after fixation with the fixation/permeabilization buffer, with Foxp3- APC (clone FJK-16s). Splenocytes were incubated with 50 µL of appropriate monoclonal antibodies (eBioscience, US) at 4°C for 30 minutes. Erythrocytes were then lysed using red blood cell lysis buffer (BioLegend, US). After washing with phosphate-buffered saline (PBS), the samples were fixed using a fixation solution (BioLegend, US) and analyzed by flow cytometry (Cytomix FC-500, Beckman Coulter, USA, with the CXP software). The cell population gate was determined based on forward and side scatter and cell size. 10,000 cells were recorded per gate.

2.9 Di-BSA-induced active protection in immunized mice

BALB/c mice were intraperitoneally immunized with the di-BSA conjugate adsorbed or non-adsorbed on aluminum hydroxide on days 0 and 14 (twenty animals per conjugate). The same animals were intraperitoneally challenged after 2 weeks with 10^5 colony-forming units *of S. pneumoniae* type 3/0.5 mL. Non-immunized control mice (twenty animals per conjugate) were also exposed to the bacteria. Mortality rates were determined at seven days post-infection.

2.10 Antibodies against double-stranded DNA

The analysis of antibodies against ds DNA in the immune sera of mice was conducted using ELISA. Salmon sperm DNA (Behringer GmbH, Germany), dissolved in a carbohydrate buffer solution at a concentration of 20 g/mL, was adsorbed onto the bottom of the wells. The plates were incubated for 2 hours at 37°C and then for additional 18 hours at 6°C. The serums were analyzed using dilutions of 1:10 to 1:1280. As secondary antibodies, secondary rabbit anti-mouse peroxidase conjugated IgG (Rockland Immunochemicals Inc., Pottstown, PA) was utilized (100 µL). After adding tetramethylbenzidine for 15 minutes, the reaction was terminated with 1 M sulfuric acid. Results were obtained utilizing a multi-channel automatic photometer (TiterTek Multiscan MC from Flow Laboratories, England), with excitation at 490nm. Serums from non-immunized mice, as well as mice immunized with either Prevnar-13 or BSA adjuvanted or non-adjuvanted with aluminum hydroxide, were used as controls.

2.11 Statistical analysis

Between-group comparisons were performed using Mann-Whitney rank sum tests for independent samples. Fisher exact tests were conducted to evaluate survival of mice after pneumococcal challenge. *P* values ≤0.05 were considered to indicate statistical significance. Statistical analyses were performed using the Statistical data analysis software system version 10 (StatSoft Inc., Tulsa, OK, USA).

3 Results

3.1 Antibodies induced by the di-BSA conjugate

Although the di-BSA conjugate adjuvanted with aluminum hydroxide was found to be less immunogenic than adjuvanted tri- and tetra-BSA conjugates, it was still able to induce the production of opsonizing antibodies and was sufficient for the development of serotype 3-protective immunity in mice (42).

In this study, we explored the ability of the di-BSA conjugate to induce antibodies capable of binding to the CP of *S. pneumoniae* serotype 3 in ELISA after primary and booster immunization with and without the adjuvant (Figure 2). The di-BSA conjugate without adjuvant did not induce Ab production after the prime and boost immunizations and no difference was observed relative to the control. The glycoconjugate adjuvanted with aluminum hydroxide induced no Ab production after prime immunization; however, after booster injection, the level of Abs increased in seven days (21 d) and was significantly elevated up to 28 d (14 d after boost). Prevnar 13 (1.1 µg/

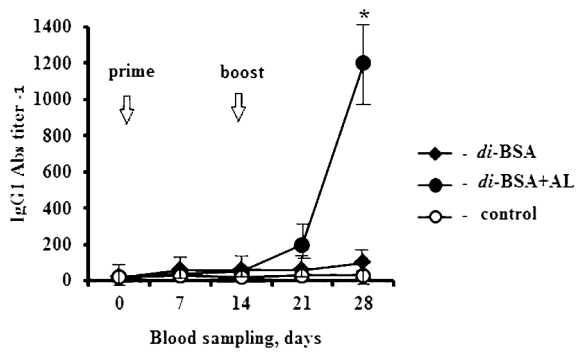


FIGURE 2
IgG1 antibody production induced by the adjuvanted and non-adjuvanted di-BSA conjugate. BALB/c mice (n = 6 per conjugate) were intraperitoneally injected with the di-BSA conjugate (20 µg/dose of carbohydrate) adjuvanted and non-adjuvanted with aluminum hydroxide, on days 0 and 14. The IgG1 Ab titer in the blood of mice was determined on days 0 (before prime immunization), days 7, 14, 21, and 28 (7 and 14 days after booster immunization, respectively), by ELISA, using CP of *S. pneumoniae* serotype 3 as the well-coating antigen. Mice (n=6) injected with saline at the same time served as a control group. AL - aluminum hydroxide. The data are presented as mean ± standard deviation (M ± SD). The Mann–Whitney rank sum test was used to determine significance, *$P < 0.05$.

dose of carbohydrate of CP of *S. pneumoniae* serotype 3) induced IgG Ab production on day 14 after boost immunization (the time of the study) at a titer of 1:800 (data not shown).

3.2 Active protection upon challenge of mice immunized with the di-BSA conjugate

Mice immunized with the di-BSA conjugate and di-BSA conjugate adjuvanted with aluminum hydroxide were challenged with *S. pneumoniae* serotype 3 on day 28 (14 d after booster immunization). All control mice injected with saline and 18 out of 20 mice immunized with the non-adjuvanted di-BSA conjugate died on the second day after the challenge (Figure 3).

The non-adjuvanted di-BSA conjugate that failed to induce Ab production also did not elicit any protection against challenge with *S. pneumoniae* serotype 3. However, the same conjugate administered to mice with aluminum hydroxide induced protection against *S. pneumoniae* serotype 3. Thus, aluminum hydroxide is indispensable for inducing protective immunity to the disaccharide conjugate. Prevnar 13 (1.1 μg/dose of carbohydrate of CP of *S. pneumoniae* serotype 3) protected all mice (n = 6) from the challenge (42).

3.3 Cytokine production in mice

To evaluate cytokine production, mice were intraperitoneally injected with the di-BSA conjugate adjuvanted or non-adjuvanted with aluminum hydroxide at a single dose of 20 μg (carbohydrate content). Serum cytokine levels were determined before injection of the glycoconjugate (d 0) and on days 1, 7, 15, and 21 (1 and 7 days after boost immunization, respectively) (Figure 4).

After prime immunization, the non-adjuvanted di-BSA conjugate induced an increase in the levels of IL-1α, IL-1β, IL-6, IL-13, IL-17A, IL-21, IFNγ, and TNFα compared with that in the control (0 d). After booster immunization with the conjugate, IL-5, IL-10, and IL-22 production was induced in addition to these cytokines. The concentration of IL-4 did not increase in any of the study periods.

After prime immunization, the di-BSA conjugate adjuvanted with aluminum hydroxide stimulated higher production of IL-1α, IL-1β, IL-4, IL-5, IL-6, IL-10, IL-13, IL-17A, IL-21, IL-22, IFNγ, and TNFα compared with the conjugate without the adjuvant. After booster immunization, all cytokines were found to be produced at high levels. When the conjugate was administered with the adjuvant, a very high level of IL-17A production was noted at all time points. In contrast, when mice were immunized with the conjugate without the adjuvant, the IL-17A level gradually decreased even after booster immunization. Regardless of the presence of the adjuvant, the levels of IL-2 and IL-12p70 did not increase during all follow-up periods. Free CP of *S. pneumoniae* serotype 3 (5 μg/mouse) elevated only the level of IFNγ (from 23.1 to 50.8 pg) after double immunization (data not shown). CP-CRM_{197} (Prevnar 13) is able to induce the production of IL-1, IL-2, IL-4, IL-5, IL-6, IL-10, IL-12, IL-17, IFNγ, and TNFα (47, 48). Free aluminum hydroxide did not elicit cytokine production when administered at the same time points (data not shown).

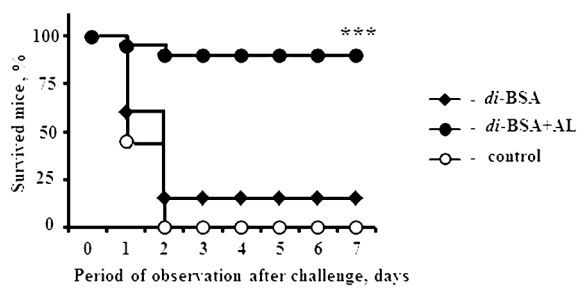


FIGURE 3
Protective activity of the adjuvanted and non-adjuvanted di-BSA conjugate. BALB/c mice (n = 20 per conjugate and control group) intraperitoneally injected with the di-BSA conjugate (20 μg/dose of carbohydrate) adjuvanted and non-adjuvanted with aluminum hydroxide on days 0 and 14 were challenged with 10^5 colony-forming units of *S. pneumoniae* serotype 3 on day 28. Mice injected with saline were used as a control. AL - aluminum hydroxide. The results of two experiments are summarized. The difference between mice immunized with the adjuvanted di-BSA conjugate and non-adjuvanted/non-immunized mice (control) is shown. Fisher exact test; ***$P < 0.001$.

3.4 Expression of cell-surface molecules on splenic mononuclear cells

After first immunization with the di-BSA conjugate adjuvanted and non-adjuvanted with aluminum hydroxide, the number of $CD45^+/CD3^+$ T cells and $CD45^+/CD4^+$ T helper cells increased compared with that in the control. After booster immunization, regardless of the presence of adjuvant, there was no difference relative to the control (Figure 5).

After primary and booster immunization with the adjuvanted di-BSA conjugate, the number of $CD45^+/CD8^+$ cytotoxic T cells (CTLs) increased compared with that in the control. The non-adjuvanted conjugate did not induce any change in the number of CTLs during the entire observation period. An interesting result was revealed in relation to γδ T cells. One day after prime immunization of mice with the adjuvanted and non-adjuvanted di-BSA conjugate, the number of γδ T cells increased compared with that in the control and decreased to the initial levels on day 7. However, after booster immunization with the adjuvanted conjugate, the number of TCR^+ γδ T cells increased on day 15 (1 d after boost), reaching high values on day 21 (7 d after boost). In contrast, in the absence of aluminum hydroxide, their values did not differ from the control level. After booster immunization with the di-BSA conjugate, the number of $CD45^+/CD19^+$ B cells increased only following booster immunization in the presence of aluminum hydroxide. After injection of the non-adjuvanted conjugate, the level of $CD45^+/CD19^+$ B cells did not differ from that in the control. The number of $CD5^+$ B1 increased on day 1 after the first immunization with adjuvanted and non-adjuvanted conjugate compared with that in the control and then decreased on day 7. Booster immunization with the adjuvanted di-BSA conjugate led to an increase of number of $CD5^+$ B1 cells on day 15 (1 d after boost) compared with that in the control, and on day

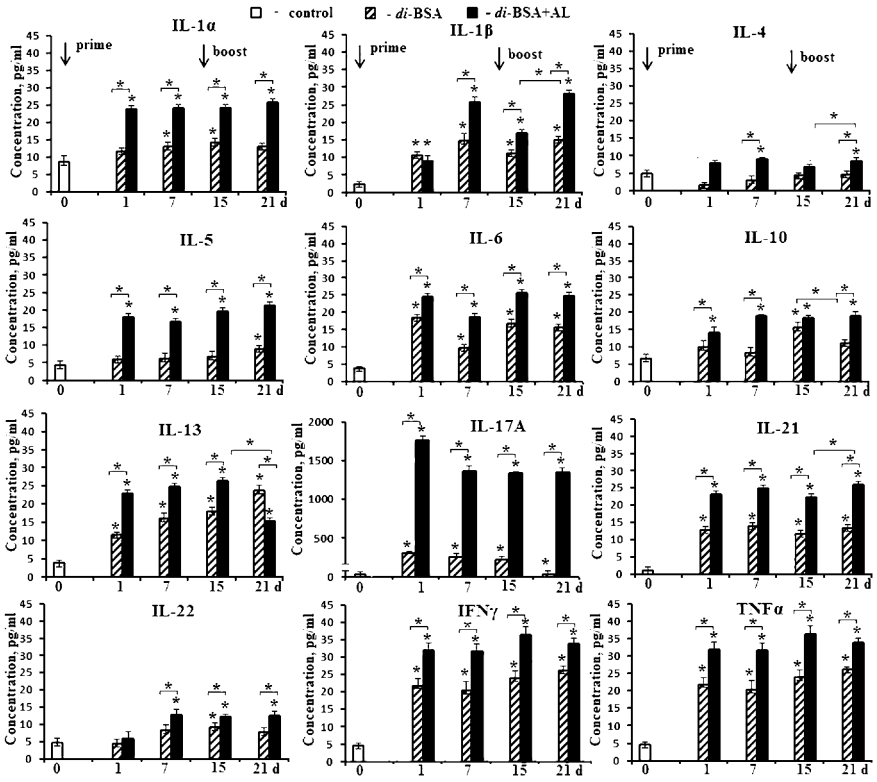


FIGURE 4
Cytokine production in mice induced by the adjuvanted and non-adjuvanted di-BSA conjugate. BALB/c mice were immunized with the di-BSA conjugate (20 μg of carbohydrate per mouse) adjuvanted or non-adjuvanted with aluminum hydroxide (n = 24 for each conjugate). Control mice (n = 6) were injected with saline 24 hours before the start of immunization (0 d). Serum was collected from mice (n=6 for each time point) after immunization. Cytokine levels were analyzed using flow cytometry. No increase in IL-2 or IL-12 p70 levels was observed in any of the time points (data not shown). The data is presented as the mean ± SD. Mann-Whitney rank sum tests were used to determine significant differences between control and other experimental groups; *P <0.05.

21 (7 d after boost) relative to the non-adjuvanted conjugate. The administration of the adjuvanted and non-adjuvanted conjugate increased the number of $CD16^+/CD32^+$ natural killer cells (NK) and $CD3^+/CD16^+/CD32^+$ natural killer T cells (NKT) after primary and booster immunization. The adjuvanted and non-adjuvanted di-BSA conjugate led to increase in the number of cells expressing CD25+ and the IL-2 receptor and $CD4^+/CD25^+/Foxp3^+$ T regulatory cells (Treg). The number of cells expressing MHC II^+ increased only after booster immunization—to a greater extent on day 21 (7 d after boost)—and was higher than that in the case of conjugate administration without aluminum hydroxide.

Prevnar 13, containing a CRM_{197}-CP of *S. pneumoniae* serotype 3 conjugate, induced similar changes on day 28 (14 d after booster immunization) in the number of cells expressing cell-surface molecules. Specifically, there was an increased number of (TCR^+) γδ T cells, $CD45^+/CD8^+$ CTLs, $CD5^+$ B1 cells, $CD45^+/CD19^+$ B cells, $CD4^+/CD25^+/Foxp3^+$ Tregs, cells expressing $CD25^+$, and cells expressing MHC II+. The number of NK- and NKT-cells did not differ from that in the control.

An elevation in Ab production and protection against *S. pneumoniae* serotype 3 was detected only after double immunization with the adjuvanted di-BSA conjugate. This finding suggests that the cells whose number showed a large increase after booster immunization (TCR^+ γδT cells and $CD5^+$ B1 cells), against the background of an increase in the number of activated cells expressing MHC II^+, play a crucial role in the protective activity of the conjugate.

3.5 Antibodies against double-stranded DNA

No difference was observed in the level of Abs against ds DNA relative to the control at the dilution of 1:80 in sera of mice immunized with the di-BSA conjugate adsorbed and non-adsorbed on aluminum hydroxide, Prevnar 13, BSA, and free aluminum hydroxide (Figure 6).

4 Discussion

In contrast to the conjugate without adjuvant, the di-BSA conjugate adjuvanted with aluminum hydroxide, induced production of IgG1 antibodies and protected mice against *S. pneumoniae* serotype 3 after prime-boost immunization. The role of adjuvants in enhancing the adaptive immune response to antigens, including semisynthetic glycoconjugates corresponds to the data of other authors (49–51).

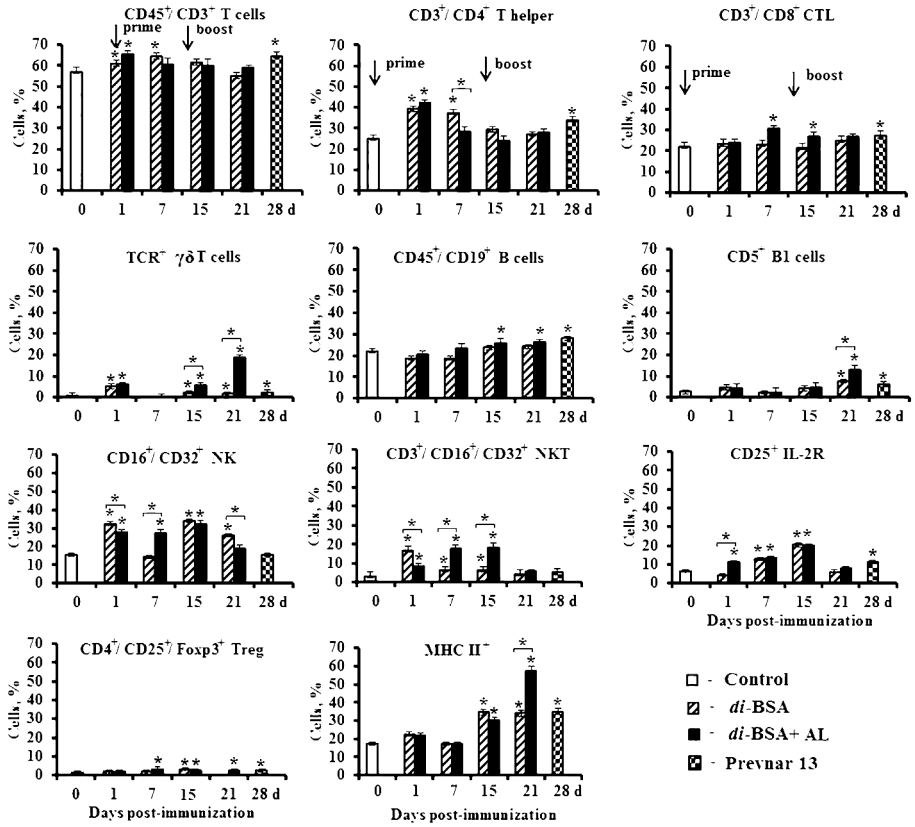


FIGURE 5
The number of splenocytes expressing membrane molecules in mice immunized with the di-BSA conjugate with and without adjuvant. BALB/c mice were immunized with the di-BSA conjugate (20 μg/dose of carbohydrate per mouse) adjuvanted or non-adjuvanted with aluminum hydroxide and with Prevnar 13 (1.1 μg/dose of carbohydrate of CP *S. pneumoniae* serotype 3 per mouse) adjuvanted with aluminum phosphate. Splenocytes were isolated from mice (n = 6 for each conjugate and each time point) on the indicated days after immunization. Control mice (n = 6) were injected with saline 24 hours before the start of immunization (0 d). Spleen cell suspensions were stained using antibodies against mouse CD3e-FITC (clone 145–2C11), CD4-FITC (clone GK1.5), CD8a-FITC (clone 53–6.7), γδT (clone γδ TCR-PE, eBioGL3), CD19-FITC (eBio1D3), CD5-PE (clone 53–7.3), NK.1.1 (clone PK136) CD25-PE (PC61.5), and MHCII-PE (I-EK) (clone 14–4-45). Treg: FITC anti-mouse CD4 (clone GK1.5). Staining with anti-Foxp3-APCconjugated Ab (clone FJK-16s) was performed according to the manufacturer's protocol. The results were determined using flow cytometry. The data are shown as the mean ± SD. Mann-Whitney rank sum tests were used to calculate significant differences between control and other experimental groups; *$P < 0.05$.

The concentrations of IL-1α, IL-1β, IL-4, IL-5, IL-6, IL-10, IL-13, IL-17A, IL-21, IL-22, IFNγ, and TNFα in mice sera in response to the di-BSA conjugate adjuvanted with aluminum hydroxide were higher compared with those in response to the non-adjuvanted glycoconjugate. Free aluminum hydroxide is known to stimulate the production of IL-1β and IL-18, and, when administered with antigens, the spectrum of cytokines expands (52–58). IFNγ, IL-17A, and IL-22 (a member of the Th17 cytokine family) plays a role in the early stages of controlling *S. pneumoniae* infections (59–67). IL-17 has an important function in protecting against bacterial carriage and lung infection (59, 65, 68–71). The di-BSA conjugate adjuvanted with aluminum hydroxide induced a very high level of IL-17A after the prime and boost immunizations, while the conjugate without adjuvant caused a weak production of IL-17A that gradually decreased after the booster injection. A high level of Th2 cytokines (IL-4 and IL-5) was revealed in mice immunized with the adjuvanted di-BSA conjugate. Th2 cytokines promote switching from IgM to IgG, which is associated with high production of IgG1 antibodies (72). The conjugate without adjuvant did not elicit IL-4 production, only weakly stimulated the production of IL-5 even after boost immunization, and did not induce the antibody response. Prevnar 13 is known to induce the production of Th1/

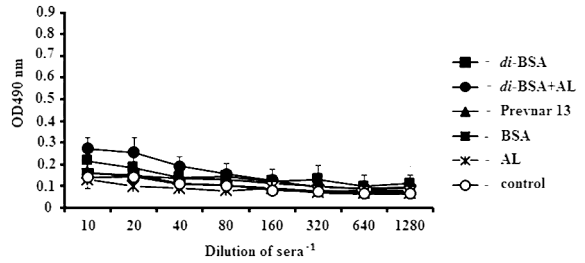


FIGURE 6
IgG antibodies to double-stranded DNA in immunized mice, analyzed by ELISA. ds DNA was used as the well-coating antigen. Sera to each conjugate, BSA, aluminum hydroxide, and control (non-immunized mice) (n = 6 for each antigen) was added to each well in dilutions from 1:10 to 1:1280. AL - aluminum hydroxide; control - mice injected with saline. After prime-boost immunization, autoantibodies to ds DNA, which target the cell nuclei, were not detected.

Th2 and Th17 cytokines (47, 48). In our previous studies, we have shown that Prevnar 13 induced anti-CP *S. pneumoniae* type 3 IgG1-antibodies and protected immunized mice from the challenge with *S. pneumoniae* type 3 (42).

The di-BSA conjugate and CPs, including that of *S. pneumoniae* serotype 3, are not Toll-like receptor (TLR) ligands (46). Purified CP from *S. pneumoniae* can bind to macrophages through the carbohydrate-recognition domains on the mannose receptor, leading to the production of proinflammatory cytokines such as IL-1, IL-6, and TNFα, as well as chemokines (73). Another receptor, the C-type lectin, also known as carbohydrate-binding protein, SIGN-R1, is expressed by macrophages, particularly in the marginal zone of the mouse spleen. This receptor is able to bind carbohydrates from several different serotypes of *S. pneumoniae* (73). Other carbohydrate-recognition receptors of macrophages remain to be identified (74). It is likely that macrophages play a significant role in the initial stage of the immune response to the di-BSA conjugate (36, 59, 74–77).

Regardless of the presence of the adjuvant, the number of $CD4^+$ T helper cells involved in the adaptive immune response to the antigen increased only after the first immunization with the di-BSA conjugate. The number of $CD4^+$ T cells after booster immunization with the BSA-conjugated synthetic hexasaccharide related to *S. pneumoniae* serotype 14 CP adsorbed on aluminum hydroxide did not differ from that in the control either (46). However, the number of $CD4^+$ T helper cells increased on day 14 after booster immunization in mice immunized with CP of *S. pneumoniae* serotype 3 conjugated to CRM_{197} and adsorbed on aluminum phosphate (Prevnar 13). This result may be attributable to the multicomponent composition of the vaccine and the presence of a small amount of bacterial impurities remaining even after purification of CPs. The number of $CD8^+$ cytotoxic cells (CTL) in response to the disaccharide conjugate and Prevnar 13 increased.

Both the adjuvanted di-BSA conjugate and Prevnar 13 significantly increased the number of (TCR^+) γδ T cells among the splenocytes after booster immunization. γδ T cells play a crucial role in prevention of pneumococcal infection owing to their ability to recognize unprocessed non-peptide antigens (41). A large number of γδ T cell ligands remain unknown to date (78, 79). In mice, most γδ T cells are found in the body's barrier tissues, with a small proportion in the blood and spleen (46, 80–83). The activation of γδ T cells through TCRs can be mediated by non-classical MHC molecules (e.g., T10/T22 and members of the CD1 family) and MHC-unrelated molecules (e.g., viral glycoproteins and butyrophilin 3A1) (79, 84–87). Putatively, γδ T cells bind the oligosaccharide portion of the glycoconjugate without processing in antigen-presenting cells (APCs) in combination with MHC-like molecules activate cytokine production. γδ T cells produce a large variety of cytokines and exhibit potent cytotoxic activity against pathogens through apoptosis-inducing receptors (FAS and TRAIL), as well as cytolytic proteins such as perforin and granzyme (88, 89). Furthermore, γδ T cells can function as professional APCs that require surface interactions with opsonized cells (90). The di-BSA conjugate has been shown to induce the formation of opsonizing antibodies (42). Certain subsets of γδ T cells express CD4. These cells have a Th1 or Th2 phenotype and produce IL-2, IL-4, IL-17A, IFNγ, and TNF (70). The di-BSA conjugate induced the production of IFNγ and TNFα (Th1 cytokines); IL-4, IL-5, and IL-13 (Th2 cytokines); IL-17A (Th17 cytokines); IL-21 (Th2 and Th17 subsets); and IL-22 (Th1 and Th17 subsets). γδ T cells play a crucial role in immune protection against extracellular respiratory bacteria (41, 91, 92). The potential role of γδ T cells in pneumococcal infection has only been investigated in animal models using *S. pneumoniae* serotype 3 (41). During infection, the number of γδ T cells can significantly increase, accounting for up to 50% of all peripheral lymphocytes (93, 94). In the mouse model, γδ T cells accumulate and become activated in the lungs during *S. pneumoniae* infection (95, 96). Mice with a lack of γδ T cells exhibit a higher bacterial load in their lungs and lower survival rates compared to control mice (66, 95, 97). The absence of γδ T cells is associated with impaired secretion of MIP-2, TNFα, and IL-17, as well as a poor recruitment of neutrophils (66, 95, 97). In addition, γδ T cells produce IFNγ during *S. pneumoniae* infections of serotypes 3 and 1. Along with their early role in defense against *S. pneumoniae*, γδ T cells participate in the resolution stage of pneumococcal pneumonia, eliminating inflammatory mononuclear phagocytes (98). Therefore, γδ T cells are essential for the host's defense against *S. pneumoniae* (66, 95).

The di-BSA conjugate adjuvanted with aluminum hydroxide and Prevnar 13 adsorbed on aluminum phosphate induced a significant increase $CD5^+$ B1 cells after booster immunization. $CD5^+$ B1 cells are mainly located in the peritoneal and pleural cavities, but very small amounts were also found in the spleen (99, 100). $CD5^+$ B1 cells are activated primarily by T-independent antigens (101, 102) and play an important role in protecting against pneumococcal infections (103) This role may be attributed to their production of natural antibodies as well as possible participation in the T-dependent immune response (102, 104–109). The B cell receptor (BCR) is involved in the phagocytosis of bacteria by B1 cells (110). $CD5^+$ B1, isolated from the spleens of mice, primarily induce IL-17 production by T cells (111). B1 cells present antigen to antigen-specific T cells and induce more efficient proliferation than conventional $CD19^+$ B cells (107, 108). After immunization with the di-BSA conjugate, the number of $CD19^+$ B cells in the blood increased, regardless the presence of the adjuvant. The number of $CD19^+$ B cells increased during all observation periods. Ovalbumin-presenting B1 cells were found to express a higher level of MHC class II compared to naïve B1 cells.

Immunization with either the adjuvanted or the non-adjuvanted di-BSA conjugate increases the number of natural killer (NK) cells and natural killer T (NKT) cells. NK cells, through the production of IFNγ, participate in the early immune response to pulmonary *S. pneumoniae* infection. NKT cells have a key role in protecting against pneumococcal infection. When mice lacking NKT cells were infected with *S. pneumoniae* serotype 3, they exhibited a higher mortality rate and bacterial load in their lungs compared to wild-type mice. It has been suggested that IFNγ derived from NKT cells has a critical function in protecting mice against pneumococcal pneumonia. Using *S. pneumoniae* serotype 1, it has also been found that NKT cells are an important innate immune effector in clearing pneumococci from the body. NKT cells can indirectly or directly assist B cells in mounting antibody responses and have a crucial role in the production of antibodies

against pneumococcus and in the switch of classes in response to the administration of pneumococcal vaccines (112–114).

IL-17A, γδ T, and $CD5^+$ B1 cells can also contribute to autoimmune diseases (115). In response to infection or immunization, autoreactive clones of B1 cells can be produced in the body's own tissues (109, 116–118). The expansion of autoreactive clones of B cells is controlled by IL-10, leaving the BCR in a state of anergy. After booster immunization, there was an increase in the number of $CD4^+/CD25^+/FoxP3^+$ T regulatory cells (Tregs) on the background of interleukin-10 (IL-10) production, which regulates the development of the immune response. After prime-boost immunization with the di-BSA conjugate or Prevnar 13, no formation of autoantibodies against ds DNA targeting cell nuclei was detected.

5 Conclusion

The key effectors of the immune response in mice following immunization with aluminum hydroxide adjuvanted di-BSA conjugate, associated with antibody response and protection from infection by *S. pneumoniae* serotype 3, were IL-17A, γδ T, and $CD5^+$ B1 cells, with an increase in the number of MHC II-expressing cells after booster immunization. The roles of non-conventional γδ T cells, B1 cells, and production of IL-17A upon pneumococcal immunization with the semisynthetic glycoconjugate may provide an in-depth understanding of post-vaccination defense mechanisms, enabling the development of novel efficient therapies and improvement of existing vaccine formulations.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material. Further inquiries can be directed to the corresponding author/s.

Ethics statement

The animal study was approved by Mechnikov Research Institute for Vaccines and Sera Ethics Committee. The study was conducted in accordance with the local legislation and institutional requirements.

Author contributions

EK: Conceptualization, Investigation, Writing – original draft. NA: Investigation, Methodology, Writing – review & editing. AZ: Investigation, Methodology, Writing – review & editing. EA: Investigation, Methodology, Writing – review & editing. NY: Investigation, Methodology, Writing – review & editing. ES: Investigation, Methodology, Writing – review & editing. DY: Investigation, Methodology, Writing – review & editing. YT: Investigation, Methodology, Writing – review & editing. NN: Conceptualization, Funding acquisition, Project administration, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by the Russian Science Foundation (grant no. 19-73-30017-P).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. Rodgers GL, Arguedas A, Cohen R, Dagan R. Global serotype distribution among *Streptococcus pneumoniae* isolates causing otitis media in children: potential implications for pneumococcal conjugate vaccines. *Vaccine*. (2009) 27:3802–10. doi: 10.1016/j.vaccine.2009.04.021

2. O'Brien KL, Wolfson LJ, Watt JP, Henkle E, Deloria-Knoll M, McCall N, et al. Burden of disease caused by *Streptococcus pneumoniae* in children younger than 5 years: global estimates. *Lancet*. (2009) 374:893–902. doi: 10.1016/S0140–6736(09)61204–6

3. Martens P, Worm SW, Lundgren B, Konradsen HB. Benfield T Serotype-specific mortality from invasive *Streptococcus pneumoniae* disease revisited. *BMC Infect Dis*. (2004) 4:21. doi: 10.1186/1471-2334-4-21

4. Harboe ZB, Thomsen RW, Riis A, Valentiner-Branth P, Christensen JJ, Lambertsen L, et al. Pneumococcal serotypes and mortality following invasive pneumococcal disease: a population-based cohort study. *PLoS Med*. (2009) 6: e1000081. doi: 10.1371/journal.pmed.1000081

5. Weinberger DM, Harboe ZB, Sanders EA, Ndiritu M, Klugman KP, Rückinger S, et al. Association of serotype with risk of death due to pneumococcal pneumonia: a meta-analysis. *Clin Infect Dis*. (2010) 51:692–9. doi: 10.1086/655828

6. Grabenstein JD, Musey LK. Differences in serious clinical outcomes of infection caused by specific pneumococcal serotypes among adults. *Vaccine*. (2014) 32:2399–405. doi: 10.1016/j.vaccine.2014.02.096

7. Inverarity D, Lamb K, Diggle M, Robertson C, Greenhalgh D, Mitchell TJ, et al. Death or survival from invasive pneumococcal disease in Scotland: associations with serogroups and multilocus sequence types. *J Med Microbiol*. (2011) 60:793–802. doi: 10.1099/jmm.0.028803-0

8. PLoSker GL. 13-valent pneumococcal conjugate vaccine: a review of its use in infants, children, and adolescents. *Pediatr Drugs*. (2013) 15:403–23. doi: 10.1007/s40272-013-0047-z

9. Namkoong H, Ishii M, Funatsu Y, Kimizuka Y, Yagi K, Asami T, et al. Theory and strategy for pneumococcal vaccines in the elderly. *Hum Vaccin Immunother*. (2016) 12:336–43. doi: 10.1080/21645515.2015.1075678

10. Gransden WR, Eykyn SJ, Phillips I. Pneumococcal bacteraemia: 325 episodes diagnosed at St Thomas's Hospital. *Br Med J (Clin Res Ed)*. (1985) 290:505–8. doi: 10.1136/bmj.290.6467.505

11. Inostroza J, Vinet AM, Retamal G, Lorca P, Ossa G, Facklam RR, et al. Influence of patient age on *Streptococcus pneumoniae* serotypes causing invasive disease. *Clin Diagn Lab Immunol*. (2001) 8:556–9. doi: 10.1128/CDLI.8.3.556-559.2001

12. Scott JA, Hall AJ, Dagan R, Dixon JM, Eykyn SJ, Fenoll A, et al. Serogroup-specific epidemiology of *Streptococcus pneumoniae*: associations with age, sex, and geography in 7,000 episodes of invasive disease. *Clin Infect Dis*. (1996) 22:973–81. doi: 10.1093/clinids/22.6.973

13. España PP, Uranga A, Ruiz LA, Quintana JM, Bilbao A, Aramburu A, et al. Evolution of serotypes in bacteremic pneumococcal adult pneumonia in the period 2001–2014, after introduction of the pneumococcal conjugate vaccine in Bizkaia (Spain). *Vaccine*. (2019) 37:3840–8. doi: 10.1016/j.vaccine.2019.05.052

14. Lee S, Lee K, Kang Y, Bae S. Prevalence of serotype and multidrug-resistance of S. pneumoniae respiratory tract isolates in 265 adult and 36 children in Korea, 2002–2005. *Microb Drug Resist*. (2010) 16:135–42. doi: 10.1089/mdr.2009.0114

15. Jansen AG, Hak E, Veenhoven RH, Damoiseaux RA, Schilder AG, Sanders EA. Pneumococcal conjugate vaccines for preventing otitis media. *Cochrane Database Syst Rev*. (2009) 2:CD001480. doi: 10.1002/14651858.CD001480.pub3

16. Shiramoto M, Hanada R, Juergens C, Shoji Y, Yoshida M, Ballan B, et al. Immunogenicity and safety of the 13-valent pneumococcal conjugate vaccine compared to the 23-valent pneumococcal polysaccharide vaccine in elderly Japanese adults. *Hum Vaccin Immunother*. (2015) 11:2198–206. doi: 10.1080/21645515.2015.1030550

17. Andrews NJ, Waight PA, Burbidge P, Pearce E, Roalfe L, Zancolli M, et al. Serotype-specific effectiveness and correlates of protection for the 13-valent pneumococcal conjugate vaccine: a postlicensure indirect cohort study. *Lancet Infect Dis*. (2014) 14:839–46. doi: 10.1016/S1473-3099(14)70822-9

18. Schuerman L, Prymula R, Henckaerts I, Poolman J. ELISA IgG concentrations and opsonophagocytic activity following pneumococcal protein D conjugate vaccination and relationship to efficacy against acute otitis media. *Vaccine*. (2007) 25:1962–8. doi: 10.1016/j.vaccine.2006.12.008

19. Prymula R, Peeters P, Chrobok V, Kriz P, Novakova E, Kaliskova E, et al. Pneumococcal capsular polysaccharides conjugated to protein D provide protection against otitis media caused by both *Streptococcus pneumoniae* and non-typable *Haemophilus influenzae*: a randomized double blind efficacy study. *Lancet*. (2006) 367:740–8. doi: 10.1016/S0140-6736(06)68304-9

20. Yu X, Sun Y, Frasch C, Concepcion N, Nahm MH. Pneumococcal capsular polysaccharide preparations may contain non-C-polysaccharide contaminants that are immunogenic. *Clin Diagn Lab Immunol*. (1999) 6:519–24. doi: 10.1128/CDLI.6.4.519–524.1999

21. Kochetkov NK, Nifant'ev NE, Backinowsky LV. Synthesis of the capsular polysaccharide of *Streptococcus pneumoniae* type 14. *Tetrahedron*. (1987) 43:3109–21. doi: 10.1016/S0040–4020(01)86852–6

22. Sorieul C, Dolce M, Romano MR, Codée J, Adamo R. Glycoconjugate vaccines against antimicrobial resistant pathogens. *Expert Rev Vaccines*. (2023) 22:1055–78. doi: 10.1080/14760584.2023.2274955

23. del Bino L, Østerlid KE, Wu D-Y, Nonne F, Romano MR, Codée J, et al. Synthetic glycans to improve current glycoconjugate vaccines and fight antimicrobial resistance. *Chem Rev*. (2022) 122:15672–716. doi: 10.1021/acs.chemrev.2c00021

24. Krylov VB, Nifantiev NE. Synthetic carbohydrate based anti-fungal vaccines. *Drug Discovery Today: Technol*. (2020) 35–36:35–43. doi: 10.1016/j.ddtec.2020.11.002

25. Anish C, Schumann B, Pereira CL, Seeberger PH. Chemical biology approaches to designing defined carbohydrate vaccines. *Chem Biol*. (2014) 21:38–50. doi: 10.1016/j.chembiol.2014.01.002

26. Gening ML, Maira-Litran T, Kropec A, Skurnik D, Grout M, Tsvetkov YE, et al. Synthetic β(1→6)-linked N-acetylated and non-acetylated oligoglucosamines used to produce conjugate vaccines for bacterial pathogens. *Infect Immun*. (2010) 78:764–72. doi: 10.1128/IAI.01093–09

27. Parameswarappa SG, Reppe K, Geissner A, Ménová P, Govindan S, Calow ADJ, et al. A semi-synthetic oligosaccharide conjugate vaccine candidate confers protection against *Streptococcus pneumoniae* serotype 3 infection. *Cell Chem Biol*. (2016) 23:1407–16. doi: 10.1016/j.chembiol.2016.09.016

28. Seeberger PH, Pereira CL, Khan N, Xiao G, Diago-Navarro E, Reppe K, et al. A semi-synthetic glycoconjugate vaccine candidate for Carbapenem-resistant *Klebsiella pneumoniae*. *Angew Chem Int Ed Eng*. (2017) 56:13973–8. doi: 10.1002/anie.201700964

29. Broecker F, Hanske J, Martin CE, Baek JY, Wahlbrink A, Wojcik F, et al. Multivalent display of minimal *Clostridium difficile* glycan epitopes mimics antigenic properties of larger glycans. *Nat Commun*. (2016) 7:11224. doi: 10.1038/ncomms11224

30. Broecker F, Martin CE, Wegner E, Mattner J, Baek JY, Pereira CL, et al. Synthetic lipoteichoic acid glycans are potential vaccine candidates to protect from *Clostridium difficile* infections. *Cell Chem Biol*. (2016) 23:1014–22. doi: 10.1016/j.chembiol.2016.07.009

31. Solovev AS, Denisova EM, Kurbatova EA, Kutsevalova OY, Boronina LG, Ageevets VA, et al. Synthesis of methylphosphorylated oligomannosides structurally related to lipopolysaccharide O-antigens of *Klebsiella pneumoniae* serotype O3 and their application for detection of specific antibodies in rabbit and human sera. *Org Biomol Chem*. (2023) 21:8306–19. doi: 10.1039/D3OB01203D

32. van der Put RMF, Smitsman C, de Haan A, Hamzink M, Timmermans H, Uittenbogaard J, et al. The first-in-human synthetic glycan-based conjugate vaccine candidate against *Shigella*. *ACS Cent Sci*. (2022) 8:449–60. doi: 10.1021/acscentsci.1c01479

33. Laverde D, Romero-Saavedra F, Argunov DA, Enotarpi J, Krylov VB, Kalfopoulou E, et al. Synthetic Oligomers Mimicking Capsular Polysaccharide Diheteroglycan are Potential Vaccine Candidates against Encapsulated Enterococcal Infections. *ACS Infect Dis*. (2020) 6:1816–26. doi: 10.1021/acsinfecdis.0c00063

34. Aguilar-Betancourt A, González-Delgado CA, Cinza-Estévez Z, Martínez-Cabrera J, Véliz-Ríos G, Alemán-Zaldívar R, et al. Safety and immunogenicity of a combined hepatitis B virus-*Haemophilus influenzae* type B vaccine comprising a synthetic antigen in healthy adults. *Hum Vaccin*. (2008) 4:54–9. doi: 10.4161/hv.4.1.5257

35. Tsvetkov Y, Gening ML, Kurbatova EA, Akhmatova NK, Nifantiev NE. Oligosaccharide ligand tuning in design of third generation carbohydrate pneumococcal vaccines. *Pure Appl Chem*. (2017) 89:1403–1411. doi: 10.1515/pac-2016–1123

36. Gening ML, Kurbatova EA, Nifantiev NE. Synthetic analogs of *Streptococcus pneumoniae* capsular polysaccharides and immunogenic activities of glycoconjugates. *Russ J Bioorganic Chem*. (2021) 47:1–25. doi: 10.1134/S1068162021010076

37. Gening ML, Kurbatova EA, Tsvetkov YE, Nifantiev NE. Development of approaches to a conjugated carbohydrate vaccine of the third generation against *Streptococcus pneumoniae*: the search for optimal oligosaccharide ligands. *Russ Chem Rev*. (2015) 84:1100–13. doi: 10.1070/RCR4574

38. Micoli F, Romano MR, Carboni F, Adamo R, Berti F. Strengths and weaknesses of pneumococcal conjugate vaccines. *Glycoconjugate J*. (2023) 40:135–48. doi: 10.1007/s10719–023-10100–3

39. Jansen WT, Snippe H. Short-chain oligosaccharide protein conjugates as experimental pneumococcal vaccines. *Indian J Med Res*. (2004) 119:7–12.

40. Weishaupt MW, Matthies S, Hurevich M, Pereira CL, Hahm HS, Seeberger PH. Automated glycan assembly of a S. pneumoniae serotype 3 CPS antigen. *Beilstein J Org Chem*. (2016) 12:1440–6. doi: 10.3762/bjoc.12.139

41. Ivanov S, Paget C, Trottein F. Role of non-conventional T lymphocytes in respiratory infections: the case of the pneumococcus. *PLoS Pathog*. (2014) 10:e1004300. doi: 10.1371/journal.ppat.1004300

42. Kurbatova EA, Akhmatova NK, Zaytsev AE, Akhmatova EA, Egorova NB, Yastrebova NE, et al. and Nifantiev NE Higher cytokine and opsonizing antibody production induced by bovine serum albumin (BSA)-conjugated tetrasaccharide related to *Streptococcus pneumoniae* type 3 capsular polysaccharide. *Front Immunol*. (2020) 11:578019. doi: 10.3389/fimmu.2020.578019

43. Tsvetkov YE, Yashunsky DV, Sukhova EV, Nifantiev NE, Kurbatova EA. Synthesis of oligosaccharides structurally related to fragments of *Streptococcus pneumoniae* type 3 capsular polysaccharide. *Russ Chem Bull*. (2017) 66:111–22. doi: 10.1007/s11172-017-1708-9

44. Kurbatova EA, Akhmatov EA, Akhmatova NK, Egorova NB, Yastrebova NE, Romanenko EE, et al. The use of biotinylated oligosaccharides related to fragments of capsular polysaccharides from *Streptococcus pneumoniae* serotypes 3 and 14 as a tool for assessment of the level of vaccine-induced antibody response to neoglycoconjugates. *Russ Chem Bull*. (2017) 65:1608–16. doi: 10.1007/s11172–016-1488–7

45. Ananikov VP, Eremin DB, Yakukhnov SA, Dilman AD, Levin VV, Egorov MP, et al. Organic and hybrid systems: from science to practice. *Mendeleev Commun*. (2017) 27:425–38. doi: 10.1016/j.mencom.2017.09.001

46. Akhmatova NK, Kurbatova EA, Akhmatov EA, Egorova NB, Logunov DY, Gening ML, et al. The effect of a BSA conjugate of a synthetic hexasaccharide related to the fragment of capsular polysaccharide of Streptococcus pneumoniae type 14 on the activation of innate and adaptive immune responses. *Front Immunol*. (2016) 7:1–11. doi: 10.3389/fimmu.2016.00248

47. Lai Z, Schreiber JR. Outer membrane protein complex of meningococcus enhances the antipolysaccharide antibody response to pneumococcal polysaccharide-CRM197 conjugate vaccine. *Clin Vaccine Immunol*. (2011) 18:724–29. doi: 10.1128/CVI.00053–11

48. Karasartova D, Gazi U, Tosun O, Gureser AS, Sahiner IT, Dolapci M, et al. Anti-pneumococcal vaccine-induced cellular immune responses in post-traumatic splenectomized individuals. *J Clin Immunol*. (2017) 37:388–96. doi: 10.1007/s10875–017-0397–3

49. Lefeber DJ, Benaissa-Trouw B, Vliegenthart JF, Kamerling JP, Jansen WT, Kraaijeveld K, et al. Th1-directing adjuvants increase the immunogenicity of

oligosaccharide-protein conjugate vaccines related to *Streptococcus pneumoniae* type 3. *Infect Immun*. (2003) 71:6915–20. doi: 10.1128/iai.71.12.6915–6920

50. Reppe K, Tschernig T, Luhrmann A, van Laak V, Grote K, Zemlin MV, et al. Immunostimulation with macrophage-activating lipopeptide-2 increased survival in murine pneumonia. *Am J Respir Cell Mol Biol*. (2009) 40:474–81. doi: 10.1165/rcmb.2008-0071OC

51. Witzenrath M, Pache F, Lorenz D, Koppe U, Gutbier B, Tabeling C, et al. The NLRP3 inflammasome is differentially activated by pneumolysin variants and contributes to host defense in pneumococcal pneumonia. *J Immunol*. (2011) 187:434–40. doi: 10.4049/jimmunol.1003143

52. He P, Zou Y, Hu Z. Advances in aluminum hydroxide-based adjuvant research and its mechanism. *Hum Vaccin Immunother*. (2015) 11:477–88. doi: 10.1080/21645515.2014.1004026

53. Gupta RK, Siber GR. Adjuvants for human vaccines—current status, problems and future prospects. *Vaccine*. (1995) 13:1263–76. doi: 10.1016/0264-410X(95)00011-O

54. Moingeon P, Haensler J, Lindeberg A. Towards the rational design of Th1 adjuvants. *Vaccine*. (2001) 19:4363–72. doi: 10.1016/S0264-410X(01)00193-1

55. Williams A, Flavell RA, Eisenbarth SC. The role of NOD-like Receptors in shaping adaptive immunity. *Curr Opin Immunol*. (2010) 22:34–40. doi: 10.1016/j.coi.2010.01.004

56. Franchi L, Núñez G. The NLRP3 inflammasome is critical for alum-mediated IL-1β secretion but dispensable for adjuvant activity. *Eur J Immunol*. (2008) 38:2085–9. doi: 10.1002/eji.200838549

57. Li H, Nookala S, Re F. Aluminum hydroxide adjuvants activate caspase-1 and induce IL-1beta and IL-18 release. *J Immunol*. (2007) 178:5271–6. doi: 10.4049/jimmunol.178.8.5271

58. Li H, Willingham SB, Ting JP-Y, Re F, Edge C. Inflammasome activation by Alum and Alum's adjuvant effect are mediated by NLRP3. *J Immunol*. (2008) 181:17–21. doi: 10.4049/jimmunol.181.1.17

59. Zhang Z, Clarke TB, Weiser JN. Cellular effectors mediating Th17- dependent clearance of pneumococcal colonization in mice. *J Clin Invest*. (2009) 119:1899–909. doi: 10.1172/JCI36731

60. Yamamoto N, Kawakami K, Kinjo Y, Miyagi K, Kinjo T, Uezu K, et al. Essential role for the p40 subunit of interleukin-12 in neutrophil-mediated early host defense against pulmonary infection with *Streptococcus pneumoniae*: involvement of interferon-gamma. *Microbes Infect*. (2004) 6:1241–9. doi: 10.1016/j.micinf.2004.08.007

61. Sun K, Salmon SL, Lotz SA, Metzger DW. Interleukin-12 promotes gamma interferon-dependent neutrophil recruitment in the lung and improves protection against respiratory *Streptococcus pneumoniae* infection. *Infect Immun*. (2007) 75:1196–202. doi: 10.1128/IAI.01403-06

62. Nakamatsu M, Yamamoto N, Hatta M, Nakasone C, Kinjo T, Miyagi K, et al. Role of interferon-gamma in Valpha14+ natural killer T cell-mediated host defense against *Streptococcus pneumoniae* infection in murine lungs. *Microbes Infect*. (2007) 9:364–74. doi: 10.1016/j.micinf.2006.12.003

63. Yamada M, Gomez JC, Chugh PE, Lowell CA, Dinauer MC, Dittmer DP, et al. Interferon-gamma production by neutrophils during bacterial pneumonia in mice. *Am J Respir Crit Care Med*. (2011) 183:1391–401. doi: 10.1164/rccm.201004-0592OC

64. Weber SE, Tian H. Pirofski LA CD8+ cells enhance resistance to pulmonary serotype 3 *Streptococcus pneumoniae* infection in mice. *J Immunol*. (2011) 186:432–42. doi: 10.4049/jimmunol.1001963

65. Lu YJ, Gross J, Bogaert D, Finn A, Bagrade L, Zhang Q, et al. Interleukin-17A mediates acquired immunity to pneumococcal colonization. *PLoS Pathog*. (2008) 4: e1000159. doi: 10.1371/journal.ppat.1000159

66. Ma J, Wang J, Wan J, Charboneau R, Chang Y, Barke RA, et al. Morphine disrupts interleukin-23 (IL-23)/IL-17-mediated pulmonary mucosal host defense against *Streptococcus pneumoniae* infection. *Infect Immun*. (2010) 78:830–7. doi: 10.1128/IAI.00914-09

67. van Maele L, Carnoy C, Cayet D, Ivanov S, Porte R, Deruy E, et al. Activation of type 3 innate lymphoid cells and interleukin 22 secretion in the lungs during *Streptococcus pneumoniae* infection. *J Infect Dis*. (2014) 210:493–503. doi: 10.1093/infdis/jiu106

68. Kadioglu A, Coward W, Colston MJ, Hewitt CR. Andrew PW CD4-T lymphocyte interactions with pneumolysin and pneumococci suggest a crucial protective role in the host response to pneumococcal infection. *Infect Immun*. (2004) 72:2689–97. doi: 10.1128/IAI.72.5.2689-2697.2004

69. Malley R, Trzcinski K, Srivastava A, Thompson CM, Anderson PW, Lipsitch M, et al. CD4+ T cells mediate antibody-independent acquired immunity to pneumococcal colonization. *Proc Natl Acad Sci USA*. (2005) 102:4848–53. doi: 10.1073/pnas.0501254102

70. Trzciński K, Thompson CM, Srivastava A, Basset A, Malley R, Lipsitch M. Protection against nasopharyngeal colonization by *Streptococcus pneumoniae* is mediated by antigen-specific CD4+ T cells. *Infect Immun*. (2008) 76:2678–84. doi: 10.1128/IAI.00141–08

71. Wright AK, Bangert M, Gritzfeld JF, Ferreira DM, Jambo KC, Wright AD, et al. Experimental human pneumococcal carriage augments IL-17A-dependent T cell defence of the lung. *PLoS Pathog*. (2013) 9:e1003274. doi: 10.1371/journal.ppat.1003274

72. Coffman RL, Savelkoul HF, Lebman DA. Cytokine regulation of immunoglobulin isotype switching and expression. *Semin Immunol*. (1989) 1:55–63.

73. Zamz S, Martinez-Pomares L, Jones H, Taylor PR, Stillion RJ, Gordon S, et al. Recognition of bacterial capsular polysaccharides and lipopolysaccharides by the macrophage mannose receptor. *J Biol Chem*. (2002) 277:41613–23. doi: 10.1074/jbc.M207057200

74. Paterson GK, Mitchell TJ. Innate immunity and the pneumococcus. *Microbiology*. (2006) 152:285–93. doi: 10.1099/mic.0.28551–0

75. Kadioglu A, Weiser JN, Paton JC, Andrew PW. The role of *Streptococcus pneumoniae* virulence factors in host respiratory colonization and disease. *Nat Rev Microbiol*. (2008) 6:288–301. doi: 10.1038/nrmicro1871

76. van der Poll T, Opal SM. Pathogenesis, treatment, and prevention of pneumococcal pneumonia. *Lancet*. (2009) 374:1543–56. doi: 10.1016/S0140-6736(09)61114-4

77. Koppe U, Suttorp N, Opitz B. Recognition of *Streptococcus pneumoniae* by the innate immune system. *Cell Microbiol*. (2012) 14:460–66. doi: 10.1111/j.1462-5822.2011.01746.x

78. Ferreira LM. Gammadelta T cells: innately adaptive immune cells? *Int Rev Immunol*. (2013) 32:223–48. doi: 10.3109/08830185.2013.783831

79. Tanaka Y, Sano S, Nieves E, De Libero G, Rosa D, Modlin RL, et al. Nonpeptide ligands for human gamma delta T cells. *ProcNatl Acad Sci USA*. (1994) 91:8175–9. doi: 10.1073/pnas.91.17.8175

80. Sperling AI, Cron RQ, Decker DC, Stern DA, Bluestone JA. Peripheral T cell receptor γδ variable gene repertoire maps to the T cell receptor loci and is influenced by positive selection. *J Immunol*. (1992) 149:3200–207. doi: 10.4049/jimmunol.149.10.3200

81. Strominger JL. Developmental biology of T cell receptors. *Science*. (1989) 244:943–50. doi: 10.1126/science.2658058

82. Allison JP. Havran WL.The immunobiology of T cells with invariant gamma delta antigen receptors. *Annu Rev Immunol*. (1991) 9:679–705. doi: 10.1146/annurev.iy.09.040191.003335

83. Havran WL, Allison JP. Developmentally ordered appearance of thymocytes expressing different T cell antigen receptors. *Nature*. (1988) 335:443–5. doi: 10.1038/335443a0

84. Bonneville M, O'Brien RL, Born WK. Gammadelta T cell effector functions: a blend of innate programming and acquired plasticity. *Nat Rev Immunol*. (2010) 10:467–78. doi: 10.1038/nri2781

85. Kalyan S, Kabelitz D. Defining the nature of human gammadelta T cells: a biographical sketch of the highly empathetic. *Cell Mol Immunol*. (2013) 10:21–9. doi: 10.1038/cmi.2012.44

86. Vantourout P, Hayday A. Six-of-the-best: unique contributions of gammadelta T cells to immunology. *Nat Rev Immunol*. (2013) 13:88–100. doi: 10.1038/nri3384

87. Martin B, Hirota K, Cua DJ, Stockinger B, Veldhoen M. Interleukin-17-producing gammadelta T cells selectively expand in response to pathogen products and environmental signals. *Immunity*. (2009) 31:321–30. doi: 10.1016/j.immuni.2009.06.020

88. Dieli F, Troye-Blomberg M, Ivanyi J, Fournie JJ, Krensky AM, Bonneville M, et al. Granulysin-dependent killing of intracellular and extracellular Mycobacterium tuberculosis by Vgamma9/Vdelta2 T lymphocytes. *J Infect Dis*. (2001) 184:1082–5. doi: 10.1086/323600

89. Qin G, Mao H, Zheng J, Sia SF, Liu Y, Chan PL, et al. Phosphoantigen-expanded human gammadelta T cells display potent cytotoxicity against monocytederived macrophages infected with human and avian influenza viruses. *J Infect Dis*. (2009) 200:858–65. doi: 10.1086/605413

90. Himoudi N, Morgenstern DA, Yan M, Vernay B, Saraiva L, Wu Y, et al. Human gammadelta T lymphocytes are licensed for professional antigen presentation by interaction with opsonized target cells. *J Immunol*. (2012) 188:1708–16. doi: 10.4049/jimmunol.1102654

91. Zheng J, Liu Y, Lau YL, Tu W. gammadelta-T cells: an unpolished sword in human anti-infection immunity. *Cell Mol Immunol*. (2013) 10:50–7. doi: 10.1038/cmi.2012.43

92. Cheng P, Liu T, Zhou WY, Zhuang Y, Peng LS, Zhang JY, et al. Role of gammadelta T cells in host response against *Staphylococcus aureus*-induced pneumonia. *BMC Immunol*. (2012) 13:38. doi: 10.1186/1471-2172-13-38

93. Das H, Groh V, Kuijl C, Sugita M, Morita CT, Spies T, et al. MICA engagement by human Vgamma2Vdelta2 T cells enhances their antigen-dependent effector function. *Immunity*. (2001) 15:83–93. doi: 10.1016/S1074-7613(01)00168-6

94. Bertotto A, Gerli R, Spinozzi F, Muscat C, Scalise F, Castellucci G, et al. Lymphocytes bearing the gamma delta T cell receptor in acute *Brucella melitensis* infection. *Eur J Immunol*. (1993) 23:1177–80. doi: 10.1002/eji.1830230531

95. Nakasone C, Yamamoto N, Nakamatsu M, Kinjo T, Miyagi K, Uezu K, et al. Accumulation of gamma/delta T cells in the lungs and their roles in neutrophil-mediated host defense against pneumococcal infection. *Microbes Infect*. (2007) 9:251–8. doi: 10.1016/j.micinf.2006.11.015

96. Kirby AC, Newton DJ, Carding SR, Kaye PM. Evidence for the involvement of lung-specific gammadelta T cell subsets in local responses to *Streptococcus pneumoniae* infection. *Eur J Immunol*. (2007) 37:3404–13. doi: 10.1002/eji.200737216

97. Cao J, Wang D, Xu F, Gong Y, Wang H, Song Z, et al. Activation of IL-27 signalling promotes development of postinfluenza pneumococcal pneumonia. *EMBO Mol Med*. (2014) 6:120–40. doi: 10.1002/emmm.201302890

98. Kirby AC, Newton DJ, Carding SR, Kaye PM. Pulmonary dendritic cells and alveolar macrophages are regulated by gammadelta T cells during the resolution of S. pneumoniae-induced inflammation. *J Pathol*. (2007) 212:29–37. doi: 10.1002/path.2149

99. Deenen GJ, Kroese FG. Kinetics of peritoneal B-1a cells (CD5 B cells) in young adult mice. *Eur J Immunol*. (1993) 23:12–6. doi: 10.1002/eji.1830230104

100. Kroese FG, Ammerlaan WA, Deenen GJ. Location and function of B-cell lineages. *Ann N Y Acad Sci*. (1992) 651:44–58. doi: 10.1111/j.1749-6632.1992.tb24592.x

101. Martin F, Oliver AM, Kearney JF. Marginal zone and B1 B cells unite in the early response against T-independent blood-borne particulate antigens. *Immunity*. (2001) 14:617–29. doi: 10.1016/S1074–7613(01)00129–7

102. Margry B, Wieland WH, van Kooten PJ, van Eden W, Broere F. Peritoneal cavity B-1a cells promote peripheral CD4+ T-cell activation. *Eur J Immunol*. (2013) 43:2317–26. doi: 10.1002/eji.201343418

103. Sindhava VJ, Bondada S. Multiple regulatory mechanisms control B-1 B cell activation. *Front Immunol*. (2012) 3:372. doi: 10.3389/fimmu.2012.00372

104. Sato T, Ishikawa S, Akadegawa K, Ito T, Yurino H, Kitabatake M, et al. Aberrant B1 cell migration into the thymus results in activation of CD4 T cells through its potent antigen-presenting activity in the development of murine lupus. *Eur J Immunol*. (2004) 34:3346–58. doi: 10.1002/eji.200425373

105. Vigna AF, Godoy LC, Rogerio de Almeida S, Mariano M, Lopes JD. Characterization of B-1b cells as antigen presenting cells in the immune response to gp43 from *Paracoccidioides brasiliensis in vitro*. *Immunol Lett*. (2002) 83:61–6. doi: 10.1016/S0165–2478(02)00070–6

106. Wang Y, Rothstein TL. Induction of Th17 cell differentiation by B-1 cells. *Front Immunol*. (2012) 3:281. doi: 10.3389/fimmu.2012.00281

107. Zimecki M, Whiteley PJ, Pierce CW, Kapp JA. Presentation of antigen by B cells subsets. I. Lyb-5+ and Lyb-5- B cells differ in ability to stimulate antigen specific T cells. *Arch Immunol Ther Exp (Warsz)*. (1994) 42:115–23.

108. Zimecki M, Kapp JA. Presentation of antigen by B cell subsets. II. The role of CD5 B cells in the presentation of antigen to antigen-specific T cells. *Arch Immunol Ther Exp (Warsz)*. (1994) 42:349–53.

109. Berland R, Wortis HH. Origins and functions of B-1 cells with notes on the role of CD5. *Annu Rev Immunol*. (2002) 20:253–300. doi: 10.1146/annurev.immunol.20.100301.064833

110. Gao J, Ma X, Gu W, Fu M, An J, Xing Y, et al. Novel functions of murine B1 cells: active phagocytic and microbicidal abilities. *Eur J Immunol*. (2012) 42:982–92. doi: 10.1002/eji.201141519

111. Zhong X, Gao W, Degauque N, Bai C, Lu Y, Kenny J, et al. Reciprocal generation of Th1/Th17 and T(reg) cells by B1 and B2 B cells. *Eur J Immunol*. (2007) 37:2400–4. doi: 10.1002/eji.200737296

112. Kobrynski LJ, Sousa AO, Nahmias AJ, Lee FK. Cutting edge: antibody production to pneumococcal polysaccharides requires CD1 molecules and CD8+ T cells. *J Immunol*. (2005) 174:1787–90. doi: 10.4049/jimmunol.174.4.1787

113. Miyasaka T, Akahori Y, Toyama M, Miyamura N, Ishii K, Saijo S, et al. Dectin-2-dependent NKT cell activation and serotype-specific antibody production in mice immunized with pneumococcal polysaccharide vaccine. *PLoS One*. (2013) 8:e78611. doi: 10.1371/journal.pone.0078611

114. Miyasaka T, Aoyagi T, Uchiyama B, Oishi K, Nakayama T, Kinjo Y, et al. A possible relationship of natural killer T cells with humoral immune response to 23-valent pneumococcal polysaccharide vaccine in clinical settings. *Vaccine*. (2012) 30:3304–10. doi: 10.1016/j.vaccine.2012.03.007

115. Su D, Shen M, Li X, Sun L. Roles of γδ T cells in the pathogenesis of autoimmune diseases. *Clin Dev Immunol*. (2013) 2013:985753. doi: 10.1155/2013/985753

116. Choi YS, Baumgarth N. Dual role for B-1a cells in immunity to influenza virus infection. *J Exp Med*. (2008) 205:3053–64. doi: 10.1084/jem.20080979

117. Haas KM, Poe JC, Steeber DA, Tedder TF. B-1a and B-1b cells exhibit distinct developmental requirements and have unique functional roles in innate and adaptive immunity to *S. pneumoniae*. *Immunity*. (2005) 23:7–18. doi: 10.1016/j.immuni.2005.04.011

118. Popi AF, Longo-Maugéri IM, Mariano M. An Overview of B-1 cells as antigen-presenting cells. *Front Immunol*. (2016) 7:138. doi: 10.3389/fimmu.2016.00138

frontiers | Frontiers in Immunology

TYPE Original Research
PUBLISHED 06 June 2024
DOI 10.3389/fimmu.2024.1347926

OPEN ACCESS

EDITED BY
Zaza Mtine Ndhlovu,
Ragon Institute, United States

REVIEWED BY
Eduardo L. V. Silveira,
University of São Paulo, Brazil
Ravi K. Patel,
University of California, San Francisco, United States

*CORRESPONDENCE
Tim R. Mosmann
tim_mosmann@urmc.rochester.edu

RECEIVED 01 December 2023
ACCEPTED 20 May 2024
PUBLISHED 06 June 2024

CITATION
Mosmann TR, Rebhahn JA, De Rosa SC, Keefer MC, McElrath MJ, Rouphael NG, Pantaleo G, Gilbert PB, Corey L, Kobie JJ and Thakar J (2024) SWIFT clustering analysis of intracellular cytokine staining flow cytometry data of the HVTN 105 vaccine trial reveals high frequencies of HIV-specific CD4+ T cell responses and associations with humoral responses. *Front. Immunol.* 15:1347926. doi: 10.3389/fimmu.2024.1347926

SWIFT clustering analysis of intracellular cytokine staining flow cytometry data of the HVTN 105 vaccine trial reveals high frequencies of HIV-specific CD4+ T cell responses and associations with humoral responses

Tim R. Mosmann[1*], Jonathan A. Rebhahn[1], Stephen C. De Rosa[2], Michael C. Keefer[3], M. Juliana McElrath[2], Nadine G. Rouphael[4], Giuseppe Pantaleo[5,6], Peter B. Gilbert[2], Lawrence Corey[2], James J. Kobie[7] and Juilee Thakar[8]

[1]David H. Smith Center for Vaccine Biology and Immunology, University of Rochester Medical Center, Rochester, NY, United States, [2]Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Center, Seattle, WA, United States, [3]Department of Medicine, University of Rochester School of Medicine & Dentistry, Rochester, NY, United States, [4]Hope Clinic of the Emory Vaccine Center, Division of Infectious Diseases, Emory University, Atlanta, GA, United States, [5]Service of Immunology and Allergy, Department of Medicine, Lausanne University Hospital and University of Lausanne, Lausanne, Switzerland, [6]Swiss Vaccine Research Institute, Lausanne University Hospital and University of Lausanne, Lausanne, Switzerland, [7]Department of Medicine, University of Alabama at Birmingham, Birmingham, AL, United States, [8]Department of Microbiology and Immunology, University of Rochester Medical Center, Rochester, NY, United States

Introduction: The HVTN 105 vaccine clinical trial tested four combinations of two immunogens - the DNA vaccine DNA-HIV-PT123, and the protein vaccine AIDSVAX B/E. All combinations induced substantial antibody and CD4+ T cell responses in many participants. We have now re-examined the intracellular cytokine staining flow cytometry data using the high-resolution SWIFT clustering algorithm, which is very effective for enumerating rare populations such as antigen-responsive T cells, and also determined correlations between the antibody and T cell responses.

Methods: Flow cytometry samples across all the analysis batches were registered using the swiftReg registration tool, which reduces batch variation without compromising biological variation. Registered data were clustered using the SWIFT algorithm, and cluster template competition was used to identify clusters of antigen-responsive T cells and to separate these from constitutive cytokine producing cell clusters.

Results: Registration strongly reduced batch variation among batches analyzed across several months. This in-depth clustering analysis identified a greater proportion of responders than the original analysis. A subset of antigen-responsive clusters producing IL-21 was identified. The cytokine patterns in

10.3389/fimmu.2024.1347926

each vaccine group were related to the type of vaccine – protein antigens tended to induce more cells producing IL-2 but not IFN-γ, whereas DNA vaccines tended to induce more IL-2+ IFN-γ+ CD4 T cells. Several significant correlations were identified between specific antibody responses and antigen-responsive T cell clusters. The best correlations were not necessarily observed with the strongest antibody or T cell responses.

Conclusion: In the complex HVTN105 dataset, alternative analysis methods increased sensitivity of the detection of antigen-specific T cells; increased the number of identified vaccine responders; identified a small IL-21-producing T cell population; and demonstrated significant correlations between specific T cell populations and serum antibody responses. Multiple analysis strategies may be valuable for extracting the most information from large, complex studies.

KEYWORDS

HIV - human immunodeficiency virus, vaccine trial, reanalysis, algorithmic flow cytometry analysis, T cell response, T cell antibody correlation

Introduction

The HIV Vaccine Trials Network (HVTN) 105 phase I trial (ClinicalTrials.gov NCT02207920) was designed to build on the encouraging results of the RV144 "Thai Trial" HIV vaccine efficacy trial which demonstrated modest protection from HIV infection (1). RV144 immunization included AIDSVAX B/E consisting of clade B MN gp120 and clade E A244 gp120 proteins in alum given following priming immunizations with a canarypox vector vaccine. HVTN 105 investigated the preventative vaccine strategy of priming with DNA-HIV-PT123 which consisted of 3 plasmids encoding clade C ZM96 gag, clade C ZM96 gp140, and clade C CN54 pol-nef followed by boosting with AIDSVAX B/E in four treatment groups of healthy HIV-1 negative individuals at low risk of HIV acquisition to determine which strategy would best elicit favorable HIV-specific antibody and T cell responses (2). DNA vaccines are thermostable, are relatively straightforward to manufacture, and provide more flexibility for vaccine design through formulation of multiple plasmids containing different HIV components and/or adjuvants in a single injection.

The HVTN 105 trial administered intramuscular injections at 0, 1, 3, and 6 months (M). T1 received protein at M0 and M1 and DNA at M3 and M6; T2 received DNA at M0 and M1 and protein at M3 and M6; T3 received DNA at M0, M1, M3, and M6 with protein co-administered at M3 and M6; and T4 received protein and DNA co-administered at each vaccination visit (Figure 1).

The primary immunogenicity analysis was conducted 2 weeks following the final vaccination and evaluation of durability of the immune response was conducted at 6 months following the final vaccination. The previous analysis of humoral responses showed the groups receiving protein at M0 and M1, T1 and T4 had a >85% IgG response rate for ZM96.C and A244.AE after the second vaccination, however, the response rate for T1 was not sustained after subsequent vaccinations, likely a consequence of boosting with DNA only (2). After the final vaccination there was an 80% response rate for T2 and 100% for T3 and T4. Importantly, 2 weeks following the final vaccination, binding-IgG responses to the HIV V1V2 antigens that were identified as potential inverse correlates of risk (A244.AE V1V2 and 1086.C V1V2) from RV144 (3) were observed in 96% or more vaccinees in groups T2, T3, and T4. Over time geometric mean response magnitudes were similar across HIV antigens (vaccine-matched vs. consensus HIV envelopes, V1V2 antigens).

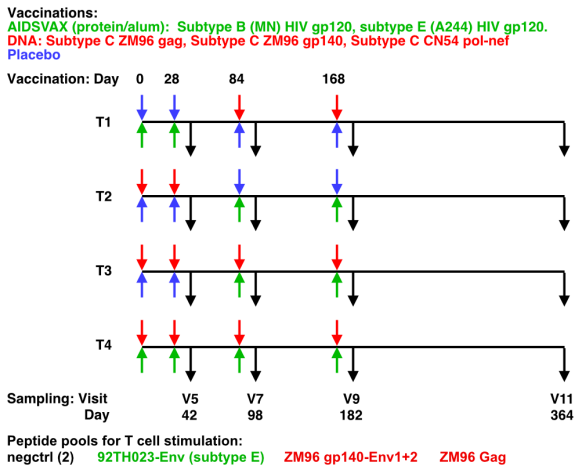


FIGURE 1
HVTN105 protocol. 26 participants per group were immunized with the indicated vaccines in the left (upper) and right (lower) deltoid muscles, at each of the time points indicated. Black arrows indicate PBMC sampling times.

HIV-1–specific CD4+ and CD8+ T cell responses were examined by flow cytometry, using a validated 17-color intracellular cytokine staining (ICS) assay, two weeks after each boost as well as 12 months after enrollment. The peptide pools evaluated were vaccine matched (ZM96 gp140-Env1, ZM96 gp140-Env2, 92TH023-Env, and ZM96 Gag), covering Env and Gag. In the previous analysis, which was done with rigorous manual templated gating of the T cell populations, vaccine-induced CD4+ T cell responses were detected in all groups. There were minimal differences found across groups, although a trend of higher CD4+ responses in T3 was observed. However, this trend of higher CD4+ responses was significant when the polyfunctionality score was assessed (4). The two prominent polyfunctional CD4+ populations were the four-function IFN-γ^+IL-2$^+$TNF-α^+CD40L$^+$ and the three function IL-2$^+$TNF-α^+CD40L$^+$.

The massive size and dimensionality of flow cytometry data is challenging for comprehensive manual analysis approaches, including its subjective and time-consuming nature, and the concern that novel and overlapping cell populations may be underappreciated with *a priori* gating strategies. Even at peak, there was an overall modest CD4+ T cell response rate to HIV Env (36%-60%) and low response to HIV Gag (0%-40%). We therefore re-analyzed the HVTN flow cytometry data, using the high-resolution SWIFT clustering algorithm (5, 6) that was originally developed to resolve rare cytokine-producing T cell subsets. We included batch registration (7) to reduce differences between batches that might obscure biological differences. Our goal was to increase the resolution of the heterogeneity of the T cell response, and to define responders more clearly.

It is anticipated that an effective HIV vaccine will require both optimal T cell and humoral immunity to confer protection. Given the inter-dependence of CD4+ T cell and antibody responses, we conducted correlative analysis of the clustered antigen-responsive T cell subsets with existing plasma antibody data sets to identify possible novel associations.

The re-analysis of flow data was consistent with the previous analysis, but yielded further discoveries of increased frequencies of participants with T cell responses; T cell sub-populations expressing IL-21; qualitatively different responses induced by DNA vs protein vaccines; and correlations between particular T cell subsets and subsequent antibody responses.

Methods

Data source

FCS files and de-identified metadata from intracellular cytokine staining (ICS) analysis of CD4+ and CD8+ T cell responses was provided from the HVTN from HVTN 105 a Phase 1 preventative vaccine trial (ClinicalTrials.gov NCT02207920). Primary ICS analysis was previously reported (2). Details regarding the study design, participants, sample and data acquisition are included in the primary study manuscript (2). Briefly, participants were randomly assigned to 1 of 4 groups with an allocation ratio of 1:1:1:1 (Figure 1). Participants received different combinations of AIDSVAX B/E, DNA-HIV-PT123, and placebo, administered intramuscularly. AIDSVAX B/E consisted of 300 µg of subtype B (MN) HIV gp120 glycoprotein and 300 µg of subtype A/E (A244) HIV gp120 glycoprotein adsorbed onto aluminum hydroxide gel adjuvant and administered into the right deltoid muscle. DNA-HIV-PT123 contained a mixture of 3 DNA plasmids: (a) clade C ZM96 gag, (b) clade C ZM96 gp140, and (c) clade C CN54 pol-nef, delivered at a total dose of 4 mg administered into the deltoid muscle via needle and syringe. Serum for humoral assays was obtained from serum-separating tubes (SSTs) and frozen at –80°C. Peripheral blood mononuclear cells (PBMCs) for cellular assays were isolated and cryopreserved from within 6 hours of venipuncture, as described previously (8). Flow cytometry was used to examine HIV-1–specific T cell responses using a validated intracellular cytokine staining (ICS) assay. The peptide pools evaluated were vaccine matched (ZM96 gp140-Env1, ZM96 gp140-Env2, 92TH023-Env, and ZM96 Gag), covering Env and Gag. Previously cryopreserved PBMCs were stimulated with the synthetic peptide pools. As a negative control, cells were not stimulated. Serum HIV-1–specific IgG, IgG3, IgG4, and IgA responses were measured with a custom HIV-1–binding antibody multiplex assay (BAMA) as previously described (9, 10) using gp120 proteins and V1V2 antigens detailed previously (11).

Data transformation

The set of fluorescent dimensions $\mathcal{F}$ in $z^{\mathcal{C}}$ were transformed using the "log-like" inverse hyperbolic sine, sinh^{-1}, in conjunction with a set of F-dimensional cofactors $[\alpha_1, \alpha_2,\ldots,\alpha_F]$ for each dimension $j \in \mathcal{F}$. Each vector $z_j^{\mathcal{C}}$ was divided by its corresponding cofactor α_j prior to transformation, which effectively removed the artifactual bimodality introduced by the raw sinh^{-1} transformation.

To determine a suitable set of cofactors, each vector $z_j^{\mathcal{C}}$ was first transformed by sinh^{-1} (Equation 1) and its intensity histogram was examined.

$$\sinh^{-1} z_j^{\mathcal{C}} = \ln\left(z_j^{\mathcal{C}} + \sqrt{1 + z_j^{\mathcal{C}} 2}\right) \tag{1}$$

Each α_j was defined as the hyperbolic sine, sinh, of half the magnitude of the distance between the positive P^+ and negative P^- peaks (Equation 2) nearest zero in the intensity histogram of each $\sinh^{-1} z_j^{\mathcal{C}}$,

$$\begin{aligned} \alpha_j^{\mathcal{T}} &= \frac{P^+ - P^-}{2} + 1 \\ \alpha_j &= \sinh\ \alpha_j^{\mathcal{T}} = \frac{e^{\alpha_j^{\mathcal{T}}} - e^{-\alpha_j^{\mathcal{T}}}}{2} \end{aligned} \tag{2}$$

and because P^+ and P^- were defined in the transformed space, sinh was required to convert values back to the raw data space. The cofactors were then applied as follows (Equation 3),

$$z^{\mathcal{T}} = \ln\left(\frac{z_{ij}^{\mathcal{C}}}{\alpha_j} + \sqrt{1 + \left(\frac{z_{ij}^{\mathcal{C}}}{\alpha_j}\right)^2}\right) \tag{3}$$

Note that scatter dimensions are typically not $\sinh^{-1}$ transformed, but for convenience we refer to the full data (scatter included) after $\sinh^{-1}$ transformation simply as $z^{\mathcal{T}}$.

Removal of saturated events

To identify saturated events, all raw data vectors Zj were transformed by (Equation 3) with $\alpha_j = 100$ to yield $z_j^{\mathcal{T}}$. Then each $z_j^{\mathcal{T}}$ was allocated to 1024 uniformly-spaced bins, denoted *s_bin*. Each minimum bin was defined by (Equation 4),

$$s\_bin_{1j} = \sinh^{-1}\left(\frac{-2\left[\frac{\log_2 R_j}{2} + 2\right]}{100}\right) \tag{4}$$

and each maximum bin was defined by (Equation 5),

$$s\_bin_{1024j} = \sinh^{-1}\left(\frac{R_j}{80}\right) \tag{5}$$

where R_j was the channel-specific keyword-value range parameter ($PnR) from the TEXT section of the FCS file. To determine the saturated event threshold h_j, we first examined a window w_j of the top-most 61 bins (Equation 6),

$$w_j = s\_bin_{[964,965,\ldots,1024]j} \tag{6}$$

Then the median and robust standard deviation of the differences between consecutive bins were used to identify bins that contained extreme differences (Equation 7),

$$\begin{aligned} w_j^D &= diff(w_j) \\ w_j^M &= median(w_j^D) \\ w_j^\sigma &= 1.4826 \times median(|\, w_j^D - w_j^M \,|) \\ w_j^X &= w_j^D > (w_j^M + 2w_j^\sigma) \end{aligned} \tag{7}$$

where w^D was the difference between consecutive bins $w_b - w_{b\text{-}1}$ for $b \in \{2, 3, \ldots, 61\}$, w^M was the median difference, w^σ was the robust standard deviation of differences, and w^X was a vector of 1's and 0's that indicated the presence or absence (respectively) of extreme differences. If no extreme differences were found, the examination window was shifted by -1 bin, $w = s\_bin_{[963,964,\ldots,1023]}$, and re-examined. This process was performed iteratively until at least 1 extreme difference was found. Then the lowest s_bin_{Xj} that contained an extreme difference was identified by (Equation 8),

$$X_j = min(argmax(w_j^D \circ w_j^X)) \tag{8}$$

and its corresponding histogram value $v_j^{\mathcal{T}}$ was inverse-transformed back to a raw intensity by (Equation 9),

$$\begin{aligned} v_j^{\mathcal{T}} &= histogram\_value(s\_bin_{Xj}) \\ v_j &= 100 \times \sinh \quad v_j^{\mathcal{T}} \end{aligned} \tag{9}$$

The saturated event threshold h_j was set to the raw intensity v_j (or 80% of the maximum data range, whichever was higher) as follows (Equation 10),

$$h_j = max(v_j,\ 0.8 \times R_j) \tag{10}$$

Finally, all events with raw intensities above the saturated event threshold were removed.

Removal of time defects

To identify time defect events, corrected fluorescence data were sorted by time. Then each z_j^C was allocated to B *n*on-uniformly-spaced bins, denoted *t_bin*, and each contained the same *bin_size* number of events as follows (Equation 11),

$$\begin{aligned} bin\_size &= \begin{cases} 1000, & N < 100,000 \\ 10,000, & N > 1,000,000 \\ \frac{N}{100}, & otherwise \end{cases} \\ B &= \left[\frac{N}{bin\_size}\right] \end{aligned} \tag{11}$$

Then the median event value m was determined for each bin. The vector of bin median event values within each dimension j were Z-score standardized by (Equation 12),

$$\begin{aligned} \tilde{m}_j &= \frac{1}{B}\sum_{b=1}^{B} m_{bj} \\ m_j^\sigma &= \sqrt{\frac{1}{B-1}\sum_{b=1}^{B} |\, m_{bj} - \tilde{m}_j \,|} \\ m_j^Z &= \frac{m_j - \tilde{m}_j}{m_j^\sigma} \end{aligned} \tag{12}$$

Then any bins containing time defects were defined by (Equation 13),

$$\mathcal{D}_j = \left|m_j^Z\right| > 3 \tag{13}$$

and all events within each $t\_bin_{\mathcal{D}j}$ were removed.

Censored saturated events and time defects

The censoring process (described above) identified and removed:

1. raw fluorescent events that saturated above the limits of detection (saturated events).

2. corrected fluorescent events that contributed to inconsistent signals over time (time defects).

Following the removal of saturated events and time defects, new FCS files were generated from the remaining data. Supplementary Figure 1 shows the number of cells per sample before and after censoring for all samples.

New compensation matrices

The quality of compensation matrices was assessed in FlowJo, and any sub-optimal compensation values were manually corrected. The optimized compensation matrices were inserted into the FCS files.

Modified channel names

All marker-fluor combinations were consistent across the entire dataset. However, some FCS files contained channel names that did not match other files. Any mismatched channel names were corrected, and new FCS files were generated.

Batch registration

To remove variation due to experimental batches, while maintaining as much biological variation as possible, swiftReg (7) was used to register batches. This approach first registered each batch separately to the same reference batch, then applied the resulting batch-specific shifts to all individual samples in that batch.

To do this, first a SWIFT cluster template was produced from a concatenate of antigen-stimulated samples from a single reference batch, and similar concatenates from each of the other batches were registered by NDCR to the reference template. The resulting batch registration template contained batch-specific maps of cluster movement vectors that specify the value-adjustments necessary to bring that batch's clusters into alignment with reference clusters. All individual samples in each batch were then registered using these batch-specific cluster movement vectors. This process generated new batch-registered FCS files.

Debris removal

To enhance detection of rare, biologically-significant populations and reduce computational burden, all batch-registered samples were randomly sub-sampled and combined into a single concatenated FCS file that was then clustered by SWIFT. The resulting SWIFT cluster template was used to identify debris clusters in FSC-A and SSC-A, as well as non-CD4 T cells. New FCS files were generated from non-debris CD4+ events.

Expanded select channel data

Detection of positive markers was selectively enhanced by smoothly increasing intensity values about a user-specified inflection point. The smooth increase was achieved by multiplying intensity values within a channel by a sigmoid function (Supplementary Figure 2) as follows (Equation 14),

$$\begin{aligned} r &= [-3.00, -3.01, -3.02, \ldots, 3.00] \\ x &= (r+1) \times L \\ y &= normcdf\left(\tfrac{6}{P} r\right) \times (10^{w} - 1) + 1 \\ s_j &= interp1(x, \quad y, \quad z_j^{C}, \quad option) \\ z_j^{E} &= s_j \circ z_j^{C} \end{aligned} \tag{14}$$

where r was a 1×601 vector of values between -3.00 and 3.00 with intervals of size 0.01, P was the degree of overlap in the expanded region (default $P = 0.5$), L was the user-specified inflection point where expansion occurred in that channel, W was the width of the expanded region in decades (default $\mathcal{W} = 1$), s_j was a vector of scaling values that were multiplied with z_j^{C} element-wise to produce expanded data z_j^{E}, *normcdf* is a MATLAB function that returned a cumulative standard normal distribution, and *interp*1 is a MATLAB function that returned interpolated values of the function $y = f(x)$ at specific query points z_j^{C} by spline interpolation (*option* = 'spline').

Aggregation of data

Because the Env-1-ZM96 and Env-2-ZM96 peptide pools constituted the non-overlapping peptides covering the Env-ZM96 Env sequence, the total Env-ZM96 response for each sample was calculated by combining Env-1-ZM96 + Env-2-ZM96 and subtracting negctrl1. Note that Negctrl1 was subtracted here once to account for the additional background contribution of combining raw Env-1-ZM96 + Env-2-ZM96 counts. The 92TH023-ENV samples were stimulated with peptides covering the whole 92TH023-ENV sequence (2).

The cell counts for AnyEnvNeg1 were then defined as the maximum of the cell counts for 92TH023-ENV or Env-ZM96 + Env-2-ZM96, minus the background from negctrl1. Because the 92TH023-ENV and ZM96 ENV sequences have some homology, it is very likely that some peptides, presented by the MHC alleles of some participants, will be cross-reactive between the two ENV peptide sets. However, the extent of cross-reaction cannot be estimated from this dataset, and so we used a conservative definition of the "Any-Env" response as the larger of the response against either ENV sequence. This uses the conservative assumption that all T cells cross-reacted, and therefore the total response is revealed by the higher of the two anti-Env responses.

The number of CD4+ T cells producing IL-2+ and/or IFNγ+ was expressed as a percentage of the total live cells in the corresponding sample. Percentages below 0.005 were thresholded to a minimum of 0.005.

Identification of responders

To identify responders, the variance of cell counts was first stabilized across clusters. Cluster-specific scaling factors were defined as half the median of cell counts across all negctrl samples for each cluster, with low scaling factors thresholded to a minimum of 10. All counts were then transformed by inverse hyperbolic sine (asinh) after division by the cluster-specific scaling factor. Transformed stimulated counts (TSC) were obtained by subtracting each cluster's transformed background count from its pairwise transformed stimulated count (for values above the threshold, this is analogous to a log ratio).

For each sample group defined by Treatment, Stimulation, Visit, and Cluster, the standardized pairwise background variances (SPBV) between Negctrl1 and Negctrl2 were defined as the square root of the sample-mean of the squares of their pairwise differences.

Then for each sample, the p-values (with a Null hypothesis of "no difference") were determined by applying the normal distribution survival function (https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.norm.html) to the ratio of TSC over SPBV. This was performed separately for the stimulated sample with each background (Neg1 or Neg2). The final reported probability of a sample being a responder was then 1 minus the mean of its two p-values. All samples with a final probability of ≥98% were considered to be responders.

Evaluation of antibody responses associated with T cell responses

Antibody levels measured by binding antibody multiplex assay (BAMA) for HVTN 105 were obtained from HVTN. To compare the CD4 responses to related antibody levels, the Spearman correlation coefficients were calculated for related antigens using log transformed antibody abundances. Specifically, CD4 responses upon ZM96 gp140-Env1 and ZM96 gp140-Env2 stimulations were compared with antibody responses to 96ZM651.D11gp120.avi, gp41, gp70–96ZM651.02 V1v2 antigens. CD4 responses to 92TH023-Env were correlated with antibody response to A244 gp120 gDneg/293F/mon, AE.A244 V1V2 Tags/293F and gp41 antigens. Finally, CD4 responses to ZM96 gag were compared with antibody responses to p24.

Results

Sample pre-processing

The HVTN 105 dataset comprised 3,200 .FCS files representing 24 batches, with accompanying compensation matrices for each batch. In general, the Visit 5 (V5), V7 and V9 PBMC samples for one participant were all analyzed in the same batch, whereas the V11 PBMC samples were analyzed in separate sets of batches. As described previously (2), if PBMC samples did not meet quality control criteria, those samples were re-analyzed in a subsequent batch, resulting in duplicate analyses. After curation according to these rules, the complete dataset potentially comprised four vaccine groups each containing 26 participants, eight *in vitro* antigen stimulations, and four time points, for a total of 3,328 flow cytometry samples. The study design did not include a placebo group receiving no HIV antigens, and the T cell data did not include a baseline sample, i.e. before vaccination. Therefore the important negative controls are the pairs of "negctrl" samples that did not receive *in vitro* stimulation with any antigens. Our re-analysis focused on six of the eight antigen stimulations: two negative control samples (negctrl2 and negctrl2); E92TH023_ENV; Env_1_ZM96; Env_2_ZM96; and Gag_ZM96. This resulted in a total of 2,496 potential samples. Due to some dropouts and missing negctrl replicates, the final number of flow cytometry samples in our re-analysis was 2,393. The samples, batches, repeated samples and final analyzed samples are shown in detail in Supplementary Figure S3.

A consensus .FCS file was produced by concatenating sub-samples of all HIV Ag-stimulated samples from all batches at the V9 time point. V9 was chosen because Visit 9 was the pre-determined immunogenicity time point for the HVTN105 trial, and for most groups and antigen stimulations *in vitro*, this was also the strongest response (see below). V9 samples were therefore enriched for the rare, activated T cells, facilitating capture of these cell populations in the cluster template. Because this concatenate included samples representing all HIV antigen stimulations, this is an objective way to include potential cell phenotypes induced by any of the HIV antigens in any treatment group.

This concatenate was clustered using SWIFT to establish a high-resolution cluster template of all cell sub-populations. All samples were assigned to the resulting cluster template, establishing the number of cells in each cluster, in each sample. All cluster membership information was then condensed to two dimensions using UMAP (Uniform Manifold Approximation and Projection) (12). The results in Figure 2A show batches encoded by colors, stimulations by symbols, and visit number by symbol size. The strongest contribution to diversity was clearly the batch - most members of each batch are clustered together, and the batches are substantially resolved. This is particularly true for the V11 batches on the right, that were analyzed in a different set of batches from V5, V7 and V9. The presence of batch effects is not surprising in samples analyzed over a period of months - we have seen batch effects in all such datasets that we have examined. The HVTN105 batch effects were relatively minor, and so registration could be used to reduce batch effects and improve the comparison of the vaccine groups.

Batch registration

We have previously developed swiftReg (7), an automated registration tool that builds on the SWIFT clustering algorithm to perform high-resolution alignment of samples at the single-cluster level. The HVTN 105 batches were registered by producing a SWIFT cluster template from Batch 2204, producing consensus samples from each batch, and then registering each batch consensus sample to the Batch 2204 consensus cluster template. This generated, for each batch, a map of registration shifts that were then applied to each individual sample in the respective batch. This procedure registers the overall batch trends, without altering the differences between individual samples within each batch that might carry biological information.

A new SWIFT cluster template was generated from a consensus of all registered V9 HIV antigen-stimulated samples. After assignment of all registered samples to the resulting cluster template, the cell numbers per cluster were reduced to two dimensions by UMAP, and Figure 2B shows that the registered batches were intermingled. 'Micro-aggregates' of samples from the same batch were still visible - focusing on just 15 participants for clarity, each micro-aggregate comprised samples from a single participant (including different stimulations and time points). These tended to group in close proximity on the UMAP projection (Figure 2C), even though the Visit 11 samples were

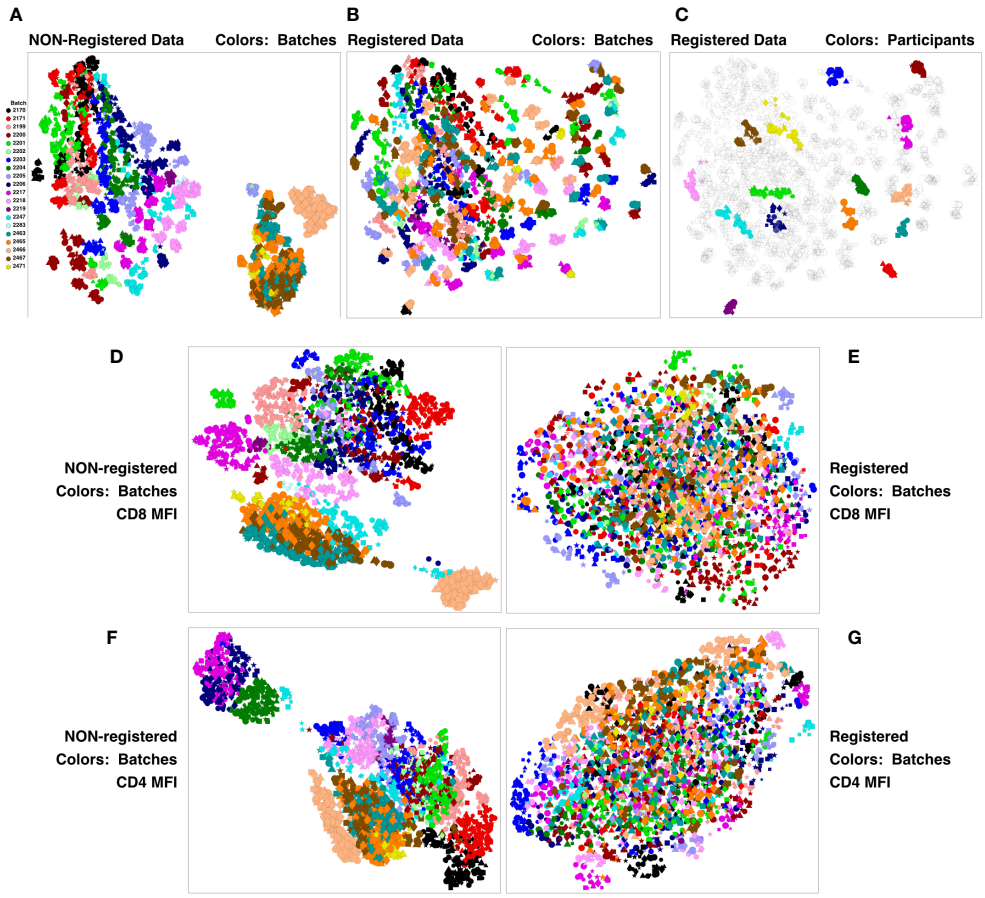


FIGURE 2
Registration minimizes batch variation and emphasizes individual stability. PBMC from the HVTN 105 vaccine trial from V5, V7, V9 and V11 (42, 98, 182 and 364 days) were analyzed by antigen stimulation, intracellular cytokine staining, and flow cytometry. A SWIFT cluster template was produced from a concatenate of HIV antigen-stimulated V9 (182 days) samples, then all individual samples were assigned to this template. All batches were then registered using swiftReg, and the registered samples were similarly analyzed by SWIFT clustering and individual sample assignment. For each of the original and registered datasets, all cluster information (sizes or MFI of individual parameters) was then condensed to two dimensions by UMAP. Each symbol represents one sample (one participant, one time point, one stimulation). Symbols: Circles, negctrl; triangles,E92TH023_ENV; stars, Env_1_ZM96; squares, Env_2_ZM96; and diamonds, Gag_ZM96. Symbol size, in increasing order, V5, V7, V9, V11. **(A-C)** UMAP plots represent all the numbers of cells/cluster information condensed down to two dimensions. **(A)** Unregistered, cluster sizes, batches colored. **(B)** Registered, cluster sizes, batches colored. **(C)** Registered, cluster sizes, 15 participants colored. **(D-G)** UMAP plots represent the UMAP condensation of the mean fluorescence intensities for each cluster of a specified marker. **(D, E)** CD4. **(F, G)** CD8. **(D, F)** Non-registered. **(E, G)**: Registered.

analyzed in different batches from the Visit 5, 7, 9 samples. Overall, the samples included time points spanning 18 months. Thus, most individuals are sufficiently diverse for SWIFT analysis of flow cytometry data to identify a unique 'fingerprint' of cell populations in different participants. We have observed this pattern in other studies (unpublished). The proximity of the registered V11 data points to the V5, V7 and V9 points from the same participant reinforces the interpretation that the HVTN 105 batch effects have been substantially reduced by the registration process. Examination of individual parameters by the same approach identified parameters, e.g., CD4 and CD8, that contributed to these batch differences (Figures 2D–G). Interestingly, the groupings of similar batches were variable between different parameters (Figure 2; Supplementary Figure 4).

Further pre-processing

We then used the model-based SWIFT multidimensional clustering algorithm (5, 6) to generate an unbiased cluster map from a sample constructed by concatenating a random subset of events from samples across all batches. The SWIFT algorithm is particularly useful for detecting rare populations (13), possibly because these were the type of samples used during SWIFT development (5, 6). Preliminary analysis of the cluster map indicated that the antigen-specific responses in many samples were small, consistent with the previous analysis (2). To maximize the sensitivity of detecting all cytokine producing sub-populations, we produced a new SWIFT cluster template from a concatenate of random subset of events from all antigen-stimulated

samples, using only the scatter, live/dead, and CD4 parameters. All samples were assigned to the resulting template. As described previously (2), the non-replicating vaccines induced almost no CD8 T cell responses, and so further analysis focused on CD4 T cells. Clusters containing CD4 T cells were selected, and all the events in this set of clusters were saved, for each flow cytometry sample, as reduced-size .FCS files for further analysis. This "cluster gating" (6) allowed subsequent analysis to focus more clearly on the cells of interest, because clustering could then be performed on a full concatenate of the entire dataset. The resulting .FCS files are more amenable to analysis by SWIFT, other automated algorithms, and manual analysis.

High-resolution clustering

A large concatenate was then produced from all cluster-gated events in all samples stimulated with HIV antigens (Env, Gag), from all groups at Visit 9, which was the time point that showed the highest responses overall. A second concatenate was produced from the corresponding negative control samples. SWIFT cluster templates were created from each of these two large concatenates, using all parameters for high-resolution clustering. The two resulting cluster templates were combined, and all individual samples from all groups, all visits, all stimulations were assigned to the resulting combined template (total clusters 2,246). This cluster competition approach (7) sharpens the differences between the two groups represented by the two templates, in this case stimulated and unstimulated cell populations. Note that each concatenate included samples from all vaccine groups, so the competition process should not affect the resolution or statistical analysis of any study group differences.

Cluster gating (6) was then used to narrow down the cell populations of interest. During cluster gating, all cells are assigned their cluster medians in all dimensions, so that the two-dimensional gating shown in Figure 3A takes advantage of all the information in all dimensions. Activated CD4 T cell clusters were identified as live, singlet, CD3+ CD4+ CD154+ TNF+ clusters (Figure 3A). Additional marker intensities for all parameters are shown in Supplementary Figure S5. These activated CD4 T cell clusters were then examined by testing the significance of differences between antigen-stimulated and negative control clusters in all participants at visit 9. A Wilcoxon test was followed by the Benjamini-Hochberg correction for multiple measurements, because of the number of clusters examined. Figure 3B shows the ratios and magnitudes of differences between antigen-stimulated and negative control cultures in a volcano plot. All clusters that were significantly increased in the antigen-stimulated samples (green shaded area) were chosen for further analysis. To facilitate comparisons with previous analysis (2), the SWIFT clusters were aggregated into four groups: IL-2+IFN-γ+, +/-, -/+ and -/- cells (15, 5, 3 and 4 clusters, respectively). The heatmaps (Figure 3C) show the marker characteristics of each cluster.

Identification of vaccine responders

The samples showing significant responses to each antigen, at each time point, were then evaluated as described in Methods, using the aggregated cluster data for all clusters producing IL-2 and/or IFN-γ. Figure 4 shows the results for each time point, each vaccine treatment, and five antigen stimulations, or combinations of stimulations: AnyEnv (Env92 or Env1/2), Env92 (Env92TH023 only), Env1/2 (Env1 plus Env2), GAG-ZM96, and the negctrl2. Background values (negctrl1) were subtracted from all antigen-stimulated values (similar conclusions were obtained if the negative controls were reversed). All samples with >98% probability of being genuine responders are shown in red. As expected, very few negctrl samples were evaluated as responders (at a confidence level of 98%, a small number of false positives are expected). As Gag antigen was only included in the DNA vaccine, Treatment 1 uniquely lacks immunization with Gag for the first blood sample evaluated, at Visit 5. Consistent with this, only Treatment group 1 lacks a response to Gag at Visit 5. At a very high confidence level of 99.9%, there were still high rates of responders (up to 88%) but no responders in any negative controls (Supplementary Figure S6). Supplementary Figure S7 shows an alternative layout of the responder data to emphasize the time course within each group.

Several combinations of vaccine treatments and times induced responses in the great majority of participants, particularly in Treatment group 3 at Visits 7 and 9. The numbers of responders were generally higher than evaluated previously (2), possibly because the extensive pre-processing and the competitive cluster templates used in our analysis provided sharper distinction between antigen-stimulated versus background cells producing cytokines. The magnitude of the net anti-HIV T cell responses was well-correlated between the original analysis and the re-analysis (Supplementary Figure S8). There is a general trend towards higher magnitudes detected by SWIFT (compared to the 1:1 reference line), possibly due to the effectiveness of high-dimensional definition of populations, as well as the sharper signal:noise discrimination by focusing on the clusters that were significantly increased by antigen stimulation.

Qualitatively different responses are associated with different vaccine modalities

The quality of the cytokine response to protein or DNA-derived immunogens was assessed between the different vaccine treatments by comparing the ratio of T cells producing IFN-γ vs. T cells producing IL-2 but not IFN-γ. The anti-Gag response is easiest to interpret, as this is induced only by the DNA vaccine. Figure 5A shows that this response is biased towards IFN-γ production, consistent with a previous report (14). The response to the ZM96 clade C peptides, primed by DNA, also showed a tendency towards an IFN-γ-biased response. In contrast, the response to clade E

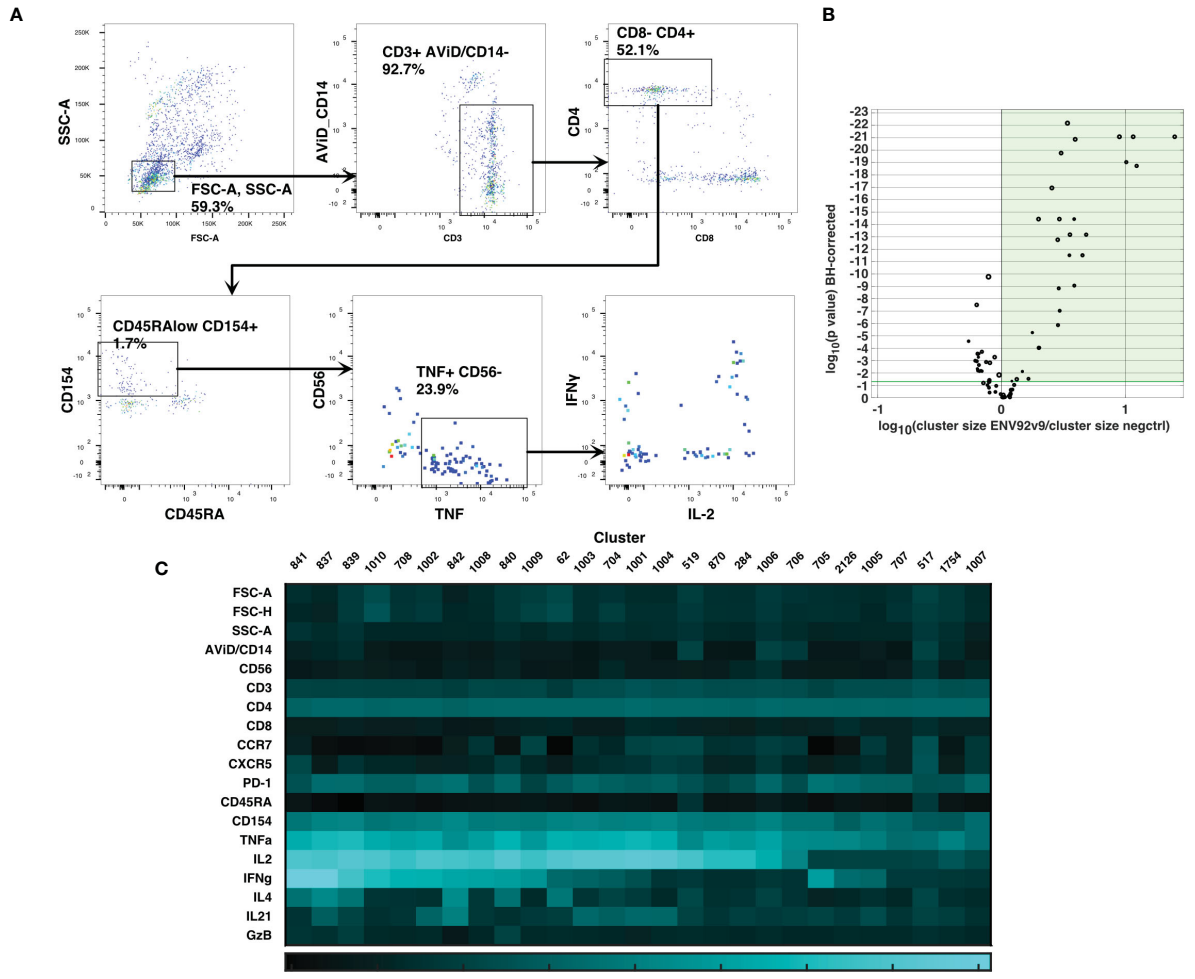


FIGURE 3
Cluster gating of cytokine-producing antigen-specific T cells. SWIFT cluster templates were produced from concatenates of antigen-stimulated samples, and control samples, and the two templates combined for competitive cluster assignment. All individual samples were assigned to the combined template. **(A)** All cells were plotted at their cluster medians in each parameter for cluster gating on bivariate plots, to identify activated CD4 T cells expressing CD154 and TNF. **(B)** For each cluster, the number of cells in a concatenate of ENV92-stimulated visit 9 samples was compared by Wilcoxon to the matched negative control sample. Each symbol indicates one cluster, and the size of the symbol is proportional to the mean number of cells per cluster. P values were adjusted according to the Benjamini-Hochberg method for multiple measures. The green shaded area indicates the clusters that were significantly increased in size by antigen stimulation. **(C)** The heatmap shows the median fluorescence intensity in each parameter (Z-scores) of the 27 significantly induced clusters from B (shaded area).

92TH023 protein immunization was biased more towards IL-2-only responses, consistent with our previous demonstration (15) that viral infections tend to induce more Th1/IFNγ responses, whereas protein vaccines tend to produce responses biased towards IL-2-producing central memory (16) cells. Figure 5B summarizes these results, including the results for the minority IL-2- IFN-γ- and IL-2-IFN-γ+ responses.

Minority cytokine responses

The flow cytometry panel included several cytokines, including IL-21 (produced by Tfh and some other cells) and IL-4 (produced by Th2 cells). Manual examination of the concatenated results suggested that antigen stimulation appeared to induce a small IL-21 response in a relatively low number of TNFa+ IL-2+ CD4 T cells. However, the IL-21 staining was weak, and did not result in a clearly separated sub-population of positive cells. As the SWIFT clustering algorithm uses a criterion of multidimensional unimodality to define individual sub-populations (6), the putative IL-21+ cells were initially difficult to identify by clustering. We therefore used a 'stretching' modification that slightly broadened the cell distribution across the expected junction between IL-21- and IL-21+ cells. Clustering the resulting data in SWIFT allowed the reproducible detection of IL-21+ clusters (Figure 6). In contrast, applying the same stretching modification to the IL-4 channel did not result in the detection of IL-4+ clusters, consistent with the manual examination of the IL-4 data (Figure 6). The IL-21+ clusters were activated memory CD4 T cells (CD154+ CD45RAlo CD4+), but interestingly, did not express the CXCR5 chemokine receptor that is characteristic of circulating T follicular helper (Tfh) cells (Figure 6), perhaps due to down-regulation of CXCR5 on the *in*

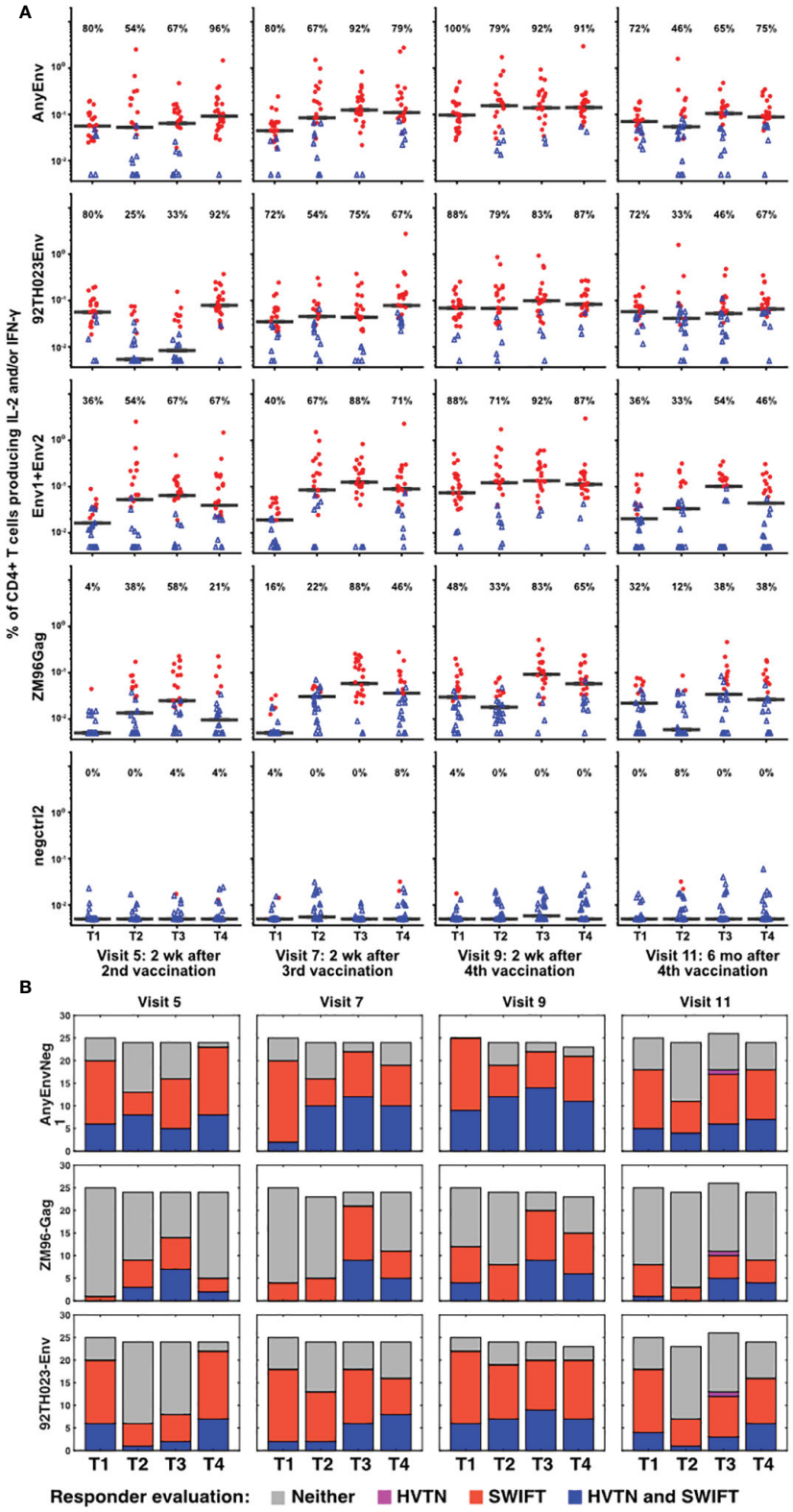


FIGURE 4
Increased numbers of vaccine responders identified by detailed analysis pipeline. **(A)** For all V5, V7, V9 and V11 samples, responders were identified as described in Methods, calculating the responses separately for 92TH023 Env; ZM96 pool 1 + pool 2 Env; ZM96 Gag; the negative control negctrl2; and Any Env (the larger of the responses to either 92TH023 or ZM96 Env1 + Env2). The values from negctrl1 were subtracted from each of these values. Red circles and blue triangles indicate responders and non-responders, respectively, and horizontal black bars indicate medians of all samples in each treatment group. The percentage of positive responses is shown above each graph. Values less than 0.005% were plotted at 0.005%. **(B)** Responder rates from the present study compared to the equivalent responder rates from the original analysis (2).

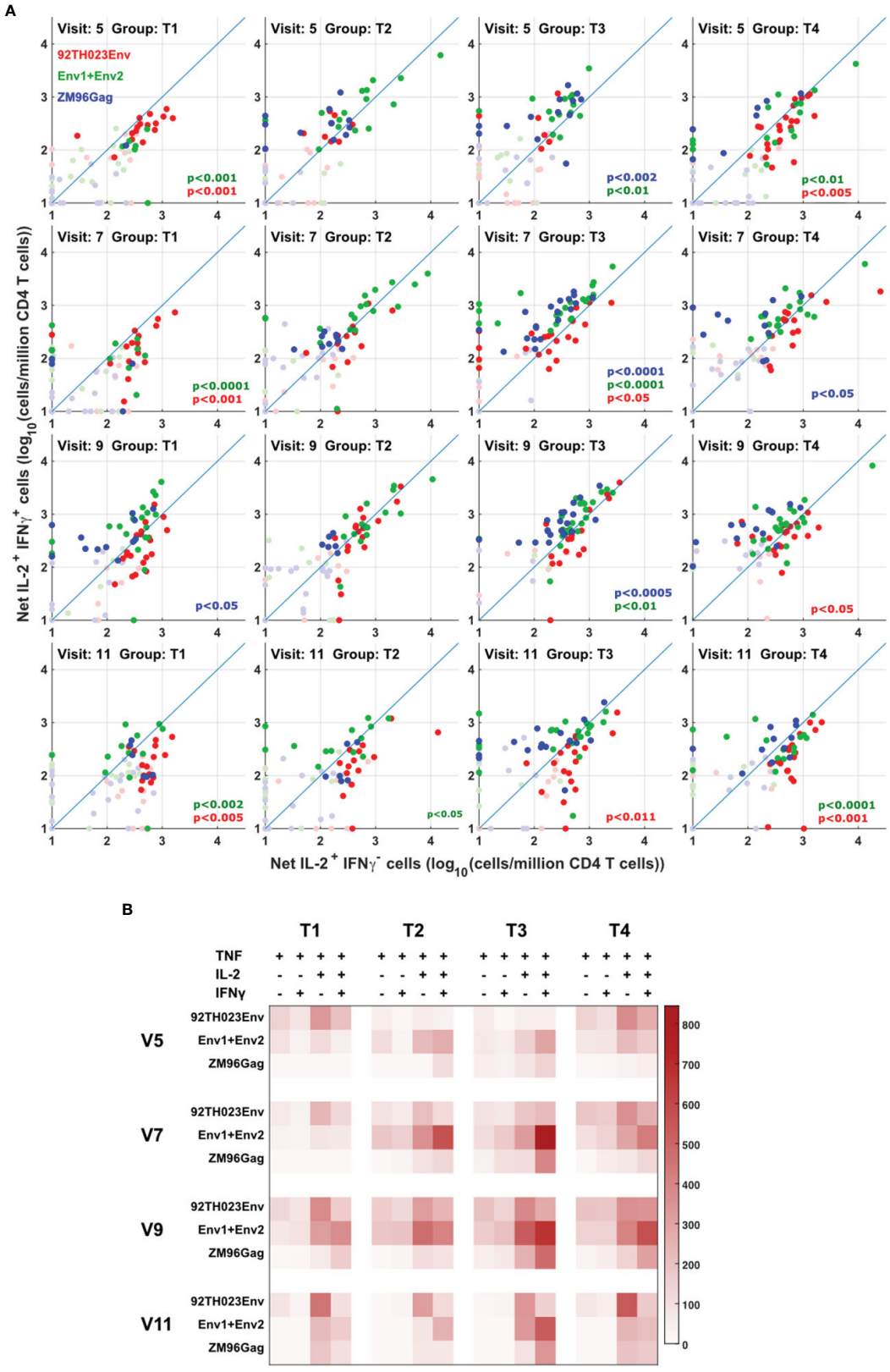


FIGURE 5
Different cytokine response patterns associated with DNA or protein vaccination. **(A)** IL-2+IFNg+ responses were compared with IL-2+IFNg- responses in all participants, all visits and for ZM-96-Gag, ZM96-Env and 92TH023-Env. Dark symbols indicate samples with positive responses (using the values for IL-2 and/or IFNγ from Figure 4) and pale symbols indicate non-responders. P values indicate the significance of the deviation from the 1:1 correlation line, with colors matching the data points. **(B)** The heatmap indicates the average number of antigen-responsive CD4 T cells per million total live CD4 T cells, for each vaccination group. Each response is divided into all combinations of IL-2 and IFNγ expression.

vitro activated cells as we have observed previously (S. De Rosa, unpublished). In contrast to the IL-2+ IFN-γ+ versus IL-2+ IFN-γ- skewing described above, the IL-21+ cells were observed in all treatment groups, and did not show obvious biases towards particular antigens or immunization strategies (Figure 7).

Correlations of SWIFT CD4+ T cell clusters with HIV-specific plasma antibody responses

Binding antibody (Ab) responses were the major correlates of risk (CoR) identified in the RV144 Trial (3). Subsequently, at V9, 2 weeks after the final immunization we assessed the relationship of IL-2/IFN-γ and IL-21 defined clusters with the contemporary HIV-specific plasma Abs (Figure 8). Overall T1 (Figures 8E, J) had low antibody responses at this timepoint compared to the other groups as expected due to the boosting immunizations with DNA alone. T2 overall exhibited the greatest number of significant correlations with the IgG response, primarily associated with responses to AE.A244 (Figure 8B) which most closely matches the protein component of the vaccine regimen. T3 and T4 overall had significant associations relatively balanced between AE.A244 (Figures 8C, D) and 96ZM651 (Figure 8H) Ab responses which most closely matches the DNA component of the vaccine regimen, and is consistent with T3 and T4 receiving 4 doses of DNA. T3 had the greatest number of significant correlations between IL-21+ and Ab responses (Figures 8C, H), consistent with T3 having the overall greatest IL-21+ response. Supplementary Figure S9 shows the magnitude of IgG responses for the T cell responders identified by the original analysis or the new SWIFT analysis.

Both total IgG specific for the V1V2 region of gp120 and IgG3 specific for V1V2 were inverse CoR in RV144 (3). IL-2-IFN-γ+ cells were significantly correlated with IgG AE.A244 V1V2 in T4 (Figure 8D), with IL-2+IFN-γ+ and IL-2-IFN-γ also significantly correlating with IgG AE.A244 V1V2 in T2 (Figure 8B). IL-2+IFN-γ+ also, and IL-21+ also significantly correlated with IgG gp70–96ZM51 V1V2 in T3 (Figure 8H). V1V2 responses of the specific

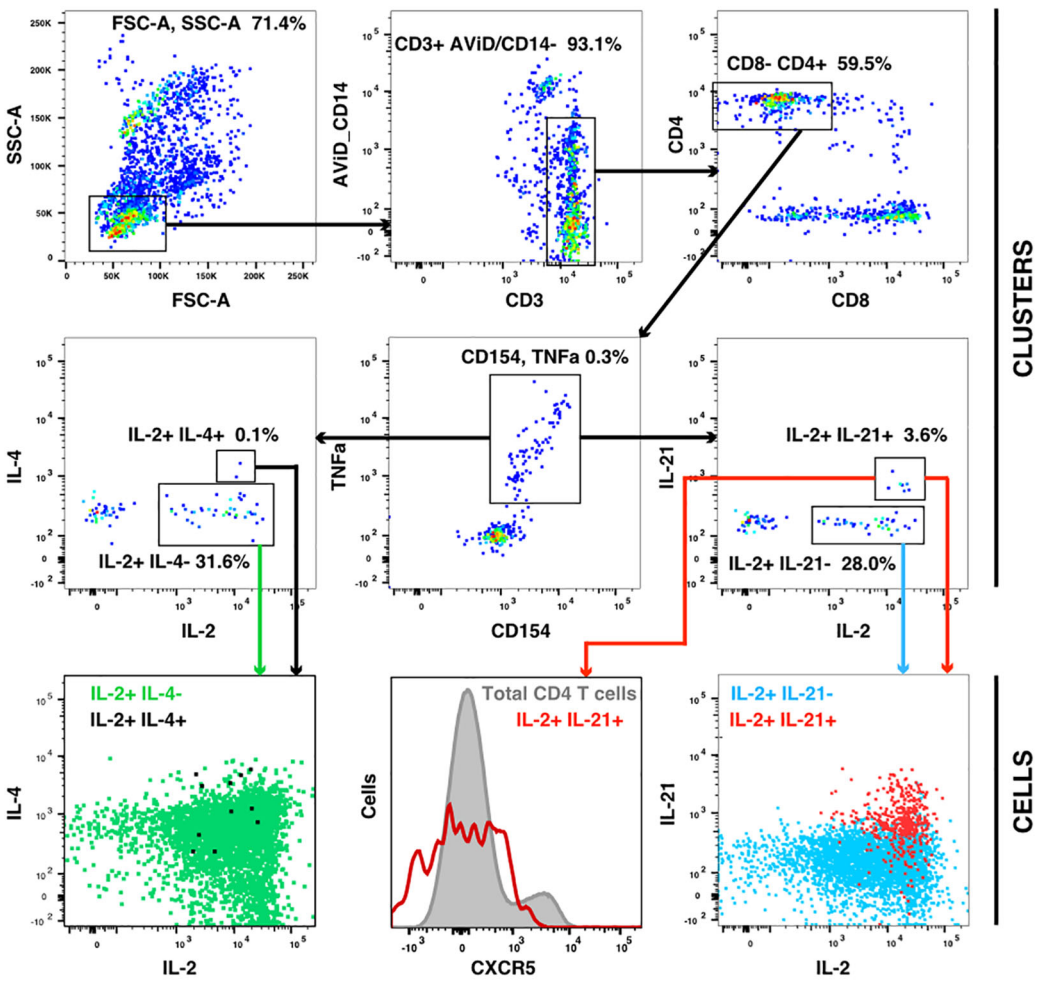


FIGURE 6
IL-21 responses after vaccination. A concatenate (10 million cells) of random samples of all HIV antigen stimulated samples at V9 was assigned to the cluster template used in Figure 3, and cluster gating was used to identify all CD4+ CD154+ TNF+ T cells (center panel). Cluster gating was used to further identify IL-2+ IL-4+ and IL-2+ IL-4- clusters (second row, left) and IL-2+ IL-21+ and IL-2+ IL-21- clusters (second row, right). In the top and middle panels, each dot represents one cluster. The bottom row shows plots of individual cells in the four sets of clusters.

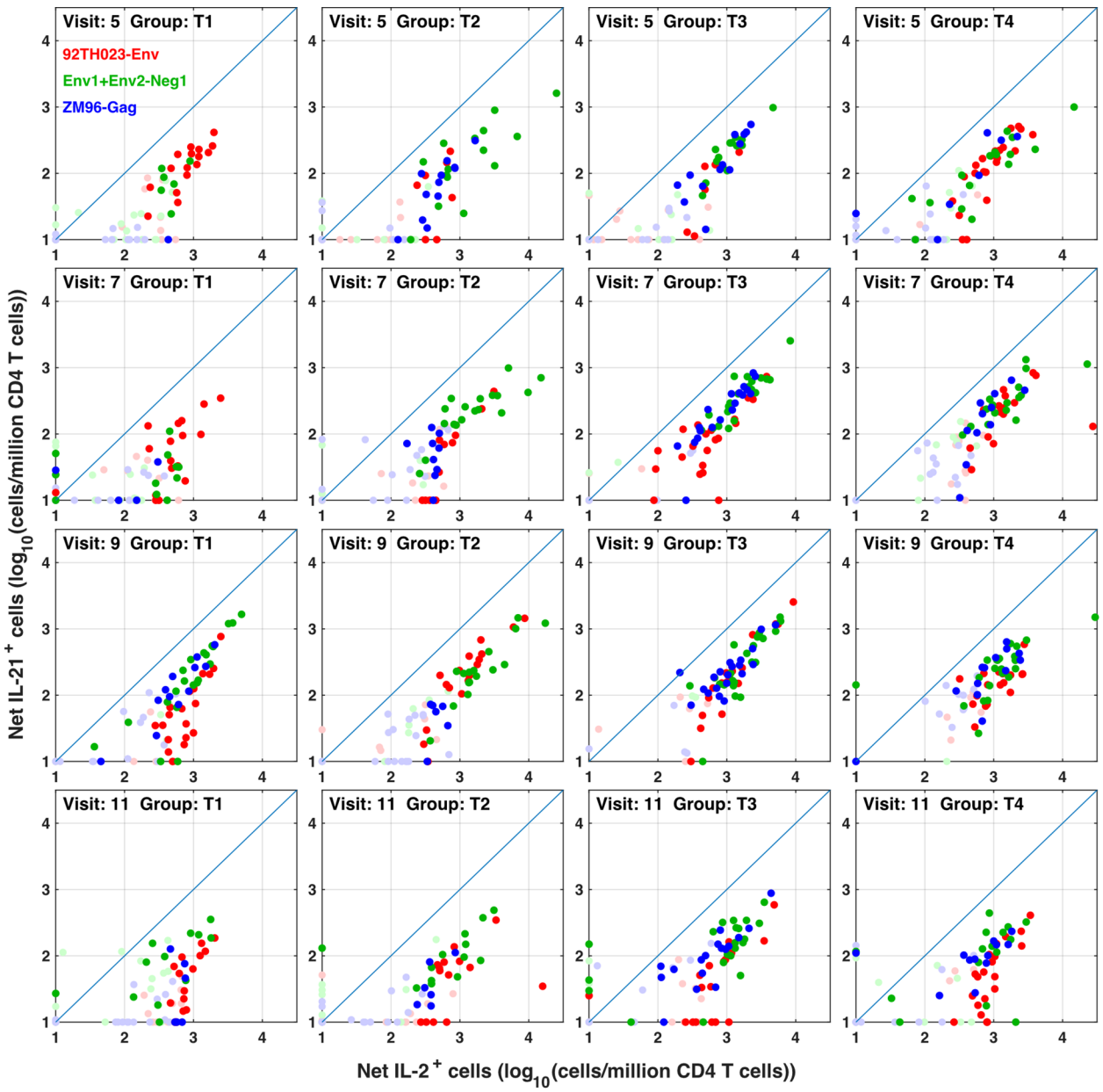


FIGURE 7
IL-21 responses to different immunogens. CD4 T cells producing IL-2 (with or without IFN γ) were compared with IL-21+ responses in all participants, all visits and for ZM-96-Gag, ZM96-Env and 92TH023-Env antigen stimulations. Dark symbols indicate samples with positive responses (using the values for IL-2 and/or IFNγ from Figure 4) and pale symbols indicate non-responders.

IgG subclass, IgG3 which is known to be a potent mediator of Fc-effector functions such as antibody dependent cellular cytotoxicity were inverse CoR in RV144, and only IL-2+IFN-γ+ in T1 was significantly correlated with IgG3 gp70–96ZM51 V1V2 (Figure 8F). IgA specific for gp120 overall as well as the V1V2 region was a CoR in RV144, suggested to compete with the binding of protective IgG3 (3, 17). Only T2 and T3 had measurable IgA responses to AE.A244 gp120 or AE.A244 V1V2 (Figure 8E), with IL-2-IFN-γ- in T2 significantly correlating with IgA AE.A244 gp120 and IgA AE.A244V1V2 (Figure 8B). IL-21+ in T3 was significantly correlated with both IgA AE.A244 gp120 and IgA AE.A244V1V2 (Figure 8C). Overall these results indicate subtleties in the association of CD4+ T cell responses and plasma Ab responses, that are impacted by vaccine regimen and may provide insight into efficacy outcomes.

Discussion

A substantial preventative vaccine trial such as HVTN 105 generates a large dataset of immunological results, which provides a valuable resource for continued analysis using different approaches. This trial was chosen for analysis, although a phase I trial, because it reiterated the general prime-boost approach of the only preventive

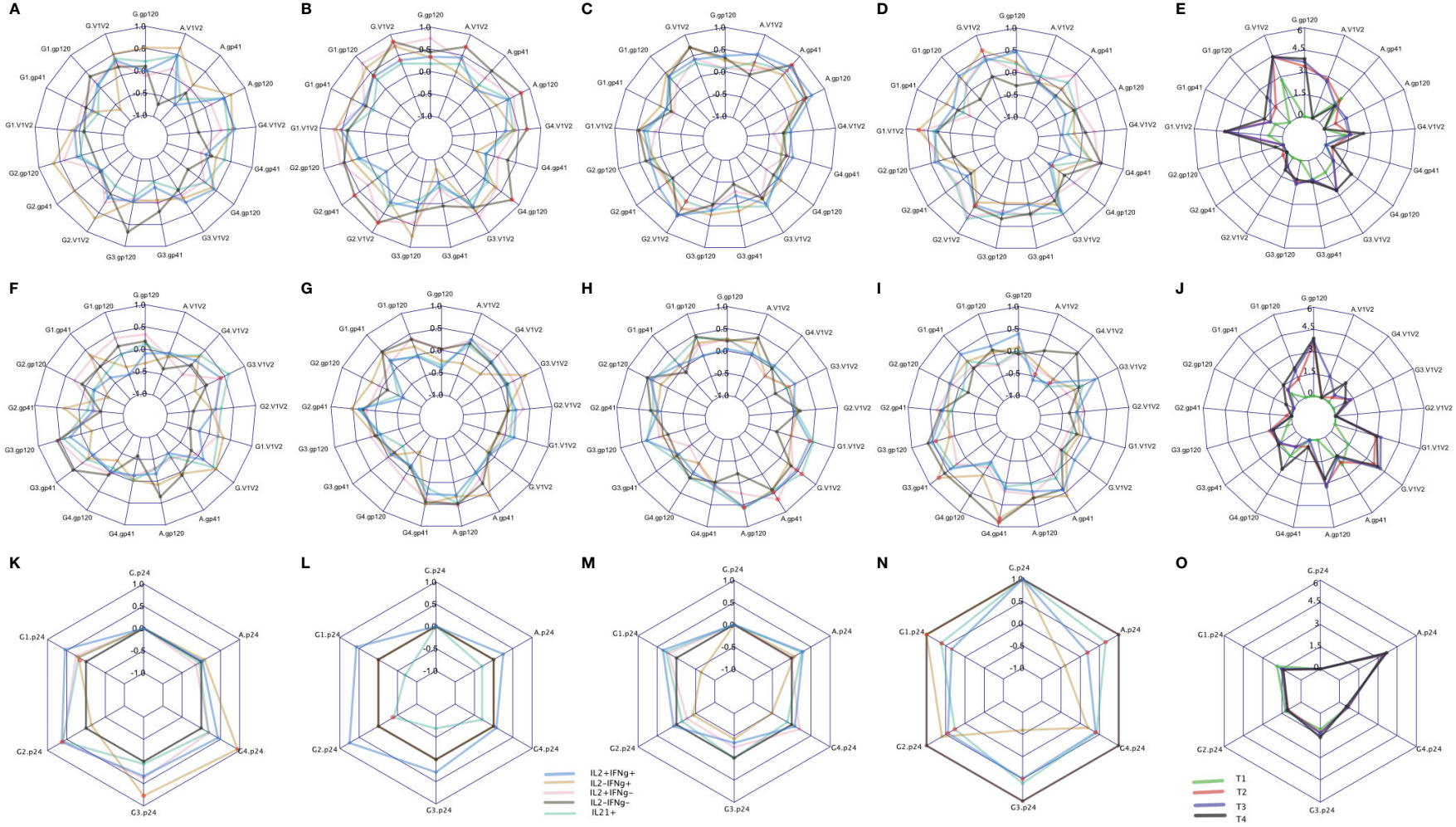


FIGURE 8
CD4 responses associated with antibody levels measured by binding antibody (BAMA) assay. The spider plots **(A–D, F–I, K–N)** show correlation coefficients between CD4 clusters (blue: IL2+IFNγ+, yellow: IL2-IFNγ+, pink: IL2+IFNγ-, grey: IL2-IFNγ- and aquamarine: IL21+) and antibody levels across four treatment groups. The spider plots **(E, J, O)** show log-transformed magnitudes of antibody levels (T1: Green, T2: red, T3: blue, T4: Black). The top, middle and bottom plots show BAMA correlations to CD4 responses to Env92 (gp120-A244 gp120 gDned/293F/mon, V1V2- A244 V1V2 Tags/293F; A,B, C,D), Env1 + 2 (gp12–96ZM 651.D11gp120.avi, V1V2–96ZM651.02 V1V2; F,G,H,I) and Gag (K,L,M,N). Red dots indicate significant correlations with adjusted p-value<0.05.

10.3389/fimmu.2024.1347926

vaccine trial to show any degree of efficacy, RV144 ("The Thai Trial"), but with priming by a more flexible DNA vaccine platform. We have re-analyzed the flow cytometry T cell response data using a detailed clustering approach, and also evaluated the correlations between different T cell responses and the levels of different isotypes and specificities of antibodies. This resulted in the detection of higher numbers of responders; revealed preferential induction of central versus effector T cell responses by different immunogens; and showed that the best correlations between T cell and antibody responses did not necessarily match the strongest responses.

The SWIFT clustering algorithm is highly effective for detecting small cell sub-populations in flow cytometry data (6, 18). This sensitivity may be related to the extensive use of antigen-stimulated PBMC datasets during SWIFT development, resulting in an algorithm that is well-suited to the detection of small cytokine-producing T cell responses of human PBMC, e.g., in the HVTN 105 dataset.

An additional advantage of the SWIFT analysis pipeline is the registration tool swiftReg (7), which can register batches of data to minimize batch effects while preserving biological variation and group differences. The HVTN 105 trial was large, and the flow cytometry data analysis was performed in many batches. Although stringent protocols ensured that the batch variation was smaller than in many other studies, it is almost impossible to completely prevent batch effects in experiments conducted over several months, and so the swiftReg tool was helpful in minimizing batch variation to allow the analysis to focus more sharply on the vaccine group differences. As swiftReg produces new .FCS files containing registered data, registration can also be a useful step in data processing pipelines using alternate clustering approaches.

Compared to the initial analysis (2) the high-resolution SWIFT analysis detected substantially higher numbers of responding participants for all antigens. A major contribution to this increase may have been due to our sharpened discrimination of responders from non-responders using competitive template assignment (19). In this approach, SWIFT cluster templates were produced from two concatenates, of antigen-stimulated and negative control samples. These two templates were then combined and all samples assigned to the joint template. Some cytokine-secreting clusters preferentially captured background responses, so by focusing only on clusters that were significantly higher in antigen-stimulated samples, we were able to sharpen the identification of antigen-responding cells and improve signal:noise ratios. This probably contributed to the higher number of responders detected, while the overall pattern of the response was similar, e.g., group T3 had higher responder frequencies in both analyses.

Several issues have to be considered for the potential T cell cross-reactions between different antigens used in the HVTN105 study. The predictions for anti-Gag responses are relatively straightforward, because Gag antigens were encoded by the DNA vaccine, but not included in the protein vaccine. Thus Gag responses should be attributable only to Gag-ZM96 priming and boosting. Consistent with this prediction, significant numbers of Gag responders were only observed in groups that had received the DNA vaccine prior to the sample draw. In addition, Gag responses are simpler to interpret because the immunogen and the *in vitro* challenge peptides were fully matched.

In contrast, three different Env sequences were included in the vaccines. The DNA vaccine expressed the clade C ZM96 gp140 protein, whereas the protein AIDSVAX vaccine contained both the clade B gp120 MN and clade E gp120 A244 proteins. Thus, the DNA and protein vaccines should stimulate partially overlapping T cell repertoires specific for Env, and a second immunization with the other vaccine type (protein to DNA, or DNA to protein) should induce a mixture of memory responses to cross-reactive epitopes, and naïve responses to non-cross-reactive epitopes.

In vitro testing of T cell anti-Env responses was performed with three peptide pools: Two vaccine-matched peptide pools covered the N-terminal and C-terminal regions of the ZM96 clade C gp140 protein, and a third pool contained peptides of the clade E 92TH023 protein, i.e. the same clade as the AIDSVAX clade E Env A244 protein, but with only about 90% homology between the protein sequences. However, the two proteins contain long stretches of completely homologous sequences, so there should be substantial but not complete cross-reaction between the immunizing and testing clade E Env epitopes. Responses to the immunizing clade B MN env protein would be expected to have lower cross-reactivity to either the clade C or Clade E test antigens, and so may not have contributed significantly to the overall *in vitro* T cell results. Because the extent of cross-reaction between the clade C- and clade E-specific T cells in this study was unknown, we made the conservative assumption that the "any env" response was taken as the maximum of the ZM96 and 92TH023 responses, i.e., assuming complete cross-reaction, as in the previous analysis (2).

The quality of the T cell response, i.e. the cytokine patterns produced by antigen-specific T cells, was influenced by the type of vaccine. In line with previous studies (14, 15) the AIDSVAX protein vaccine preferentially induced CD4 T cells producing IL-2 but not IFNγ, whereas the DNA vaccine induced more IL-2+ IFNγ+ T cells. The IL-2+ cells may be central memory T cells (Tcm) (16) that have high proliferative potential and can differentiate into effector cells (16, 20), whereas the IL-2+ IFN-γ+ T cells are effector memory cells. While both T cell populations are potentially valuable for future protection, the Tcm may have higher potential over longer times (21).

In addition to the evaluation of the major cytokines TNF, IFN-γ and IL-2, the flow cytometry analysis also measured IL-21-producing cells. Although the staining for IL-21 was not strong, there appeared to be an IL-21+ population that expressed high levels of TNF and IL-2, and variable amounts of IFN-γ. IL-21 is produced commonly, although not exclusively, by CXCR5+ Tfh cells in lymph nodes (22, 23). However, the IL-21+ cells in the HVTN105 study were generally CXCR5-. Although this might suggest that these were not circulating Tfh-like cells (24, 25) it is also possible that CXCR5 expression was lost during *in vitro* stimulation.

Assessment of the SWIFT-defined CD4+ T cell clusters' association with the plasma Ab response to HVTN 105, revealed that although polyfunctional TNF-alpha+ IL-2+ IFN-γ+ effector memory cells dominated the CD4+ T cell response in T3 and T4, a

subdominant IFN-γ producing population, IFN-γ+IL-2- cells in T3 and T4 correlated with IgG AE.A244 V1V2 (an inverse CoR in RV144), suggesting that the magnitude of a specific CD4+ T cell cluster is not the sole determinant of correlation with the Env-specific Ab response. The consequences of associations between CD4+ T cell cytokine producing subsets and protective antibody responses to HIV remain uncertain, however intriguing findings regarding this relationship continue to emerge.

A limitation of this study was that although HVTN 105 used the same protein immunogen as RV144, AIDSVAX B/E, unlike RV144, HVTN 105 was not an efficacy trial. Subsequently, the differences observed in response rates or phenotypes of CD4+ T cells observed between groups in HVTN 105 either in this re-analysis or the primary analysis (2) cannot infer association with vaccine efficacy.

The recent HVTN 702 efficacy trial conducted in South Africa, which was an iteration of RV144 with Clade C immunogens consisting of priming with a canarypox-based env/gag/pro immunogen and boosting with the addition of a Env protein immunogen, unfortunately resulted in similar infection rates in placebo and vaccine recipients (26). *Post-hoc* analysis revealed that among individuals that had high IgG AE.A244V1V2 responses, CD4+ T cell polyfunctional score was associated with lower risk of HIV acquisition. And conversely, among individuals that had low IgG AE.A244V1V2 responses, the CD4+ T cell polyfunctional score was associated with a higher risk of HIV acquisition. These findings highlight the increasing need to better define and monitor the nuanced relationship between the CD4+ T cell response to HIV vaccines and the protection that may be conferred by antibody responses, and we suggest that advanced flow cytometry analysis approaches, such as SWIFT, can enhance resolution of the HIV-specific T cell response.

Software availability

The SWIFT and swiftReg software packages are freely available for download at: http://www.ece.rochester.edu/projects/siplab/Software/SWIFT.html.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Materials, further inquiries can be directed to the corresponding author/s.

Ethics statement

Ethical approval was not required for the studies involving humans because the samples came from a previously published clinical trial (ClinicalTrials.gov NCT02207920). The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and institutional requirements because samples were de-identified prior to the analyses.

Author contributions

TM: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. JR: Writing – original draft, Software, Methodology, Investigation, Formal analysis, Conceptualization. SR: Writing – review & editing, Resources, Investigation. MK: Writing – review & editing, Project administration, Investigation, Funding acquisition. MM: Writing – review & editing, Resources, Investigation, Funding acquisition. NR: Writing – review & editing, Resources, Investigation, Funding acquisition. GP: Resources, Investigation, Writing – review & editing. PG: Writing – review & editing, Resources, Investigation. LC: Writing – review & editing, Resources, Investigation. JK: Writing – review & editing, Writing – original draft, Supervision, Project administration, Funding acquisition, Conceptualization. JT: Writing – review & editing, Writing – original draft, Supervision, Software, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by the National Institute of Allergy and Infectious Diseases through awards UM1 AI068614 [HVTN LOC], UM1 AI068635 [HVTN SDMC], UM1 AI068618 [HVTN LC] to SR, 5UM1AI069452 to JK, 5UM1AI069511 to MK, and the National Institute of Dental and Craniofacial Research (R01DE027245 to JK).

Acknowledgments

We thank the HVTN 105 trial team and participants.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer RP declared a past co-authorship with the author NR to the handling editor.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

Author disclaimer

The content is solely the responsibility of the authors and does not necessarily represent the official views of any of the funding sources.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2024.1347926/full#supplementary-material

References

1. Rerks-Ngarm S, Pitisuttithum P, Nitayaphan S, Kaewkungwal J, Chiu J, Paris R, et al. Vaccination with ALVAC and AIDSVAX to prevent HIV-1 infection in Thailand. *N Engl J Med*. (2009) 361:2209–20. doi: 10.1056/NEJMoa0908492

2. Rouphael NG, Morgan C, Li SS, Jensen R, Sanchez B, Karuna S, et al. DNA priming and gp120 boosting induces HIV-specific antibodies in a randomized clinical trial. *J Clin Invest*. (2019) 129:4769–85. doi: 10.1172/JCI128699

3. Haynes BF, Gilbert PB, McElrath MJ, Zolla-Pazner S, Tomaras GD, Alam SM, et al. Immune-correlates analysis of an HIV-1 vaccine efficacy trial. *N Engl J Med*. (2012) 366:1275–86. doi: 10.1056/NEJMoa1113425

4. Lin L, Finak G, Ushey K, Seshadri C, Hawn TR, Frahm N, et al. COMPASS identifies T-cell subsets correlated with clinical outcomes. *Nat Biotechnol*. (2015) 33:610–6. doi: 10.1038/nbt.3187

5. Naim I, Datta S, Rebhahn J, Cavenaugh JS, Mosmann TR, Sharma G. SWIFT-scalable clustering for automated identification of rare cell populations in large, high-dimensional flow cytometry datasets, Part 1: Algorithm design. *Cytomet A*. (2014) 85:408–21. doi: 10.1002/cyto.a.22446

6. Mosmann TR, Naim I, Rebhahn J, Datta S, Cavenaugh JS, Weaver JM, et al. SWIFT-scalable clustering for automated identification of rare cell populations in large, high-dimensional flow cytometry datasets, Part 2: Biological evaluation. *Cytomet A*. (2014) 85:422–33. doi: 10.1002/cyto.a.22445

7. Rebhahn JA, Quataert SA, Sharma G, Mosmann TR. SwiftReg cluster registration automatically reduces flow cytometry data variability including batch effects. *Commun Biol*. (2020) 3:218. doi: 10.1038/s42003-020-0938-9

8. Bull M, Lee D, Stucky J, Chiu YL, Rubin A, Horton H, et al. Defining blood processing parameters for optimal detection of cryopreserved antigen-specific responses for HIV vaccine trials. *J Immunol Methods*. (2007) 322:57–69. doi: 10.1016/j.jim.2007.02.003

9. Yates NL, Liao HX, Fong Y, deCamp A, Vandergrift NA, Williams WT, et al. Vaccine-induced Env V1-V2 IgG3 correlates with lower HIV-1 infection risk and declines soon after vaccination. *Sci Transl Med*. (2014) 6:228ra239. doi: 10.1126/scitranslmed.3007730

10. Tomaras GD, Yates NL, Liu P, Qin L, Fouda GG, Chavez LL, et al. Initial B-cell responses to transmitted human immunodeficiency virus type 1: virion-binding immunoglobulin M (IgM) and IgG antibodies followed by plasma anti-gp41 antibodies with ineffective control of initial viremia. *J Virol*. (2008) 82:12449–63. doi: 10.1128/JVI.01708-08

11. Pollara J, Hart L, Brewer F, Pickeral J, Packard BZ, Hoxie JA, et al. High-throughput quantitative analysis of HIV-1 and SIV-specific ADCC-mediating antibody responses. *Cytomet A*. (2011) 79:603–12. doi: 10.1002/cyto.a.21084

12. Becht E, McInnes L, Healy J, Dutertre CA, Kwok IWH, Ng LG, et al. Dimensionality reduction for visualizing single-cell data using UMAP. *Nat Biotechnol*. (2018). 27:38–44. doi: 10.1038/nbt.4314

13. Pedersen NW, Chandran PA, Qian Y, Rebhahn J, Petersen NV, Hoff MD, et al. Automated analysis of flow cytometry data to reduce inter-Lab variation in the detection of major histocompatibility complex multimer-Binding T cells. *Front Immunol*. (2017) 8:858. doi: 10.3389/fimmu.2017.00858

14. Raz E, Tighe H, Sato Y, Corr M, Dudler JA, Roman M, et al. Preferential induction of a Th1 immune response and inhibition of specific IgE antibody formation by plasmid DNA immunization. *Proc Natl Acad Sci U.S.A*. (1996) 93:5141–5. doi: 10.1073/pnas.93.10.5141

15. Divekar AA, Zaiss DM, Lee FE, Liu D, Topham DJ, Sijts AJ, et al. Protein vaccines induce uncommitted IL-2-secreting human and mouse CD4 T cells, whereas infections induce more IFN-gamma-secreting cells. *J Immunol*. (2006) 176:1465–73. doi: 10.4049/jimmunol.176.3.1465

16. Sallusto F, Lenig D, Forster R, Lipp M, Lanzavecchia A. Two subsets of memory T lymphocytes with distinct homing potentials and effector functions. *Nature*. (1999) 401:708–12. doi: 10.1038/44385

17. Tomaras GD, Ferrari G, Shen X, Alam SM, Liao HX, Pollara J, et al. Vaccine-induced plasma IgA specific for the C1 region of the HIV-1 envelope blocks binding and effector function of IgG. *Proc Natl Acad Sci U.S.A*. (2013) 110:9019–24. doi: 10.1073/pnas.1301456110

18. Bernard NF, Pederson K, Chung F, Ouellet L, Wainberg MA, Tsoukas CM. HIV-specific cytotoxic T-lymphocyte activity in immunologically normal HIV-infected persons. *Aids*. (1998) 12:2125–39. doi: 10.1097/00002030-199816000-00007

19. Rebhahn JA, Roumanes DR, Qi Y, Khan A, Thakar J, Rosenberg A, et al. Competitive SWIFT cluster templates enhance detection of aging changes. *Cytomet A*. (2016) 89:59–70. doi: 10.1002/cyto.a.22740

20. Sad S, Mosmann TR. Single IL-2-secreting precursor CD4 T cell can develop into either Th1 or Th2 cytokine secretion phenotype. *J Immunol*. (1994) 153:3514–22. doi: 10.4049/jimmunol.153.8.3514

21. Wu CY, Kirman JR, Rotte MJ, Davey DF, Perfetto SP, Rhee EG, et al. Distinct lineages of T(H)1 cells have differential capacities for memory cell generation in *vivo*. *Nat Immunol*. (2002) 3:852–8. doi: 10.1038/ni832

22. Chtanova T, Tangye SG, Newton R, Frank N, Hodge MR, Rolph MS, et al. T follicular helper cells express a distinctive transcriptional profile, reflecting their role as non-Th1/Th2 effector cells that provide help for B cells. *J Immunol*. (2004) 173:68–78. doi: 10.4049/jimmunol.173.1.68

23. Bryant VL, Ma CS, Avery DT, Li Y, Good KL, Corcoran LM, et al. Cytokine-mediated regulation of human B cell differentiation into Ig-secreting cells: predominant role of IL-21 produced by CXCR5+ T follicular helper cells. *J Immunol*. (2007) 179:8180–90. doi: 10.4049/jimmunol.179.12.8180

24. Herati RS, Silva LV, Vella LA, Muselman A, Alanio C, Bengsch B, et al. Vaccine-induced ICOS(+)CD38(+) circulating Tfh are sensitive biosensors of age-related changes in inflammatory pathways. *Cell Rep Med*. (2021) 2:100262. doi: 10.1016/j.xcrm.2021.100262

25. Koutsakos M, Wheatley AK, Loh L, Clemens EB, Sant S, Nussing S, et al. Circulating TFH cells, serological memory, and tissue compartmentalization shape human influenza-specific B cell immunity. *Sci Transl Med*. (2018) 10. doi: 10.1126/scitranslmed.aan8405

26. Gray GE, Bekker LG, Laher F, Malahleha M, Allen M, Moodie Z, et al. Vaccine efficacy of ALVAC-HIV and bivalent subtype C gp120-MF59 in adults. *N Engl J Med*. (2021) 384:1089–100. doi: 10.1056/NEJMoa2031499

frontiers | Frontiers in Immunology

TYPE Original Research
PUBLISHED 08 July 2024
DOI 10.3389/fimmu.2024.1383753

OPEN ACCESS

EDITED BY
Helder Nakaya,
University of São Paulo, Brazil

REVIEWED BY
Gregory C. Ippolito,
The University of Texas at Austin, United States
Sofia Kossida,
Université de Montpellier, France
Matthew Zirui Tay,
A*STAR Infectious Disease Labs, Singapore

*CORRESPONDENCE
Dominic F. Kelly
dominic.kelly@paediatrics.ox.ac.uk

RECEIVED 08 February 2024
ACCEPTED 11 June 2024
PUBLISHED 08 July 2024

CITATION
Richardson E, Bibi S, McLean F, Schimanski L, Rijal P, Ghraichy M, von Niederhäusern V, Trück J, Clutterbuck EA, O'Connor D, Luhn K, Townsend A, Peters B, Pollard AJ, Deane CM and Kelly DF (2024) Computational mining of B cell receptor repertoires reveals antigen-specific and convergent responses to Ebola vaccination. *Front. Immunol.* 15:1383753. doi: 10.3389/fimmu.2024.1383753

Computational mining of B cell receptor repertoires reveals antigen-specific and convergent responses to Ebola vaccination

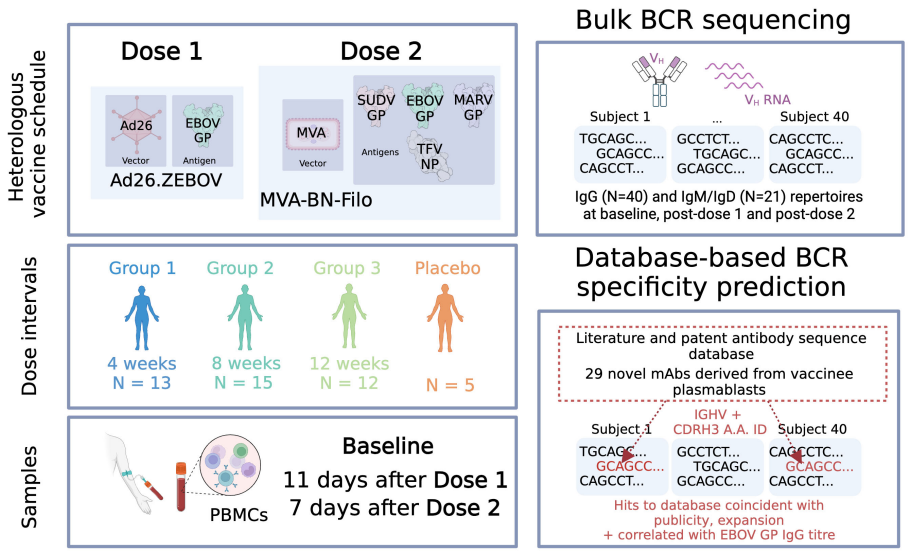
Eve Richardson[1,2,3], Sagida Bibi[2], Florence McLean[2], Lisa Schimanski[4], Pramila Rijal[4], Marie Ghraichy[5], Valentin von Niederhäusern[5], Johannes Trück[5], Elizabeth A. Clutterbuck[2], Daniel O'Connor[2], Kerstin Luhn[6], Alain Townsend[4], Bjoern Peters[3], Andrew J. Pollard[2], Charlotte M. Deane[1] and Dominic F. Kelly[2,7]*

[1]Department of Statistics, University of Oxford, Oxford, United Kingdom, [2]Oxford Vaccine Group, Department of Pediatrics, University of Oxford, Oxford, United Kingdom, [3]La Jolla Institute for Immunology, La Jolla, CA, United States, [4]Weatherall Institute for Molecular Medicine, University of Oxford, Oxford, United Kingdom, [5]Divisions of Allergy and Immunology, University Children's Hospital and Children's Research Center, University of Zurich (UZH), Zurich, Switzerland, [6]Janssen Vaccines and Prevention, Leiden, Netherlands, [7]NIHR Oxford Biomedical Research Centre, Oxford University Hospitals NHS Foundation Trust, Oxford, United Kingdom

Outbreaks of Ebolaviruses, such as Sudanvirus (SUDV) in Uganda in 2022, demonstrate that species other than the Zaire ebolavirus (EBOV), which is currently the sole virus represented in current licensed vaccines, remain a major threat to global health. There is a pressing need to develop effective pan-species vaccines and novel monoclonal antibody-based therapeutics for Ebolavirus disease. In response to recent outbreaks, the two dose, heterologous Ad26.ZEBOV/MVA-BN-Filo vaccine regimen was developed and was tested in a large phase II clinical trial (EBL2001) as part of the EBOVAC2 consortium. Here, we perform bulk sequencing of the variable heavy chain (VH) of B cell receptors (BCR) in forty participants from the EBL2001 trial in order to characterize the BCR repertoire in response to vaccination with Ad26.ZEBOV/MVA-BN-Filo. We develop a comprehensive database, EBOV-AbDab, of publicly available Ebolavirus-specific antibody sequences. We then use our database to predict the antigen-specific component of the vaccinee repertoires. Our results show striking convergence in VH germline gene usage across participants following the MVA-BN-Filo dose, and provide further evidence of the role of IGHV3–15 and IGHV3–13 antibodies in the B cell response to Ebolavirus glycoprotein. Furthermore, we found that previously described Ebola-specific mAb sequences present in EBOV-AbDab were sufficient to describe at least one of the ten most expanded BCR clonotypes in more than two thirds of our cohort of vaccinees following the boost, providing proof of principle for the utility of computational mining of immune repertoires.

KEYWORDS

vaccination, BCR - B cell receptor, BCR-Seq, Ebola (EBOV), monoclonal abs, prediction model



GRAPHICAL ABSTRACT

Introduction

Ebolaviruses are highly infectious zoonotic filoviruses which can cause severe hemorrhagic fever in humans, referred to as Ebolavirus disease (EVD). EVD can have mortality rates of up to 90% (1). There are six species currently classified within the Ebolavirus genus: *Zaire ebolavirus* (EBOV), *Sudan ebolavirus* (SUDV), *Bundibugyo ebolavirus* (BDBV), *Tai Forest ebolavirus* (TFV), *Reston virus* (RESTV) and the most recently described *Bombali ebolavirus* (BOMV). All but Reston and Bombali virus have been associated with severe disease in humans (2, 3). Only three species (EBOV, SUDV, BDBV) have caused outbreaks, with EBOV and SUDV in particular responsible for tens of thousands of deaths in over thirty separate outbreaks in West and equatorial Africa since 1976 (4, 5). Outbreaks continue to occur with regularity, and there have been three distinct Ebolavirus outbreaks in the Democratic Republic of Congo (DRC) between May 2018 and November 2020, an outbreak in Guinea in 2021, and again in the DRC between April and June of 2022. The most recent outbreak was in Uganda from September 2022 to January of 2023.

The 2013–16 Ebola virus outbreak in the DRC was the largest to date, causing in excess of 28,000 cases and 11,000 deaths (6). This epidemic expedited human safety and efficacy testing of Ebola vaccine candidates (7, 8) and the first Ebola virus vaccine, ERVEBO, was approved for use in 2019. ERVEBO is a replication-competent vesicular stomatitis virus (rVSV) based vaccine, and is currently the only FDA-approved vaccine used to immunize at-risk individuals during active outbreaks. ERVEBO is monovalent, only containing the surface glycoprotein of *Zaire ebolavirus*, and efficacy has only been demonstrated for this species. In addition to ERVEBO, a heterologous two-dose vaccination regimen using an adenovirus viral vector expressing Zaire ebolavirus glycoprotein (Ad26.ZEBOV) and an Ankara vector based vaccine expressing the Zaire, Ebola and Sudan ebolavirus glycoproteins along with Tai Forest virus nucleoprotein (MVA-BN-Filo), showed safety and immunogenicity in clinical trials and was licensed for prophylactic use in the European Union in 2020 (9–14). Both vaccines are licensed as monovalent vaccines against *Zaire ebolavirus*.

B cells isolated from convalescent human participants and vaccinees (with both ERVEBO and ChAD3.EBOV/MVA-BN-Filo) have been an important source of therapeutic monoclonals for Ebolavirus. Two monoclonal antibody (mAb)-based immunotherapeutics, Inmazeb and Ebanga/mAb114, are currently FDA approved for the treatment of EVD, however, these only confer moderate protection (15–17). Ebanga/mAb114 was discovered in memory B cells of a survivor of the 1995 Kikwit EVD outbreak (18) and the two component mAbs of MBP134AF were discovered in a survivor of the 2014 EVD outbreak (19, 20). These mAbs are among hundreds discovered in EVD survivors (18, 19, 21–29). Most recently, Chen and colleagues conducted a large-scale sequencing study of a survivor of the 2014 EVD outbreak in Nigeria, estimating over 20,000 EBOV GP-specific clonal lineages within the memory B cell repertoire in just this single participant (30). Among antibody discovery efforts in vaccinees, Rjial and colleagues identified 82 anti-EBOV GP monoclonals from the memory B cells and plasmablasts of participants vaccinated with the ChAD3.EBOV/MVA-BN-Filo vaccine in 2019, while Ehrhardt and colleagues identified 94 anti-EBOV GP monoclonals from rVSV vaccinees (31, 32). As part of the Viral Hemorrhagic Fever Immunotherapeutics Consortium, Saphire and colleagues studied 171 mAbs (of which 102 were human-derived) in the context of the epitopes targeted, neutralization and protection in a mouse model (33). Survivor-derived mAbs and derivatives thereof currently constitute the majority of current immunotherapies for Ebolavirus. While several vaccinee-derived mAbs have demonstrated protection in mice and NHPs, none are currently in the clinic.

In addition to acting as a source of monoclonal antibodies, B cells and their receptor repertoires provide an important window into the response to Ebolavirus infection and vaccination. A

recurrent theme in B cell receptor (BCR) repertoire studies in Ebolavirus and in infectious disease more generally, is the concept of public clonotypes, i.e., groups of related BCR sequences observed in multiple independent participants. Studies of the BCR repertoire in convalescence, and of EBOV-GP specific monoclonal antibodies have highlighted a number of public responses, including usage of IGHV3–13 in antibodies which target the GP1 region of the glycoprotein, IGHV3–15/IGLV1–40-encoded antibodies which target the receptor binding region (RBR), and IGHV1–69 and IGHV1–2 antibodies, which may be important in the early antiviral response (25, 30). Sequencing of four individuals vaccinated with ERVEBO identified a number of public clonotypes shared between the four vaccinees (32). Recently, Chen et al. curated a database of EBOV-specific antibodies from 12 either vaccinated or infected individuals across five studies (30). However, there are currently no publicly available databases where these sequences are compiled.

In the present work, we examined the B cell receptor repertoire response of participants in the Ad26.ZEBOV/MVA-BN-Filo trial. We generated bulk BCR repertoires from forty-five individuals enrolled in the trial, split into three groups according to timing of dose 2 administration, at baseline, after the monovalent dose 1 and the multivalent dose 2. We used our database to computationally annotate the likely antigen-specific component of these repertoires.

Materials and methods

Compilation of EBOV-AbDab

Publications describing Ebolavirus specific monoclonal antibodies were identified from the Immune Epitope database (a database of experimental B and T-cell epitope data by searching for B cell assay data with Ebolavirus as the Epitope Organism. EBOV-specific sequences from patents were retrieved from PLAbDab via searching for the word Ebolavirus (34, 35). Germline gene assignment and identification of CDR3s for the identified antibodies were calculated using IgBLAST and the appropriate IMGT database (human, mouse or macaque) (36–38). In the absence of available nucleotide sequence data, we curated amino acid sequence and used IgBLAST-aa to assign IGV genes and ANARCI to assign IGJ genes (36, 39). In the absence of nucleotide or amino acid sequences, germline genes and CDRH3 and CDRL3 sequences were collected as reported in the original publications. Binding data and neutralization data was collected where available for each antibody as well as, if available, binding to sGP. To create a non-Ebolavirus specific baseline for our antibody specificity predictions, we used two databases: Human CoV-AbDab, filtering for human antibodies based on the Heavy V Gene attribute (dated 13/6/23) and the IEDB (dated 13/6/23), after removing Ebolavirus-specific mAb sequences (34, 40). This resulted in 10,741 and 2,022 entries respectively. We also compared IGHV and IGKLV gene frequencies to a database of HIV antibodies, CATNAP (41). We filtered for IGHV and IGKLV genes and CDRH3s resulting in 394 entries.

Isolation of mAbs from plasmablasts

Antibodies were isolated by FACS sorting, PCR and antibody variable gene cloning of a single B cell plasmablast from six vaccinated human individuals using the previously described methods (Rijal et al., 2019). Briefly, PBMC were incubated with a cocktail of antibodies to CD3 (PB; UCHT1; BD PharMingen), CD20 (APC-H7; 2H7; BD PharMingen), CD19 (FITC; H1B19; BD PharMingen), CD27 (PE-Cy7, M-T271; BD PharMingen), CD38 (PE-Cy5, HIT2; BD PharMingen) and IgG (BV605, G18–145; BD PharMingen). For some sorts, Ebola GP protein (10 µg/mL) and a known biotin-labeled anti-MLD antibody (10 µg/mL) were used to sort antigen specific B cell plasmablasts. Single cells with the phenotype of $CD3^{-}$ $CD20^{-/low}$, $CD19^{+}$, $CD27^{++}$, $CD38^{++}$, IgG^{+} were sorted on a FACS Aria III cell sorter (BD Biosciences). Single cells were sorted into 96-well PCR plates containing lysis buffer followed by single cell RT-PCR. Nested PCR was slightly modified to existing methods. Overlapping bases (approx. 20 nucleotides) were added on to existing 5′ and 3′ primers without interfering the restriction sites, which could be used as a back-up, to enable digestion free Gibson cloning. PCR products were purified in a QIAGEN 96-well system and the inserts were assembled with restriction enzyme-digested plasmids in the Gibson mix (NEB). Two µL of assembled product was used to transform 10 µL DH5α E. Coli (NEB, C2987) in 96-well plates. Three colonies for each heavy and light chain were grown in a 96-well plate format and purified using QIAGEN Turbo 96 miniprep kit. Plasmids were eluted using 100 µL TE buffer.

Expression and purification of antibody

Antibodies were expressed in ExpiCHO cells (Thermo Fisher) by co-transfection with heavy and light plasmids. Antibodies were purified from harvested cell supernatant using MabSelect SuRe (GE Healthcare, 17–5438-01). The column was washed with Tris buffered saline (TBS) and eluted with sodium citrate buffer pH 3.0 – 3.4. Elution pools were neutralized with 2 M Tris/HCl pH 8.0 and absorbance read at 280 nm. Samples were then buffer exchanged into PBS pH 7.4 using 10ml Zeba spin desalting columns, 7K MWCO (Thermo Fisher 89893).

EBL2001 vaccine trial

EBL2001 was a heterologous two-dose randomized, double-blind, placebo-controlled, phase 2 trial of a new Ebolavirus vaccine, performed by the EBOVAC2 consortium (9, 10, 12). Dose 1 of the vaccination regimen is a replication-deficient adenovirus type 26 vector-based vaccine (Ad26.ZEBOV), encoding Zaire Ebola virus glycoprotein, and the dose 2 vaccination is a non-replicating, recombinant, modified Vaccinia ankara (MVA) vector-based vaccine, encoding glycoproteins from Zaire Ebola virus, Sudan virus, and Marburg virus, and nucleoprotein from the Tai Forest virus. Four hundred twenty three participants were enrolled and

randomly assigned to the three different regimes (Groups 1, 2 and 3). Dose 1 administration consisted of either Ad26.ZEBOV or placebo, then this was followed by either MVA-BN-Filo or placebo as dose 2 at 28 (group 1), 56 (group 2), or 84 (group 3) days later.

Samples for BCR sequencing

Peripheral blood was taken from 45 participants enrolled in the EBL2001 trial. Forty participants received the Ad26.ZEBOV dose 1 and MVA-BN-Filo dose 2, while five had received a placebo at both doses. Subjects were selected according to sample availability. Thirteen participants were from interval regimen group 1, 15 from interval regimen group 2, and 12 from interval regimen group 3. Samples were taken prior to vaccination, referred to as Baseline, 11 days following dose 1 referred to as Post-dose 1, and 7 days post-dose 2 referred to as Post-dose 2. 42 of these 45 participants (38/40 vaccinees; 4/5 control participants) were white, with the remainder of Asian (1), mixed (1) or Unknown ethnicity. Twenty-five participants were female and 20 were male. The average age was 42.5, 39, 37.4 and 36.6 years in Group 1, 2, 3 and the Placebo cohort each.

BCR sequencing

PBMCs were isolated via Ficoll-Paque density centrifugation. RNA was extracted using Qiagen RNeasy kit. RT-PCR was performed separately with either IgG, IgA and IgE (all 45 participants) or IgM and IgD primers (21 participants), incorporating unique molecular identifiers (UMIs). V_H cDNA was amplified using a mix of IGHV region primers and Illumina adapter primers as per previous work (42). Samples were multiplexed via combinatorial dual indexing.

Processing of BCR-seq data

BCR-seq data was processed using the Immcantation toolkit (v. 4.4.0) (43, 44). Samples were demultiplexed using the i5 and i7 Illumina indices. A quality filter was applied using *FilterSeq* with a quality cut-off of 30; paired-end reads were joined and merged, and consensuses built according to their UMIs. IgBlast was used to perform germline gene assignment using the *AssignGenes* wrapper with a standard IMGT human germline database, and isotype subtype annotated was performed using *stampy* (36, 45). Sequences were grouped into clonotypes within participant and time points, across time points within the same participant using the *DefineClones* module, with a junctional amino acid identity threshold of 90%. There are multiple clonotype definitions in use: we selected 90% as intermediate in the common range of 80 – 100%. To combat possible index hopping despite dual indexing, the presented analyses consider only UMIs supported by at least two reads or sequences supported by at least two reads. Where sequence or clone abundance is mentioned, this refers to the number of unique UMIs. Without the sequence count filter, we obtained on average 19,257.1 ± 2,832.9 and 99,781.5 ± 10,607.4 sequences per sample; applying this filter resulted in 5,404.2 ± 623.5 sequences per sample and 40,670.8 ± 3,616.7 unique sequences per sample respectively.

Participant EBOV-GP IgG titers

Humoral immunogenicity assessments were carried out with serum from participants and Total IgG Ebola virus glycoprotein-specific binding antibody concentrations were measured by use of an Ebola virus glycoprotein Filovirus Animal Non-Clinical Group ELISA at Q^2 Solutions Laboratories (San Juan Capistrano, CA, USA). Data and methods previously published in Pollard et al. (2021) (9). IgG titers for EBOV GP were measured at baseline and 21 days post-dose 2 for 42/45 participants (37 vaccinee and 5 control).

Competition ELISA

mAb114 and mAb040 which bind non-overlapping epitopes on the Ebola glycoprotein, (the receptor binding region and the glycan cap respectively), were biotinylated using the EZ-Link™ Sulfo-NHS-Biotinylation Kit Biotin-labeled mAb was mixed with unconjugated blocker mAb in a 50-fold excess and they were let to compete for binding to the EBOV-GP on the cell surface. The binding by the biotin-labelled mAb was detected using streptavidin-HRP and TMB peroxidase substrate (Seracare, Cat No. 5120–0076). The reaction was stopped with 1M H_2SO_4 and the absorbance at 450 nm was read using the CLARIOstar plate reader. The data is shown as a percentage of biotinylated mAb binding compared to maximal binding (non-overlapping mAb blocker).

Calculation of immune repertoire parameters

Custom Python scripts were used to calculate parameters such as Gini index, median IGHV identity (% nucleotide identity to the assigned IGHV allele) and to identify expanded and convergent clonotypes. The formula for Gini index is as per *Formula 1*. For IGHV fold changes, fold changes are calculated with a pseudocount of 1.

$$G = \sum_{i=1}^{n}(2i - n - 1)x_i / n\sum_{i=1}^{n}x_i$$

Formula 1: Gini index

Statistical methods

Non-parametric methods are used, i.e. for paired tests, Wilcoxon rank-sum test (implemented with *scipy.stats.wilcoxon*),

and for non-paired tests, Mann-Whitney U-test (implemented with *scipy.stats.mannwhitneyu*) (46). Multiple testing correction is performed via the Benjamini-Hochberg method within *statsmodels.multitests.multipletests* (47). For the IGHV gene comparison to reduce the number of tests performed, repeated measures ANOVA is used prior to *post-hoc* Mann-Whitney U-testing (using *statsmodels.stats.anova.AnovaRM*). For the correlation analysis, Spearman's rank correlation coefficient was calculated with *scipy.stats.spearmanr*, and ordinary least squares regression was performed on log-transformed variables with *statsmodels.formula.api*'s *ols* function.

Antigen-specificity prediction

We predicted Ebolavirus-specificity of EBL2001 participant V_H sequences via shared IGHV and amino acid identity over length-matched CDRH3. In the main text, we used a 70% CDRH3 amino acid identity threshold but explore 80% or 90% CDRH3 identity thresholds in Supplementary Materials. Using clonal relatives of known antibodies to predict antigen specificity is a common approach and was validated previously in a transgenic model (48). This method was implemented in Python and is published as a Python package, *clone_search_ab*.

Results

Compilation of a database of anti-EBOV antibody and nanobody sequences

To collate current knowledge on antibodies against Ebolavirus antigens, we compiled a reference database of antibodies with known specificity for Ebolavirus proteins, collected from academic publications and patents which were identified using the Immune Epitope Database (IEDB) and the Patent/Literature Antibody Database (PLAbDab) (34, 49). In addition, we included the sequences of 29 previously unpublished mAbs (Rijal et al, *in prep*). These novel mAbs were generated from plasmablasts sorted on EBOV-GP from six participants in the same trial who had received an Ad26.ZEBOV dose 1 and a MVA-BN-Filo Ebola dose 2. Altogether, this resulted in a database of 1,019 antibodies and 6 nanobodies, with the encoding IGHV or IGKLV gene and CDRH3 or CDRL3 sequences provided as a minimum. The workflow for our database curation can be seen in Figure 1A.

The majority of the antibodies and nanobodies (939/1,025) targeted the glycoprotein (GP), with 13 targeting the nucleoprotein (NP), 11 targeting the matrix protein VP40, and a single antibody each targeting VP35 and VP30. The majority of the database was of human origin (981/1,025) with the remainder of antibodies of

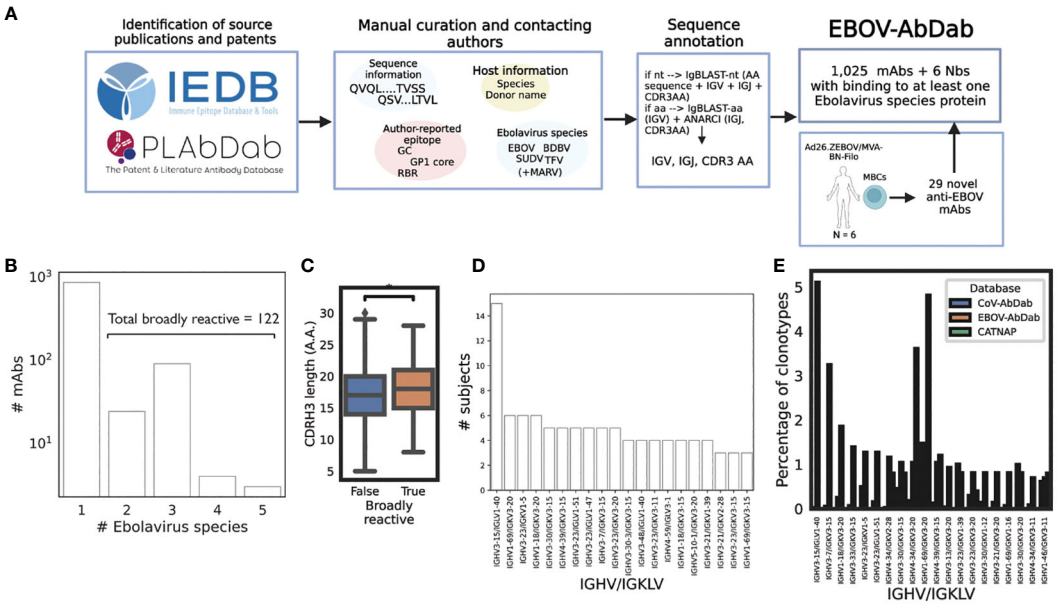


FIGURE 1
Curation of publicly-available Ebolavirus antibody sequences reveals common gene combinations. We manually curated a database of Ebolavirus-binding mAbs (N = 1,025) and Nbs (nanobodies, N = 6) from the workflow described in panel **(A)** Curated and annotated sequence information from the literature with labels such as viral species, protein and epitope, were combined with 29 novel mAbs derived from six Ad26.ZEBOV/MVA-BN-Filo vaccines post-dose 2 to produce a comprehensive database. We curated viral species **(B)**; among the human subset of the data, we identified 122 entries which displayed binding to more than one Ebolavirus. While we are careful to compare these broadly reactive mAbs with mAbs which only have one Ebolavirus label as the absence of data is not equivalent to negative data, we noted a significantly longer CDRH3 length in the broadly reactive subset ($p = 0.03$) **(C)**. We then analyzed the database to identify public antibodies with respect to IGHV/IGKLV gene pairings **(D)** and noted exceptional publicity of the IGHV3–15/IGLV1–40 lineage of antibodies. To put these frequencies into the context of independent viral antibody databases, we compared clonotype frequency within the database to CoV-AbDab and CATNAP (a database of HIV antibodies), and note that IGHV3–15/IGLV1–40 antibodies are rare in other anti-viral antibodies being 174x more common in EBOV-AbDab than CoV-AbDab, and not observed among HIV antibodies **(E)**.

murine (36) and macaque (2) origin and all nanobodies derived from llamas (6).

One of the information fields we collected was the Ebolavirus species known to be targeted by each mAb (EBOV, TFV, BDBV, SUDV, as well as non-Ebolavirus MARV). We curated this information if it was available, but do not distinguish absence vs. negative (i.e., if a mAb is labeled as "EBOV", this does not mean that it does not bind to BDBV, simply that this has not been observed). Focusing on the human subset of the data, we identified 122 entries which bind to at least two Ebolaviruses, which we refer to as broadly reactive (Figure 1B). We are careful not to draw too firm conclusions with respect to this label, however we do note that the average CDRH3 length among clonotypes within this "broadly reactive" category is significantly greater than in antibodies with confirmed binding to a single species ($p = 0.03$, Mann-Whitney U-test) (Figure 1C).

As we are beginning to understand the role of particular IGV genes in determining immunodominance, and since much of the Ebolavirus mAb literature is understood within the context of these genes, e.g. IGHV1–69 and the mucin-like domain (MLD) or IGHV3–15/IGLV1–40 mAbs and the RBR (25, 50, 51) we analyzed our database with respect to these IGV gene pairings. As we collected author-reported donor labels (e.g., EVD5 or Subject 45), we looked at how many donors each gene pairing was identified in. IGHV3–15/IGLV1–40 mAbs were discovered in fifteen participants with the next most public pairing being observed in six participants (IGHV1–69/IGKV3–20, IGHV3–23/IGKV1–5 and IGHV1–18/IGKV3–20) (Figure 1D). We then calculated the frequency of these pairings based on unique clonotypes (IGHV/IGKLV and 90% amino acid identity in the CDRH3) and compared this frequency to that observed in a much larger, independent viral antibody database (CoV-AbDab) (Figure 1E). While IGHV3–15/IGLV1–40 mAbs constitute around 5% of clonotypes within EBOV-AbDab, they constitute just 0.03% of CoV-AbDab, i.e. are 174x more frequent in EBOV-AbDab than CoV-AbDab. There are a further 55 IGHV/IGKLV gene pairings which are at least 10x more frequent in EBOV-AbDab than CoV-AbDab. The differential frequency of IGHV3–15/IGLV1–40 mAbs is primarily driven by the frequency of IGHV3–15 (being 4.9x more frequent vs IGLV1–40 being 1.2 more frequent).

A novel lineage of IGHV3–15/IGLV1–40 mAbs and rediscovery of a known one

We generated 29 novel mAbs from memory B cells of Ad26.ZEBOV/MVA-BN-Filo vaccinees. We noted the frequency of IGHV3–15/IGLV1–40 mAbs (eight mAbs in four clonotypes, two clonotypes in each donor). We examined two lineages of IGHV3–15/IGLV1–40 antibodies from one donor (Donor 58; EBO-1 and EBO2–5) and a single mAb from Donor 35 (EBO11). We measured the competition of these mAbs with mAbs114 (RBR) and mAb040 (GC). These mAbs competed for binding to EBOV GP with mAb114 making it probable that all three lineages target the RBR (Figure 2A). The EBO2–5 lineage is visualized via dendrograms in Figure 2B (VH) and Figure 2C (VL). All six tested mAbs are shown aligned via IMGT numbering in Figure 2D, alongside two separate independent IGHV3–15/IGLV1–40 mAb lineages from the literature - 6666 and 6662 derived from ChAdOx.ZEBOV/MVA-BN-Filo vaccinees and 5T0180 derived from rVSV vaccinees (31, 32). As described by Cohen-Dvashi and colleagues, there is evidence of relative conservation of germline-encoded paratope residues in both the VH and VL, but significant diversity in the CDRH3 (50).

We wanted to contextualize our novel mAbs within our broader database of IGHV3–15/IGLV1–40 antibodies. Given the conservation of the CDRL3, we focused on the non-conserved CDRH3 (Figure 2F). Hierarchical clustering of non-length matched CDRH3 amino acid identity reveals two subclusters: one is the lineage we describe here (represented by EBO-2–5) which has maximally 62% CDRH3 identity to any previously described IGHV3–15/IGLV1–40 mAb thus representing a novel lineage which we refer to as the Donor 58 lineage. The second subcluster shows CDRH3 homology to mAbs isolated from rVSV vaccinees (3T0245 and to a lesser extent 3T0253) and ChAdOx.ZEBOV/MVA-BN-Filo vaccinees (6662 and 6666). We refer to this as the *6666-like clonotype* (as the 6666 CDRH3 is the central CDRH3 in terms of sequence identity).

BCR repertoire sequencing suggests the proliferation of B cells carrying non-mutated IgG BCRs following the monovalent dose 1, with evidence for increasingly mutated BCRs with increasing dose 1-dose 2 interval

When B cells are activated by an antigen stimulus, AID is switched on and the B cell undergoes class switching from IgM/IgD to IgG, A and E and B cell clones responding to the antigen stimulus will accumulate mutations in the variable region of the BCR as the evolve to have higher affinity binding for their epitope. Furthermore, B cells that take on the plasmablast (PB) phenotype rapidly proliferate in a process known as clonal expansion. Finding clonally-related, class-switched or mutated BCR sequences is indicative of antigen exposure.

Following dose 1, we noted a significant increase in the proportion of the IgG repertoire that was unmutated from an average of 1.0 ± 0.4% at baseline to 3.5 ± 0.9% post-prime in the non-placebo group ($p << 0.001$, Wilcoxon test) (Figure 3A) suggesting an increase in frequency of class-switched but non-affinity matured BCRs. Following dose 2, the Group 1 participants still had elevated non-mutated BCRs (3.3 ± 1.4%) relative to both the placebo and Group 2 and Group 3 participants (0.7 ± 0.2%) at this time point ($p = 0.001$ and 0.002 for Group 1 vs. Group 2 and 3 respectively). For the Group 2 and Group 3 participants, the proportion of the repertoire that was non-mutated decreased to comparable levels to baseline and the placebo group (average 1.6 ± 0.6% in the non-placebo, vs. 1.7 ± 3.8% in the placebo) (Figure 3B). Group 1 had the shortest interval regimen of 4 weeks (vs. 8 and 12 weeks respectively). In the longer interval groups, the significant reduction in the proportion of the IgG repertoire that is non-

10.3389/fimmu.2024.1383753

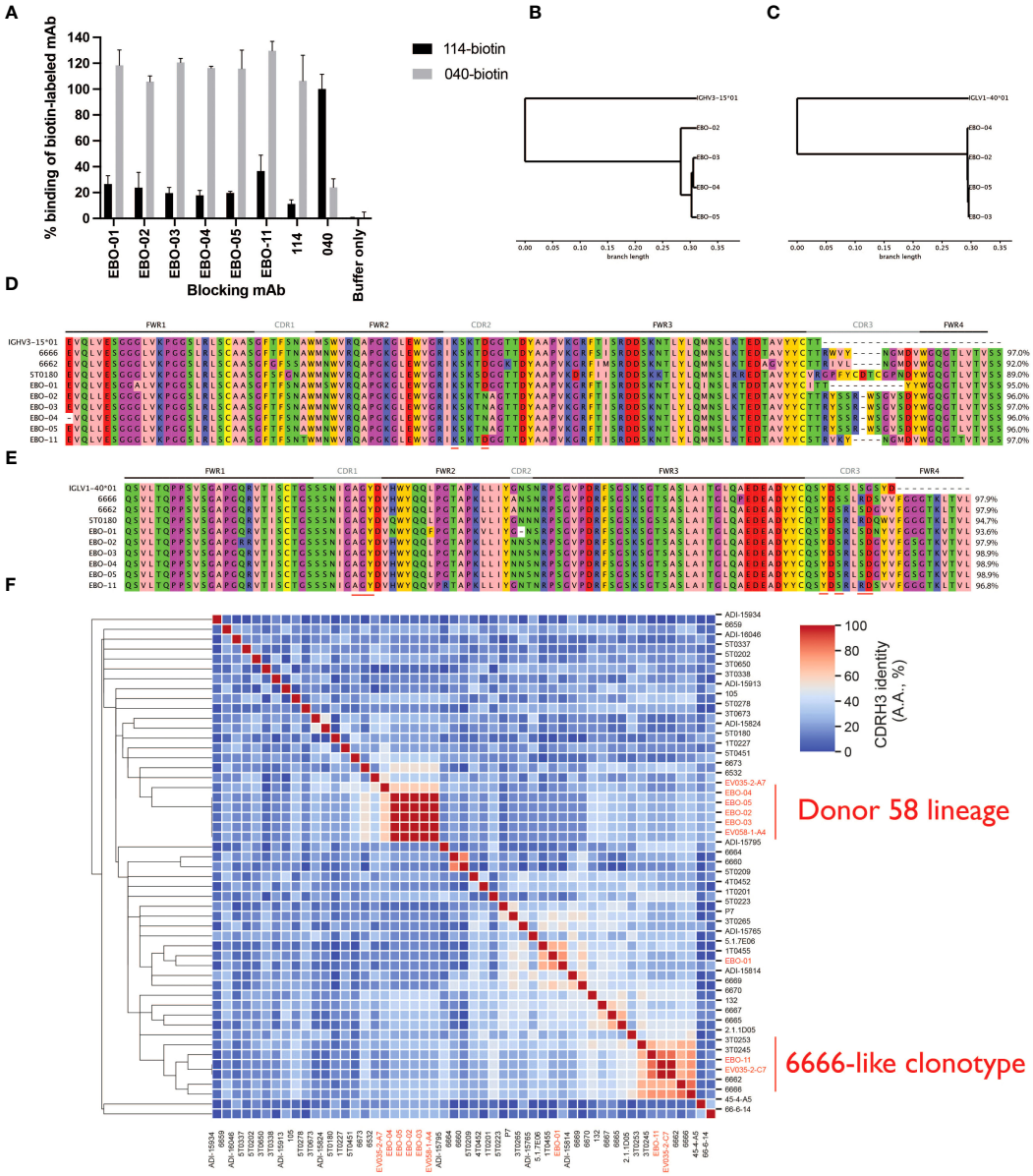


FIGURE 2
Novel IGHV3–15/IGLV1–40 mAbs fall into two groups according to their CDRH3s, the Donor 58 lineage and 6666-like clonotype eight IGHV3–15 mAbs were recovered from two subjects, five from a single subject. We tested six mAbs, EBO-01 to EBO-05, from one subject, and EBO-11 from the other subject, for competition with mAb114, which binds the RBR, and mAb040 which binds a non-overlapping epitope on the glycan cap, on EBOV GP. All six IGHV4-15 mAbs competed with mAb114 suggesting an epitope on the RBR **(A)**. EBO-02 to EBO-05 likely derived from the same clonal expansion: UPGMA dendrograms calculated based on the nucleotide sequences are shown for the VH **(B)** and VL **(C)** with the originating IGHV3–15*01 and IGLV1–40*01 as the outgroup. Panels **(D, E)** show the IMGT-gapped amino acid sequence alignments; red bars on the bottom indicate the paratope residues which are conserved across all three structures solved by Cohen-Dvashi et al. (2020). 5T0180, one of these mAbs, is also included, as are 6666 and 6662 which are RBR-binding mAbs discovered in ChAd.ZEBOV/MVA-BN-Filo vaccinees. Amino acid identity across the IGHV-encoded region is displayed. Germline D61 (IMGT) in CDRH2 which is reported to be a paratope residue is substituted for asparagine in EBO-02, -03 and -04 mAbs while K57 is retained. The new lineage lacks the S113R substitution observed in 5T0180, 6666 and EBO-11. We examined the CDRH3s of our IGHV3–15/IGLV1–40 mAbs within the context of all IGHV3–15/IGLV1–40 mAbs within EBOV-AbDab, with the novel mAbs highlighted in red **(F)**. One subset of our novel mAbs represent one subcluster with 100% CDRH3 identity and maximally 55% CDRH3 identity to any previously described mAb (Donor 58 lineage). We identify a separate subcluster which we refer to as the 6666-like lineage (as 6666's CDRH3 is central) with greater CDRH3 homology.

mutated relative to post-dose 1 for Groups 2 and 3 is consistent with circulating B cells being generated from memory B cells (MBCs) that have had longer to undergo the process of affinity maturation and selection within the germinal center (>8 weeks for groups 2/3 versus 4 weeks for group 1).

To focus on the somatically-mutated, responding clonotypes that were likely to have undergone clonal expansion, we examined the 100 largest, somatically-mutated clonotypes in each repertoire. Measuring IGHV identity (percentage identity to the assigned IGHV allele) in this mutated subset provides a separate insight into how mutated an

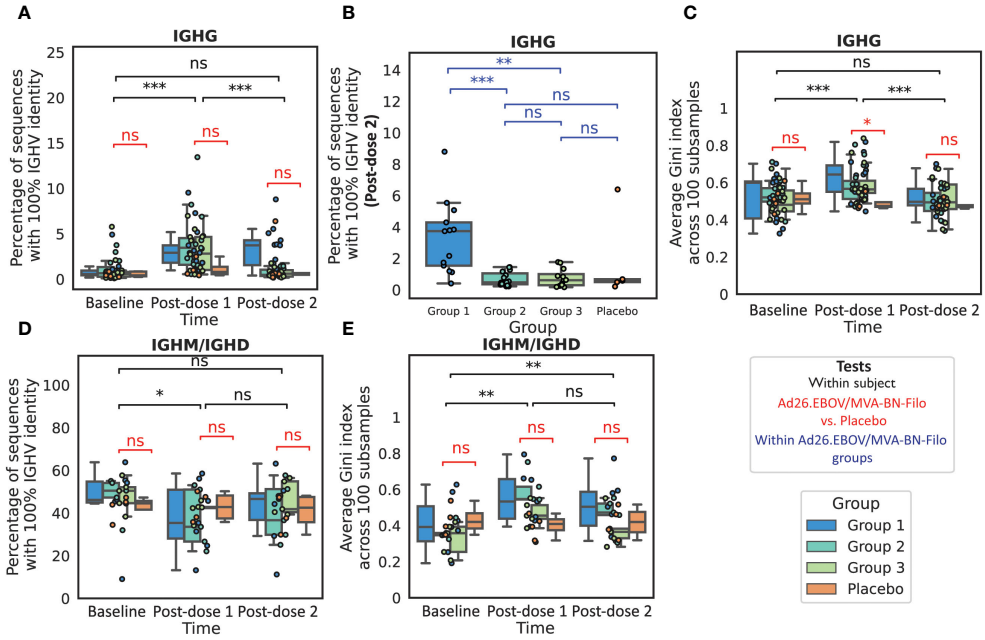


FIGURE 3
The IgG and IgM repertoires exhibit features of antigen exposure following vaccination We noted a significant increase in the proportion of the sequenced IgG repertoire that was non-mutated, defined over the IGHV region **(A)**, post-dose 1 relative to baseline, resulting in a significantly higher proportion of non-mutated sequences in the vaccinees than the placebo group. There was no significant increase post-dose 2 when grouping all boost interval cohorts, however we found that the proportion of the repertoire that was non-mutated was significantly higher in Group 1, which had the shortest dose 1-dose 2 interval of 4 weeks, than in Group 2 (8 weeks), Group 3 (12 weeks) or the Placebo group **(B)**. We next looked at the repertoire polarity in terms of the Gini index (higher Gini indices reflect increased polarization) averaged over 100 subsamples to the minimal number of sequences in the comparison, and noted a significant increase in Gini index from baseline to post-dose 1 in the IgG repertoires followed by a significant decline post-dose 2 to comparable polarization as observed at baseline **(C)**. In IGHM repertoires (with a reduced cohort of 21 subjects with 17 vaccines and four placebo), we noted a small but significant decrease in the proportion of the non-mutated repertoire post-dose 1, however the values were not significantly lower than observed in the placebo group **(D)**. The Gini index was significantly higher than at baseline in the IgM repertoires at both time points **(E)**, however the values observed in the vaccinee repertoires were again not significantly elevated in comparison to the Placebo group. (*, **, ***, ns: significant at the 5%, 1% and 0.01% level, and $p \geq 0.05$).

average responding BCR is, vs. the total repertoire. There was a significant increase in the median IGHV identity for the Groups 1–3 combined after both dose 1 and dose 2 ($p << 0.001$) compared to baseline, with significantly higher median IGHV identities than the placebo at these time points ($p = 0.003$ and 0.003 respectively) (Supplementary Figure 1A). There was no significant difference in median IGHV identity for the 100 largest clonotypes *between* post-dose 1 and post-dose 2 with an average IGHV identity of 94.4 ± 0.7% and 94.6 ± 0.7% respectively ($p = 0.72$). There was a small but significant difference ($p = 0.04$) between Group 1 and Group 2 in the average IGHV identity in the 100 largest clonotypes post-boost, with Group 1 having a slightly higher average IGHV identity (95.2 ± 1.7%, vs. 94.5 ± 1.2% in Group 2) (Supplementary Figure 1B). In summary, these results suggest a post-dose 1 repertoire dominated by recently generated B cells with low or absent SHM. The total post-dose 2 repertoire has a comparable frequency of predicted memory BCRs to baseline, but with lower median IGHV identity; Group 1, with the shortest boost interval, has significantly more non-mutated sequences post-boost than Group 2 or Group 3, and mutated sequences tend to have slightly higher IGHV identity, suggesting that boost interval affects the nature of the B cell memory recall.

We next assessed repertoire polarity via the Gini index which is the area under the curve relating rank and cumulative abundance, averaged over 100 subsamples to the minimum repertoire size in the comparison (Figures 3C, E). In the IgG repertoires, we noted that while there was a significant increase in Gini index (repertoire polarity) from baseline to post-dose 1 (from 0.52 ± 0.03 to 0.60 ± 0.03), there was a significant decrease from post-dose 1 to post-dose 2 (0.51 ± 0.03, $p << 0.001$) such that the expansion was comparable to baseline and the placebo at this time point (0.48 ± 0.03, $p = 0.61$). We would expect to find a comparable if not greater degree of clonal expansion post-dose 2 than post-dose 1, given that the post-dose 1 time point is at the tail end of the expected PB peak. We speculate that this could indicate a more polyclonal response engendered by the multivalent dose 2 than the monovalent dose 1.

In a subset of our cohort (N = 21), we performed IgM/IgD sequencing in addition to IgG sequencing. This is intended to provide a window into the naive repertoire, which is the non-mutated subset of the IgM/IgD repertoire, as well as IgM memory. We noted a small but significant reduction in the naive repertoire (non-mutated IgM/IgD) following dose 1, but not dose 2, from 26.3 ± 4.6% to 19.4 ± 4.6% ($p = 0.01$ and 0.91 respectively) (Figure 3D), however this was not significant relative to the placebo group (24.2 ± 15.3%) ($p = 0.7$). We noted a significant increase in the repertoire clonality from both baseline to post-dose 1 and to post-dose 2 (Figure 3E).

We looked into the longitudinal persistence of clones observed at baseline, post-dose 1 and post-dose 2. We first noted that on average

the clonal overlap between baseline and post-vaccination repertoires was slightly lower than observed in the placebo group, but not significantly so. Focusing on the post-dose 2 repertoire, we found that there was a comparable proportion of clonotypes retained from post-dose 1 in the post-dose 2 repertoires of Group 1, 2 and 3 participants to one another and the placebo group. In the light of the higher abundance of non-mutated IgG sequences at post-dose 2 in Group 1, we specifically focused on the naive to mutated clonotype transition and found that while this appeared to be slightly greater in Group 1 than Group 2 or Group 3, the effect was not statistically significant. There was also no significant difference in the proportion of IgM clonotypes at a prior time point that were observed class-switched at the following time point.

On analyzing the relative proportions of isotype subtype frequencies among the IgG repertoires we found that the proportion of the repertoire occupied by IgG1 increased significantly post-dose 1, from a mean of 50.6 ± 4.2% to 64.8 ± 3.7%, and then again at post-dose 2 to 72.4 ± 3.4%. There were compensatory decreases in IgG2 (39.4 ± 4.3% to 24.0 ± 3.4% to 19.4 ± 2.8%) and IgG4 (1.6 ± 0.6% to 0.8 ± 0.3% to 0.6 ± 0.3%). The vast majority of clonotypes in this post-prime IgG1 increase were novel, in that they did not appear prior to vaccination (96.2 ± 1.1%). There were no significant changes in the isotype subtype frequencies in control participants.

We examined IGHV gene frequencies and while we noted several IGHV genes with significant changes in frequency throughout the course of vaccination among vaccinees in IgG and IgM repertoires, none of these changes were significant relative to those observed in the placebo group, post correction for multiple testing (Supplementary Figures 2, 3). While the IGHV frequencies in IgM/IgD and IgG repertoires within participants at the same time point were reasonably well correlated, with average Spearman correlation coefficients of 0.95 ± 0.01, 0.94 ± 0.01 and 0.95 ± 0.01 at baseline, post-dose 1 and post-dose 2 respectively, the changes in IGHV genes observed in the IgG repertoire were not mirrored in the IgM repertoires (Supplementary Figure 3B). IgG and IgM repertoires were least correlated at post-prime, indicating divergence in the repertoires coincident with the aforementioned predicted PB peak (Supplementary Figure 3C).

These observations suggest an antigen-specific response in vaccinees both post-dose 1 and post-dose 2. While clonal expansion is a reliable marker for antigen-specificity, we decided to use computational immune repertoire mining to refine our prediction of the antigen-specific component of the response.

A database method for the prediction of the EBOV GP-specific IgG repertoire

In order to predict the component of the repertoire that is likely to bind to one of the vaccine antigens, we used our database of Ebolavirus sequences to search for clonal relatives likely to share the same specificity. Clonal relatives were defined as sharing the same IGHV and 70% amino acid identity across the length-matched CDRH3. Predicted Ebolavirus-binding heavy chain sequences are referred to as "Ebolavirus hit sequences". As a control, we compared these results to clonotype predictions using a non-Ebolavirus antibody database built from the non-Ebolavirus-specific, human subset of the Immune Epitope Database (IEDB), as well as the human subset of a separate Coronavirus database (CoV-AbDab). The trial occurred prior to the COVID-19 pandemic, so there is not expected to be any systematic increase in the hit rate to the Coronavirus database. Furthermore, the vaccinees are UK-based and all lacked any IgG titer to the Ebolavirus GP antigen at baseline, therefore there is not expected to be an appreciable hit rate to the Ebolavirus antibody database at baseline.

We measured the proportion of sequences in the repertoire that were hits to the database, which is a function of both the number of hit clonotypes and their abundance. In line with our expectations, we observed a significant increase in the proportion of IgG Ebolavirus hit sequences in the vaccinees' repertoires post-dose 2 (Figure 4A): at baseline, a mean of 0.06± 0.04% of sequences, and maximally 0.68%, were predicted to bind to Ebolavirus. There was a significant increase to 0.51 ± 0.25% post-dose 1 (maximally 4.1%) followed by a significant increase to 3.6 ± 1.2% (maximally 16.6%) post-dose 2. We did not observe any significant changes in the proportion of sequences predicted to bind to non-Ebolavirus antigens (Figures 4B, C). There was no signal in the placebo group, with a mean of 0.07 ± 0.05%, and 0.08 ± 0.11% and 0.03 ± 0.04%, and maximum of 0.10%, 0.18% and 0.06% of IgG sequences predicted to be EBOV-reactive, at baseline, post-dose 1 and post-dose 2 respectively (Figure 4A). There were no significant differences between the different dose 1-dose 2 interval groups in the proportion of the repertoire mapping to the database (Figures 4D–F). We note the same significance intervals, though with hit rates on average 3.6 or 36.0 times lower, using CDRH3 amino acid identity thresholds of 80% and 90% in addition to the 70% threshold used in the main figures (Supplementary Figure 4). We did not observe this signal in the IgM repertoires, with comparably very low hit rates and no significant difference in the proportion of IgM Ebolavirus hit sequences in the vaccinees repertoires post-dose 1 or post-dose 2 compared to baseline (Supplementary Figure 5).

We next looked at the diversity of these predicted hit sequences by looking at the originating clonotypes. We found a significant increase in the number of unique hit clonotypes post-dose 1 and post-dose 2 from on average 1.6 ± 0.6 at baseline to 6.1 ± 2.0 post-prime, and 23.3 ± 5.9 post-boost (Supplementary Figure 6). We found that 45% of clonotypes post-prime were also found within the same participant post-boost, i.e., these predicted antigen-specific sequences that arose during the first vaccination were also observed following the multivalent dose 2. By contrast, only 17.5% of hit clonotypes post-dose 2 were found at the preceding time point, i.e., the majority of post-dose 2 clonotypes were derived from lineages absent at the post-dose 1 time point in the same participant.

To examine whether these novel clonotypes arose through somatic hypermutation of existing hit antibodies, we looked at the hits on the basis of IGHV origin (Figure 5). On average, hit sequences derived from 5.4 ± 0.9 different IGHV genes prior to vaccination, 7.3 ± 1.2 post-dose 1, and 10.1 ± 1.1 post-dose 2, revealing significant diversification in the genetic origins of the predicted antigen-specific component of the BCR repertoire within participants following the multivalent dose 2. Our sequences mapped to 166 mAbs within the EBOV-AbDab database (out of 981 human mAbs); these were encoded by 31 and 36 IGHV genes

10.3389/fimmu.2024.1383753

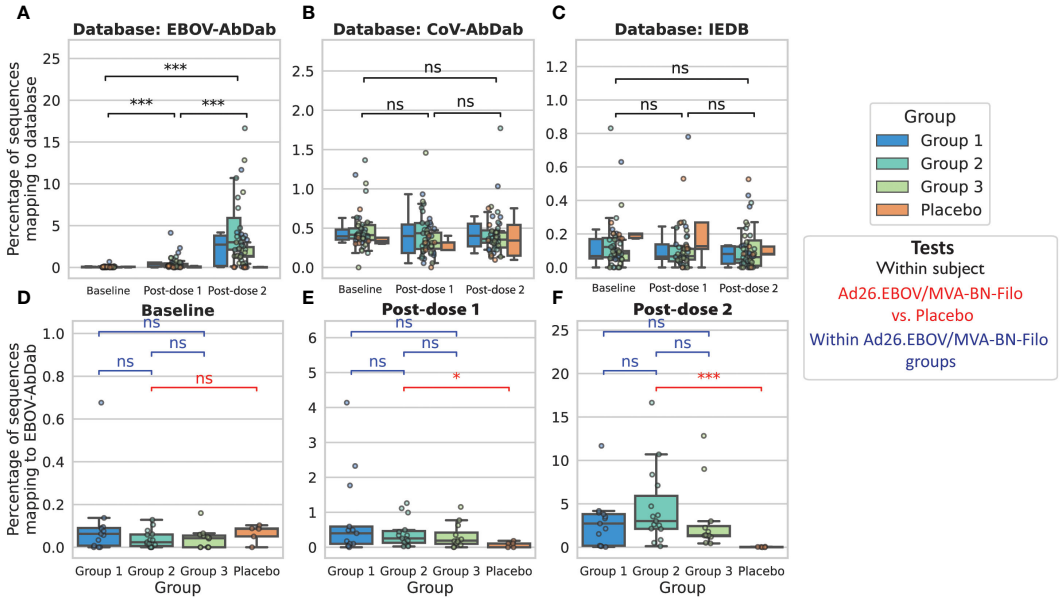


FIGURE 4
Predicted Ebolavirus-specific antibody sequences significantly increase in frequency post-dose 1 and post-dose 2 from baseline, while predicted Coronavirus and other antigen-specific sequences do not significantly change in frequency We used either our curated Ebolavirus antibody database, EBOV-AbDab **(A)**, a Coronavirus-specific antibody database (CoV-AbDab) **(B)** or a non-Ebolavirus database of antibodies to diverse antigens (IEDB) **(C)** to predict the subset of the IgG repertoire that is specific to an antigen in question, and found a significant increase in the percentage of sequences mapping to EBOV-AbDab (referred to as "hits") throughout the course of vaccination, particularly post-dose 2, while there was no significant change in the percentage of sequences mapping to CoV-AbDab or the IEDB. This indicates that these are likely antigen-specific BCRs. For statistical testing, black bars show paired tests between time points **(A-C)**, while red bars show tests between vaccinees and the control group, and blue bars between groups of vaccinees **(D-F)**. There was no significant difference between the placebo group and Ad26.EBOV/MVA-BN-Filo vaccinees prior to vaccination (Mann-Whitney U-test; p = 0.25; red bar in panel **(D)**. Following dose 1, there was a significant increase in the proportion of EBOV-AbDab hits (p<< 0.001), and a further significant increase from post-dose 1 to post-dose 2 (p<< 0.001; black bars in panel **(A)**, resulting in significantly higher percentages of EBOV-AbDab hit sequences in the Ad26.EBOV/MVA-BN-Filo vaccinees vs. the placebo group post-dose 1 (p = 0.03) and post-dose 2 (p<< 0.001) (red bars, **(E, F)**. There were no significant differences between the different dose interval groups at any time point (blue bars, **(D-F)**. There were no significant differences in the hit rates to any database in the IgM/IgD repertoires (Supplementary Figure 5). (*, ***, ns: significant at the 5%, 1% and 0.01% level, and p ≥ 0.05).

post-prime and post-boost respectively with the majority of IGHV genes observed at both time points (29).

Finally, we looked at the hits in the context of the breadth of reactivity to different Ebolavirus species in our database. Of 121 human mAbs with the "broadly reactive" label, there were hits to 27 in the post-dose 1 and post-dose 2 repertoires combined. There were hits to 13 mAbs post-dose 1 and 25 post-dose 2 of which 11 were shared (Figure 5D) indicating that the significant diversification of hit sequences observed post-dose 2 results in more sequences predicted to be broadly reactive appearing in the vaccinee repertoires.

Predicted EBOV-specific sequences are found within expanded and public clones

Ebolavirus hit sequences post-dose 2 were on average found in larger clonotypes than the repertoire average: the mean size of a clonotype containing a hit sequence post-dose 2 had 73.8 ± 27.5 members, in contrast with the repertoire-wide mean of 18.6 ± 3.7. For 35 of 40 vaccinees, the mean clonotype size was larger for hit sequences than the repertoire-wide mean, while for 24 of 40 participants at least one of the ten largest clonotypes contained hit sequences (including 12 participants for which the largest clonotype mapped to the database): for ⅔ of our cohort of vaccinees, our database approach was sufficiently powerful to be able to map at least one of the ten most expanded IgG clonotypes post-dose 2 to characterized mAbs. In one participant, four of the ten largest clonotypes had a hit to our database. Our coverage of the vaccinees' most expanded clonotypes post-dose 2 demonstrates the strength of database-based specificity prediction.

These hit clonotypes were also exceptional with regards to their publicity (Figure 6). Focusing on the 100 largest clonotypes per subject at each time point, we noted 50 clonotypes that were found in more than one subject post-dose 1 or post-dose 2 (bars in blue). Of these 50 public clonotypes, 6 post-dose 1 and 14 post-dose 2 mapped back to our EBOV-specific database (bars in gray) (Figures 6A–C). Post-dose 1, the most public clonotype which was observed in 20 participants, was an IGHV1–2/IGHJ4 clonotype that did not match to our EBOV-AbDab database nor to any antibody in the IEDB or CoV-AbDab. This post-dose 1 clonotype significantly reduced in frequency post-dose 2 (Figure 6D). The second most public clonotype post-dose 1 was observed in 12 participants, corresponding to an IGHV3–15/IGHJ6 clonotype with hits to the 6666-like clonotype mAbs that we highlighted in EBOV-AbDab. This clonotype significantly increased in both publicity (Figure 6C) and within-participant

10.3389/fimmu.2024.1383753

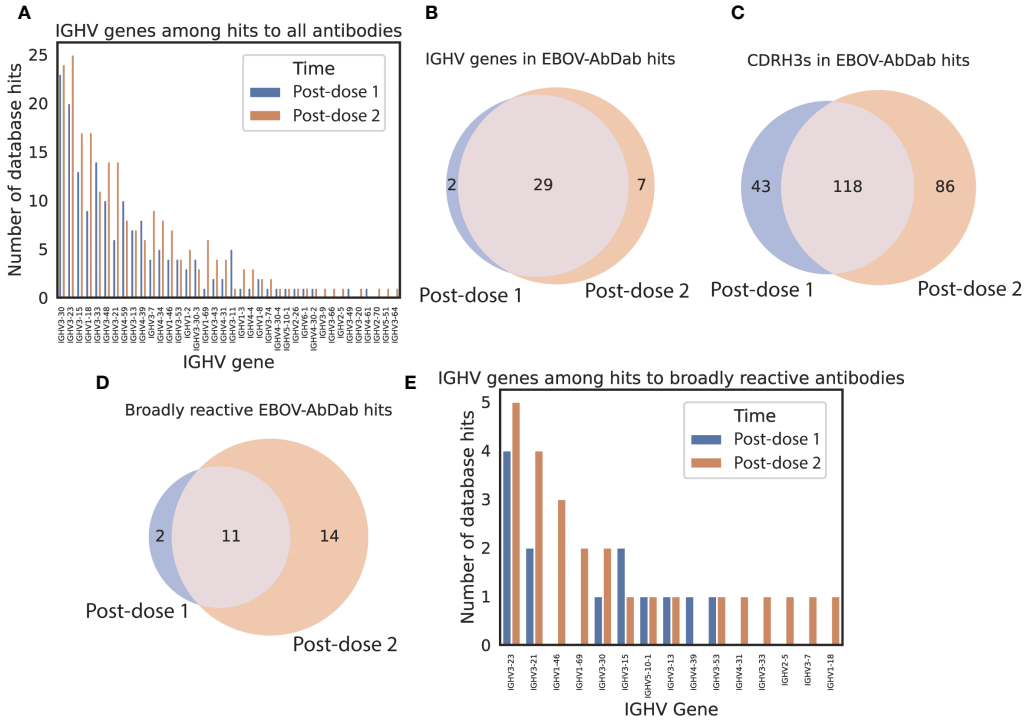


FIGURE 5
predicted hit sequences derive from diverse IGHV origins, and more predicted broadly reactive sequences appear following the second dose EBOV-AbDab hits derive from 38 IGHV genes **(A)**, the majority of which are seen at both time points **(B)**. There are 1.5x as many unique CDRH3s found among post-dose 2 hits than post-dose 1 hits **(C)**. We note that there are hits to twice as many broadly reactive mAbs post-boost than at post-dose 1, indicating that the dose interval may be conducive to developing broadly neutralizing mAbs **(D)**. The broadly neutralizing mAbs derive from 15 IGHV genes **(E)**.

frequency post-dose 2 (Figure 6E), being observed within the 100 largest clonotypes of 32 participants in our cohort of 40 vaccinees. Focusing on this lineage, we noted lower IGHV identities post-dose 2; interestingly, the Group 1 participants had significantly fewer mutations in this lineage (Figure 6E). Permutation test on a per-participant basis on the subset of subjects which had the lineage both post-dose 1 and post-dose 2, revealed a significant decrease in IGHV identity within the majority of participants (Figure 6F).

Figure 6G shows the presence/absence of each hit present in public (in top 100) clonotypes in each participant with any predicted hits post-dose 1; the most public clonotype is clearly the 6666-like clonotype, which is present at a greater frequency than the other set of mAbs we highlighted, our novel lineage discovered within Donor 58 (Figure 6H). Figures 6J, K show the same results post-dose 2, where the number of public clonotypes can be seen to be larger, with again the 6666-like clonotype standing out for its frequency and the number of participants in which it is observed.

The proportion of predicted Ebolavirus-specific sequences correlates with fold-change in anti-EBOV IgG titer

We found a significant correlation between the proportion of Ebolavirus hit sequences in the repertoire 7 days post-boost and the anti-EBOV IgG titer 21 days post-dose 2, after adjusting for an established group effect ($p = 0.001$, $R^2 = 0.51$) (Figure 7). The total Spearman's rho coefficient, not accounting for the different groups, was 0.54 ($p = 0.0006$). There was no significant correlation with the proportion of Ebolavirus hit sequences at the post-dose 1 time point ($p = 0.56$) nor with the Gini index at either time point ($p = 0.86$ and 0.12 for post-dose 1 and post-dose 2 time points respectively) (Supplementary Figure 7).

Discussion

Vaccine development is supported by improvements in our understanding of the humoral immune response to both natural infection and vaccination. This includes gaining insights into epitope immunodominance, the genetic composition of the BCRs targeting those epitopes, how vaccine-induced immunity may generalize to novel variants, and how particular populations respond differently (52). Repertoire sequencing's utility in this context is its view of the repertoire in depth, particularly in the case of bulk VH sequencing where tens to hundreds of thousands of cells can be sequenced. A limitation, in comparison to the wealth of possible single-cell assays, is the loss of the native pairing information which would allow expression and testing of BCRs of interest, as well as the loss of transcriptional or cell surface marker information which would inform on B cell phenotype. Monoclonal antibodies from sorted cells provide information about antigen-

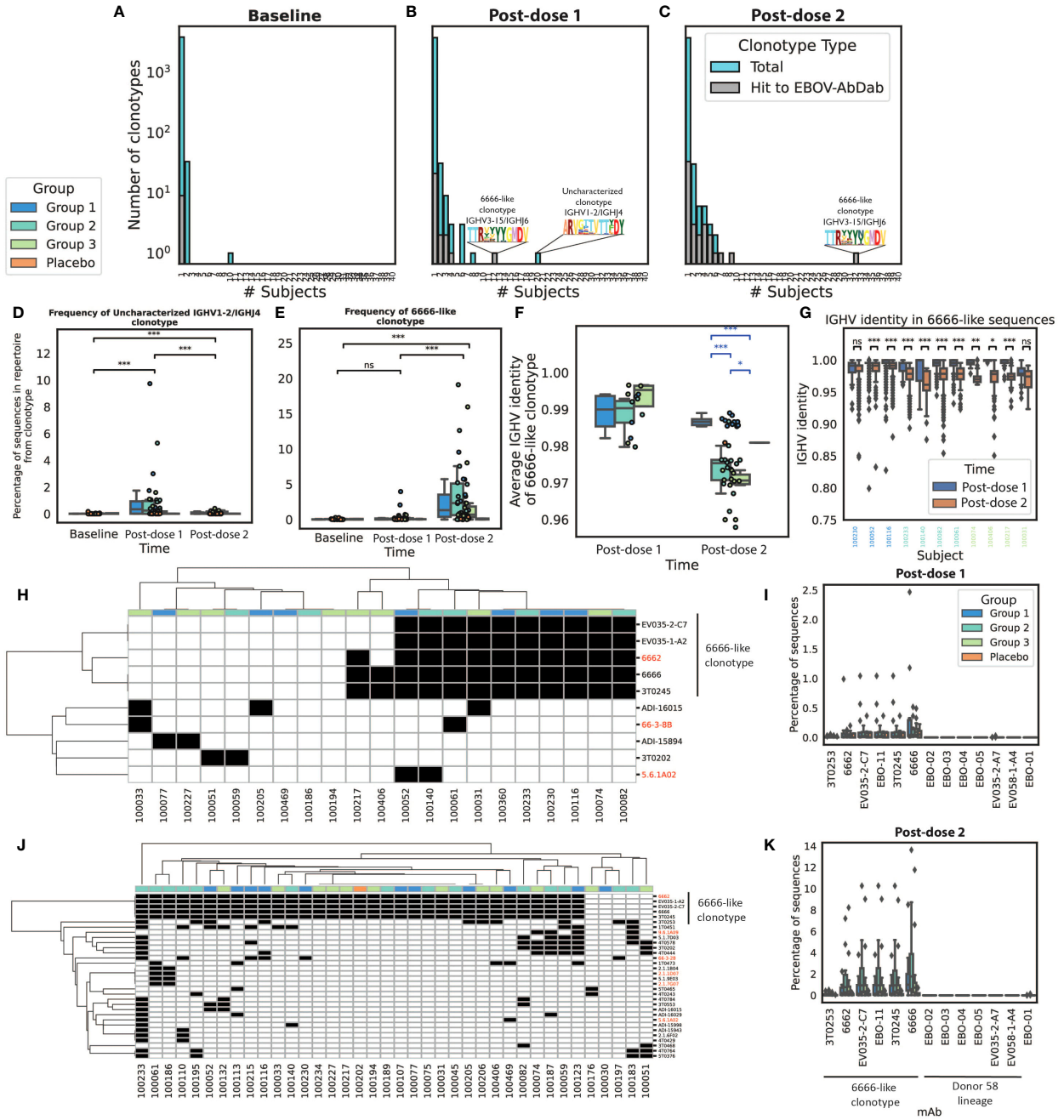


FIGURE 6
Many of the most public clonotypes are predicted to be antigen-specific by our method, with the most notable being the 6666-like lineage which increases in frequency and has reduced IGHV identity post-boost we explored convergence among the 100 largest clonotypes in each IgG repertoire. We noted limited convergence at baseline with the exception of an IGHV3–7 lineage found in ten subjects **(A)**. Post-dose 1, there were 50 clonotypes which were seen in at least two subjects of which six (20%) contained hits to EBOV-AbDab. Unfortunately, the most public clonotype observed in 20 subjects was not a hit to our database (referred to as the uncharacterized IGHV1–24/IGHJ4 clonotype), however the next most public clonotype was a 6666-like clonotype which we had already noted for its publicity within the database itself (**B**). Post-dose 2, there were also 50 public clonotypes, of which 14 (28%) were hits to the database; most notably, the 6666-like clonotype was observed within the 100 most abundant clonotypes of 32/40 vaccinees **(C)**. The uncharacterized IGHV1–24/IGHJ4 clonotype is not only public post-dose 1 but significantly increases in frequency, decreasing again post-dose 2 to comparable levels as at baseline **(D).** By contrast, the 6666-like clonotype significantly increases from baseline to post-dose 2 ($p << 0.001$) but is not significantly increased in frequency post-dose 1 ($p = 0.11$) **(E)**. We focused on this 6666-like clonotype to look for evidence of somatic hypermutation. Interestingly, we found post-dose 2 that this clonotype was significantly less mutated in Group 1, with the shortest boost interval **(F)**. We looked at this lineage on a per-subject basis in the eleven subjects in which there were hit sequences at both post-vaccination timepoints; using a permutation test, we identified four subjects for which there was sufficient evidence that sequences were more mutated post-dose 2 **(G)**. Focusing further on the convergent hit clonotypes, it can be seen that at both post-dose 1 **(H)** and post-dose 2 **(J)** the 6666-like lineage is the most public hit clonotype (red labels correspond to broadly neutralizing antibodies; black boxes indicate presence of the lineage within the 100 largest clonotypes). Of the two lineages we noted in Figure 2, the 6666-like lineage is significantly higher frequency after both dose 1 **(I)** and dose 2 **(K)**, and more public than the Donor 58 lineage. (*, **, ***, ns: significant at the 5%, 1% and 0.01% level, and $p \geq 0.05$).

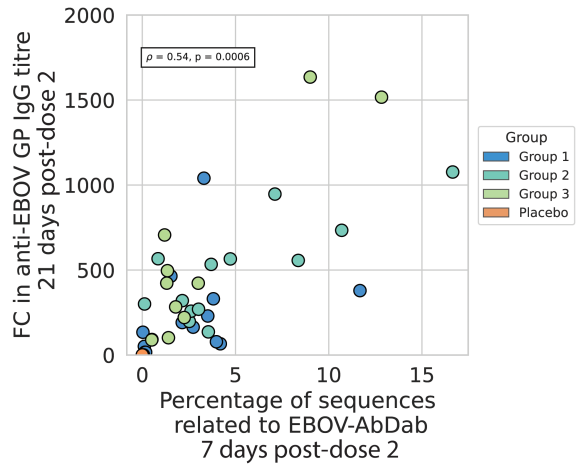


FIGURE 7
The percentage of sequences related to EBOV-AbDab post-dose 2 correlates positively with the fold change in anti-EBOV GP titer from baseline to 21 days post-dose 2 the percentage of sequences related to EBOV-AbDab is positively correlated with the fold-change in anti-EBOV GP IgG titer 21 days post-dose 2, with a Spearman's correlation coefficient of 0.54, p<< 0.001. Given that the fold change is higher on average for Group 2 and Group 3 than Group 1, we added Group as another variable in an OLS regression on the log-transformed values; the total regression had an R^2 of 0.51 with the hit rate variable positively associated with a correlation coefficient of 0.43 and p-value of 0.001. There was no significant correlation with hit rate post-dose 1 (Supplementary Figure 7).

specificity and functionality, but offer a limited window into the diversity of the immune response. Here, computational immune repertoire mining allowed us to somewhat combine the strengths of these two techniques. This database-based technique, validated in a transgenic model system in previous work, has been used in previous studies as validation of antigen-specificity of public clonotypes, for example in the study by Galson and colleagues in which the Coronavirus antibody database (CoV-AbDab) was used to provide evidence of antigen-specificity of convergent clonotypes (48, 53). With the ongoing expansion of available immune repertoire sequence data and monoclonal antibody discovery, we envisage that this approach will become increasingly useful.

In EBL2001 vaccinees we found repertoire polarization following dose 1 in both the IgG and IgM repertoires, a significant increase in the proportion of non-mutated IgG sequences and decrease in the proportion of non-mutated IgM sequences. Among the IgG repertoires, we noted a significant increase in the frequency of the IGHG1 subclass and compensatory decrease in the frequency of the IGHG2 and IGHG4 subclasses. The post-dose 1 B-cell repertoire signature is indicative of clonal expansion and class switching consistent with a plasmablast peak. There was a notable lack of these signatures following the MVA-BN-Filo dose 2, which is consistent with transcriptomic data in which genes related to B cell activation that are clearly upregulated seven days after the Ad26.ZEBOV dose 1 are not significantly upregulated (relative to baseline) following the MVA-BN-Filo dose 2 (54).

The most notable property of the post-dose 2 IgG repertoires was the significantly elevated proportion of sequences predicted to bind to the Ebolavirus glycoprotein according to our database method, which were found disproportionately in expanded and public clonotypes. The most exceptional publicity we observed was in the 6666-like lineage, which was within the 100 largest clonotype post-boost in 32 participants, and which we had already noted as the most public lineage of antibodies in our reference database. An increase in IGHV3–15 frequency was observed by BCR-seq in primary vaccination with ERVEBO by Erhardt and colleagues (32) as well as via RNA-seq by Blengio and colleagues (54). IGHV3–15 thus plays a clear role in the B cell response to Ebolavirus vaccination, from its abundance in monoclonals isolated from at least fifteen EVD survivors and vaccinees, to its appearance in bulk BCR-seq data in both our own Ad26.ZEBOV/MVA-BN-Filo cohort and Erhardt and colleague's ERVEBO cohort, and finally in bulk RNA-seq data.

However, it is not clear that the role this class of antibody plays is equal in both infection and vaccination. Davis and colleagues performed bulk BCR sequencing in a number of EVD survivors and IGHV3–15 was not noteworthy; rather, IGHV3–13 was identified as appearing convergently in their monoclonals isolated from two EVD survivors (55). In Chen and colleagues' 2023 study, IGHV1–69 and IGHV1–2 were the highest frequency among ~10,000 EBOV GP specific clonal lineages sequenced in a single EVD survivor (30). The simplest hypothesis for this discrepancy is that the EVD survivor B cells tend to be from the memory population and collected months post-infection, vs. the plasmablast sequences that we are most likely sampling eleven- and seven days post-prime and post-boost, and maximally 3 months after primary vaccination. The presence of IGHV3–15 among MBC-derived mAbs, as well as the reappearance of more somatically-mutated IGHV3–15 lineages post-boost, indicates that cells expressing this lineage of antibody do indeed enter the memory compartment.

There are further more complex differences among these studies that lie in the broader immunological context of vaccination vs. infection. There is clearly a major role for IGHV1–69 and IGHV1–2 in antiviral B-cell responses more generally to influenza, HIV and hepatitis C virus which could indicate that there is some induction method for these genes that is secondary in vaccination (56). Alternatively, this discrepancy could be immunogenetic. Our Ad26.ZEBOV/MVA-BN-Filo cohort is primarily Caucasian, while Ebolavirus is endemic to West Africa. The role of immunogenetics in the expressed repertoire is only now beginning to be understood due to historic difficulties in resolving the immunoglobulin locus at high-throughput (57–59).

We identified a correlation between the proportion of the repertoire that is predicted to bind to Ebolavirus post-boost and the anti-EBOV IgG titer, however we did not sequence the serum antibodies via Ab/Ig-seq to verify that there was any overlap with the sequences we predicted as antigen-specific. Some have reported little overlap between BCR-seq and the serum, and this was found to be true of the MBC repertoire and serum antibodyome in an Ebolavirus survivor (30, 60, 61). However, the repertoire and

serum overlap should be sensitive to when the two experiments (cellular vs. serum) are performed, as well as to the immunological context – we identified a correlation between the BCR repertoire at seven (cellular) and 21 days (serum) post-boost, in the context of four protein antigens in a non-replication competent viral vector. We would not expect findings in the context of natural infection, with significantly longer or shorter intervals between the two experiments, to generalize to our own findings. Jackson and colleagues found a positive correlation between change in clonality index (comparable to Gini index) and fold-change in anti-HA titer following influenza vaccination (62). With a cohort of just five participants, Trück and colleagues identified a correlation between predicted Hib (*Haemophilis influenzae*)-specific CDR3 sequences and anti-Hib avidity index (63). Our cohort of 45 participants provides stronger statistical evidence that BCR repertoire features can correlate with IgG titer. Whether our predicted antigen-specific antibodies contribute to humoral immunity is a key question that could be addressed via Ab/Ig-seq.

Given the extremely large possible combinatorial diversity of the antibody response, it was surprising to us that existing Ebola-specific mAb sequences were sufficient to describe at least one of the ten most expanded clonotypes post-dose 2 in more than 2/3 of our cohort of vaccinees. Public clonotypes have been identified in BCR-seq data from diverse infection and vaccination contexts (influenza, dengue, HIV-1, coronavirus, Hepatitis C) (30, 62, 64–69) and appear to be the rule rather than the exception, mediated by common physicochemical motifs encoded by both the germline genes and shared somatic hypermutation (70, 71). We observed significantly more public clonotypes following the secondary immunization – we do not yet know whether this is a common feature of the BCR repertoire in primary vs. secondary exposure, or a result of the different viral vectors. Given the ubiquity of public clonotypes, BCR repertoire data should be understood within the context of previously described antibodies to antigens of interest. It would be interesting to consider how large a database of monoclonals would need to be to completely cover the most expanded clonotypes in a cohort of a given size. Continued antibody discovery efforts combined with standardization and deposition of antibody sequence and epitope data in public databases will be critical to the success of these methods.

Data availability statement

The original contributions presented in the study are publicly available. This data can be found here: https://zenodo.org/records/10631605. The original codes presented in the study are found here: https://github.com/erichardson97/CloneSearch.

Ethics statement

The studies involving humans were approved by the French national Ethics Committee (CPP Ile de France III; 3287), the French Medicine Agency (150646A-61), the UK Medicines and Healthcare Products Regulatory Agency (MHRA), and the UK National Research Ethics Service (South Central, Oxford; A 15/SC/0211). The study was done according to the current Declaration of Helsinki and the Good Clinical Practice guidelines. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

ER: Conceptualization, Data curation, Formal analysis, Methodology, Software, Visualization, Writing – original draft, Writing – review & editing. SB: Conceptualization, Data curation, Formal analysis, Methodology, Investigation, Project administration, Writing – original draft, Writing – review & editing. FM: Writing – review & editing. LS: Data curation, Formal analysis, Methodology, Investigation, Writing – original draft, Writing – review & editing. PR: Data curation, Formal analysis, Methodology, Investigation, Writing – original draft, Writing – review & editing. MG: Data curation, Investigation, Writing – review & editing. VV: Data curation, Investigation, Writing – review & editing. JT: Data curation, Investigation, Writing – review & editing. EC: Data curation, Investigation, Methodology, Project administration, Writing – review & editing. DO'C: Investigation, Conceptualization, Writing – review & editing. KL: Writing – review & editing. AT: Writing – review & editing. BP: Writing – review & editing. AP: Writing – review & editing, Funding acquisition, Supervision. CD: Conceptualization, Writing – original draft, Writing – review & editing, Supervision, Funding acquisition. DK: Conceptualization, Writing – original draft, Writing – review & editing, Supervision, Funding acquisition.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. The study was coordinated by the EBOVAC2 multi-partner research consortium, which received funding from the Innovative Medicines Initiative 2 Joint Undertaking (grant number, 115861) as part of the IMI Ebola+ Program. This Joint Undertaking receives support from the EU's Horizon 2020 Framework Program for Research and Innovation and the European Federation of Pharmaceutical Industries and Associations. The EBOVAC2 members include Janssen Vaccines & Prevention BV, which is part of the Janssen Pharmaceutical Companies of Johnson & Johnson (the study sponsor), Centre Muraz, INSERM, INSERM Transfert SA, London School of Hygiene & Tropical Medicine, and the University of Oxford. ER was supported by an Oxford MRC-iCASE studentship (MR/R015708/1). PR was supported by the Chinese Academy of Medical Sciences (CAMS) Innovation Fund for Medical Science (CIFMS), China (grant no. 2018-I2M-2-002). This project has been funded in whole or in part with Federal funds from the National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services, under Contract No. 75N93019C00001.

Conflict of interest

Author AP was a member of WHO's SAGE until 2022, and he is chair of the UK Department of Health and Social Care's Joint Committee on Vaccination and Immunisation. KL was a full-time employee of Janssen Pharmaceuticals at the time of the study and reported stock or stock options in Janssen Pharmaceuticals.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest

Janssen Vaccines & Prevention B.V., the vaccination study sponsor, was involved in the writing of the manuscript and the decision to publish the results. The Oxford Vaccine Group was contracted by Janssen Pharmaceuticals to help conduct the vaccination study from which data was collected.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2024.1383753/full#supplementary-material

References

1. Van Kerkhove MD, Bento AI, Mills HL, Ferguson NM, Donnelly CA. A review of epidemiological parameters from Ebola outbreaks to inform early public health decision-making. *Sci Data*. (2015) 2:150019. doi: 10.1038/sdata.2015.19

2. Goldstein T, Anthony SJ, Gbakima A, Bird BH, Bangura J, Tremeau-Bravard A, et al. Discovery of a new ebolavirus (Bombali virus) in molossid bats in Sierra Leone. *Nat Microbiol*. (2018) 3:1084–9. doi: 10.1038/s41564-018-0227-2

3. Martell HJ, Masterson SG, McGreig JE, Michaelis M, Wass MN. Is the Bombali virus pathogenic in humans? *Bioinformatics*. (2019) 35:3553–8. doi: 10.1093/bioinformatics/btz267

4. Malvy D, McElroy AK, de Clerck H, Günther S, van Griensven J. Ebola virus disease. *Lancet*. (2019) 393:936–48. doi: 10.1016/S0140-6736(18)33132-5

5. Misasi J, Sullivan NJ. Immunotherapeutic strategies to target vulnerabilities in the Ebolavirus glycoprotein. *Immunity*. (2021) 54:412–36. doi: 10.1016/j.immuni.2021.01.015

6. Garske T, Cori A, Ariyarajah A, Blake IM, Dorigatti I, Eckmanns T, et al. Heterogeneities in the case fatality ratio in the West African Ebola outbreak 2013–2016. *Philos Trans R Soc B Biol Sci*. (2017) 372:20160308. doi: 10.1098/rstb.2016.0308

7. Whitty CJM. The contribution of biological, mathematical, clinical, engineering and social sciences to combatting the West African Ebola epidemic. *Philos Trans R Soc B Biol Sci*. (2017) 372:20160293. doi: 10.1098/rstb.2016.0293

8. Osterholm M, Moore K, Ostrowsky J, Kimball-Baker K, Farrar J. The Ebola Vaccine Team B: a model for promoting the rapid development of medical countermeasures for emerging infectious disease threats. *Lancet Infect Dis*. (2016) 16:e1–9. doi: 10.1016/S1473-3099(15)00416-8

9. Pollard AJ, Launay O, Lelievre JD, Lacabaratz C, Grande S, Goldstein N, et al. Safety and immunogenicity of a two-dose heterologous Ad26.ZEBOV and MVA-BN-Filo Ebola vaccine regimen in adults in Europe (EBOVAC2): a randomized, observer-blind, participant-blind, placebo-controlled, phase 2 trial. *Lancet Infect Dis*. (2021) 21:493–506. doi: 10.1016/S1473-3099(20)30476-X

10. Ishola D, Manno D, Afolabi MO, Keshinro B, Bockstal V, Rogers B, et al. Safety and long-term immunogenicity of the two-dose heterologous Ad26.ZEBOV and MVA-BN-Filo Ebola vaccine regimen in adults in Sierra Leone: a combined open-label, non-randomized stage 1, and a randomized, double-blind, controlled stage 2 trial. *Lancet Infect Dis*. (2022) 22:97–109. doi: 10.1016/S1473-3099(21)00125-0

11. Afolabi MO, Ishola D, Manno D, Keshinro B, Bockstal V, Rogers B, et al. Safety and immunogenicity of the two-dose heterologous Ad26.ZEBOV and MVA-BN-Filo Ebola vaccine regimen in children in Sierra Leone: a randomized, double-blind, controlled trial. *Lancet Infect Dis*. (2022) 22:110–22. doi: 10.1016/S1473-3099(21)00128-6

12. Anywaine Z, Barry H, Anzala O, Mutua G, Sirima SB, Eholie S, et al. Safety and immunogenicity of 2-dose heterologous Ad26.ZEBOV, MVA-BN-Filo Ebola vaccination in children and adolescents in Africa: A randomized, placebo-controlled, multicenter Phase II clinical trial. *PloS Med*. (2022) 19:e1003865. doi: 10.1371/journal.pmed.1003865

13. Zabdeno. *European medicines agency* (2024). Available online at: https://www.ema.europa.eu/en/medicines/human/EPAR/zabdeno.

14. Mvabea. *European medicines agency* (2024). Available online at: https://www.ema.europa.eu/en/medicines/human/EPAR/mvabea.

15. Pascal KE, Dudgeon D, Trefry JC, Anantpadma M, Sakurai Y, Murin CD, et al. Development of clinical-stage human monoclonal antibodies that treat advanced ebola virus disease in nonhuman primates. *J Infect Dis*. (2018) 218:S612–26. doi: 10.1093/infdis/jiy285

16. Levine MM. Monoclonal antibody therapy for ebola virus disease. *N Engl J Med*. (2019) 381:2365–6. doi: 10.1056/NEJMe1915350

17. Rijal P, Donnellan FR. A review of broadly protective monoclonal antibodies to treat Ebola virus disease. *Curr Opin Virol*. (2023) 61:101339. doi: 10.1016/j.coviro.2023.101339

18. Corti D, Misasi J, Mulangu S, Stanley DA, Kanekiyo M, Wollen S, et al. Protective monotherapy against lethal Ebola virus infection by a potently neutralizing antibody. *Science*. (2016) 351:1339–42. doi: 10.1126/science.aad5224

19. Bornholdt ZA, Turner HL, Murin CD, Li W, Sok D, Souders CA, et al. Isolation of potent neutralizing antibodies from a survivor of the 2014 Ebola virus outbreak. *Science*. (2016) 351:1078–83. doi: 10.1126/science.aad5788

20. Wec AZ, Bornholdt ZA, ShiHua H, Herbert AS, Goodwin E, Wirchnianski AS, et al. Development of a human antibody cocktail that deploys multiple functions to confer pan-Ebolavirus protection. *Cell Host Amp Microbe*. (2019) 25:39–48.E5. doi: 10.1016/j.chom.2018.12.004

21. Maruyama T, Parren PWHI, Sanchez A, Rensink I, Rodriguez LL, Khan AS, et al. Recombinant human monoclonal antibodies to ebola virus. *J Infect Dis*. (1999) 179:S235–9. doi: 10.1086/514280

22. Meissner F, Maruyama T, Frentsch M, Hessell AJ, Rodriguez LL, Geisbert TW, et al. Detection of antibodies against the four subtypes of ebola virus in sera from any species using a novel antibody-phage indicator assay. *Virology*. (2002) 300:236–43. doi: 10.1006/viro.2002.1533

23. Lee JE, Fusco ML, Hessell AJ, Oswald WB, Burton DR, Saphire EO. Structure of the Ebola virus glycoprotein bound to an antibody from a human survivor. *Nature*. (2008) 454:177–82. doi: 10.1038/nature07082

24. Flyak AI, Shen X, Murin CD, Turner HL, David JA, Fusco ML, et al. Cross-reactive and potent neutralizing antibody responses in human survivors of natural ebolavirus infection. *Cell*. (2016) 164:392–405. doi: 10.1016/j.cell.2015.12.022

25. Murin CD, Gilchuk P, Ilinykh PA, Huang K, Kuzmina N, Shen X, et al. Convergence of a common solution for broad ebolavirus neutralization by glycan cap-directed human antibodies. *Cell Rep*. (2021) 35:108984. doi: 10.1016/j.celrep.2021.108984

26. Gilchuk P, Kuzmina N, Ilinykh PA, Huang K, Gunn BM, Bryan A, et al. Multifunctional pan-ebolavirus antibody recognizes a site of broad vulnerability on the ebolavirus glycoprotein. *Immunity*. (2018) 49:363–374.e10. doi: 10.1016/j.immuni.2018.06.018

27. Williamson LE, Flyak AI, Kose N, Bombardi R, Branchizio A, Reddy S, et al. Early human B cell response to ebola virus in four U.S. Survivors of infection. *J Virol*. (2019) 93(8):e01439-18. doi: 10.1128/jvi.01439–18

28. Gilchuk P, Guthals A, Bonissone SR, Shaw JB, Ilinykh PA, Huang K, et al. Proteo-genomic analysis identifies two major sites of vulnerability on ebolavirus glycoprotein for neutralizing antibodies in convalescent human plasma. *Front Immunol*. (2021) 12:706757. doi: 10.3389/fimmu.2021.706757

29. Milligan JC, Davis CW, Yu X, Ilinykh PA, Huang K, Halfmann PJ, et al. Asymmetric and non-stoichiometric glycoprotein recognition by two distinct antibodies results in broad protection against ebolaviruses. *Cell*. (2022) 185:995–1007.e18. doi: 10.1016/j.cell.2022.02.023

30. Chen EC, Gilchuk P, Zost SJ, Ilinykh PA, Binshtein E, Huang K, et al. Systematic analysis of human antibody response to ebolavirus glycoprotein shows high prevalence of neutralizing public clonotypes. *Cell Rep*. (2023) 42:112370. doi: 10.1016/j.celrep.2023.112370

31. Rijal P, Elias SC, MaChado SR, Xiao J, Schimanski L, O'Dowd V, et al. Therapeutic monoclonal antibodies for ebola virus infection derived from vaccinated humans. *Cell Rep*. (2019) 27:172–186.e7. doi: 10.1016/j.celrep.2019.03.020

32. Ehrhardt SA, Zehner M, Krähling V, Cohen-Dvashi H, Kreer C, Elad N, et al. Polyclonal and convergent antibody response to Ebola virus vaccine rVSV-ZEBOV. *Nat Med*. (2019) 25:1589–600. doi: 10.1038/s41591-019-0602-4

33. Saphire EO, Schendel SL, Fusco ML, Gangavarapu K, Gunn BM, Wec AZ, et al. Systematic analysis of monoclonal antibodies against ebola virus GP defines features that contribute to protection. *Cell*. (2018) 174:938–952.e13. doi: 10.1016/j.cell.2018.07.033

34. Vita R, Mahajan S, Overton JA, Dhanda SK, Martini S, Cantrell JR, et al. The immune epitope database (IEDB): 2018 update. *Nucleic Acids Res*. (2019) 47:D339–43. doi: 10.1093/nar/gky1006

35. Abanades B, Olsen TH, Raybould MIJ, Aguilar-Sanjuan B, Wong WK, Georges G, et al. The Patent and Literature Antibody Database (PLAbDab): an evolving reference set of functionally diverse, literature-annotated antibody sequences and structures. *Nucleic Acids Res*. (2024) 52:D545–51. doi: 10.1093/nar/gkad1056

36. Ye J, Ma N, Madden TL, Ostell JM. IgBLAST: an immunoglobulin variable domain sequence analysis tool. *Nucleic Acids Res*. (2013) 41:W34–40. doi: 10.1093/nar/gkt382

37. Lefranc MP, Giudicelli V, Ginestoux C, Jabado-Michaloud J, Folch G, Bellahcene F, et al. IMGT®, the international ImMunoGeneTics information system®. *Nucleic Acids Res*. (2009) 37:D1006–12. doi: 10.1093/nar/gkn838

38. Manso T, Folch G, Giudicelli V, Jabado-Michaloud J, Kushwaha A, Nguefack Ngoune V, et al. IMGT® databases, related tools and web resources through three main axes of research and development. *Nucleic Acids Res*. (2022) 50:D1262–72. doi: 10.1093/nar/gkab1136

39. Dunbar J, Deane CM. ANARCI: antigen receptor numbering and receptor classification. *Bioinformatics*. (2016) 32:298–300. doi: 10.1093/bioinformatics/btv552

40. Raybould MIJ, Kovaltsuk A, Marks C, Deane CM. CoV-AbDab: the coronavirus antibody database. *Bioinformatics*. (2021) 37:734–5. doi: 10.1093/bioinformatics/btaa739

41. Yoon H, Macke J, West APJr, Foley B, Bjorkman PJ, Korber B, et al. CATNAP: a tool to compile, analyze and tally neutralizing antibody panels. *Nucleic Acids Res*. (2015) 43:W213–9. doi: 10.1093/nar/gkv404

42. Ghraichy M, Galson JD, Kovaltsuk A, von Niederhäusern V, Pachlopnik Schmid J, Recher M, et al. Maturation of the human immunoglobulin heavy chain repertoire with age. *Front Immunol*. (2020) 11:1734. doi: 10.3389/fimmu.2020.01734

43. Vander Heiden JA, Yaari G, Uduman M, Stern JNH, O'Connor KC, Hafler DA, et al. pRESTO: a toolkit for processing high-throughput sequencing raw reads of lymphocyte receptor repertoires. *Bioinformatics*. (2014) 30:1930–2. doi: 10.1093/bioinformatics/btu138

44. Gupta NT, Vander Heiden JA, Uduman M, Gadala-Maria D, Yaari G, Kleinstein SH. Change-O: a toolkit for analyzing large-scale B cell immunoglobulin repertoire sequencing data. *Bioinformatics*. (2015) 31:3356–8. doi: 10.1093/bioinformatics/btv359

45. Lunter G, Goodson M. Stampy: A statistical algorithm for sensitive and fast mapping of Illumina sequence reads. *Genome Res*. (2011) 21:936–9. doi: 10.1101/gr.111120.110

46. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat Methods*. (2020) 17:261–72. doi: 10.1038/s41592-019-0686-2

47. Seabold S, Perktold J. (2010). Statsmodels: econometric and statistical modeling with python, in: *Proc 9th Python Sci Conf*, . pp. 92–6.

48. Richardson E, Galson JD, Kellam P, Kelly DF, Smith SE, Palser A, et al. A computational method for immune repertoire mining that identifies novel binders from different clonotypes, demonstrated by identifying anti-pertussis toxoid antibodies. *mAbs*. (2021) 13:1869406. doi: 10.1080/19420862.2020.1869406

49. Abanades B, Olsen TH, Raybould MIJ, Aguilar-Sanjuan B, Wong WK, Georges G, et al. The Patent and Literature Antibody Database (PLAbDab): an evolving reference set of functionally diverse, literature-annotated antibody sequences and structures. *Nucleic Acids Res*. (2024) 52(D1):D545–51. doi: 10.1093/nar/gkad1056

50. Cohen-Dvashi H, Zehner M, Ehrhardt S, Katz M, Elad N, Klein F, et al. Structural basis for a convergent immune response against ebola virus. *Cell Host Microbe*. (2020) 27:418–27. doi: 10.1016/j.chom.2020.01.007

51. Yu X, Hastie KM, Davis CW, Avalos RD, Williams D, Parekh D, et al. The evolution and determinants of neutralization of potent head-binding antibodies against Ebola virus. *Cell Rep*. (2023) 42(11):113366. doi: 10.1016/j.celrep.2023.113366

52. Fink K. Can we improve vaccine efficacy by targeting T and B cell repertoire convergence? *Front Immunol*. (2019) 10:110. doi: 10.3389/fimmu.2019.00110

53. Galson JD, Schaetzle S, Bashford-Rogers RJM, Raybould MIJ, Kovaltsuk A, Kilpatrick GJ, et al. Deep sequencing of B cell receptor repertoires from COVID-19 patients reveals strong convergent immune signatures. *Front Immunol*. (2020) 11:605170. doi: 10.3389/fimmu.2020.605170

54. Blengio F, Hocini H, Richert L, Lefebvre C, Durand M, Hejblum B, et al. Identification of early gene expression profiles associated with long-lasting antibody responses to the Ebola vaccine Ad26.ZEBOV/MVA-BN-Filo. *Cell Rep*. (2023) 42:113101. doi: 10.1016/j.celrep.2023.113101

55. Davis CW, Jackson KJL, McElroy AK, Halfmann P, Huang J, Chennareddy C, et al. Longitudinal analysis of the human B cell response to ebola virus infection. *Cell*. (2019) 177:1566–82. doi: 10.1016/j.cell.2019.04.036

56. Chen F, Tzarum N, Wilson IA, Law M. VH1–69 antiviral broadly neutralizing antibodies: genetics, structures, and relevance to rational vaccine design. *Curr Opin Virol*. (2019) 34:149–59. doi: 10.1016/j.coviro.2019.02.004

57. Rodriguez OL, Safonova Y, Silver CA, Shields K, Gibson WS, Kos JT, et al. Genetic variation in the immunoglobulin heavy chain locus shapes the human antibody repertoire. *Nat Commun*. (2023) 14:4419. doi: 10.1038/s41467-023-40070-x

58. Castro Dopico X, Mandolesi M, Karlsson Hedestam GB. Untangling associations between immunoglobulin genotypes, repertoires and function. *Immunol Lett*. (2023) 259:24–9. doi: 10.1016/j.imlet.2023.05.003

59. deCamp AC, Corcoran MM, Fulp WJ, Willis JR, Cottrell CA, Bader DLV, et al. Human immunoglobulin gene allelic variation impacts germline-targeting vaccine priming. *NPJ Vaccines*. (2024) 9:58. doi: 10.1038/s41541-024-00811-5

60. Quý KL, Chernigovskaya M, Stensland M, Singh S, Leem J, Revale S, et al. Benchmarking and integrating human B-cell receptor genomic and antibody proteomic profiling [Internet]. *bioRxiv*. (2023), 2023.11.01.565093. doi: 10.1101/2023.11.01.565093v1

61. Lavinder JJ, Wine Y, Giesecke C, Ippolito GC, Horton AP, Lungu OI, et al. Identification and characterization of the constituent human serum antibodies elicited by vaccination. *Proc Natl Acad Sci*. (2014) 111:2259–64. doi: 10.1073/pnas.1317793111

62. Jackson KJL, Liu Y, Roskin KM, Glanville J, Hoh RA, Seo K, et al. Human responses to influenza vaccination show seroconversion signatures and convergent antibody rearrangements. *Cell Host Microbe*. (2014) 16:105–14. doi: 10.1016/j.chom.2014.05.013

63. Trück J, Ramasamy MN, Galson JD, Rance R, Parkhill J, Lunter G, et al. Identification of antigen-specific B cell receptor sequences using public repertoire analysis. *J Immunol*. (2015) 194:252–61. doi: 10.4049/jimmunol.1401405

64. Skinner NE, Ogega CO, Frumento N, Clark KE, Paul H, Yegnasubramanian S, et al. Convergent antibody responses are associated with broad neutralization of hepatitis C virus. *Front Immunol*. (2023) 14:1135841. doi: 10.3389/fimmu.2023.1135841

65. Wang Y, Yuan M, Lv H, Peng J, Wilson IA, Wu NC. A large-scale systematic survey reveals recurring molecular features of public antibody responses to SARS-CoV-2. *Immunity*. (2022) 55:1105–1117.e4. doi: 10.1016/j.immuni.2022.03.019

66. Waltari E, Nafees S, McCutcheon KM, Wong J, Pak JE. AIRRscape: An interactive tool for exploring B-cell receptor repertoires and antibody responses. *PloS Comput Biol*. (2022) 18:e1010052. doi: 10.1371/journal.pcbi.1010052

67. Stewart A, Sinclair E, Ng JCF, O'Hare JS, Page A, Serangeli I, et al. endemic: B cell repertoire analysis reveals unique anti-viral responses to SARS-coV-2, ebola and respiratory syncytial virus. *Front Immunol*. (2022) 13:807104. doi: 10.3389/fimmu.2022.807104

68. Parameswaran P, Liu Y, Roskin KM, Jackson KKL, Dixit VP, Lee JY, et al. Convergent antibody signatures in human dengue. *Cell Host Microbe*. (2013) 13:691–700. doi: 10.1016/j.chom.2013.05.008

69. Robbiani DF, Bozzacco L, Keeffe JR, Khouri R, Olsen PC, Gazumyan A, et al. Recurrent potent human neutralizing antibodies to zika virus in Brazil and Mexico. *Cell*. (2017) 169:597–609.e11. doi: 10.1016/j.cell.2017.04.024

70. Shrock EL, Timms RT, Kula T, Mena EL, West AP, Guo R, et al. Germline-encoded amino acid–binding motifs drive immunodominant public antibody responses. *Science*. (2023) 380:eadc9498. doi: 10.1126/science.adc9498

71. Abu-Shmais AA, Vukovich MJ, Wasdin PT, Suresh YP, Rush SA, Gillespie RA, et al. Convergent sequence features of antiviral B cells. *bioRxiv*. (2023), 2023.09.06.556442. doi: 10.1101/2023.09.06.556442v1

Frontiers in Immunology

Explores novel approaches and diagnoses to treat immune disorders.

The official journal of the International Union of Immunological Societies (IUIS) and the most cited in its field, leading the way for research across basic, translational and clinical immunology.

Discover the latest Research Topics

See more →

Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact