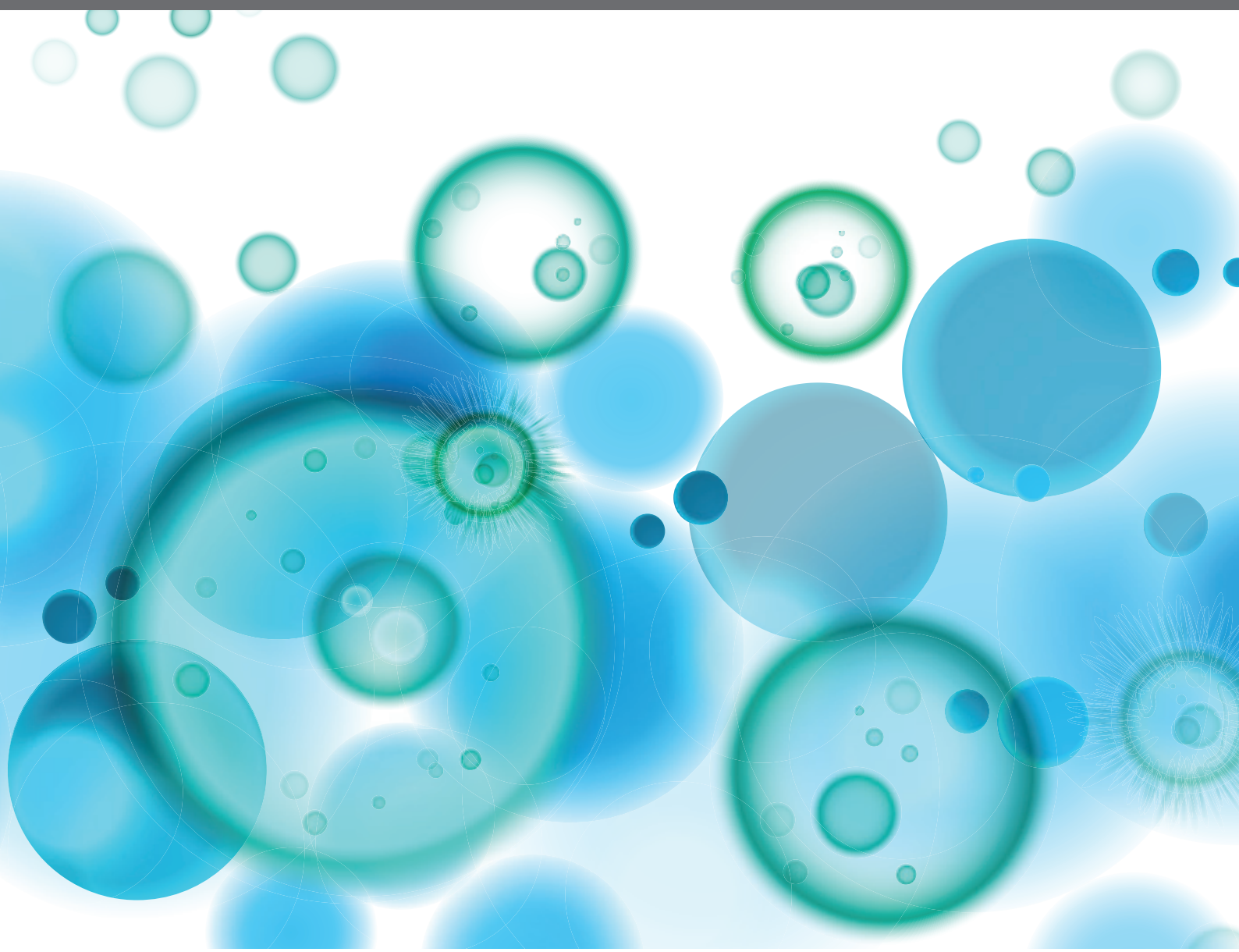
SYSTEMS IMMUNOLOGY – LANDSCAPING IMMUNE REGULATORY NETWORKS

EDITED BY: Aridaman Pandit, Jurjen Tel, Federica Eduati, Rob J. De Boer, Florian Wimmers and David Gomez-Cabrero

PUBLISHED IN: Frontiers in Immunology



frontiers Research Topics

ISSN 1664-8714
ISBN 978-2-88974-860-0
DOI 10.3389/978-2-88974-860-0

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 – LANDSCAPING IMMUNE REGULATORY NETWORKS

Topic Editors:
Aridaman Pandit, University Medical Center Utrecht, Netherlands
Jurjen Tel, Eindhoven University of Technology, Netherlands
Federica Eduati, Eindhoven University of Technology, Netherlands
Rob J. De Boer, Utrecht University, Netherlands
Florian Wimmers, Stanford University, United States
David Gomez-Cabrero, NavarraBiomed, Spain

Citation: Pandit, A., Tel, J., Eduati, F., De Boer, R. J., Wimmers, F., Gomez-Cabrero, D., eds. (2022). Systems Immunology – Landscaping Immune Regulatory Networks. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88974-860-0

Table of Contents

05 ***Novel Insights Into Rheumatoid Arthritis Through Characterization of Concordant Changes in DNA Methylation and Gene Expression in Synovial Biopsies of Patients With Differing Numbers of Swollen Joints***
Andrew Y. F. Li Yim, Enrico Ferrero, Klio Maratou, Huw D. Lewis, George Royal, David F. Tough, Chris Larminie, Marcel M. A. M. Mannens, Peter Henneman, Wouter J. de Jonge, Marleen G. H. van de Sande, Danielle M. Gerlag, Rab K. Prinjha and Paul P. Tak

20 ***Innate Immune Cells in the Esophageal Tumor Microenvironment***
Kele Cui, Shouxin Hu, Xinyu Mei and Min Cheng

33 ***Corrigendum: Innate Immune Cells in the Esophageal Tumor Microenvironment***
Kele Cui, Shouxin Hu, Xinyu Mei and Min Cheng

34 ***Phenotypical Diversification of Early IFNα-Producing Human Plasmacytoid Dendritic Cells Using Droplet-Based Microfluidics***
Laura C. Van Eyndhoven, Eleni Chouri, Nikita Subedi and Jurjen Tel

46 ***Comparison of Two Strategies to Generate Antigen-Specific Human Monoclonal Antibodies: Which Method to Choose for Which Purpose?***
Anna M. Ehlers, Constance F. den Hartog Jager, Tineke Kardol-Hoefnagel, Miriam M.D. Katsburg, André C. Knulst and Henny G. Otten

61 ***Integration of* Immunome *With Disease-Gene Network Reveals Common Cellular Mechanisms Between IMIDs and Drug Repurposing Strategies***
Abhinandan Devaprasad, Timothy R. D. J. Radstake and Aridaman Pandit

79 ***Transcriptional and Microenvironmental Landscape of Macrophage Transition in Cancer: A Boolean Analysis***
Ugo Avila-Ponce de León, Aarón Vázquez-Jiménez, Meztli Matadamas-Guzman, Rosana Pelayo and Osbaldo Resendis-Antonio

96 ***A Stochastic Intracellular Model of Anthrax Infection With Spore Germination Heterogeneity***
Bevelynn Williams, Martín López-García, Joseph J. Gillard, Thomas R. Laws, Grant Lythe, Jonathan Carruthers, Thomas Finnie and Carmen Molina-París

119 ***Literature Mining and Mechanistic Graphical Modelling to Improve mRNA Vaccine Platforms***
Lorena Leonardelli, Giuseppe Lofano, Gianluca Selvaggio, Silvia Parolo, Stefano Giampiccolo, Danilo Tomasoni, Enrico Domenici, Corrado Priami, Haifeng Song, Duccio Medini, Luca Marchetti and Emilio Siena

125 ***Sustained Drug Treatment Alters the Gut Microbiota in Rheumatoid Arthritis***
Liyan Mei, Zhihua Yang, Xiaolin Zhang, Zehao Liu, Maojie Wang, Xiaodong Wu, Xiumin Chen, Qingchun Huang and Runyue Huang

137 ***Continuous Modeling of T CD4 Lymphocyte Activation and Function***
David Martínez-Méndez, Luis Mendoza, Carlos Villarreal and Leonor Huerta

152 ***Corona Virus Disease 2019 (COVID-19) as a System-Level Infectious Disease With Distinct Sex Disparities***
Modjtaba Emadi-Baygi, Mahsa Ehsanifard, Najmeh Afrashtehpour, Mahnaz Norouzi and Zahra Joz-Abbasalian

169 ***Integrative Lipidomics and Metabolomics for System-Level Understanding of the Metabolic Syndrome in Long-Term Treated HIV-Infected Individuals***
Sofie Olund Villumsen, Rui Benfeitas, Andreas Dehlbæk Knudsen, Marco Gelpi, Julie Høgh, Magda Teresa Thomsen, Daniel Murray, Henrik Ullum, Ujjwal Neogi and Susanne Dam Nielsen

frontiers in Immunology

ORIGINAL RESEARCH
published: 22 April 2021
doi: 10.3389/fimmu.2021.651475

OPEN ACCESS

Edited by:
Aridaman Pandit,
University Medical Center Utrecht, Netherlands

Reviewed by:
W Tao,
University Medical Center Utrecht, Netherlands
Saikat Chowdhury,
University of Texas MD Anderson Cancer Center, United States

***Correspondence:**
Rab K. Prinjha
rabinder.prinjha@gsk.com
Paul P. Tak
tak.paulpeter@gmail.com

†Present Address:
Enrico Ferrero,
Autoimmunity Transplantation and Inflammation Bioinformatics, Novartis Institutes for BioMedical Research, Basel, Switzerland;
Paul P. Tak,
Candel Therapeutics, Needham, MA, United States

‡These authors share first authorship

§These authors share last authorship

Specialty section:
This article was submitted to Molecular Innate Immunity, a section of the journal Frontiers in Immunology

Received: 09 January 2021
Accepted: 25 March 2021
Published: 22 April 2021

Citation:
Li Yim AYF, Ferrero E, Maratou K, Lewis HD, Royal G, Tough DF, Larminie C, Mannens MMAM, Henneman P, de Jonge WJ, van de Sande MGH, Gerlag DM, Prinjha RK and Tak PP (2021) Novel Insights Into Rheumatoid Arthritis Through Characterization of Concordant Changes in DNA Methylation and Gene Expression in Synovial Biopsies of Patients With Differing Numbers of Swollen Joints. Front. Immunol. 12:651475. doi: 10.3389/fimmu.2021.651475

Novel Insights Into Rheumatoid Arthritis Through Characterization of Concordant Changes in DNA Methylation and Gene Expression in Synovial Biopsies of Patients With Differing Numbers of Swollen Joints

Andrew Y. F. Li Yim[1,2‡], Enrico Ferrero[1†‡], Klio Maratou[1], Huw D. Lewis[1], George Royal[1], David F. Tough[1], Chris Larminie[1], Marcel M. A. M. Mannens[2], Peter Henneman[2], Wouter J. de Jonge[3,4], Marleen G. H. van de Sande[5,6], Danielle M. Gerlag[1], Rab K. Prinjha[1§] and Paul P. Tak[1,5,7,8*†§]*

[1] R&D GlaxoSmithKline, Stevenage, United Kingdom, [2] Department of Clinical Genetics, Genome Diagnostics Laboratory, Amsterdam Reproduction & Development, Amsterdam University Medical Centers, University of Amsterdam, Amsterdam, Netherlands, [3] Tytgat Institute for Liver and Intestinal Research, Amsterdam Gastroenterology & Metabolism, Amsterdam University Medical Centers, University of Amsterdam, Amsterdam, Netherlands, [4] Department of Surgery, University Clinic of Bonn, Bonn, Germany, [5] Department of Rheumatology and Clinical Immunology, Amsterdam Rheumatology and Immunology Center, Amsterdam Institute for Infection & Immunity, Amsterdam University Medical Centers, University of Amsterdam, Amsterdam, Netherlands, [6] Department of Experimental Immunology, Amsterdam Institute for Infection & Immunity, Amsterdam University Medical Centers, University of Amsterdam, Amsterdam, Netherlands, [7] Department of Rheumatology, Ghent University, Ghent, Belgium, [8] Department of Medicine, University of Cambridge, Cambridge, United Kingdom

In this study, we sought to characterize synovial tissue obtained from individuals with arthralgia and disease-specific auto-antibodies and patients with established rheumatoid arthritis (RA), by applying an integrative multi-omics approach where we investigated differences at the level of DNA methylation and gene expression in relation to disease pathogenesis. We performed concurrent whole-genome bisulphite sequencing and RNA-Sequencing on synovial tissue obtained from the knee and ankle from 4 auto-antibody positive arthralgia patients and thirteen RA patients. Through multi-omics factor analysis we observed that the latent factor explaining the variance in gene expression and DNA methylation was associated with Swollen Joint Count 66 (SJC66), with patients with SJC66 of 9 or more displaying separation from the rest. Interrogating these observed differences revealed activation of the immune response as well as dysregulation of cell adhesion pathways at the level of both DNA methylation and gene expression. We observed differences for 59 genes in particular at the level of both transcript expression and DNA methylation. Our results highlight the utility of genome-wide multi-omics profiling of synovial samples for improved understanding of changes associated with disease spread in arthralgia and RA patients, and point to novel candidate targets for the treatment of the disease.

Keywords: rheumatoid arthritis, arthralgia, DNA methylation, transcriptomics, multi-omics analyses, synovial biopsies, target identification

INTRODUCTION

Rheumatoid arthritis (RA) is a complex, multifactorial, and chronic autoimmune disease that primarily affects the synovial tissue in joints (1). It affects about 1% of the population and manifests with significant unmet medical need (2). Investigating pathogenesis during different stages of disease, or across a spectrum of disease severity, is critical to optimize appropriate therapeutic interventions that affect disease progression (3, 4).

Diagnosis of established RA coincides with the development of painful and swollen joints, although circulating auto-antibodies can be detected up to 10 years before diagnosis (5, 6). Clinically manifested joint swelling reflects synovial tissue inflammation (synovitis), which is characterized by infiltration into the synovium of multiple immune cell types [including T cells, B cells, macrophages, plasma cells, dendritic cells, natural killer (NK) cells, and mast cells (7)], with up to 18 distinct infiltrating cell populations being reported on in a recent single cell transcriptomics analysis (8). As disease progresses, synovial fibroblasts adopt an increasingly aggressive and invasive phenotype, promoting further inflammation and joint damage together with other processes induced by the inflammatory environment, such as the differentiation of bone-resorbing osteoclasts (9, 10). Disease progression in early RA is often associated with the involvement of an increasing number of inflamed joints, but the mechanisms responsible for this spread of disease are poorly understood. Moreover, differences in the rate of disease manifestation and variability of response to therapy indicate that different pathophysiological mechanisms are implicated in disease development and progression compared to disease etiology (11).

An increasing body of evidence indicates that epigenetic modifications play an important role in the regulation of RA pathogenesis (12). Several array-based studies have reported widespread differences in DNA methylation among peripheral blood cells from RA patients and controls, suggesting that epigenetic modifications in circulating cells associate with disease (13, 14). However, wider conclusions may be limited by the unknown correlation of these effects to synovial cells directly at the site of inflammation. Epigenetic modifications have also been implicated in modulating the function of synovial fibroblasts in RA, through comparisons of DNA methylation patterns of cultured cells isolated from RA and osteoarthritis patients (15–20). Such studies have identified DNA methylation patterns that distinguish RA from other forms of arthritis, along with regulatory elements and biomarkers related to the pathological phenotype of RA. However, to identify novel candidate genes for therapeutic interventions that affect disease progression, it is important to study samples from patients across different stages of disease. Two such studies using cultured fibroblast-like synoviocytes found small but statistically significant patterns of hypomethylation in patients with longstanding RA compared to those with early RA, suggesting that the DNA methylome could be associated with the transformation of synovial fibroblasts into invasive cells capable of joint destruction and the resulting disease progression (21, 22).

Here, we gain insights into RA heterogeneity by combining whole-genome DNA methylome and transcriptome analyses from the same synovial biopsies and stratifying patients according to the number of swollen joints, a clinical parameter reflective of disease evolution. We find that swollen joint count based on 66 joints (SJC66), which reflects the amount and spread of inflamed synovial tissue, is associated with major changes in gene transcription and DNA methylation at promoters. By cross-interrogating differentially methylated genomic regions and their associated genes, we reveal novel candidate loci associated with the spread of the disease across joints.

MATERIALS AND METHODS

Sample Description

Synovial tissue was collected *via* a mini-arthroscopic procedure as described previously (23) from patients at the Amsterdam University Medical Center, University of Amsterdam. A total of 17 samples were obtained from three cohorts. The first cohort contained individuals that had either arthralgia and/or a positive family history for RA, but without arthritis (as determined by a clinician), and that were positive for IgM Rheumatoid Factor (IgM-RF) and/or Anti-Citrullinated Protein Antibodies (ACPA) (Pre-synoviomics; n = 4) (24). The second cohort consisted of individuals that at inclusion were Disease Modifying Anti-Rheumatic Drug (DMARD)-naïve with early arthritis, as defined by a disease duration of less than 1 year (Synoviomics; n = 9) (25, 26). The third cohort contained samples from patients with established RA on active treatment with a disease duration of more than one year who had at least one swollen joint suitable for synovial tissue sampling (Synoviomics II; n = 4) (27). For all analyses, samples were prepared simultaneously to mitigate batch effects.

All subjects provided written informed consent and the collection and use of the samples received Institutional Review Board review and approval. Characteristics of patients included in this study are listed in **Table S1**.

RNA-Sequencing and Whole Genome Bisulphite Sequencing

Flash-frozen synovial tissue biopsies were utilized to simultaneously isolate RNA and DNA using an AllPrep DNA/RNA Mini kit (Qiagen), with QIAshredder spin columns (Qiagen) used to disrupt the tissue. RNA samples were quantified and their integrity assessed using Qubit RNA Broad Range Assay Kit (Thermo Fisher Scientific) and an Agilent 2100 Bioanalyzer RNA 6000 Nano Kit (Agilent Technologies), respectively. Depending on sample yield, DNA samples were quantified using Qubit DNA BR or Qubit DNA HS kits (Thermo Fisher Scientific).

RNA-Seq libraries were generated from 150 ng of total RNA. The TruSeq® Stranded Total RNA LT was used with a Ribo-Zero™ Human/Mouse/Rat kit (Illumina), following the 'Low Sample' protocol except for two modifications. Firstly, the time of the 'Elution 2 – Frag – Prime' program was reduced from 8 to

6 min to increase the length of the RNA fragments. Secondly, 11 instead of 15 cycles were used to enrich the DNA fragments. Libraries were quantified with a KAPA Library Quantification Kit (KAPA Biosystems) on a QuantStudio 12 K Flex Real-Time PCR System (Thermo Fisher Scientific). Five samples were multiplexed per lane and libraries were clustered and sequenced using HiSeq® PE Cluster Kit v3 – cBot™ and HiSeq® SBS Kit v3 kits (Illumina). Paired-end sequencing (2 x 76 bp) was performed using a HiSeq 1500 (Illumina) to a depth of ~40 M reads per sample.

Whole Genome Bisulphite Sequencing (WGBS) libraries were generated using an EpiGnome Methyl-Seq Library Preparation Kit (Epicentre, now Illumina) from 100 ng of sample DNA. Bisulphite conversion was performed using an EZ DNA Methylation-Lightning Kit (Zymo Research). Each bisulphite conversion reaction contained 500 pg of unmethylated lambda DNA (Promega), which was used as a control to verify that bisulphite conversion efficiencies were at least 98%. Libraries were quantified using a KAPA Library Quantification Kit (KAPA Biosystems) on a 7900HT Real-Time PCR System (Thermo Fisher Scientific). Six samples were multiplexed per lane and libraries were clustered and sequenced using HiSeq® PE Cluster Kit v4 – cBot™ and HiSeq® SBS Kit v4 kits (Illumina). Paired-end sequencing (2 x 125 bp) was performed using a HiSeq 1500 (Illumina) to a depth of ~600M reads per sample. To provide sufficient coverage each batch was sequenced over 2 high output runs.

Exploratory Analyses of DNA Methylation and Gene Expression

To concurrently explore the DNA methylation and gene expression data, Multi-Omics Factor Analysis (MOFA) was applied (v1.0.0) (28). In short, MOFA performs unsupervised dimensionality reduction simultaneously across multiple data modalities from the same sample through a small number of inferred latent factors, enabling the detection of co-ordinated changes between the different data modalities (28). Here, we used the 5,000 most variable CpGs and genes as input and with the number of latent factors set to 9, the tolerance to 0.1 and the factor threshold to 0.02.

Gene Expression Data Analysis

Quality assessment of the raw reads was performed using FastQC (v0.11.2) (29). Adapters were removed from reads using Scythe (v0.991) (30) and sequences were quality-trimmed using Sickle (v1.33) (31) using a quality threshold of 20. Alignment to the GRCh38 build of the human genome was performed using Kallisto (v0.44.0) (32).

All subsequent analyses were performed in R (v3.5.0) (33). Gene-level counts were generated from the transcript abundances using tximport (v1.12.0) (34). Allosome-associated genes were removed to mitigate obvious sex effects. Differential Gene Expression (DGE) analyses were conducted using DESeq2 (v1.22.2) (35) where $SJC66_{high}$ was compared with $SJC66_{low}$ whilst correcting for sex, age and DMARD usage using the following design formula:

$$Gene\ Expression \sim sex + age + DMARD\ usage + SJC66_{dichotomized}$$

DMARD usage was a binary variable defined by the usage of any medication: conventional DMARDs (cDMARD, including methotrexate) or biological DMARDs (bDMARD). As Kallisto provided abundance levels for individual transcripts, Gene Differential Expression (GDE) analyses were also conducted to identify genes where particular transcripts were differentially expressed (36). In short, differential expression analyses was performed using DESeq2 and the resulting *p*-values were combined using the Lancaster aggregation method found in aggregation (v1.0.1) (36), where observations were weighted by the base expression. DGEs and GDEs were defined as genes with a false discovery rate (FDR)-adjusted *p*-value less than 0.05.

DNA Methylation Data Analysis

Quality assessment of the raw reads was performed using FastQC (v0.11.2) (29). Adapter and quality trimming was performed using Skewer (v0.1.123) (37) and a quality filter of 20. To assess bisulphite conversion rates, Bismark (v0.14.1) (38) was used to align the reads to the genome of the phage lambda, and again for alignment to the GRCh38 build of the human genome. Post-alignment filtering of unmapped reads, reads aligning at multiple locations and reads with a mapping score lower than 10 was carried out using SAMtools (39).

All subsequent analyses were performed in R (v3.5.0). CpG loci located on the allosomes were removed to mitigate the sex effect. The differential methylation analyses were performed using dmrseq (v1.2.5) (40), where we contrasted $SJC66_{high}$ with $SJC66_{low}$ whilst correcting for sex, age and DMARD usage using the following design formula:

$$Methylation \sim sex + age + DMARD\ usage + SJC66_{dichotomized}$$

Differentially Methylated Regions (DMRs) were annotated using ChipPeakAnno (v3.16.1) (41) to genes if the DMR was located within 5,000 bp upstream or 1,000 bp downstream of the gene as obtained from Gencode (v29) (42).

Integrated DNA Methylation and Gene Expression Analysis

Integrated analyses were based on the DMRs and GDEs found through the separate DNA methylation and gene expression analyses. The overlap between DMRs and GDEs were called Genes displaying both Differential Expression and Methylation (GDEMs). For each GDEM the median percentage methylation was calculated for all constituent CpGs per sample and correlated with the $\log_2$ transformed expression counts to obtain the Pearson correlation coefficient. Confidence intervals (95%) were calculated through 1000 bootstraps for each GDEM. In short, 17 samples were randomly drawn from the original samples with replacement, whereupon the Pearson correlation coefficient was calculated. This process was repeated 1000 times to generate the empirical distribution function, which was then used to estimate the confidence intervals (43). The aforementioned bootstrapping approach was performed using the boot (v1.3) package (44). For inferential purposes, *p*-values

were calculated by means of a permutation approach. In short, per GDEM, 1000 sets of consecutive CpGs equal to the length of the observed DMR were sampled and correlated with the gene expression signal as described above, after which the proportion of correlation coefficients higher than the observed correlation coefficient was calculated yielding the *p*-value.

Functional Enrichment, Cell Type Enrichment and Protein-Protein Interaction Network Analyses

Gene set enrichment analyses were performed using fgsea (v1.8.0) (45) using the Metabase pathways terms (46) as reference. Metabase pathway terms with an FDR-adjusted *p*-value less than 0.05 were considered significant.

Cell proportions were imputed using xCell (v1.1.0) (47), where transcripts per million were used to estimate the proportion of each of the 64 immune and stromal cell types. Subsequent linear regressions were performed to calculate the *p*-values to assess statistical significance. Again, we compared $SJC66_{high}$ with $SJC66_{low}$ also correcting for age, sex and DMARD usage.

Protein-protein interaction (PPI) network analyses were performed using the STRING (v11) database (48), in order to identify whether a set of genes was over-represented for interactions. In short, the PPI analysis returned networks of genes where the encoded proteins interacted, co-expressed or co-evolved with one another, based on text mining, curated databases, and experimental data.

Data Availability

The datasets generated and analyzed for this study can be found in the ArrayExpress repository under accession number E-MTAB-6638 and E-MTAB-6684 for WGBS and RNA-Seq, respectively. All code is hosted on GitHub at https://github.com/enricoferrero/BTCURE.

RESULTS

Swollen Joint Count Is Associated With the Latent Factor Explaining Variance in Gene Expression and DNA Methylation

We profiled the DNA methylome and transcriptome of 17 synovial tissue samples (**Table S1**) using whole genome bisulphite sequencing (WGBS) and RNA-sequencing (RNA-Seq). We initially attempted to link gene expression changes to disease duration as well as to cross-patient variations in the Disease Activity Score of 28 joints (DAS28) variables, but these analyses resulted in weak and non-biologically relevant signals, and were ultimately deemed inconclusive for this set of samples (data not shown). Principal Component Analysis (PCA) indicated that variation in both DNA methylation and gene expression were independently correlated with the swollen joint count in 66 joints (SJC66; **Figures 1A, B**). Comparison of the first principal component of both the DNA methylation and gene expression data suggested agreement with samples 33, 19, 3, 12A and 25 broadly separating from the other samples for both modalities (**Figure 1C**). While sample 33 appeared to be an outlier based on the DNA methylation data the removal thereof did not alter the correlation substantially (**Figure S1**). To further explore DNA methylation and gene expression at a genome-wide level in an integrative fashion, we performed variance decomposition using multi-omics factor analysis (MOFA) (28). MOFA infers a set of latent factors that capture sources of variability across different measured -omic modalities of the same samples. We found that most variance (approx. 70%) was better explained by gene expression as compared to DNA methylation (**Figure 1D**). Further decomposition of the variance identified 8 latent factors, with LF1 explaining 40% of the variance in gene expression, whereas variance in methylation was more evenly distributed amongst all eight latent factors. Focusing on LF1, we observed a marked separation between samples with SJC66 of 9 and above compared to those with SJC66 of 8 or less (**Figures 1E** and **S2A, B**). While most samples were obtained from the knee, two were obtained from the ankle. Interrogation of the first latent factor did not indicate any correlation with the source of the sample (**Figure S2C**). Importantly, sex was also strongly associated with LF1 (*p*-value = 8.8E-03; **Figure S2D**) where samples with SJC66 of 9 and above were mostly males and most samples with SJC66 equal to or fewer than 8 were mostly females. Therefore, subsequent comparative analyses accounted for the imbalance in sex by removal of allosome-associated genes as well as including sex as covariate in linear models. females. Therefore, subsequent comparative analyses accounted for the imbalance in sex by removal of allosome-associated genes as well as including sex as covariate in linear models.

Differences in DNA Methylation and Gene Expression in Patients With High and Lower Numbers of Swollen Joints Are Associated With Immune Response and Cell Adhesion Pathways

Having observed genome-wide differences through exploratory analyses in both DNA methylation and gene expression data that associated with SJC66, we next investigated which regions and genes were differentially methylated and expressed. At this point we investigated DNA methylation and gene expression separately. While samples with SJC66 = 0 were included for the aforementioned exploratory analyses, they were excluded from subsequent comparative analyses as they were medically not a homogeneous group consisting of very early arthritis as well as late arthritis with no swelling following medication. The swollen samples were stratified according to the separation observed in the exploratory analyses (**Figure 1E**) and we compared $SJC66_{high}$ (SJC66 ≥ 9) with $SJC66_{low}$ (SJC66 < 8) correcting for age, sex and DMARD usage. Comparative methylation analysis identified 3,536 DMRs (**Figure 2A** and **Table S2**), where 2,140 were hypomethylated and 1,396 were hypermethylated. The majority of DMRs were located within 1 Kb of a TSS or in distal regions (**Figure 2B**). Notably, the most statistically significant DMRs spanned regions larger than 10 Kb (**Table S2**).

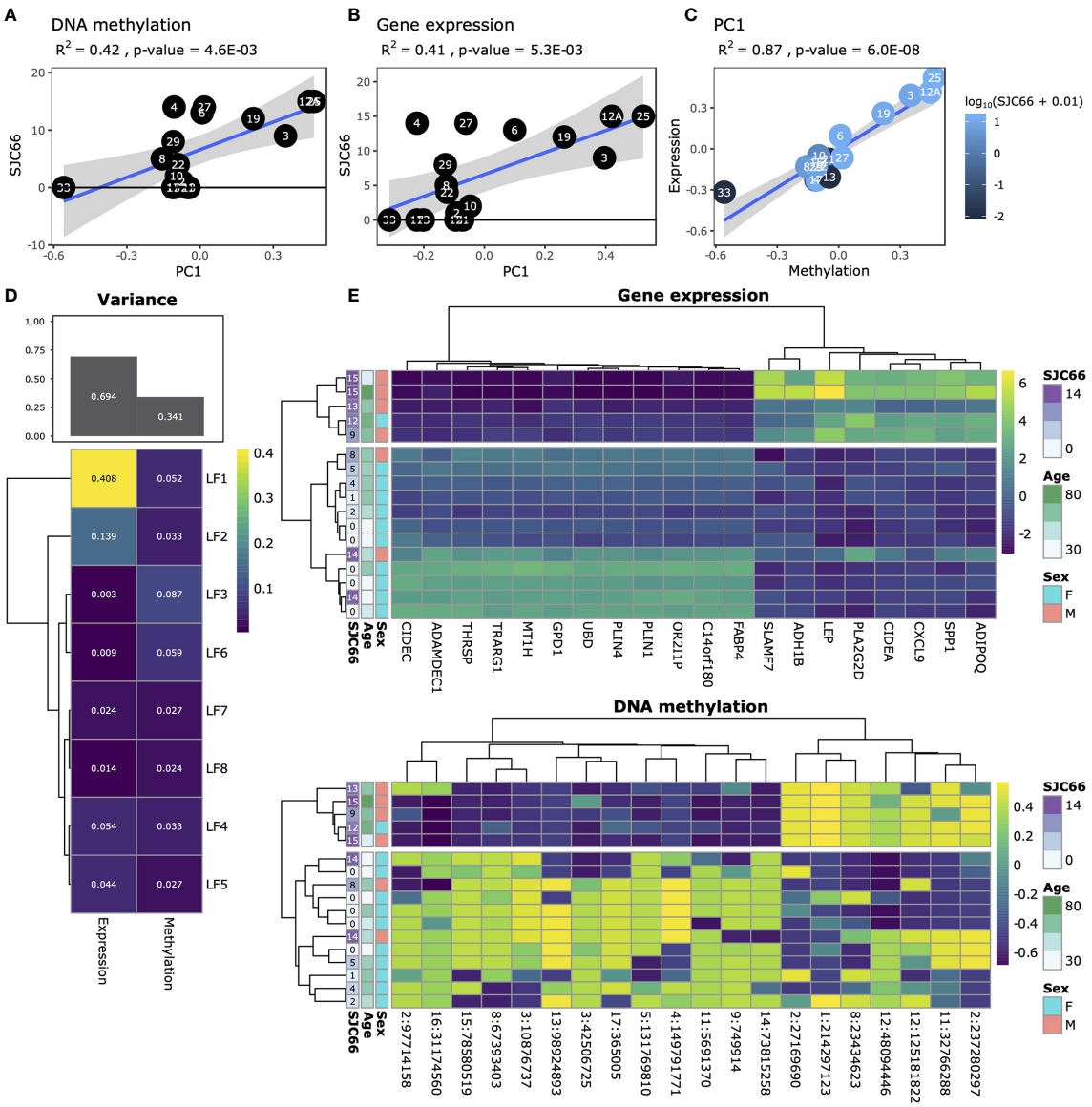


FIGURE 1 | Exploratory analysis. Analysis of SJC66 regressed onto the first principal component of **(A)** DNA methylation and **(B)** gene expression annotated with the Pearson r-squared (R^2) and the *p*-value. **(C)** Comparison of the first principal components for DNA methylation on the x-axis and gene expression together coloured for SJC66. Trendlines represent the mean and 95% confidence intervals. **(D)** Results obtained from multi-omics factor analysis, where the figure at the top depicts a bar chart representing the percentage variance explained by gene expression and DNA methylation, which has been decomposed into the separate latent factors (LF) at the bottom. **(E)** Decomposition of the variance observed for LF1 into the loadings assigned by MOFA to the top 20 weighted features for gene expression (top; colours represent $\log_2$(count)) and DNA methylation (bottom; colours represent percentage methylation).

For example, the top two DMRs were located within the promoter regions of the *MIRLET7B* and *MIR3619* genes (*MIRLET7BHG*; 22:46,072,724-46,090,732; **Figure 2C**) and microRNA 10B (*MIR10B*; 2:176,147,225-176,162,267; **Figure 2D**) and exceeded 15 Kb in length. While the *MIRLET7BHG*-associated DMR was hypermethylated, the *MIR10B*-associated DMR was hypomethylated when comparing $SJC66_{high}$ with $SJC66_{low}$. Functional analysis of the DMRs identified several pathways with evidence of differential methylation. Among the top 10, we observed that the NGF/TrkA MAPK pathway, immune response pathways including antigen presentation by MHC class II, macrophage and dendritic cell phenotype shifts, as well as CCR3 signalling in eosinophils, were hypermethylated (**Figure 2E** and **Table S3**).

Comparing $SJC66_{high}$ with $SJC66_{low}$ at the gene expression level identified 142 DGEs, of which 106 up-regulated and 36 down-regulated (**Figure 3A** and **Table S4**). The top 2 DGEs, chemokine ligand 13 (*CCL13*) (**Figure 3B**) and C-Type Lectin Domain Containing 10A (*CLEC10A*) (**Figure 3C**), were both found to be more highly expressed in the $SJC66_{high}$ samples.

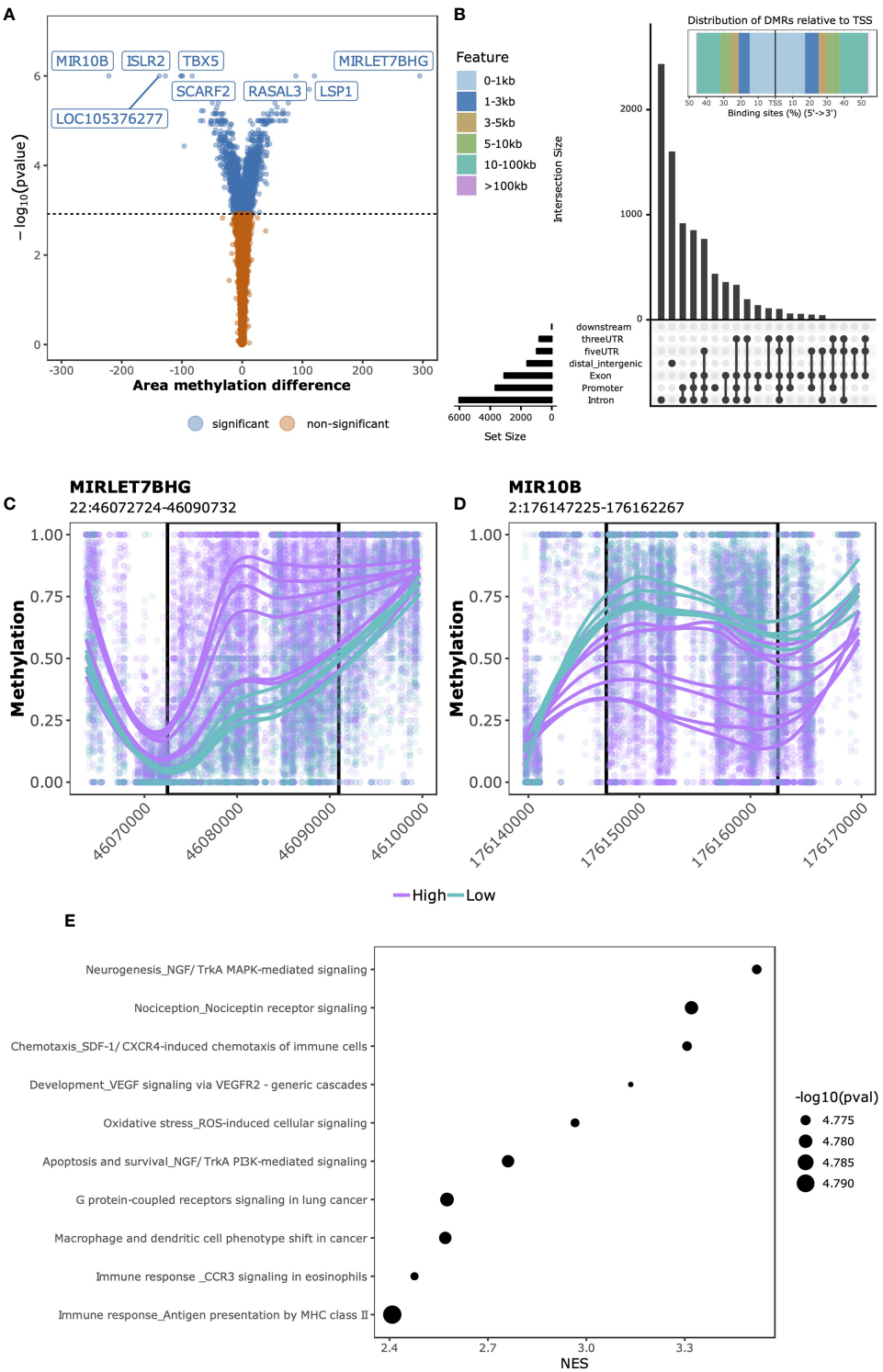


FIGURE 2 | $SJC66_{high}$ versus $SJC66_{low}$ differential methylation analysis. **(A)** Volcano plots depicting the $-\log_{10}$(*p*-value) on the y-axis relative to the difference in % methylation when comparing the $SJC66_{high}$ with $SJC66_{low}$ samples on the x-axis. Colours represent the non-significant (orange) and significant regions (blue). Statistical significance is further depicted by a dashed horizontal line. **(B)** At the top: distribution of the DMRs relative to the transcription start site (TSS) and at the bottom: upset plot indicating the distribution of the DMRs across the different genetic features. **(C, D)** DMRs found at genomic co-ordinates **(C)** 22:46,072,724-46,090,732 and **(D)** 2:176,147,225-176,162,267 are located in the promoters of MIRLET7BHG and MIR10B, respectively. DNA methylation is visualized as the percentage methylation (y-axis) with a smoothed trendline per sample. **(E)** The 10 most enriched MetaCore gene sets associated to the promoter-bound DMRs. Tick marks represent gene ranks relative to the direction of the methylation effect and is summarized by the normalized expression score (NES).

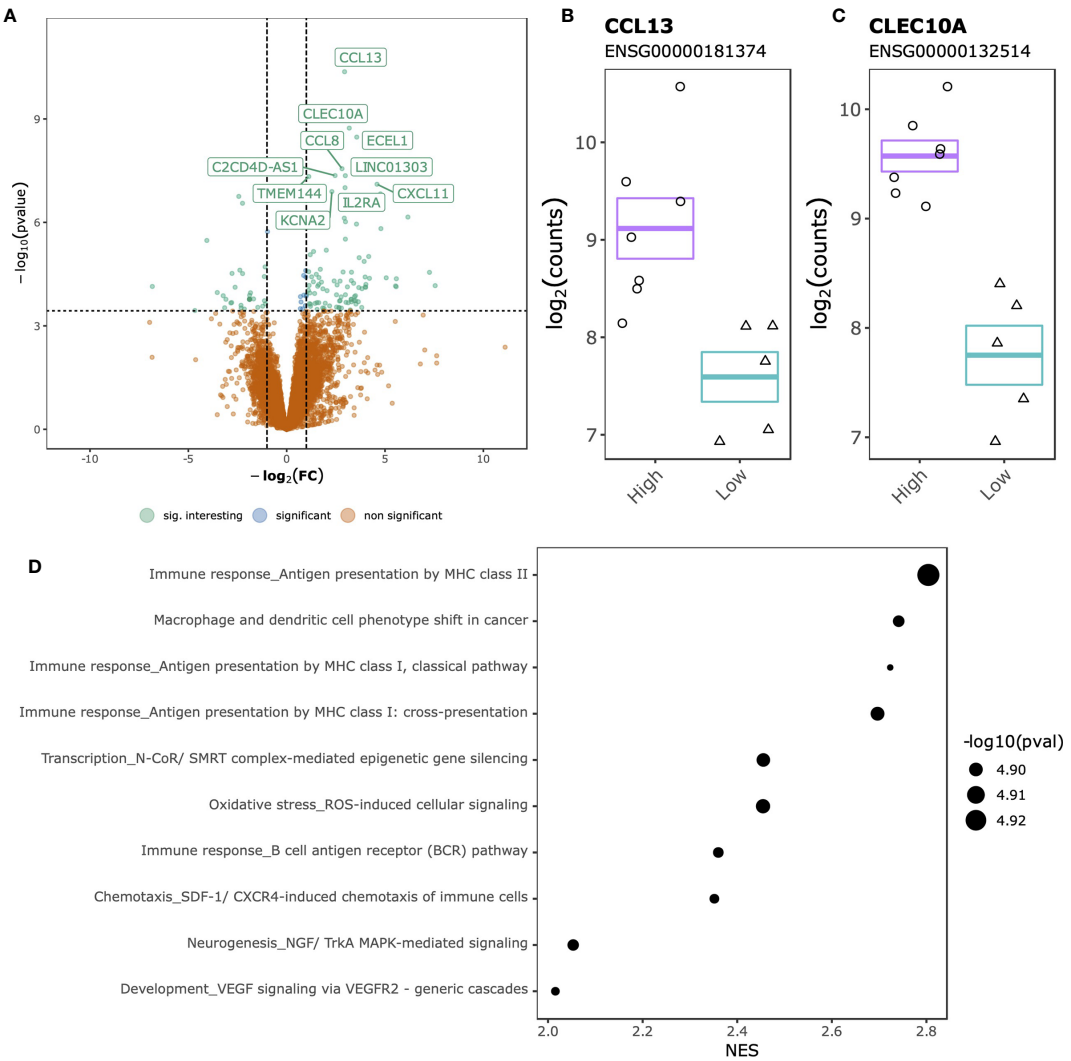


FIGURE 3 | SJC66$_{high}$ versus SJC66$_{low}$ differential expression analysis. **(A)** Volcano plots depicting the -log$_{10}$(*p*-value) on the y-axis relative to the difference in log$_2$ fold change when comparing the SJC66$_{high}$ with SJC66$_{low}$ samples on the x-axis. Colours represent the non-significant genes (orange), the significant genes (blue) and the significant genes with a log$_2$ fold change of larger than 1 (green; "sig. interesting"). Statistical significance is further depicted by a dashed horizontal line. Mean and standard error bars superposed onto a jitterplot for the two most differentially expressed genes **(B)** CCL13 and **(C)** CLEC10A. **(D)** The 10 most enriched MetaCore gene sets associated to the DGEs. Tick marks represent gene ranks relative to the direction of the expression effect and are summarized by the normalized expression score (NES).

Functional analyses revealed a striking similarity with the differential methylation results, where genes associated with antigen presentation by MHC class II as well as the NGF/TrkA MAPK pathways were overexpressed (**Figure 3D** and **Table S5**).

Transcripts Associated With Differentially Methylated and Expressed Genes Display Concordant Expression and Identify Nodal Points for Key Interactions

Having observed pathway-concordant differences in DNA methylation and gene expression, we were interested in identifying specific genes that were both differentially methylated and expressed. Although DNA methylation has traditionally been associated with gene-level expression, emerging evidence shows that it also regulates alternative splicing (49–51). Therefore, we complemented our Differential Gene Expression (DGE) analysis, which masks transcript-level dynamics, with Gene Differential Expression (GDE) analysis, which identifies genes that display transcript-level differences, by combining the *p*-values of individual transcripts associated with a single gene (36). We identified 290 genes that displayed perturbations in the expression of their transcripts (GDEs; **Table S6**). Combining the GDEs with the DMRs yielded 97 unique DMRs associated with 59 unique genes, which we termed Genes Differentially Expressed and Methylated (GDEMs) (**Figures 4A, B**). We observed that most transcripts, those

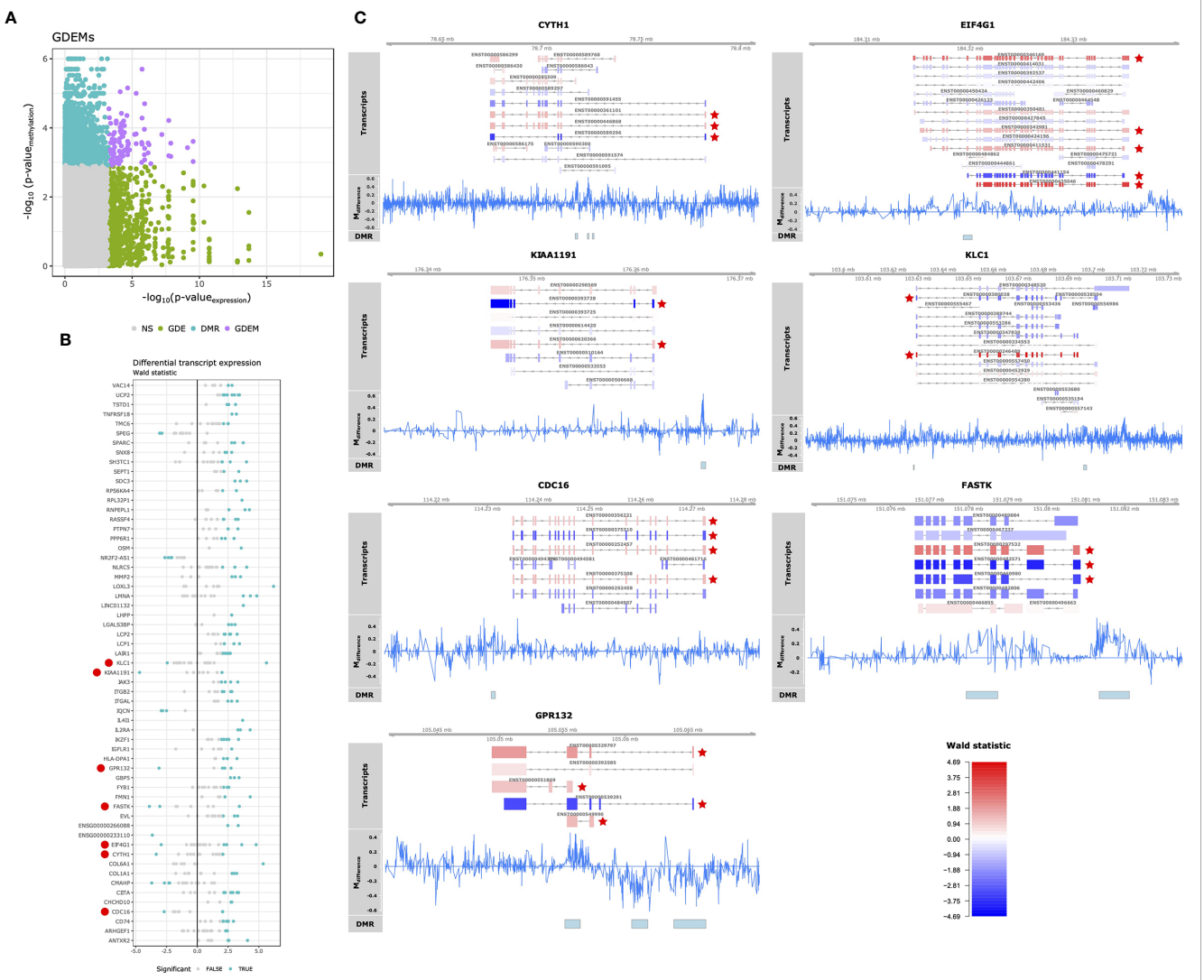


FIGURE 4 | Differential methylation and alternative splicing. **(A)** Comparison of the -$\log_{10}$(*p*-values) obtained from gene differential expression analysis and differential methylation analysis on the x- and y-axis, respectively. Grey, green, blue and purple represent the genes: unchanged for methylation and gene expression ("NS"), differentially expressed ("GDE"), differentially methylated ("DMR"), differentially methylated and expressed ("GDEM"), respectively. **(B)** Visualization of the Wald statistic (a statistic calculated by DESeq2 representing the effect size relative to the variance) calculated for the individual transcripts (represented in dots) associated with each of the GDEMs. Grey and blue dots represent non-significant and significantly differentially expressed transcripts, respectively. Large red circles left of the gene symbols indicate genes with transcripts that display opposite effect sizes. **(C)** Summary visualization of the DNA methylation and gene expression signals for GDEMs: CYTH1, EIF4G11, KIAA1191, KLC1, CDC18, FASTK and GPR132. The "Transcripts" track represents individual transcripts coloured by the Wald statistic. Stars in red indicate differentially expressed transcripts. The "$M_{difference}$" track represents the difference in percentage methylation when comparing $SJC66_{high}$ with $SJC66_{low}$. The "DMR" track represents the locations of the DMRs as found by dmrseq.

associated with 52 GDEMs, were concordantly differentially expressed. Transcripts of the remaining 7 GDEMs were found to display opposite directions of expression, namely Cytohesin 1 (*CYTH1*), Eukaryotic Translation Initiation Factor 4 Gamma 1 (*EIF4G1*), a putative monooxygenase (*KIAA1191*), Kinesin Light Chain 1 (*KLC1*), Cell Division Cycle 16 (*CDC16*), Fas-Activated Serine/Threonine Kinase (*FASTK*) and G protein coupled receptor 132 (*GPR132;* **Figure 4B**). Investigation of the methylation status of the gene overlaid onto the transcript location indicated that the observed DMRs for *KIAA1191*, *KLC1*, *CDC16*, *FASTK* and *GPR132* were located in the promoter shared by most transcripts (**Figure 4C**). The mechanisms that could cause opposite direction of expression in these genes are currently unclear although studies in Arabidopsis have identified a methylation reader complex that can enhance rather than suppress gene transcription in the presence of methylation (52). Our work therefore supports the rationale for further validation and mechanistic studies.

Functionally, we observed that the 59 GDEMs were primarily over-represented for immune response-associated pathways,

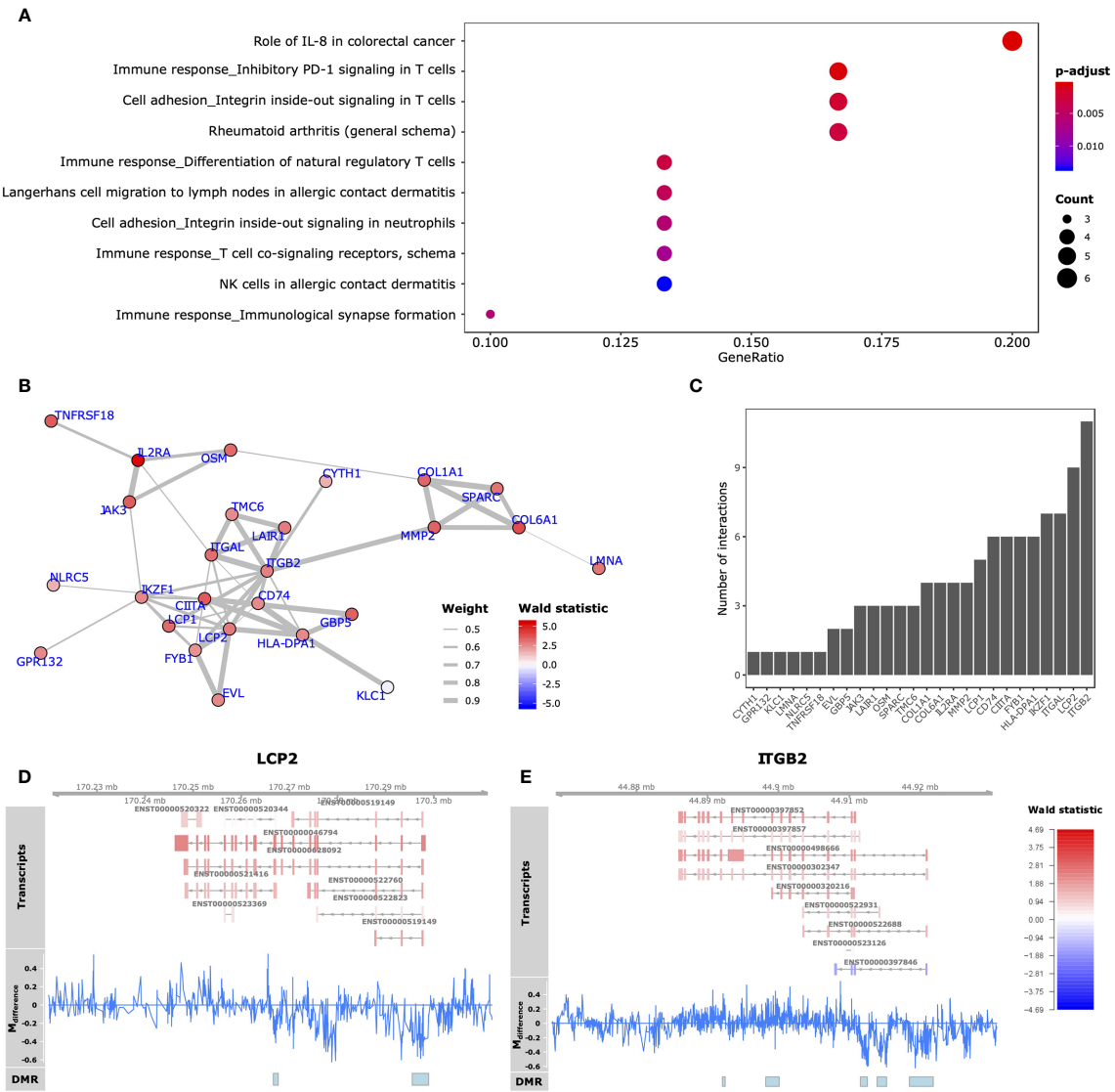


FIGURE 5 | Functional analyses GDEMs. **(A)** The 10 most overrepresented MetaCore gene sets found for the GDEMs depicting the ratio of pathway-associated GDEMs relative to the total number of pathway-associated genes. The size and colour intensity are proportional to the number of pathway-associated GDEMs and the -$\log_{10}$(adjusted p-values), respectively. **(B)** GDEMs were analysed for known interactions by querying the STRING database. Represented here is the protein-protein interaction network, where the thickness of the edge is proportional to the STRINGdb evidence score and the colour proportional to the Wald statistic, respectively. **(C)** Bargraph representing the number of interactions per protein. Summary visualization of the DNA methylation and gene expression signals for GDEMs: **(D)** LCP2, **(E)** ITGB2. The "Transcripts" track represents individual transcripts coloured by the Wald statistic. The "$M_{difference}$" track represents the difference in percentage methylation when comparing $SJC66_{high}$ with SJC66low. The "DMR" track represents the locations of the DMRs as found by dmrseq.

specifically T-lymphocyte associated pathways (**Figure 5A**). We interrogated which of the GDEMs encoded known interaction partners by querying the STRING database for documented interactions (48). Almost half (46%) of the GDEMs encoded for interacting proteins with the most connected GDEMs being *ITGB2* (11 interactions) and *LCP2* (9 interactions) (**Figure 5C**). Both *ITGB2* and *LCP2* were differentially methylated at multiple locations, with the largest visible differences occurring within the TSS and downstream thereof (**Figures 5D, E**). Transcript-wise, *ITGB2*-transcripts ENST00000498666 and ENST00000397852 as well as *LCP2* transcript ENST00000046794 were most differentially expressed. Expression Quantitative Trait Methylation (eQTM) analyses confirmed strong inverse correlations between the differences in methylation in the promoter regions of *ITGB2* (21:44918461-44921815) and *LCP2* (5:170295513-170298924) with transcripts ENST00000498666 (r = -0.9; *p*-value = 1E-04) and ENST00000046794 (r = -0.9; *p*-value = 1E-04), respectively (**Table 1**). While the association between the *ITGB2* promoter DMR and ENST00000397852 expression was non-significant (*p*-value = 0.2353), the correlation

TABLE 1 | Differentially methylated and expressed genes.

Gene	$GDE_{p\text{-value}}$	DMR	DMR_{stat}	$DMR_{p\text{-value}}$	EnsT	$DTE_{p\text{-value}}$	$eQTM_r$	$eQTM_{p\text{-value}}$
ENSG00000233110	2.53E-04	4:185586913-185587794	-9.94	2.78E-04	ENST00000411847.1	2.53E-04	0.88 [0.74;0.95]	2.54E-02
CD74	2.00E-04	5:150410549-150413278	-9.48	3.59E-04	ENST00000009530.11	1.77E-02	-0.88 [-0.96;-0.7]	1.00E-04
CD74	2.00E-04	5:150410549-150413278	-9.48	3.59E-04	ENST00000518797.5	1.23E-02	-0.9 [-0.96;-0.76]	1.00E-04
CIITA	2.93E-10	16:10877103-10878257	-10.23	2.47E-04	ENST00000637439.1	5.06E-03	-0.92 [-0.96;-0.78]	1.00E-04
CIITA	2.93E-10	16:10877103-10878257	-10.23	2.47E-04	ENST00000637439.1	5.06E-03	-0.92 [-0.96;-0.78]	1.00E-04
CIITA	2.93E-10	16:10877103-10878257	-10.23	2.47E-04	ENST00000575513.1	8.19E-04	-0.88 [-0.96;-0.66]	8.10E-03
CIITA	2.93E-10	16:10877103-10878257	-10.23	2.47E-04	ENST00000575513.1	8.19E-04	-0.88 [-0.96;-0.66]	8.10E-03
CIITA	2.93E-10	16:10877103-10878257	-10.23	2.47E-04	ENST00000572665.1	6.67E-03	-0.86 [-0.94;-0.62]	8.80E-03
CIITA	2.93E-10	16:10877103-10878257	-10.23	2.47E-04	ENST00000572665.1	6.67E-03	-0.86 [-0.94;-0.62]	8.80E-03
CIITA	2.93E-10	16:10877103-10878257	-10.23	2.47E-04	ENST00000576601.1	3.12E-02	-0.89 [-0.96;-0.67]	2.51E-02
CIITA	2.93E-10	16:10877103-10878257	-10.23	2.47E-04	ENST00000576601.1	3.12E-02	-0.89 [-0.96;-0.67]	2.51E-02
CIITA	2.93E-10	16:10877103-10878257	-10.23	2.47E-04	ENST00000381835.9	2.59E-02	-0.86 [-0.95;-0.52]	4.06E-02
CIITA	2.93E-10	16:10877103-10878257	-10.23	2.47E-04	ENST00000381835.9	2.59E-02	-0.86 [-0.95;-0.52]	4.06E-02
CMAHP	1.89E-04	6:25232404-25233193	10.46	2.24E-04	ENST00000377993.8	2.12E-04	-0.7 [-0.87;-0.25]	1.00E-04
CMAHP	1.89E-04	6:25232404-25233193	10.46	2.24E-04	ENST00000493257.5	2.08E-02	-0.68 [-0.84;-0.28]	1.00E-04
COL6A1	1.18E-06	21:45978416-45978817	8.14	8.44E-04	ENST00000361866.7	8.83E-08	0.58 [0.07;0.8]	2.20E-03
COL6A1	1.18E-06	21:45983116-45984355	15.46	2.00E-05	ENST00000361866.7	8.83E-08	0.45 [-0.19;0.78]	1.80E-03
CYTH1	1.53E-04	17:78716750-78718002	12.83	7.40E-05	ENST00000589296.5	8.32E-04	-0.42 [-0.78;0.23]	4.12E-02
EIF4G1	4.11E-07	3:184319325-184320192	7.83	1.09E-03	ENST00000342981.8	2.97E-02	0.32 [-0.28;0.72]	1.00E-04
EIF4G1	4.11E-07	3:184319325-184320192	7.83	1.09E-03	ENST00000342981.8	2.97E-02	0.32 [-0.28;0.72]	1.00E-04
EIF4G1	4.11E-07	3:184319325-184320192	7.83	1.09E-03	ENST00000346169.6	3.07E-04	0.68 [0.28;0.9]	1.00E-04
EIF4G1	4.11E-07	3:184319325-184320192	7.83	1.09E-03	ENST00000346169.6	3.07E-04	0.68 [0.28;0.9]	1.00E-04
EIF4G1	4.11E-07	3:184319325-184320192	7.83	1.09E-03	ENST00000411531.5	1.71E-02	0.7 [0.33;0.9]	1.00E-04
EIF4G1	4.11E-07	3:184319325-184320192	7.83	1.09E-03	ENST00000411531.5	1.71E-02	0.7 [0.33;0.9]	1.00E-04
EIF4G1	4.11E-07	3:184319325-184320192	7.83	1.09E-03	ENST00000435046.6	1.78E-06	0.53 [-0.05;0.8]	1.00E-04
EIF4G1	4.11E-07	3:184319325-184320192	7.83	1.09E-03	ENST00000435046.6	1.78E-06	0.53 [-0.05;0.8]	1.00E-04
EVL	2.46E-05	14:100132409-100133139	9.08	4.52E-04	ENST00000556921.1	3.54E-02	0.89 [0.73;0.97]	3.04E-02
FASTK	1.23E-06	7:151081348-151082120	11.88	1.10E-04	ENST00000460980.5	2.32E-03	-0.21 [-0.76;0.52]	1.00E-04
FMN1	1.83E-04	15:33154437-33154950	-7.96	9.88E-04	ENST00000320930.7	2.49E-02	-0.82 [-0.94;-0.59]	1.26E-02
GPR132	3.80E-05	14:105055190-105056414	9.09	4.48E-04	ENST00000329797.7	6.12E-03	0.9 [0.62;0.99]	1.00E-04
GPR132	3.80E-05	14:105063870-105066425	-11.52	1.36E-04	ENST00000549990.1	2.60E-02	-0.38 [-0.72;0.44]	3.57E-02
IKZF1	6.58E-06	7:50305276-50312367	-11.26	1.55E-04	ENST00000471793.1	1.04E-02	-0.8 [-0.92;-0.33]	1.00E-04
IKZF1	6.58E-06	7:50305276-50312367	-11.26	1.55E-04	ENST00000492119.1	3.29E-02	-0.88 [-0.94;-0.7]	1.00E-04
IKZF1	6.58E-06	7:50305276-50312367	-11.26	1.55E-04	ENST00000413698.5	1.30E-02	-0.9 [-0.95;-0.72]	2.55E-02
IL4I1	2.32E-04	19:49895253-49897628	-7.97	9.82E-04	ENST00000391826.6	2.32E-04	-0.96 [-0.98;-0.74]	1.00E-04
ITGB2	7.75E-05	21:44898213-44900097	8.96	4.89E-04	ENST00000498666.5	5.07E-03	0.91 [0.67;0.98]	1.00E-04
ITGB2	7.75E-05	21:44911555-44912533	-11.65	1.24E-04	ENST00000397852.5	5.15E-03	-0.94 [-0.98;-0.72]	1.00E-04
ITGB2	7.75E-05	21:44911555-44912533	-11.65	1.24E-04	ENST00000498666.5	5.07E-03	-0.93 [-0.98;-0.81]	2.00E-03
ITGB2	7.75E-05	21:44911555-44912533	-11.65	1.24E-04	ENST00000522688.5	3.04E-02	-0.9 [-0.97;-0.63]	3.80E-03
ITGB2	7.75E-05	21:44911555-44912533	-11.65	1.24E-04	ENST00000320216.10	6.78E-03	-0.79 [-0.93;-0.57]	4.08E-02
ITGB2	7.75E-05	21:44911555-44912533	-11.65	1.24E-04	ENST00000302347.9	4.31E-02	-0.9 [-0.97;-0.62]	4.71E-02
ITGB2	7.75E-05	21:44913899-44915246	-12.34	9.60E-05	ENST00000397852.5	5.15E-03	-0.96 [-0.98;-0.88]	1.00E-04
ITGB2	7.75E-05	21:44913899-44915246	-12.34	9.60E-05	ENST00000397852.5	5.15E-03	-0.96 [-0.98;-0.88]	1.00E-04
ITGB2	7.75E-05	21:44913899-44915246	-12.34	9.60E-05	ENST00000498666.5	5.07E-03	-0.92 [-0.98;-0.79]	1.14E-02
ITGB2	7.75E-05	21:44913899-44915246	-12.34	9.60E-05	ENST00000498666.5	5.07E-03	-0.92 [-0.98;-0.79]	1.14E-02
ITGB2	7.75E-05	21:44913899-44915246	-12.34	9.60E-05	ENST00000302347.9	4.31E-02	-0.92 [-0.97;-0.72]	2.00E-02
ITGB2	7.75E-05	21:44913899-44915246	-12.34	9.60E-05	ENST00000302347.9	4.31E-02	-0.92 [-0.97;-0.72]	2.00E-02
ITGB2	7.75E-05	21:44913899-44915246	-12.34	9.60E-05	ENST00000522688.5	3.04E-02	-0.87 [-0.96;-0.61]	3.08E-02
ITGB2	7.75E-05	21:44913899-44915246	-12.34	9.60E-05	ENST00000522688.5	3.04E-02	-0.87 [-0.96;-0.61]	3.08E-02
ITGB2	7.75E-05	21:44913899-44915246	-12.34	9.60E-05	ENST00000320216.10	6.78E-03	-0.79 [-0.95;-0.45]	3.97E-02
ITGB2	7.75E-05	21:44913899-44915246	-12.34	9.60E-05	ENST00000320216.10	6.78E-03	-0.79 [-0.95;-0.45]	3.97E-02
ITGB2	7.75E-05	21:44918461-44921815	-18.22	9.00E-06	ENST00000498666.5	5.07E-03	-0.9 [-0.98;-0.55]	1.00E-04
ITGB2	7.75E-05	21:44918461-44921815	-18.22	9.00E-06	ENST00000498666.5	5.07E-03	-0.9 [-0.98;-0.55]	1.00E-04
ITGB2	7.75E-05	21:44918461-44921815	-18.22	9.00E-06	ENST00000320216.10	6.78E-03	-0.85 [-0.93;-0.69]	3.35E-02
ITGB2	7.75E-05	21:44918461-44921815	-18.22	9.00E-06	ENST00000320216.10	6.78E-03	-0.85 [-0.93;-0.69]	3.35E-02
KLC1	7.08E-05	14:103696092-103697307	8.07	9.05E-04	ENST00000246489.11	1.93E-08	0.62 [0.28;0.84]	2.73E-02
LAIR1	6.55E-07	19:54375122-54375966	-8.84	5.29E-04	ENST00000440716.5	9.38E-03	-0.93 [-0.98;-0.73]	3.80E-03
LCP2	1.90E-08	5:170295513-170298924	-13.19	6.10E-05	ENST00000046794.9	1.23E-03	-0.9 [-0.97;-0.66]	1.00E-04
LCP2	1.90E-08	5:170295513-170298924	-13.19	6.10E-05	ENST00000519149.1	8.96E-03	-0.82 [-0.9;-0.62]	1.00E-04
LCP2	1.90E-08	5:170295513-170298924	-13.19	6.10E-05	ENST00000628092.2	6.87E-03	-0.92 [-0.97;-0.74]	1.00E-04
LCP2	1.90E-08	5:170295513-170298924	-13.19	6.10E-05	ENST00000519594.5	2.89E-02	-0.87 [-0.93;-0.72]	3.70E-03
LCP2	1.90E-08	5:170295513-170298924	-13.19	6.10E-05	ENST00000522760.5	7.55E-03	-0.93 [-0.97;-0.8]	8.10E-03
OSM	4.92E-04	22:30266011-30268071	-13.63	5.10E-05	ENST00000215781.2	3.59E-04	-0.68 [-0.95;-0.13]	1.00E-04

(Continued)

TABLE 1 | Continued

Gene	$GDE_{p\text{-}value}$	DMR	DMR_{stat}	$DMR_{p\text{-}value}$	EnsT	$DTE_{p\text{-}value}$	$eQTM_r$	$eQTM_{p\text{-}value}$
OSM	4.92E-04	22:30266011-30268071	-13.63	5.10E-05	ENST00000215781.2	3.59E-04	-0.68 [-0.95;-0.13]	1.00E-04
PPP6R1	5.34E-04	19:55251330-55251959	-11.33	1.51E-04	ENST00000592242.1	4.83E-02	-0.86 [-0.94;-0.6]	4.22E-02
PPP6R1	5.34E-04	19:55251330-55251959	-11.33	1.51E-04	ENST00000412770.6	2.83E-03	-0.88 [-0.96;-0.6]	4.54E-02
RASSF4	4.87E-05	10:44977022-44977811	11	1.76E-04	ENST00000489171.5	3.42E-02	0.93 [0.76;0.98]	1.00E-04
RPS6KA4	3.11E-04	11:64359911-64360661	-10.55	2.11E-04	ENST00000334205.8	1.51E-03	-0.7 [-0.86;-0.43]	1.00E-04
RPS6KA4	3.11E-04	11:64359911-64360661	-10.55	2.11E-04	ENST00000334205.8	1.51E-03	-0.7 [-0.86;-0.43]	1.00E-04
SH3TC1	3.47E-04	4:8240533-8241022	-7.96	9.85E-04	ENST00000507891.1	6.46E-05	-0.7 [-0.83;-0.17]	1.64E-02
TMC6	6.31E-05	17:78126612-78127570	-9.6	3.35E-04	ENST00000590602.5	3.31E-02	-0.92 [-0.98;-0.82]	3.02E-02
TMC6	6.31E-05	17:78126612-78127570	-9.6	3.35E-04	ENST00000590602.5	3.31E-02	-0.92 [-0.98;-0.82]	3.02E-02
TMC6	6.31E-05	17:78126612-78127570	-9.6	3.35E-04	ENST00000591436.5	1.21E-02	-0.81 [-0.91;-0.64]	3.78E-02
TMC6	6.31E-05	17:78126612-78127570	-9.6	3.35E-04	ENST00000591436.5	1.21E-02	-0.81 [-0.91;-0.64]	3.78E-02
TMC6	6.31E-05	17:78126612-78127570	-9.6	3.35E-04	ENST00000593044.5	1.28E-02	-0.85 [-0.93;-0.67]	4.06E-02
TMC6	6.31E-05	17:78126612-78127570	-9.6	3.35E-04	ENST00000593044.5	1.28E-02	-0.85 [-0.93;-0.67]	4.06E-02
UCP2	7.66E-10	11:73979728-73982457	-9.4	3.75E-04	ENST00000545562.2	3.59E-02	-0.67 [-0.83;-0.46]	1.33E-02
UCP2	7.66E-10	11:73979728-73982457	-9.4	3.75E-04	ENST00000545562.2	3.59E-02	-0.67 [-0.83;-0.46]	1.33E-02
VAC14	2.54E-04	16:70691109-70693047	11.18	1.56E-04	ENST00000261776.9	4.82E-03	0.88 [0.55;0.96]	2.74E-02
VAC14	2.54E-04	16:70746190-70746702	9.13	4.36E-04	ENST00000261776.9	4.82E-03	0.87 [0.71;0.94]	3.48E-02

Expression quantitative trait methylation (eQTM) analysis of the GDEMs representing the correlation between methylation and transcript expression. Key in table legend: Gene = HGNC gene symbol, DMR = co-ordinates of the DMR (GRCh38), $DMR_{p\text{-}value}$ = p-value associated to the differential methylation analysis, EnsT = Ensembl transcript ID, $DTE_{p\text{-}value}$ = p-value associated to the differential transcript expression analysis, $eQTM_r$ = Pearson correlation coefficient for the methylation-expression correlation and the 95% confidence intervals, $eQTM_{p\text{-}value}$ = p-value associated to the methylation-expression correlation. An extended parsable table including the full eQTM analysis for all GDEMs can be found in ***Table S6****.*

coefficient remained high (r = -0.87) (**Table S7**). Nonetheless, given the centrality of *ITGB2* and *LCP2* among the GDEMs would make them interesting candidates in future targeted studies.

Estimated Cellular Composition Suggests Lower Proportion of Neuronal Cells in $SJC66_{high}$

Systematic differences in DNA methylation and gene expression could reflect changes in the cellular composition. To this end, we estimated the cellular composition using the transcriptome data as input for xCell, which is capable of estimating enrichment scores of 64 immune and stromal cell types (47). By comparing the estimated proportions from $SJC66_{high}$ with $SJC66_{low}$ we identified significant differences for 6 cell types: neurons, dendritic cells (DCs: all, conventional and immature), megakaryocytes and platelets (**Figure 6** and **Table S8**). Expectedly, higher enrichment scores for DCs (all subtypes) and platelets were estimated for the $SJC66_{high}$ samples. By contrast, lower proportions of neuron and megakaryocyte signatures were observed for the $SJC66_{high}$ samples. Notably, the difference in neuronal enrichment was found to be the most statistically significant with an almost fourfold difference when comparing $SJC66_{high}$ with $SJC66_{low}$.

DISCUSSION

In this study, we highlight insights into RA progression by combining the outputs of parallel whole-genome DNA methylome and transcriptome analyses on extracted preparations of synovial biopsies, from auto-antibody positive individuals, early arthritis patients and patients with established RA, stratified by the number of swollen joints. We observed that synovia from patients with a higher number of swollen joints (SJC66 ≥ 9) were different at the level of DNA methylation and gene expression from synovia from patients with a lower number of swollen joints. Specifically comparing $SJC66_{high}$ with $SJC66_{low}$ revealed 3536 DMRs and 142 DGEs, with both datasets primarily enriched for pathways associated with immune responses. The most significant difference in methylation was found spanning the promoter regions of *MIRLET7B* and *MIR10B*. Interestingly, mouse *miR-let-7b* has been shown to provoke arthritic joint inflammation by remodeling naïve myeloid cells into M1 macrophages *via* TLR-7 ligation (53) and can augment disease severity (54). *MIR10B* has been shown to regulate Th17 cells in patients with ankylosing spondylitis (55) but no studies have specifically associated it with RA. At the level of transcription, *CCL13* and *CLEC10A* were found to be the most differentially expressed. CCL13 (MCP-4) is an extensively studied chemokine that is thought to be involved with RA pathogenesis and disease progression (56–58). By contrast, not much is known about the role of CLEC10A in RA besides it being highly expressed on immature dendritic cells (DCs), monocyte-derived DCs and alternatively activated macrophages (59), as well as having been observed in the inflamed synovium of patients with active RA (60). Chemokines CCL13, CCL8, CXCL11, CXCL10, and CXCL9 regulate the recruitment of leukocytes into tissue and have therefore been implicated in the pathogenesis of RA (61). Differential methylation was observed in the vicinity of the promoter for *CCL13*, *CXCL11*, and *CXCL9*. Such results support a role for epigenetic/transcriptional processes in the spread of pathology to additional joints. While definitive mechanisms of joint spreading remain elusive, possible roles for immune cell migration due to chemokine expression (62, 63) could be further evaluated based on our data.

Altogether, we observed that 3% of DMRs associate with 20% of the differentially expressed GDEs. It is not surprising that not all DMRs could be linked to genes as a large number are found in distal intergenic (18.6%) or intronic (28.3%) regions, making any functional inference challenging. Of the genes that presented

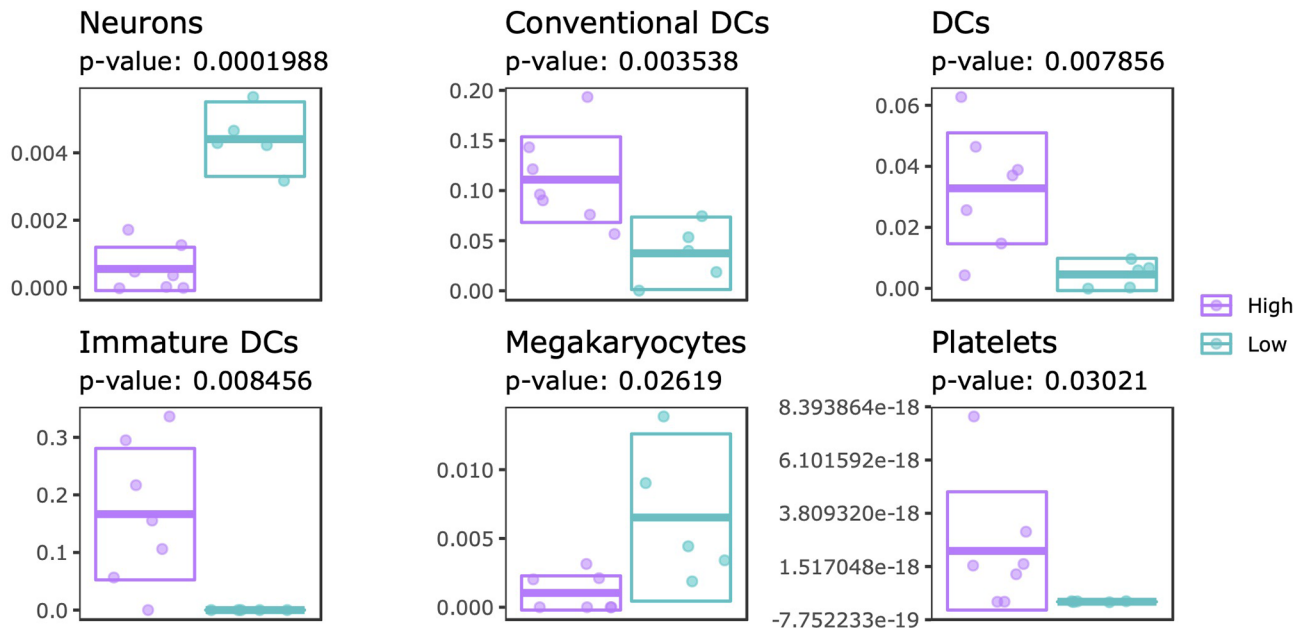


FIGURE 6 | Cellular composition estimation. Cellular proportions were estimated for 64 cell types using the xCell algorithm. Of these, 6 cell types were found to be significantly different when comparing $SJC66_{high}$ with $SJC66_{low}$. Visualized as scatter- and cross-bar plots are the significantly differentially represented cell types, namely the neurons, conventional dendritic cells (cDC), dendritic cells (DC), immature dendritic cells (iDC), megakaryocytes, and platelets.

both differentially expression and methylation, protein-protein interaction networks indicated that half encoded for interacting proteins, suggesting that the observed GDEMs function together. The most interconnected GDEMs appeared to be *ITGB2* and *LCP2*, with multiple regions of differential methylation observed surrounding both genes. While we observed transcript differential expression, most transcripts belonging to *ITGB2* and *LCP2* behaved similarly and all displayed reasonably strong inverse correlations with the DMRs located in the promoter area. *ITGB2* encodes an integrin, which would typically be involved in cell-surface mediated signalling. We observed that the gene encoding ITGB2's interaction partner integrin alpha L (*ITGAL*) was also differentially methylated and expressed. ITGB2 and ITGAL together form lymphocyte function-associated antigen (LFA-1), which interacts with intracellular adhesion molecule 1 (ICAM-1 or CD54) resulting in an enhanced immune cell influx into the synovial tissue (64, 65). Inhibiting LFA-1 has been reported to reduce inflammation and joint destruction in murine models of arthritis (66).

Functionally, the 59 GDEMs were primarily over-represented for immune response-associated pathways, specifically T-lymphocyte associated ones. It would be fascinating to understand why the DNA methylome and transcriptome between patients with a SJC66 of 9 and above relative to patients with a SJC66 of 8 present with such a sudden split instead of a gradual difference. It is clear that inflammation is likely an important factor contributing to the observed differences as patients with a high SJC66 also generally present higher levels of inflammation as expressed through using erythrocyte sedimentation rate (ESR) or concentration C-reactive protein (CRP). Our transcriptomic data indeed suggested an increased proportion of immune cells among $SJC66_{high}$ samples, as would be expected while pathology develops and cells migrate into the affected joints. Previous work has shown that clinically manifest arthritis in established RA is associated with increased infiltration of leukocytes. In synovial tissue samples from clinically involved joints, scores for infiltration by DCs are consistently higher than in clinically uninvolved joints obtained simultaneously from the same RA patients (67). Importantly, when comparing different clinically inflamed joints from the same RA patient simultaneously, leukocyte infiltration in one inflamed joint was shown to be representative of that in other inflamed joints, supporting the notion that leukocytes migrate from one joint to another (68). Indeed, there is continuous influx of leukocytes into the joints in established RA (69). We postulate that if synovial leukocytes exhibit properties that would facilitate cell migration, arthritis might spread from one inflamed joint to another. The results presented here support a disease mechanism in which, after development of clinically established RA, inflammatory and cell adhesion-associated processes play a key role in the progression of RA to greater joint involvement (70–72). Interestingly, we also observed differences of neuronal signatures suggesting a lower relative enrichment of neurons among $SJC66_{high}$ samples. In addition, enrichment analyses on the DMRs suggested hypermethylation of genes encoding nociception receptors, which are typically associated with peripheral sensory neurons (73, 74). A similar decrease in neuronal signature has previously been associated with RA severity, where the authors noted a potential role in the maladaptive response towards damage (75).

This is consistent with a more general loss of anti-inflammatory control by the nervous system in RA (76). Important to note is that our observations are based on estimates made by xCell, which can only calculate enrichment scores based on signatures rather than absolute values of cells. We are therefore unable to discern whether a population increased in size or a different population had decreased. Ideally, a similar estimate would have been generated based on the DNA methylation data, but the currently available reference datasets do not include the cell types measured available in xCell.

There are two limitations of this study, namely the fact that sex confounds the separation between $SJC66_{high}$ and $SJC66_{low}$ and the limited sample size. While we have sought to mitigate the confounding effect of sex by removing genes and CpGs located on the allosomes as well as by including sex as a covariate in our analyses, we acknowledge that we cannot fully eliminate the possibility that a sex effect is present. Accordingly, validation studies would be necessary where the DMEGs are verified using a larger, independent cohort while controlling for an interaction effect between sex and SJC66. The observed differences in transcript expression could be validated using a quantitative PCR approach with primers designed specifically against particular transcripts. Similarly, for validating the DMRs, targeted bisulphite-sequencing using primers for the regions of interest would be a cost-effective approach.

In conclusion, our study constitutes an exploratory analysis of whole genome DNA methylation and gene expression data performed on primary synovial tissue material from auto-antibody positive arthralgia patients without arthritis as well as patients with early and established RA patients. Where previous studies investigated cells from patients with RA versus disease controls and were potentially limited by their use of cultured cells, we focus on an integrative analysis of epigenetic marks and alternative splicing associated with swelling spread, providing novel insights into the mechanisms of disease progression towards more severe phenotypes. Nonetheless, further validation is necessary if the identified target genes are to be used for monitoring or treatment of the swelling and associated inflammatory processes in the joints of RA patients.

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.ebi.ac.uk/arrayexpress/, E-MTAB-6638; https://www.ebi.ac.uk/arrayexpress/, E-MTAB-6684.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by AMC medisch ethisch toetsingscommissie under MEC 02/152, MEC 05/107, and MEC 07/253 for the Synoviomics, Pre-synoviomics, and Synoviomics II cohorts, respectively. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

EF, HL, DT, CL, MS, DG, PT, and RP conceived the study. KM and GR carried out the laboratory experimental work. EF and CL conceived the analytical design. EF and AL performed the data analysis. HL, DT, CL, MM, PH, and WJ helped supervise the secondary data analyses. EF, AL, KM, HL, and DT led the writing of the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

AL was funded by the European Union's Horizon 2020 research and innovation program under Grant Agreement No. ITN-2014-EID-641665. This study was supported by the Dutch Arthritis Foundation (06-1-303 and 11-1-407), IMI BeTheCure (115142), and FP7 Euro-TEAM consortium (305549).

ACKNOWLEDGMENTS

We would like to thank the BTCure research consortium for providing access to the samples, Victor Neduva (Target and Pathway Validation, Target Sciences, GSK) for transferring the raw data files, Erika Cule (Research Statistics, GSK) for input into statistical methodology, and Yee Ying Chang (Clinical Genetics, Genome Diagnostics Laboratory, Amsterdam UMC, University of Amsterdam) for implementing the eQTM function in R. This manuscript has been released as a Pre-Print at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3576744 (77).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2021.651475/full#supplementary-material

REFERENCES

1. Tak PP, Bresnihan B. The pathogenesis and prevention of joint damage in rheumatoid arthritis: Advances from synovial biopsy and tissue analysis. *Arthritis Rheum* (2000) 43:2619–33. doi: 10.1002/1529-0131(200012)43:12<2619::AID-ANR1>3.0.CO;2-V
2. Taylor PC, Moore A, Vasilescu R, Alvir J, Tarallo M. A structured literature review of the burden of illness and unmet needs in patients with rheumatoid arthritis: a current perspective. *Rheumatol Int* (2016) 36:685–95. doi: 10.1007/s00296-015-3415-x
3. Okada Y, Wu D, Trynka G, Raj T, Terao C, Ikari K, et al. Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature* (2014) 506:376–81. doi: 10.1038/nature12873

4. Wijbrandts CA, Tak PP. Prediction of Response to Targeted Treatment in Rheumatoid Arthritis. *Mayo Clin Proc* (2017) 92:1129–43. doi: 10.1016/j.mayocp.2017.05.009
5. Nielen MMJ, van Schaardenburg D, Reesink HW, van de Stadt RJ, van der Horst-Bruinsma IE, de Koning MHMT, et al. Specific autoantibodies precede the symptoms of rheumatoid arthritis: a study of serial measurements in blood donors. *Arthritis Rheum* (2004) 50:380–6. doi: 10.1002/art.20018
6. Kolfenbach JR, Deane KD, Derber LA, O'Donnell CI, Gilliland WR, Edison JD, et al. Autoimmunity to peptidyl arginine deiminase type 4 precedes clinical onset of rheumatoid arthritis. *Arthritis Rheum* (2010) 62:2633–9. doi: 10.1002/art.27570
7. Tak PP, Smeets TJ, Daha MR, Kluin PM, Meijers KA, Brand R, et al. Analysis of the synovial cell infiltrate in early rheumatoid synovial tissue in relation to local disease activity. *Arthritis Rheum* (1997) 40:217–25. doi: 10.1002/art.1780400206
8. Zhang F, Wei K, Slowikowski K, Fonseka CY, Rao DA, Kelly S, et al. Defining inflammatory cell states in rheumatoid arthritis joint synovial tissues by integrating single-cell transcriptomics and mass cytometry. *Nat Immunol* (2019) 20:928–42. doi: 10.1038/s41590-019-0378-1
9. Tak PP, Zvaifler NJ, Green DR, Firestein GS. Rheumatoid arthritis and p53: how oxidative stress might alter the course of inflammatory diseases. *Immunol Today* (2000) 21:78–82. doi: 10.1016/S0167-5699(99)01552-2
10. Ospelt C. Synovial fibroblasts in 2017. *RMD Open* (2017) 3:e000471. doi: 10.1136/rmdopen-2017-000471
11. Pitzalis C, Kelly S, Humby F. New learnings on the pathophysiology of RA from synovial biopsies. *Curr Opin Rheumatol* (2013) 25:334–44. doi: 10.1097/BOR.0b013e32835fd8eb
12. Nemtsova MV, Zaletaev DV, Bure IV, Mikhaylenko DS, Kuznetsova EB, Alekseeva EA, et al. Epigenetic Changes in the Pathogenesis of Rheumatoid Arthritis. *Front Genet* (2019) 10:570. doi: 10.3389/fgene.2019.00570
13. Ospelt C, Gay S, Klein K. Epigenetics in the pathogenesis of RA. *Semin Immunopathol* (2017) 39:409–19. doi: 10.1007/s00281-017-0621-5
14. Nair N, Wilson AG, Barton A. DNA methylation as a marker of response in rheumatoid arthritis. *Pharmacogenomics* (2017) 18:1323–32. doi: 10.2217/pgs-2016-0195
15. Whitaker JW, Shoemaker R, Boyle DL, Hillman J, Anderson D, Wang W, et al. An imprinted rheumatoid arthritis methylome signature reflects pathogenic phenotype. *Genome Med* (2013) 5:40. doi: 10.1186/gm444
16. de la Rica L, Urquiza JM, Gómez-Cabrero D, Islam ABMMK, López-Bigas N, Tegnér J, et al. Identification of novel markers in rheumatoid arthritis through integrated analysis of DNA methylation and microRNA expression. *J Autoimmun* (2013) 41:6–16. doi: 10.1016/j.jaut.2012.12.005
17. Whitaker JW, Boyle DL, Bartok B, Ball ST, Gay S, Wang W, et al. Integrative omics analysis of rheumatoid arthritis identifies non-obvious therapeutic targets. *PLoS One* (2015) 10:1–14. doi: 10.1371/journal.pone.0124254
18. Ekwall A-KH, Whitaker JW, Hammaker D, Bugbee WD, Wang W, Firestein GS. The Rheumatoid Arthritis Risk Gene LBH Regulates Growth in Fibroblast-like Synoviocytes. *Arthritis Rheumatol (Hoboken NJ)* (2015) 67:1193–202. doi: 10.1002/art.39060
19. Hammaker D, Whitaker JW, Maeshima K, Boyle DL, Ekwall A-KH, Wang W, et al. LBH Gene Transcription Regulation by the Interplay of an Enhancer Risk Allele and DNA Methylation in Rheumatoid Arthritis. *Arthritis Rheumatol (Hoboken NJ)* (2016) 68:2637–45. doi: 10.1002/art.39746
20. Ham S, Bae J-B, Lee S, Kim B-J, Han B-G, Kwok S-K, et al. Epigenetic analysis in rheumatoid arthritis synoviocytes. *Exp Mol Med* (2019) 51:22. doi: 10.1038/s12276-019-0215-5
21. Ai R, Whitaker JW, Boyle DL, Tak PP, Gerlag DM, Wang W, et al. DNA Methylome Signature in Synoviocytes From Patients With Early Rheumatoid Arthritis Compared to Synoviocytes From Patients With Longstanding Rheumatoid Arthritis. *Arthritis Rheumatol* (2015) 67:1978–80. doi: 10.1002/art.39123
22. Karouzakis E, Raza K, Kolling C, Buckley CD, Gay S, Filer A, et al. Analysis of early changes in DNA methylation in synovial fibroblasts of RA patients before diagnosis. *Sci Rep* (2018) 8:7370. doi: 10.1038/s41598-018-24240-2
23. Gerlag DM, Tak PP. How to perform and analyse synovial biopsies. *Best Pract Res Clin Rheumatol* (2013) 27:195–207. doi: 10.1016/j.berh.2013.03.006
24. de Hair MJH, van de Sande MGH, Ramwadhdoebe TH, Hansson M, Landewé R, Van Der Leij C, et al. Features of the synovium of individuals at risk of developing rheumatoid arthritis : Implications for understanding preclinical rheumatoid arthritis. *Arthritis Rheumatol* (2014) 66:513–22. doi: 10.1002/art.38273
25. de Hair MJH, Harty LC, Gerlag DM, Pitzalis C, Veale DJ, Tak PP. Synovial tissue analysis for the discovery of diagnostic and prognostic biomarkers in patients with early arthritis. *J Rheumatol* (2011) 38:2068–72. doi: 10.3899/jrheum.110426
26. van de Sande MGH, de Hair MJH, Schuller Y, van de Sande GPM, Wijbrandts CA, Dinant HJ, et al. The features of the synovium in early rheumatoid arthritis according to the 2010 ACR/EULAR classification criteria. *PLoS One* (2012) 7:e36668. doi: 10.1371/journal.pone.0036668
27. Aletaha D, Neogi T, Silman AJ, Funovits J, Felson DT, Bingham CO, et al. 2010 Rheumatoid arthritis classification criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. *Ann Rheum Dis* (2010) 69S:1580–8. doi: 10.1136/ard.2010.138461
28. Argelaguet R, Velten B, Arnol D, Dietrich S, Zenz T, Marioni JC, et al. Multi-Omics Factor Analysis—a framework for unsupervised integration of multi-omics data sets. *Mol Syst Biol* (2018) 14:e8124. doi: 10.15252/msb.20178124
29. Andrews S. *FastQC: a quality control tool for high throughput sequence data*. Available at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc.
30. Buffalo V. *Scythe: a Bayesian adapter trimmer*.
31. Joshi N. *Sickle: a windowed adaptive trimming tool for FASTQ files using quality*.
32. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol* (2016) 34:525–7. doi: 10.1038/nbt.3519
33. Development Core Team R. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing (2008). Available at: http://www.r-project.org/.
34. Soneson C, Love MI, Robinson MD. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Research* (2016) 4:1521. doi: 10.12688/f1000research.7563.2
35. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* (2014) 15:550. doi: 10.1186/s13059-014-0550-8
36. Yi L, Pimentel H, Bray NL, Pachter L. Gene-level differential analysis at transcript-level resolution. *Genome Biol* (2018) 19:52. doi: 10.1186/s13059-018-1419-z
37. Jiang H, Lei R, Ding S-W, Zhu S. Skewer: a fast and accurate adapter trimmer for next-generation sequencing paired-end reads. *BMC Bioinf* (2014) 15:182. doi: 10.1186/1471-2105-15-182
38. Krueger F, Andrews SR. Bismark: A flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* (2011) 27:1571–2. doi: 10.1093/bioinformatics/btr167
39. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. Genome Project Data Processing S. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* (2009) 25:2078–9. doi: 10.1093/bioinformatics/btp352
40. Korthauer K, Chakraborty S, Benjamini Y, Irizarry RA, User G. Detection and accurate false discovery rate control of differentially methylated regions from whole genome bisulfite sequencing. *Biostatistics* (2018) 00:1–17. doi: 10.1093/biostatistics/kxy007
41. Zhu LJ, Gazin C, Lawson ND, Pagès H, Lin SM, Lapointe DS, et al. ChIPpeakAnno: a Bioconductor package to annotate ChIP-seq and ChIP-chip data. *BMC Bioinf* (2010) 11:237. doi: 10.1186/1471-2105-11-237
42. Frankish A, Diekhans M, Ferreira A-M, Johnson R, Jungreis I, Loveland J, et al. GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Res* (2019) 47:D766–73. doi: 10.1093/nar/gky955
43. Efron B, Tibshirani RJ. *An introduction to the bootstrap*. New York, N.Y.; London: Chapman & Hall (1993).
44. Canty A, Ripley B. *boot*. (2021).
45. Sergushichev AA. An algorithm for fast preranked gene set enrichment analysis using cumulative statistic calculation. *bioRxiv* (2016), 060012. doi: 10.1101/060012
46. *Clarivate. MetaCore*.
47. Aran D, Hu Z, Butte AJ. xCell: digitally portraying the tissue cellular heterogeneity landscape. *Genome Biol* (2017) 18:220. doi: 10.1186/s13059-017-1349-1
48. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, et al. STRING v11: Protein-protein association networks with increased coverage,

supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res* (2019) 47:D607–13. doi: 10.1093/nar/gky1131
49. Shayevitch R, Askayo D, Keydar I, Ast G. The importance of DNA methylation of exons on alternative splicing. *RNA* (2018) 24:1351–62. doi: 10.1261/rna.064865.117
50. Lev Maor G, Yearim A, Ast G. The alternative role of DNA methylation in splicing regulation. *Trends Genet* (2015) 31:274–80. doi: 10.1016/J.TIG.2015.03.002
51. Linker SM, Urban L, Clark SJ, Chhatriwala M, Amatya S, McCarthy DJ, et al. Combined single-cell profiling of expression and DNA methylation reveals splicing regulation and heterogeneity. *Genome Biol* (2019) 20:30. doi: 10.1186/s13059-019-1644-0
52. Harris CJ, Scheibe M, Wongpalee SP, Liu W, Cornett EM, Vaughan RM, Jacobsen SE. A DNA methylation reader complex that enhances gene transcription. *Science* (2018) 362(4619):1182–6. doi: 10.1126/science.aar7854
53. Kim SJ, Chen Z, Essani AB, Elshabrawy HA, Volin MV, Volkov S, et al. Identification of a Novel Toll-like Receptor 7 Endogenous Ligand in Rheumatoid Arthritis Synovial Fluid That Can Provoke Arthritic Joint Inflammation. *Arthritis Rheumatol* (2016) 68:1099–110. doi: 10.1002/art.39544
54. Umar S, Palasiewicz K, Van Raemdonck K, Volin MV, Romay B, Amin MA, et al. IRAK4 inhibition: a promising strategy for treating RA joint inflammation and bone erosion. *Cell Mol Immunol* (2020). doi: 10.1038/s41423-020-0433-8
55. Chen L, Al-Mossawi MH, Ridley A, Sekine T, Hammitzsch A, De Wit J, et al. MiR-10b-5p is a novel Th17 regulator present in Th17 cells from ankylosing spondylitis. *Ann Rheum Dis* (2017) 76:620–4. doi: 10.1136/annrheumdis-2016-210175
56. Hintzen C, Quaiser S, Pap T, Heinrich PC, Hermanns HM. Induction of CCL13 expression in synovial fibroblasts highlights a significant role of oncostatin M in rheumatoid arthritis. *Arthritis Rheum* (2009) 60:1932–43. doi: 10.1002/art.24602
57. Yamaguchi A, Nozawa K, Fujishiro M, Kawasaki M, Suzuki F, Takamori K, et al. CC motif chemokine ligand 13 is associated with rheumatoid arthritis pathogenesis. *Mod Rheumatol* (2013) 23:856–63. doi: 10.3109/s10165-012-0752-4
58. Rump L, Mattey DL, Kehoe O, Middleton J. An initial investigation into endothelial CC chemokine expression in the human rheumatoid synovium. *Cytokine* (2017) 97:133–40. doi: 10.1016/j.cyto.2017.05.023
59. Hoober JK. ASGR1 and Its Enigmatic Relative, CLEC10A. *Int J Mol Sci* (2020) 21:4818. doi: 10.3390/ijms21144818
60. McHugh J. Synovial macrophage populations linked to RA remission. *Nat Rev Rheumatol* (2020) 16:471. doi: 10.1038/s41584-020-0481-6
61. Miyabe Y, Miyabe C, Iwai Y, Luster AD. Targeting the Chemokine System in Rheumatoid Arthritis and Vasculitis. *JMA J* (2020) 3:182–92. doi: 10.31662/jmaj.2020-0019
62. Elemam NM, Hannawi S, Maghazachi AA. Role of Chemokines and Chemokine Receptors in Rheumatoid Arthritis. *ImmunoTargets Ther* (2020) 9:43–56. doi: 10.2147/itt.s243636
63. Yu X, Song Z, Rao L, Tu Q, Zhou J, Yin Y, et al. Synergistic induction of CCL5, CXCL9 and CXCL10 by IFN-γ and NLRs ligands on human fibroblast-like synoviocytes—A potential immunopathological mechanism for joint inflammation in rheumatoid arthritis. *Int Immunopharmacol* (2020) 82:1–6. doi: 10.1016/j.intimp.2020.106356
64. Lowin T, Straub RH. Integrins and their ligands in rheumatoid arthritis. *Arthritis Res Ther* (2011) 13:244. doi: 10.1186/ar3464
65. Tak PP, Thurkow EW, Daha MR, Kluin PM, Smeets TJ, Meinders AE, et al. Expression of adhesion molecules in early rheumatoid synovial tissue. *Clin Immunol Immunopathol* (1995) 77:236–42. doi: 10.1006/clin.1995.1149
66. Suchard SJ, Stetsko DK, Davis PM, Skala S, Potin D, Launay M, et al. An LFA-1 (alphaLbeta2) small-molecule antagonist reduces inflammation and joint destruction in murine models of arthritis. *J Immunol* (2010) 184:3917–26. doi: 10.4049/jimmunol.0901095
67. Kraan MC, Versendaal H, Jonker M, Bresnihan B, Post WJ, Hart B, et al. A symptomatic synovitis precedes clinically manifest arthritis. *Arthritis Rheum* (1998) 41:1481–8. doi: 10.1002/1529-0131(199808)41:8<1481::AID-ART19>3.0.CO;2-O
68. Kraan MC, Reece RJ, Smeets TJM, Veale DJ, Emery P, Tak PP. Comparison of synovial tissues from the knee joints and the small joints of rheumatoid arthritis patients: Implications for pathogenesis and evaluation of treatment. *Arthritis Rheum* (2002) 46:2034–8. doi: 10.1002/art.10556
69. Thurlings RM, Wijbrandts CA, Bennink RJ, Dohmen SE, Voermans C, Wouters D, et al. Monocyte scintigraphy in rheumatoid arthritis: The dynamics of monocyte migration in immune-mediated inflammatory disease. *PLoS One* (2009) 4:1–6. doi: 10.1371/journal.pone.0007865
70. Tsubaki T, Arita N, Kawakami T, Shiratsuchi T, Yamamoto H, Takubo N, et al. Characterization of histopathology and gene-expression profiles of synovitis in early rheumatoid arthritis using targeted biopsy specimens. *Arthritis Res Ther* (2005) 7:R825–36. doi: 10.1186/ar1751
71. Lequerré T, Bansard C, Vittecoq O, Derambure C, Hiron M, Daveau M, et al. Early and long-standing rheumatoid arthritis: distinct molecular signatures identified by gene-expression profiling in synovia. *Arthritis Res Ther* (2006) 11:R105. doi: 10.1186/ar2744
72. Choy E. Understanding the dynamics: pathways involved in the pathogenesis of rheumatoid arthritis. *Rheumatol (Oxford)* (2012) 51(Suppl 5):v3–11. doi: 10.1093/rheumatology/kes113
73. Dubin AE, Patapoutian A. Nociceptors: The sensors of the pain pathway. *J Clin Invest* (2010) 120:3760–72. doi: 10.1172/JCI42843
74. Miller RE, Tran PB, Obeidat AM, Raghu P, Ishihara S, Miller RJ, et al. The Role of Peripheral Nociceptive Neurons in the Pathophysiology of Osteoarthritis Pain. *Curr Osteoporos Rep* (2015) 13:318–26. doi: 10.1007/s11914-015-0280-1
75. Orange DE, Agius P, DiCarlo EF, Robine N, Geiger H, Szymonifka J, et al. Identification of Three Rheumatoid Arthritis Disease Subtypes by Machine Learning Integration of Synovial Histologic Features and RNA Sequencing Data. *Arthritis Rheumatol* (2018) 70:690–701. doi: 10.1002/art.40428
76. Koopman FA, van Maanen MA, Vervoordeldonk MJ, Tak PP. Balancing the autonomic nervous system to reduce inflammation in rheumatoid arthritis. *J Intern Med* (2017) 282:64–75. doi: 10.1111/joim.12626
77. Ferrero E, Li Yim A, Maratou K, Lewis HD, Royal G, Tough DF, et al. Novel Insights into Rheumatoid Arthritis Through Characterisation of Concordant Changes in DNA Methylation and Gene Expression in Synovial Biopsies of Patients with Differing Numbers of Swollen Joints. *SSRN Electron J* (2020). doi: 10.2139/ssrn.3576744

Conflict of Interest: AL, KM, GR, HL, DT, CL, DG, and RP were employed by GSK when this study was conducted. EF and PT were employed by Novartis and Candel Therapeutics when this study was conducted, respectively. WJ was financially supported by GSK and Mead Johnson.

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.

frontiers in Immunology

REVIEW
published: 28 April 2021
doi: 10.3389/fimmu.2021.654731

Innate Immune Cells in the Esophageal Tumor Microenvironment

Kele Cui[1,3,4*†], *Shouxin Hu*[2,3,4†], *Xinyu Mei*[5] *and Min Cheng*[2,3,4*]

[1] Department of Clinical Laboratory, The First Affiliated Hospital of USTC, Division of Life Sciences and Medicine, University of Science and Technology of China, Hefei, China, [2] Department of Geriatrics, Gerontology Institute of Anhui Province, The First Affiliated Hospital of USTC, Division of Life Sciences and Medicine, University of Science and Technology of China, Hefei, China, [3] Anhui Provincial Key Laboratory of Tumor Immunotherapy and Nutrition Therapy, Hefei, China, [4] Cancer Immunotherapy Center, The First Affiliated Hospital of USTC, Division of Life Sciences and Medicine, University of Science and Technology of China, Hefei, China, [5] Department of Thoracic Surgery, The First Affiliated Hospital of USTC, Division of Life Sciences and Medicine, University of Science and Technology of China, Hefei, China

Esophageal cancer (EC) is one of the most common mucosa-associated tumors, and is characterized by aggressiveness, poor prognosis, and unfavorable patient survival rates. As an organ directly exposed to the risk of foodborne infection, the esophageal mucosa harbors distinct populations of innate immune cells, which play vital roles in both maintenance of esophageal homeostasis and immune defense and surveillance during mucosal anti-infection and anti-tumor responses. In this review, we highlight recent progress in research into innate immune cells in the microenvironment of EC, including lymphatic lineages, such as natural killer and γδT cells, and myeloid lineages, including macrophages, dendritic cells, neutrophils, myeloid-derived suppressor cells, mast cells and eosinophils. Further, putative innate immune cellular and molecular mechanisms involved in tumor occurrence and progression are discussed, to highlight potential directions for the development of new biomarkers and effective intervention targets, which can hopefully be applied in long-term multilevel clinical EC treatment. Fully understanding the innate immunological mechanisms involved in esophageal mucosa carcinogenesis is of great significance for clinical immunotherapy and prognosis prediction for patients with EC.

Keywords: innate immune cells, crosstalk, regulation, esophageal tumor microenvironment, immunotherapy strategy

OPEN ACCESS

Edited by:
Aridaman Pandit, University Medical Center Utrecht, Netherlands

Reviewed by:
Nathalie Jacobs, University of Liège, Belgium
Antonino Bruno, MultiMedica (IRCCS), Italy
Natalia Plekhova, Pacific State Medical University, Russia

****Correspondence:***
Min Cheng
chengmin@ustc.edu.cn
Kele Cui
ckele@mail.ustc.edu.cn

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Molecular Innate Immunity, a section of the journal Frontiers in Immunology

Received: *17 January 2021*
Accepted: *13 April 2021*
Published: *28 April 2021*

Citation:
Cui K, Hu S, Mei X and Cheng M (2021) Innate Immune Cells in the Esophageal Tumor Microenvironment. Front. Immunol. 12:654731. doi: 10.3389/fimmu.2021.654731

INTRODUCTION

According to an analysis of 36 cancer types in 185 countries, esophageal cancer (EC) accounts for approximately 3.2% of incidence and 5.3% of mortality attributable to total cancers (1). Risk factors for EC include smoking, alcohol consumption, low fruit intake, and high body-mass index, and it is becoming a major disease burden worldwide (2). There are two main types of EC, esophageal adenocarcinoma (EAC) and esophageal squamous cell carcinoma (ESCC) (3). EAC is most common in developed countries (e.g., Europe and America), while ESCC mainly occurs in developing countries, including eastern Asia and Africa, and particularly China, which had the highest age-standardized incidence, mortality, and disability-adjusted life-years rates among 195 countries in 2017 (2). Due to the lack of reliable diagnostic indicators, the prognosis of patients with EC remains relatively poor; while surgical resection can prolong patient survival, rates of recurrence and metastasis remain high, with 5-year survival rates only 15%–25% (3, 4). In many patients who cannot benefit from esophagectomy, immune status determines sensitivity to radiotherapy and chemotherapy (3).

The recent realization that the involvement of innate immune system in the process of defending mucosal-associated infection and tumors has fuelled the accelerated interest in the roles of innate immune cells in the pathogenesis of EC. The esophageal mucosa harbors numerous innate immune cells, which is attributed to their quick responses when encountering foreign foodborne antigens (4, 5). The partial exposure of the esophageal mucosa to the external environment makes it vulnerable to pathogen attack, which can cause long-term inflammation that may develop into esophageal dysplasia and subsequently cancer (6). Via recognizing molecular alterations caused by microbial infections (7) or cancer cells with multiple genetic mutations (8), innate immune cells can induce effector responses such as cytotoxicity by natural killer (NK) cells and phagocytosis by macrophages. Besides, they can initiate adaptive immune responses by antigen presentation to tumor-specific $CD8^+$ T antigen-presenting cells (APCs) (9). Innate immune cells can also exert effector functions after antibody induction, including antibody-dependent cellular cytotoxicity or phagocytosis, which rely on Fc receptor (FcR) expression (9). Cancer cells can escape from anti-tumor immune responses by promoting polarization toward immunosuppressive cell phenotypes, including tumor-associated macrophages (TAMs) and dendritic cells (DCs), recruiting immunomodulatory cells, such as myeloid-derived suppressor cells (MDSCs) and regulatory T cells (Tregs), and inducing over-expression of immune checkpoint molecules by NK or T cells (10). Due to the plasticity of innate immune cells, another strategy of tumor immune escape is to enable them to orchestrate the angiogenic switch under tumor microenvironment (TME) stimuli, to support the tumor progression (5, 11). Given the important defensive and regulatory roles of innate immune cells in cancer progression, extensive attention has been focused on their pathogenic or protective functions in the microenvironment of many solid tumors (12). Clinical immunotherapy approaches based on innate immune cells, including inhibitors targeting immune checkpoints, such as PD1/PD-L1, CTLA4, TIGIT, CD96, TIM3, and LAG3, as well as bispecific antibodies or chimeric antigen receptor (CAR) T cells, to promote specific T cell responses, have been extensively studies and have potential for use as adjuvant therapies, alongside surgical resection and chemoradiotherapy, to treat cancers (9, 13). Therefore, adequate understanding of how variations of innate immunity in the TME affect EC pathogenesis is of great practical significance for clinical treatment. However, until now, the roles of innate immune cells in EC have not been comprehensively described.

Herein, we review recent progress in understanding of the roles of innate immune cells, including NK cells, γδT cells, TAMs, DCs, MDSCs, neutrophils, mast cells (MCs) and eosinophils in the TME of EC, as well as the underlying cellular and molecular mechanisms involved in tumor occurrence and progression, with the aim of providing directions for combined immunotherapy strategies and prognosis prediction.

NK CELLS

The innate immune system serves as the front line of host defense against pathogen invasion and tumor, in which NK cells play a vanguard role due to their powerful cytotoxic activity (14). NK cells express various activating and inhibitory receptors for tumor cell recognition and are the primary force in innate anti-tumor immune surveillance, playing vital roles in inhibiting cancer development at early stages, and in controlling cancer metastasis (13, 15). NK cells can initiate anti-tumor responses through directly killing tumor cells, secreting cytokines, including IFN-γ and TNF-α, and recruiting other anti-tumor immune cells (13). Studies indicated that infiltrating NK cell density in the EC TME is positively correlated with patient overall survival (OS) and favorable postoperative prognosis (16, 17). NK cells in the TME of EC can recognize and kill tumor cells *via* NKp30/B7-H6 pathway, which substantially contributes to NK cell-mediated immune responses (16).

Abbreviations: APCs, antigen-presenting cells; BE, Barrett's esophagus; CAR, chimeric antigen receptor; CCR2, CC chemokine receptor 2; DCs, dendritic cells; EC, esophageal cancer; EAC, esophageal adenocarcinoma; ESCC, esophageal squamous cell carcinoma; FcR, Fc receptor; FGF2, fibroblast growth factor 2; FGFR1, fibroblast growth factor receptor 1; HMGB1, high-mobility group box 1; HSP, heat shock protein; IDO1, indoleamine 2,3-dioxygenase 1; iNOS, inducible nitric oxide synthase-2; ILCs, innate lymphoid cells; LAMP3, lysosome-associated membrane glycoprotein 3; MCs, mast cells; MCP-1, macrophage chemotactic protein-1; MDSCs, myeloid-derived suppressor cells; M-MDSCs, mononuclear myeloid-derived suppressor cells; MMP9, matrix metalloproteinase 9; NETs, neutrophil extracellular traps; NK, natural killer; NAC, neoadjuvant chemotherapy; NLR, neutrophil to lymphocyte ratio; OS, overall survival; PMN-MDSCs, polymorphonuclear myeloid-derived suppressor cells; RCAS1, receptor-binding cancer antigen expression on SiSo cells; RLRs, RIG-I-like receptors; SIRPα, signal regulatory protein-α; TLRs, targeting toll-like receptors; Tregs, regulatory T cells; TK1, tissue kallikrein; TAMs, tumor-associated macrophages; TANs, tumor-associated neutrophils; TME, tumor microenvironment; TNF, tumor necrosis factor; TP, thymidine phosphorylase; VEGF, vascular endothelial growth factor.

However, EC patients with high B7-H6 expression have inferior survival, primarily because EC cells can secret soluble B7-H6, which competitively binds the NKp30 receptor on NK cells, thereby inhibiting NKp30-mediated killing (16). Interestingly, IL-17 secreted by $CD4^{+}Foxp3^{-}$ Th17 cells can increase NK cell numbers by stimulating EC cells to produce chemokines (CXCL9 and CXCL10), and augment NK cell activation and function by enhancing their TNF-α, IFN-γ, granzyme B, and perforin production, and the expression of activating receptors (NKp46, NKp44, NTB-A, and NKG2D) (17, 18). Nevertheless, an important mechanism by which tumor cells counteract NK cells is by promoting over-expression of immune checkpoints and down-regulation of activating receptors on NK cells, inducing their dysfunction and exhaustion (13). Up-regulated expression of the inhibitory receptor, Tim-3, on EC tumor-infiltrated NK cells is accompanied by NK cell dysfunction and exhaustion, and associated with tumor invasion depth, nodal status, and advanced clinical stage (19). Furthermore, increased PD1 expression on peripheral and tumor-infiltrating NK cells can inhibit their IFN-γ secretion and CD107a expression through the PD-1/PD-L1 pathway, and is associated with poor survival of EC patients (20) (**Figure 1**).

Cancer cells can promote a suppressive TME, which challenges anti-tumor immunity by inducing an imbalance in activating and inhibitory immune cell signaling, suppressive factor secretion, and recruitment of suppressive immune cells. In principle, $CD56^{dim}$ NK cells, which exhibit higher cytotoxicity, are more sensitive to apoptosis than $CD56^{bright}$ NK cells in the presence of physiological H_2O_2 levels; large amounts of H_2O_2 can be produced in the TME of EC, which contributes to reduced infiltration by $CD56^{dim}$ NK cells as tumors develop (21). Consistently, NK cell numbers in the circulation and omentum of patients with EAC were significantly reduced and skewed toward the $CD56^{bright}$ phenotype with increased IL-10 and reduced NKp46 and TNF-α production, exhibiting reduced toxicity and inhibited function (22). Similarly, in ESCC patients, with the down-regulation of CD16 and increased expression of CD56, the NK cell levels were also declined in the tumors and exhibited an exhausted phenotype (23, 24), demonstrating the vital roles of the suppressive microenvironment formed by cancer cells in altering NK cell phenotype and activity (**Figure 1**).

Although existing studies on EC have provided a preliminary understanding of anti-tumor effects of NK cells, there is still not enough to fully outline the roles and prognostic significance of NK cells involved in EC. Recently, a study in human hepatocellular carcinoma reported that breaking the balance between active receptor CD226 and inhibitory receptors CD96 and TIGIT lead to impaired NK cell function (25). Other than the attenuation of targeting and killing of tumor cells and acquisition of tolerogenic/immunosuppressive behavior, tumor-associated NK cells can acquire pro-angiogenic activities favoring tumor progression due to TME stimuli in various solid malignancies including prostate cancer, lung cancer and colon cancer, which inspires us to pay attention to the full-scale immune function of NK cells in EC (11, 26). The significance of investigations on the function of NK cells lies in facilitating the development of NK cell-based immunotherapy in EC. Indeed, studies from an ESCC animal model revealed that IL-18 deficiency can down-regulate local NK cell anti-tumor immunity by decreasing their IFN-γ production,

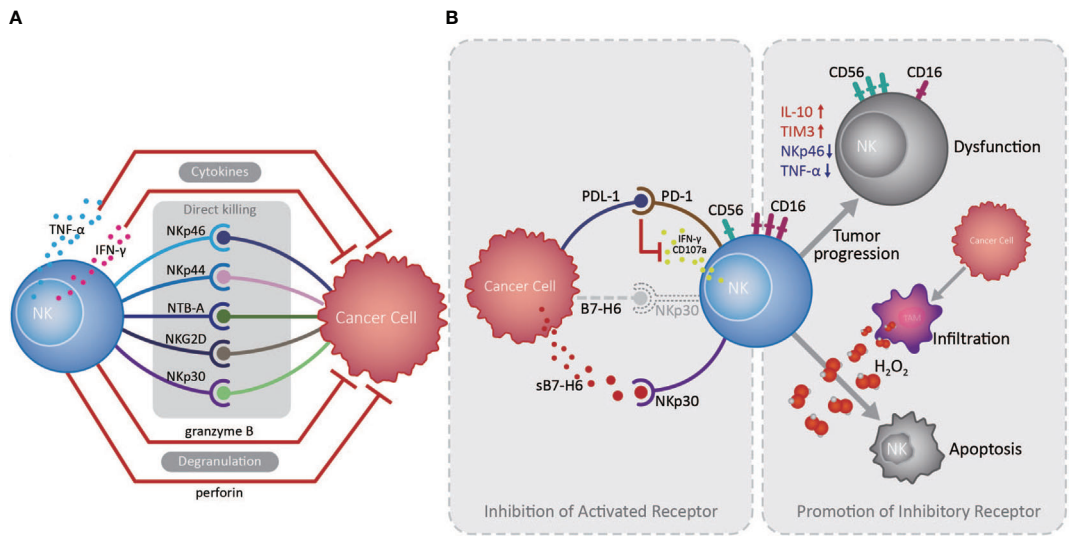


FIGURE 1 | Roles of natural killer cells in the esophageal cancer tumor microenvironment. **(A)** Natural killer (NK) cells exert essential anti-tumor functions through degranulation, cytokine release, and activated receptor expression, directly killing esophageal tumor cells in the absence of antigen recognition in the tumor microenvironment (TME). **(B)** Cancer cells can escape from NK cell immune surveillance in the TME through the PD-1/PD-L1 pathway and the competitive combination of activated NKp30 receptor, expressed on NK cells, by releasing soluble B7-H6 (indicated as sB7-H6). Otherwise, large amounts of H_2O_2 produced by tumor infiltrated macrophages in the TME are prone to induce $CD56^{dim}$ NK cell apoptosis, while $CD56^{dim}$ NK cells can also transform to a $CD56^{bright}$ phenotype during tumor development, as well as decreasing in number and becoming dysfunctional.

suggesting that exogenous IL-18 supplementation has potential to delay EC development (27). Furthermore, NK cells expanded *in vitro* have high cytotoxicity against ESCC cells expressing NKG2DLs, particularly those exhibiting an epithelial-mesenchymal transition (EMT) phenotype, raising the possibility of clinical therapy targeting these NK cells in patients with ESCC (18). However, much remains to be done before these findings can be applied to the clinical treatment of EC.

γδT CELLS

As important contributors to innate immunity, γδT cells perform complex roles, including immune surveillance, immune regulation, and effector functions (28); they can be divided into two types according to their T cell receptor δ chain: Vδ1 and Vδ2 T cells. The former exists in healthy epithelium and participates in maintaining epithelial homeostasis, whereas the latter is present on 70% of total peripheral γδT cells (29). Flow cytometry analysis of γδT cells from patients with EC revealed that the majority of peripheral circulatory γδT cells expressed the Vγ9 and Vδ2 T cell receptors and exhibited cytotoxicity against EC cells, mainly by recognizing heat shock protein (HSP) 60 and HSP70 on the tumor cell surface (30). However, $Vδ1^+$ T cells are dominant in the TME of ESCC, possibly because activated $Vδ1^+$ γδT cells in peripheral blood can adhere to ESCC cells and fibroblasts *via* adhesion molecules, including LFA-1 (CD11a), CD49d, CD49e, L-selectin, and CD103, whereas Vδ2 T cells can only use a few adhesion molecules, including LFA-1, L-selectin, and CD44v6 (31).

The role of γδT cells in EC is far from well understood. Although γδT cells make up a small population of tissue-resident lymphocytes, they constitute an important first line of defense against infections, autoimmune diseases and tumors, especially in the mucosal barrier such as the skin, lung, liver, tongue, genital tract and peritoneal cavity (32–34). However, the alterations in γδT cell subsets and functions, as well as their prognostic and diagnostic significance in the EC remains obscure. Indeed, we have been focusing on the role of tissue-resident γδT cells in lung cancer for years, and have illustrated the involvement of γδT17 cells in the effective immune surveillance of lung mucosa shaped by microbiota, as well as in the control of melanoma in the elderly (35, 36). It is worth pondering whether commensal bacteria engage in the maintenance of esophageal homeostasis and the occurrence and progression of EC, and importantly, whether the roles of γδT cells are involved. Recently, we also paid attention to the variations of tissue-resident γδT cells in surgical ESCC specimens, and hope to clarify the functions of these γδT cells in the tumor progression and its prognostic and diagnostic value in ESCC in future study.

TAMs

Accounting for up to half of the total, TAMs are the most abundant infiltrated leukocyte in tumors and have two functionally polarized phenotypes in the TME: classically activated M1, and alternatively activated M2, macrophages (37). In the initial stages of various tumors, TAMs are preferentially polarized toward the M1 phenotype, producing abundant proinflammatory cytokines, including IL-12 and TNF, and exerting anti-tumor functions (38); however, on cancer progression and changes in the TME, TAMs, driven by tumor cell- and T cell-derived cytokines, including IL-4, IL-13, and IL-10, gradually acquire a polarized M2 phenotype, expressing mannose and the scavenger receptors, CD163 and CD204, and exhibit distinct functional properties that promote angiogenesis, as well as tissue remodeling and repair (37). The downstream signaling pathways activated by the numerous proteins and molecules produced by tumor cells and TAMs in the TME can increase TAM infiltration in EC, which is correlated with unfavorable prognosis and OS (39). For example, cysteine-rich angiogenic inducer 61 (Cyr61) from tumor cells and TAMs may contribute to the increase in $CD204^+$ TAMs *via* MEK/ERK pathway activation in ESCC TME (40). Cancer cell-derived fibroblast growth factor 2 (FGF2) can facilitate TAM survival and migration through AKT/ERK signaling, activated by neural cell adhesion molecule 1-enhanced classical FGF receptor 1 (FGFR1) and intracellular FGF2/FGFR1 signaling. These tumor-infiltrating TAM are skewed toward $CD163^+$ M2 phenotype under the action of the transcription factor, GATA3, and cytokines, including IL-4, IL-6, and IL-13, and promoted an immunosuppressive TME in EAC (33). The similar event happened in ESCC, for that $CD68^+PD\text{-}1^+$ TAMs in the ESCC TME are skewed toward an M2 phenotype (41), which can lead to elevation of tumor cell PD-L1 expression and promote tumor cell invasion and migration, associating with poor OS (42). Moreover, a study of EC patients who received neoadjuvant chemotherapy (NAC) followed by surgery demonstrated that high tumor $CD163^+$ M2 macrophage infiltration is an independent predictor of response to NAC, and associated with poor prognosis and OS (43).

In fact, TAMs are involved in a variety of pathways that promote tumor progression of EC. The activation of the AKT/ERK pathway is the driving force to promote tumor cell growth, migration and invasion in EC (39, 44). This AKT/ERK pathway can be triggered by multiple factors derived from TAMs or cancer cells themselves, involving the FGF2/FGFR1 signaling we mentioned above (39), growth differentiation factor 15 induced in TAMs and derived from cancer cells (possibly through TGF-β type II receptor) (39, 44), overexpression of ANXA10 by cancer cells interacting with $CD204^+$ TAMs (41), and high CXCL8 expression in TAMs and cancer cells (through the CXCL8-CXCR1/CXCR2 axis) (42), which are closely correlated with tumor invasion depth, lymph node metastasis and poor prognosis and OS of ESCC patients.

Another mechanism that the increased $CD163^+$ TAM in the TME promote ESCC tumor progression is that they can augment angiogenesis by releasing thymidine phosphorylase (TP) under the influence of macrophage chemotactic protein-1(MCP-1) (45, 46), inducing vascular endothelial growth factor (VEGF)

expression in EC cells (45, 47), and promoting stromal cell matrix metalloproteinase 9 (MMP9) production (48). $CD163^+$ TAM distribution in tumor sites is closely related to EMT (49), possibly because their IL-1β production can enhance EMT, promoting tumor cell migration and invasion (50). Additionally, MCP-1 expression levels in the TME are positively correlated with increased stromal cell and TAM CC chemokine receptor 2 (CCR2) expression, associated with tumor invasion depth, lymph node metastasis, and distant metastasis, and predict poor prognosis in patients with ESCC (45–47).

The majority of research on TAMs in EC has concentrated on ESCC. The clearly increased numbers of TAMs in both tumor structures and stroma is significantly negatively correlated with EC patient survival (43, 51–53). Under the combined action of various factors in esophageal TME, TAMs are gradually skewed toward an M2 phenotype, which is closely associated with angiogenesis and tumor aggressiveness, and thus predicts poor prognosis in patients with EC (**Figure 2**) (48). The M2/M1 macrophage ratio in esophageal tumors can serve as a sensitive indicator predicting lymph node metastasis and patient prognosis (54). To our knowledge, there is currently no TAMs-based immunotherapy strategy in EC. Study of an N-nitrosomethylbenzy-lamine-induced ESCC animal model suggested that CCL2-CCR2 signaling activation participates in TAM recruitment into the TME, which can promote immune evasion and tumor progression through the PD-1/PD-L2 pathway, indicating potential intervention and immunotherapy strategies targeting TAMs in patients with ESCC (47).

DCs

As the main professional APCs, DCs are essential for triggering and regulating antigen-specific immune responses, and closely connect innate and acquired immunity. Human DCs are a heterogeneous population consisting of two types of conventional DC (cDC), cDC1 and cDC2, and plasmacytoid dendritic cell subsets in equilibrium, plus inflammatory DCs, which are generated in response to inflammation, and Langerhans cells (LCs), which originate from embryonic monocytes and can self-renew (55). DCs distributed in the esophageal mucosa are generally LCs, which remain in an immature immune state under normal conditions, rapidly maturing into professional APCs after capturing pathogen or tumor associated-antigens, to trigger T cell activation and immune responses (56). Mature DCs express various important markers, including CD80, CD86, and CD208. DCs at different stages of maturity are uniquely distributed in the ESCC TME. Abundant $CD1\alpha^+$ immature DCs are distributed in the cancerous epithelium, while fewer $CD208^+$ mature DCs are present in the tumor stroma, particularly the peri-tumoral region (57). In ESCC, DC density indicates the immune defense status

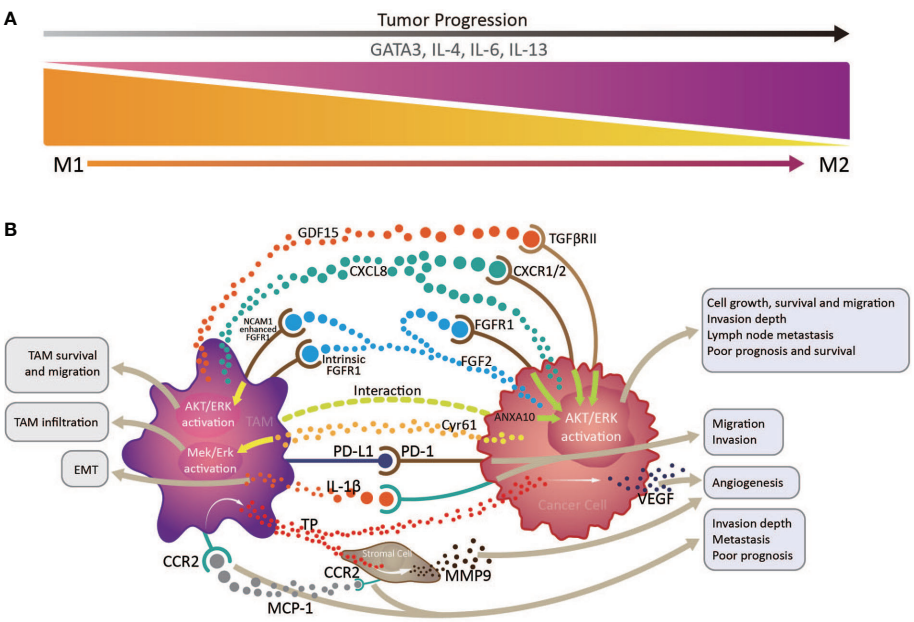


FIGURE 2 | Interactions between tumor infiltrating macrophages, cancer cells, and stromal cells *via* multiple immune factors in the esophageal cancer tumor microenvironment. **(A)** tumor infiltrating macrophages (TAMs) transform from an anti-tumor M1 phenotype to a pro-tumor M2 phenotype with tumor progression, under the influence of the transcription factor, GATA3, and cytokines, including IL-4, IL-6, and IL-13. **(B)** By producing various immune factors, or interacting with cancer cells and stromal cells to promote their release of associated components, TAMs are closely correlated with epithelial to mesenchymal transition, angiogenesis, cancer cell survival and migration, invasion depth, and lymph node metastasis, and generally predict poor prognosis and survival in patients with esophageal cancer.

of the host against the carcinoma, as patients with marked DC infiltration in tumors survive for longer than those with low DC density (58).

DC maturity in the TME is closely related to cancer progression (59). Importantly, the amount of intratumor mature DCs expressing lysosome-associated membrane glycoprotein 3 (LAMP3), is closely associated with tumor-infiltrated $CD8^+$ T cell numbers, which predict favorable prognosis in patients with ESCC (60). In clinical studies, preoperative chemoradiotherapy was shown to lead to significant increases high-mobility group box 1 (HMGB1) protein levels in the ESCC TME. HMGB1 is closely related to DC maturation and positively correlated with patient survival (61). Additionally, by stimulating cancer cells to release inflammatory chemokines (CCL2 or CCL20), IL-17A-producing cells can enhance $CD1a^+$ DC infiltration of the TME, which is correlated with favorable OS of patients with ESCC (18). These reports highlight the connections between DC maturity and EC tumorigenesis and development.

Interactions of tumor and immune cells with the cytokines/chemokines they produce in the EC TME generate complex regulatory networks, which significantly impact DC phenotype and function, thereby influencing tumor progression. DC immune functions are impaired both in the circulation and at tumor sites in patients with EC, and this is accompanied by decreased CD80 and CD86 expression (62). The reduced activity and function of these infiltrated DCs involves mutated p53 protein overexpression in tumors (63). Compared with benign Barrett's esophagus (BE), DC density is dramatically increased in adenocarcinoma (64), along with decreased C1q expression, which contributes to immune complex capture and subsequent classical complement activation pathway initiation, indicating the potential roles of DCs in EAC dysplasia and tumorigenesis (65). Esophageal tumor cells can also induce production of the tryptophan-catabolizing enzyme, indoleamine 2,3-dioxygenase 1 (IDO1) and/or expression of PD-L1 by immunosuppressive DCs, which can promote immune tolerance by inhibiting $CD8^+$ T cell infiltration and inducing immunosuppressive Tregs, and are associated with poor prognosis in EC patients (3, 66–68). The density of receptor-binding cancer antigen expression on SiSo cells (RCAS1) in esophageal tumor tissues with dramatic DC infiltration was inclined to accompany a decrease in TILs, suggesting that RCAS1 can promote tumor cell escape from immune surveillance by inducing DC-activated TIL apoptosis (69). Further, CD47, a cell transmembrane protein expressed in ESCC tumor cells, can inhibit $CD8^+$ T-cell infiltration and anti-tumor immune responses in a DC-dependent manner, by interacting with signal regulatory protein-α (SIRPα), expressed in DCs (70).

MDSCs

MDSCs are generated in the bone marrow and rapidly differentiate into macrophages, DCs, neutrophils, eosinophils, basophils, and mast cells in healthy individuals; however, when cancer occurs, MDSCs can migrate into peripheral lymphoid tissues and tumor sites, contributing to TAM formation (71, 72). MDSCs can be polymorphonuclear (PMN-MDSCs) or mononuclear (M-MDSCs). PMN-MDSCs are morphologically and phenotypically similar to neutrophils, while M-MDSCs are similar to monocytes (71); both have immunosuppressive functions, mainly targeting T cells through Arginase-1 and inducible nitric oxide synthase-2 (iNOS) (73).

In ESCC patients, circulating MDSC numbers are elevated, accompanied by high PD-L1 expression (74). Concurrently, MDSC-derived TGF-β can induce high PD-1 expression on $CD8^+$ T cells in the TME (75). Hence, MDSCs can exert immunosuppressive functions on T cells *via* the PD-1/PD-L1 pathway and are correlated with tumor burden, lymph node metastasis, and tumor stage (74). Importantly, IL-6 exerts vital roles in MDSC induction and their production of ROS, Arginase 1, and p-STAT3 (76). Circulating MDSC numbers and IL-6 levels in the TME are positively correlated with NLR, predicting poor OS in ESCC patients (77). The cell-cell junctions formed by interaction between p120ctn and E-cadherin are critical in maintaining normal esophageal epithelial homeostasis; however, p120ctn expression in the ESCC TME is decreased or absent, leading to E-cadherin degradation and NF-κB, AKT, and STAT3 phosphorylation in cancer cells, promoting cancer cell GM-CSF release, which can recruit MDSCs into the TME. NF-κB signaling activation in MDSCs up-regulates their IL-4RA expression and nitric oxide production, thereby inhibiting $CD8^+$ T cell cytotoxicity and contributing to a TME conducive to tumor cell growth (78). Further, in a conditional p120-ctn knockout mouse model of oral-EC, expression of the receptor CD38 induced by tumor-derived IL-6, IGFBP-3, and CXCL16, promoted arrest of MDSC maturation in an immature state, with stronger inhibitory functions of activated T cells through production of iNOS, among other factors, thus promoting tumor growth (79).

NEUTROPHILS

Tumor-associated neutrophils (TANs) have a different phenotype and cell/chemokine activity from circulating neutrophils (80). TANs can be functionally divided into anti-tumor N1 and cancer-promoting N2 phenotypes. TGF-β in the TME contributes to the transformation of neutrophils from N1 to N2 (80, 81). Further, neutrophil phenotype and function in the TME change with tumor development. In early stage tumors, neutrophils are only on the tumor periphery and exhibit anti-tumor effects, while in later stages, they can penetrate the tumor and demonstrate pro-tumor effects (82). Studies on neutrophils in cancer have focused on the neutrophil to lymphocyte ratio (NLR), as it is impossible to classify N1/N2 neutrophils using surface markers (83). The NLR is usually derived from routine blood tests, and may reflect changes in the TME and systemic inflammation status (84), which are independent prognostic indicators in patients with EC (85). NLR and platelet-to-lymphocyte ratio are associated with EC progression (86), and

elevated preoperative NLR is related to lymph node metastasis, deeper tumor invasion, and advanced TNM stage (87), predicting poor prognosis and OS in patients who have undergone esophagectomy (85, 88, 89). In addition to defending against microbial invasion through phagocytosis and degranulation, neutrophils can undergo apoptosis after activation, and then form neutrophil extracellular traps (NETs), fibrinoid structures comprising extracellular chromatin and granulocyte proteins, including myeloperoxidase and neutrophil elastase, which were discovered because of their pathogen-trapping function, which can promote tumor metastasis by capturing circulating tumor cells and causing their proliferation at a second site (83, 90). In EC patients without surgical stress or any other stimuli, tumors alone can induce high levels of circulating NETs, which are predictive of positive lymph node status, distant metastasis, and advanced disease stage (91).

The functions of neutrophils in the TME, influenced by various cytokines and/or chemokines, are controversial (92). By changing the esophageal microenvironment and gut bacteria, a high-fat diet can cause esophageal dysplasia, which promotes the development of BE into EAC, which involves IL-8 chemokine family activation and neutrophil recruitment, along with NK cell reduction, suggesting that increased neutrophils may inhibit NK cell-mediated tumor cell cytotoxicity and indicate poor prognosis (93). Conversely, in the ESCC TME, IL-17 [mainly produced by $CD4^+$ $Foxp3^-$ Th17 cells (17)] stimulates tumor cells chemokine (CXCL2 and CXCL3) production, causing accumulation and activation of $myeloperoxidase^+$ TANs, which increase their killing capacity by releasing cytotoxic molecules, including IFN-γ, reactive oxygen species (ROS), and TNF-related apoptosis-inducing ligand, and predict favorable prognosis in ESCC patients (94).

MCs AND EOSINOPHILS

MCs mainly localize to areas where organisms are likely to come into contact with pathogens or harmful substances, including the gastrointestinal tract, respiratory mucosa, and skin, and act as multifunctional immune cells involved in both innate and adaptive immunity in health and various disease states (95). In several human cancers, MCs recruited by stem cell factor or other mast cell activators in the TME, release angiogenic factors and proteases to promote blood vessel formation and degrade the extracellular matrix, leading to tumor cell invasion; however, they can also release ROS/TNF-α, tryptase, heparin, IL-1, IL-4, and IL-6, among other factors, to inhibit tumor cell growth and apoptosis (96). In the ESCC TME, high MC density is positively correlated with tumor angiogenesis, lymph node metastasis, invasion depth, and tumor progression (97, 98), and a predictor of poor survival in ESCC patients (97, 99). Furthermore, activated MCs in EC tissue express high levels of tissue kallikrein (TK1), which may subsequently generate mitogenic kinin, a promoter of tumor cell growth (100). Interestingly, another study found that the presence of a group of MCs able to produce IL-17 in the esophageal muscularis propria, rather than in tumor nests, is positively correlated with the level of activated $CD169^+$ macrophages and effector $CD8^+$ T cells in the same region, indicating favorable prognosis and survival (101). It is noteworthy that IL-17 released from Th17 cells in the EC TME, as we mentioned above, was involved in recruiting cells with anti-tumor effects including NK cells, neutrophils, and $CD1\alpha^+$ DCs, so whether MCs participated in this process deserves further attention.

Eosinophils usually cluster together with MCs in tissue sites under both homeostatic and inflammatory conditions (102). With similar developmental and functional patterns, such as releasing cationic proteins pre-stored in cytoplasmic granules by degranulation upon activation, they often participate in host responses to helminth infection and allergic disease in a synergistic manner (102, 103). Based on their abilities to release cytokines, eosinophils are being recognized to be also involved in local immunity, tissue homeostasis, remodeling, and repair in multiple previously unexpected tissues, especially the mucosal tissues such as gut and esophageal (104–106). For example, eosinophil infiltration is a typical feature of eosinophilic esophagitis, an allergic disease associated with epithelial barrier dysfunction and chronic type 2 inflammation (106). Other than this, the increase of eosinophils is also found in some ESCC patients (107), positively correlating with low incidence of LN metastasis in patients with early ESCC and predicting favorable OS in ESCC patients treated with concurrent chemoradiotherapy (108, 109). Conversely, eosinophils appear to be significantly reduced across all stages of dysplasia and EAC progression, indicating the loss of immune surveillance by eosinophils may contribute to BE progression toward dysplasia and cancer (110). Indeed, eosinophils play controversial roles in modulating tumor initiation and progression, for that they are both the source of anti-tumorigenic factors including TNF-a, granzyme, cationic proteins, and IL-18, and protumorigenic molecules such as pro-angiogenic factors, depending on the different immune milieu (35). Nevertheless, how eosinophils exert their functions in the occurrence and development of EC remains unclear. There is still a long way to go to understand the specific mechanism of eosinophils in EC, which will shed light on the control of EC progression and the immunotherapy based on them.

CROSSTALK AND REGULATION OF INNATE IMMUNE CELLS

The EC TME contains a various innate immune cells and associated cytokines/chemokines. By regulating or being regulated, innate immune cells and diverse other cell populations, including adaptive immune cells, stromal cells, and cancer cells, form complex regulatory networks through receptor-ligand binding and immune factor release in the TME, which influences the proliferation, migration, and invasion of cancer cells, as well as angiogenesis, thus influencing tumor growth and metastasis (**Figure 3**).

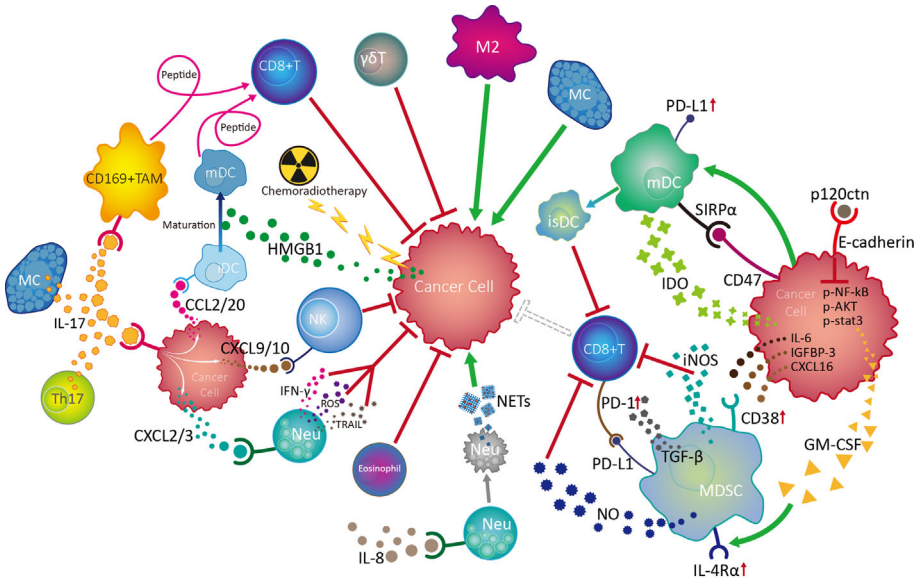


FIGURE 3 | Crosstalk and regulation of innate immune cells in the esophageal cancer tumor microenvironment. Among the multiple innate immune cells involved in esophageal cancer progression, NK and γδT cells are active in the front line of anti-tumor defences with their powerful cytotoxicity. Mature dendritic cells (mDC) play vital roles in antitumour responses by boosting the function of CD8+ effector T cells, while chemoradiotherapy can promote DC maturity by increasing HMGB1 levels in the tumor microenvironment (TME). IL-17 derived from Th17 cells and MCs can activate CD169+ tumor infiltrating macrophages (TAMs) and effector CD8+ T cells, as well as recruit NK cells, CD1a+ immature DCs (iDC) and neutrophils (Neu) into the TME by stimulating cancer cells to release various chemokines, thereby exerting antitumour effects. Although the mechanism is unknown, eosinophils are also involved in the anti-tumor process. On the contrary, neutrophils can also promote tumor development by inhibiting NK cell function in response to IL-8, as well as by forming NETs after tumor-induced apoptosis. MCs can promote tumor cell growth through the TK1/mitogenic kinin pathway. Importantly, tumor cells can escape from innate immune surveillance by promoting TAM progression to a suppressive M2 phenotype, which inhibits $CD8^+$ T cell function by transforming DCs into an immunosuppressive phenotype (isDC) and recruiting myeloid-derived suppressor cells to inhibit the cytotoxic effects of $CD8^+$ T and NK cells.

The balance of activating and inhibitory receptor expression, cytokine release, and degranulation ability endow NK cells with powerful immune surveillance and direct killing functions toward tumor cells; however, NK cells exhibit a dysfunctional phenotype and exhaustion on tumor development of EC, involving cancer cell-induced over-expression of inhibitory receptors, such as PD-1, and suppression of activated receptors, such as NKp30, tumor infiltrating macrophage-induced $CD56^{dim}$ NK cell apoptosis *via* H_2O_2, and aggregation of neutrophils under the influence of IL-8. As the most common infiltrating immune cells in the tumor milieu, TAMs are polarized into an inflammatory M1 phenotype during early EC, facilitating $CD8^+$ T cell activation and exerting anti-tumor effects. As EC develops, TAMs gradually transform into an M2 phenotype, promoting tumor cell proliferation, migration, and invasion. DC maturation in the TME can be promoted by HMGB1 stimulated by preoperative chemoradiotherapy and is accompanied by surface molecule (CD80, CD86, CD208, and LAMP3) expression with anti-tumor functions of tumor-associated antigen presentation, which facilitates effective $CD8^+$ T cell activation. Nevertheless, DC activity and function can be damaged by cancer cells overexpressing mutated p53. EC tumor cells can also escape from immune surveillance by inducing DC-triggered TIL apoptosis under the influence of RCAS1, or by promoting the DC transformation into an immunosuppressive phenotype, with PD-L1 and SIRPα expression and IDO1 production, inhibiting $CD8^+$ T cell activity. Moreover, tumors can induce high levels of circulating NETs from apoptotic neutrophils, which predict positive lymph node status, distant metastasis, and advanced stage in EC. Additionally, MDSCs in EC TME can be recruited and activated by cancer cell-derived cytokines/chemokines (IL-6, IGFBP-3, CXCL16, and GM-CSF), thereby accelerating tumor progression by inhibiting $CD8^+$ T cell activation. Controversially, MCs in EC can both promote tumor cell growth through TK1/mitogenic kinin signaling and exert anti-tumor effects, through activation of $CD169^+$ TAMs and $CD8^+$ effector cells. Importantly, IL-17 from Th17 cells and MCs in EC TME can promote cancer cell release of CXCL9/10, CXCL2/3, and CCL2/20, which can increase the infiltration and anti-tumor effects of NK cells, neutrophils, and $CD1\alpha^+$ DCs, respectively, and predict favorable prognosis for patients with esophageal tumors.

IMMUNOTHERAPY STRATEGY BASED ON INNATE IMMUNE CELLS

Treatment for EC remains less than satisfactory. Recent studies indicate that treatment with a single PD-1 inhibition agent is more effective for ESCC than EAC, while a combination of inhibitors targeting PD-1 with chemotherapy is a good strategy

for treatment of metastatic disease (68); however, partly due to a lack of reliable predictive indicators of whether patients respond effectively, the use PD-1 inhibitors, alone or with other checkpoint antibodies, has had controversial results (36). Problems remain for use of PD-L1 as a predictive biomarker, because of tumor heterogeneity, a lack of reproducibility of results, and a complex scoring system (68). Therefore, it is imperative to identify new predictive indicators and immunotherapy strategies.

Various tumor treatment regimens aim to enhance effector cell function and/or control immunosuppression (9). Hence, there is potential to treat EC by boosting innate immune functions, such as NK cell cytotoxicity, phagocytosis, and DC maturation, which subsequently activate and sustain tumor-specific $CD8^+$ T cell effects. First, the development of broad-spectrum immune checkpoint inhibitors targeting NK cells (i.e., LAG3, TIM3, and PD-1, which we review here, and NKG2A, CTLA4, TIGIT, and CD96 which require further investigation) and/or myeloid cells (i.e., SIRPα), is a promising approach that should be advanced in EC therapy. NK cells share the majority of checkpoints with T cells; therefore, inhibition of these receptors will also release various brakes on T cells and benefit both innate immunity and T-cell functions. Second, the development of anti-tumor antibodies that can bind to activating FcRs expressed on innate immune cells lacking antigen receptors, such as NK cells and macrophages, will enable them to act specifically on EC cells. Third, multiple pattern recognition receptors expressed on the surface of innate immune cells in mucosal sites ensure rapid responses to pathogenic microorganisms by recognizing pathogen-related molecular patterns. Therefore, targeting toll-like receptors (TLRs), RIG-I-like receptors (RLRs), and stimulators of interferon, to generate a 'pathogen-induced-like' innate immune responses at the tumor site, may be promising approaches in EC treatment, since the innate immune system can sense the nucleic acids of growing tumors using the pathogen and damage receptors involved in infection detection. Finally, generating engineered CAR NK cells with high anti-tumor activity and CAR macrophages which can be polarized towards an anti-tumor M1 phenotype, in addition to, or instead of, CAR T cells, may provide a route to next generation immunotherapies for EC (**Figure 4**).

In conclusion, the application of these methods to clinical treatment is based on sufficient research of innate immune functions in the EC TME and numerous preclinical trials. Indeed, combined therapy approaches may become the norm in future treatment of EC.

PERSPECTIVE

The roles of innate immune cells in mucosal tissues in maintaining regional homeostasis and in host resistance to infection and tumor has been extensively elaborated. However, partly due to specimen constraints and regional disparities in incidence (i.e., its higher prevalence in developing countries, particularly East Asia) (2), few studies have focused on innate immunity in EC, and ongoing immunotherapy of patients with esophageal tumors is almost entirely restricted to targeting of the PD1/PD-L1 pathway (68). Based on the limited available data, one limitation of this review is that we cannot comprehensively compare the similarity or difference of innate immune cells between ESCC and EAC.

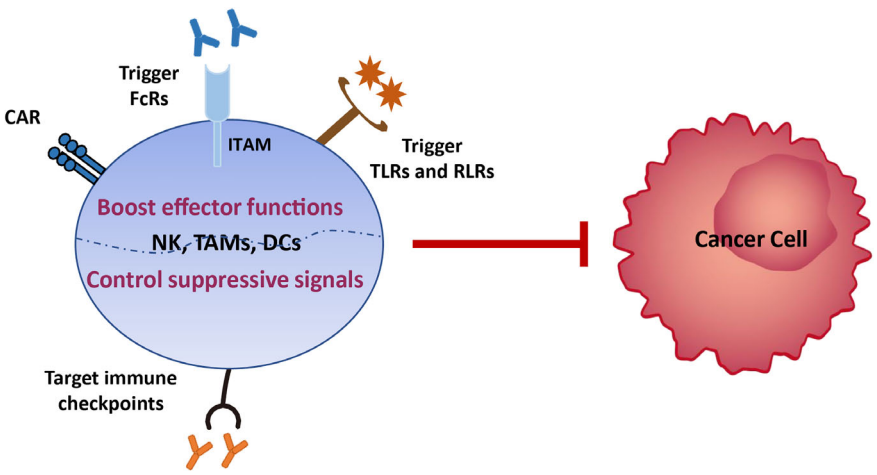


FIGURE 4 | Immunotherapy strategies based on innate immune cells. Various tumor treatment regimens can be implemented by boosting the effector functions of innate immune cells, including generating engineered CAR NK cells with high anti-tumor activity and CAR macrophages which can be polarized towards an anti-tumor M1 phenotype, developing anti-tumor antibodies that can bind to activating FcRs expressed on innate immune cells lacking antigen receptors, and enabling them to act specifically on EC cells, and targeting toll-like receptors (TLRs) and RIG-I-like receptors (RLRs), to generate a 'pathogen-induced-like' innate immune responses at the tumor site, since the innate immune system can sense the nucleic acids of growing tumors using the pathogen and damage receptors involved in infection detection. Another immunotherapy strategy is to control immunosuppression signals on innate immune cells *via* development of broad-spectrum immune checkpoint inhibitors targeting NK cells (i.e., LAG3, TIM3, and PD-1, which we review here, and NKG2A, CTLA4, TIGIT, and CD96 which require further investigation) and/or myeloid cells (i.e., SIRPα).

Although previous research provides clues to the essential roles of various innate immune cells in the EC TME, considerable further investigations of their functions in EC occurrence and development are required. It is worth noting that the inadequacy and imbalance of the previous studies may lead to incomplete evaluation of the complicated immune contexture of EC. For example, γδT and NKT cells, two typical innate immune cells which are deeply involved in anti-tumor responses to multiple cancers, have rarely been studied in EC. Recently, a group of innate lymphocytes, innate lymphoid cells (ILCs), were identified. ILCs can be subdivided into ILC1, ILC2, and ILC3, subtypes, based on cytokine production and transcription factors associated with their development (111). Alternatively, ILCs can be classified as cytotoxic (i.e., conventional NK cells) and helper ILCs, which resemble the T cell classification (i.e., $CD8^+$ cytotoxic T and $CD4^+$ T helper cells) (112). In addition to circulating cytotoxic NK cells, ILCs exhibit clear tissue tropism, preferentially localizing to barrier tissues, including the lung, intestine, and skin, and involving in inflammation and carcinogenesis (112). Whether ILCs in the esophageal mucosa participate in the development of esophageal diseases, such as BE and EC remains unknown; hence, the functions of these cells in EC warrants attention in future preclinical and clinical studies. Overall, a more comprehensive understanding of the roles of innate immune cell populations in EC and identification of better treatment targets will likely ultimately benefit patients with EC.

AUTHOR CONTRIBUTIONS

KC designed and drafted the manuscript. SH participated in writing the manuscript. XM provided and contributed fruitful discussion. MC started the study and participated in the paper writing. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by Natural Science Foundation of Anhui Province (2008085MH277), National Natural Science Foundation of China (81471552), the Anhui Provincial Project of the Key Laboratory of Tumor Immunotherapy and Nutrition Therapy (2019b12030026).

REFERENCES

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global Cancer Statistics 2018: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin* (2018) 68:394–424. doi: 10.3322/caac.21492
2. Collaborators GBDOC. The Global, Regional, and National Burden of Esophageal Cancer and its Attributable Risk Factors in 195 Countries and Territories, 1990-2017: A Systematic Analysis for the Global Burden of Disease Study 2017. *Lancet Gastroenterol Hepatol* (2020) 5:582–97. doi: 10.1016/S2468-1253(20)30007-8
3. Huang TX, Fu L. The Immune Landscape of Esophageal Cancer. *Cancer Commun (Lond)* (2019) 39:79. doi: 10.1186/s40880-019-0427-z
4. Short MW, Burgers KG, Fry VT. Esophageal Cancer. *Am Fam Physician* (2017) 95:22–8.
5. Bruno A, Pagani A, Pulze L, Albini A, Dallaglio K, Noonan DM, et al. Orchestration of Angiogenesis by Immune Cells. *Front Oncol* (2014) 4:131. doi: 10.3389/fonc.2014.00131
6. Lin EW, Karakasheva TA, Hicks PD, Bass AJ, Rustgi AK. The Tumor Microenvironment in Esophageal Cancer. *Oncogene* (2016) 35:5337–49. doi: 10.1038/onc.2016.34
7. Pradeu T, Jaeger S, Vivier E. The Speed of Change: Towards a Discontinuity Theory of Immunity? *Nat Rev Immunol* (2013) 13:764–9. doi: 10.1038/nri3521
8. Dhupar R, Van Der Kraak L, Pennathur A, Schuchert MJ, Nason KS, Luketich JD, et al. Targeting Immune Checkpoints in Esophageal Cancer: A High Mutational Load Tumor. *Ann Thorac Surg* (2017) 103:1340–9. doi: 10.1016/j.athoracsur.2016.12.011
9. Demaria O, Cornen S, Daeron M, Morel Y, Medzhitov R, Vivier E. Harnessing Innate Immunity in Cancer Therapy. *Nature* (2019) 574:45–56. doi: 10.1038/s41586-019-1593-5
10. Vesely MD, Kershaw MH, Schreiber RD, Smyth MJ. Natural Innate and Adaptive Immunity to Cancer. *Annu Rev Immunol* (2011) 29:235–71. doi: 10.1146/annurev-immunol-031210-101324
11. Bassani B, Baci D, Gallazzi M, Poggi A, Bruno A, Mortara L. Natural Killer Cells as Key Players of Tumor Progression and Angiogenesis: Old and Novel Tools to Divert Their Pro-Tumor Activities Into Potent Anti-Tumor Effects. *Cancers (Basel)* (2019) 11:461. doi: 10.3390/cancers11040461
12. Hinshaw DC, Shevde LA. The Tumor Microenvironment Innately Modulates Cancer Progression. *Cancer Res* (2019) 79:4557–66. doi: 10.1158/0008-5472.CAN-18-3962
13. Sun H, Sun C. The Rise of NK Cell Checkpoints as Promising Therapeutic Targets in Cancer Immunotherapy. *Front Immunol* (2019) 10:2354. doi: 10.3389/fimmu.2019.02354
14. Melaiu O, Lucarini V, Cifaldi L, Fruci D. Influence of the Tumor Microenvironment on NK Cell Function in Solid Tumors. *Front Immunol* (2019) 10:3038. doi: 10.3389/fimmu.2019.03038
15. Souza-Fonseca-Guimaraes F, Cursons J, Huntington ND. The Emergence of Natural Killer Cells as a Major Target in Cancer Immunotherapy. *Trends Immunol* (2019) 40:142–58. doi: 10.1016/j.it.2018.12.003
16. Wang J, Jin X, Liu J, Zhao K, Xu H, Wen J, et al. The Prognostic Value of B7-H6 Protein Expression in Human Oral Squamous Cell Carcinoma. *J Oral Pathol Med* (2017) 46:766–72. doi: 10.1111/jop.12586
17. Lv L, Pan K, Li XD, She KL, Zhao JJ, Wang W, et al. The Accumulation and Prognosis Value of Tumor Infiltrating IL-17 Producing Cells in Esophageal Squamous Cell Carcinoma. *PLoS One* (2011) 6:e18219. doi: 10.1371/journal.pone.0018219
18. Lu L, Pan K, Zheng HX, Li JJ, Qiu HJ, Zhao JJ, et al. Il-17A Promotes Immune Cell Recruitment in Human Esophageal Cancers and the Infiltrating Dendritic Cells Represent a Positive Prognostic Marker for Patient Survival. *J Immunother* (2013) 36:451–8. doi: 10.1097/CJI.0b013e3182a802cf
19. Zheng Y, Li Y, Lian J, Yang H, Li F, Zhao S, et al. TNF-Alpha-Induced Tim-3 Expression Marks the Dysfunction of Infiltrating Natural Killer Cells in Human Esophageal Cancer. *J Transl Med* (2019) 17:165. doi: 10.1186/s12967-019-1917-0
20. Liu Y, Cheng Y, Xu Y, Wang Z, Du X, Li C, et al. Increased Expression of Programmed Cell Death Protein 1 on NK Cells Inhibits NK-cell-mediated Anti-Tumor Function and Indicates Poor Prognosis in Digestive Cancers. *Oncogene* (2017) 36:6143–53. doi: 10.1038/onc.2017.209
21. Izawa S, Kono K, Mimura K, Kawaguchi Y, Watanabe M, Maruyama T, et al. H(2)O(2) Production Within Tumor Microenvironment Inversely Correlated With Infiltration of CD56(dim) NK Cells in Gastric and Esophageal Cancer: Possible Mechanisms of NK Cell Dysfunction. *Cancer Immunol Immunother* (2011) 60:1801–10. doi: 10.1007/s00262-011-1082-7
22. Conroy MJ, Fitzgerald V, Doyle SL, Channon S, Useckaite Z, Gilmartin N, et al. The Microenvironment of Visceral Adipose Tissue and Liver Alter

Natural Killer Cell Viability and Function. *J Leukoc Biol* (2016) 100:1435–42. doi: 10.1189/jlb.5AB1115-493RR
23. Watanabe M, Kono K, Kawaguchi Y, Mizukami Y, Mimura K, Maruyama T, et al. NK Cell Dysfunction With Down-Regulated CD16 and Up-Regulated CD56 Molecules in Patients With Esophageal Squamous Cell Carcinoma. *Dis Esophagus* (2010) 23:675–81. doi: 10.1111/j.1442-2050.2010.01073.x
24. Zheng Y, Chen Z, Han Y, Han L, Zou X, Zhou B, et al. Immune Suppressive Landscape in the Human Esophageal Squamous Cell Carcinoma Microenvironment. *Nat Commun* (2020) 11:6268. doi: 10.1038/s41467-020-20019-0
25. Sun H, Huang Q, Huang M, Wen H, Lin R, Zheng M, et al. Human CD96 Correlates to Natural Killer Cell Exhaustion and Predicts the Prognosis of Human Hepatocellular Carcinoma. *Hepatology* (2019) 70:168–83. doi: 10.1002/hep.30347
26. Gallazzi M, Baci D, Mortara L, Bosi A, Buono G, Naselli A, et al. Prostate Cancer Peripheral Blood Nk Cells Show Enhanced Cd9, CD49a, Cxcr4, CXCL8, Mmp-9 Production and Secrete Monocyte-Recruiting and Polarizing Factors. *Front Immunol* (2020) 11:586126. doi: 10.3389/fimmu.2020.586126
27. Li J, Qiu G, Fang B, Dai X, Cai J. Deficiency of IL-18 Aggravates Esophageal Carcinoma Through Inhibiting IFN-gamma Production by CD8(+)T Cells and NK Cells. *Inflammation* (2018) 41:667–76. doi: 10.1007/s10753-017-0721-3
28. Zou C, Zhao P, Xiao Z, Han X, Fu F, Fu L. Gammadelta T Cells in Cancer Immunotherapy. *Oncotarget* (2017) 8:8900–9. doi: 10.18632/oncotarget.13051
29. Wu D, Wu P, Qiu F, Wei Q, Huang J. Human gammadeltaT-cell Subsets and Their Involvement in Tumor Immunity. *Cell Mol Immunol* (2017) 14:245–53. doi: 10.1038/cmi.2016.55
30. Thomas ML, Samant UC, Deshpande RK, Chiplunkar SV. Gammadelta T Cells Lyse Autologous and Allogenic Esophageal Tumors: Involvement of Heat-Shock Proteins in the Tumor Cell Lysis. *Cancer Immunol Immunother* (2000) 48:653–9. doi: 10.1007/s002620050014
31. Paul S, Shilpi, Lal G. Role of Gamma-Delta (Gammadelta) T Cells in Autoimmunity. *J Leukoc Biol* (2015) 97:259–71. doi: 10.1189/jlb.3RU0914-443R
32. McKenzie DR, Comerford I, Silva-Santos B, McColl SR. The Emerging Complexity of gammadeltaT17 Cells. *Front Immunol* (2018) 9:796. doi: 10.3389/fimmu.2018.00796
33. Cheng M, Qian L, Shen G, Bian G, Xu T, Xu W, et al. Microbiota Modulate Tumoral Immune Surveillance in Lung Through a gammadeltaT17 Immune Cell-Dependent Mechanism. *Cancer Res* (2014) 74:4030–41. doi: 10.1158/0008-5472.CAN-13-2462
34. Cheng M, Hu S. Lung-Resident Gammadelta T Cells and Their Roles in Lung Diseases. *Immunology* (2017) 151:375–84. doi: 10.1111/imm.12764
35. Varricchi G, Galdiero MR, Loffredo S, Lucarini V, Marone G, Mattei F, et al. Eosinophils: The Unsung Heroes in Cancer? *Oncoimmunology* (2018) 7:e1393134. doi: 10.1080/2162402X.2017.1393134
36. Baba Y, Nomoto D, Okadome K, Ishimoto T, Iwatsuki M, Miyamoto Y, et al. Tumor Immune Microenvironment and Immune Checkpoint Inhibitors in Esophageal Squamous Cell Carcinoma. *Cancer Sci* (2020) 111:3132–41. doi: 10.1111/cas.14541
37. Mantovani A, Sozzani S, Locati M, Allavena P, Sica A. Macrophage Polarization: Tumor-Associated Macrophages as a Paradigm for Polarized M2 Mononuclear Phagocytes. *Trends Immunol* (2002) 23:549–55. doi: 10.1016/S1471-4906(02)02302-5
38. Vitale M, Cantoni C, Pietra G, Mingari MC, Moretta L. Effect of Tumor Cells and Tumor Microenvironment on NK-cell Function. *Eur J Immunol* (2014) 44:1582–92. doi: 10.1002/eji.201344272
39. Yokozaki H, Koma YI, Shigeoka M, Nishio M. Cancer as a Tissue: The Significance of Cancer-Stromal Interactions in the Development, Morphogenesis and Progression of Human Upper Digestive Tract Cancer. *Pathol Int* (2018) 68:334–52. doi: 10.1111/pin.12674
40. Shigeoka M, Urakawa N, Nishio M, Takase N, Utsunomiya S, Akiyama H, et al. Cyr61 Promotes CD204 Expression and the Migration of Macrophages Via MEK/ERK Pathway in Esophageal Squamous Cell Carcinoma. *Cancer Med* (2015) 4:437–46. doi: 10.1002/cam4.401
41. Kodaira H, Koma YI, Hosono M, Higashino N, Suemune K, Nishio M, et al. ANXA10 Induction by Interaction With Tumor-Associated Macrophages Promotes the Growth of Esophageal Squamous Cell Carcinoma. *Pathol Int* (2019) 69:135–47. doi: 10.1111/pin.12771
42. Hosono M, Koma YI, Takase N, Urakawa N, Higashino N, Suemune K, et al. CXCL8 Derived From Tumor-Associated Macrophages and Esophageal Squamous Cell Carcinomas Contributes to Tumor Progression by Promoting Migration and Invasion of Cancer Cells. *Oncotarget* (2017) 8:106071–88. doi: 10.18632/oncotarget.22526
43. Yamamoto K, Makino T, Sato E, Noma T, Urakawa S, Takeoka T, et al. Tumor-Infiltrating M2 Macrophage in Pretreatment Biopsy Sample Predicts Response to Chemotherapy and Survival in Esophageal Cancer. *Cancer Sci* (2020) 111:1103–12. doi: 10.1111/cas.14328
44. Okamoto M, Koma YI, Kodama T, Nishio M, Shigeoka M, Yokozaki H. Growth Differentiation Factor 15 Promotes Progression of Esophageal Squamous Cell Carcinoma Via TGF-beta Type II Receptor Activation. *Pathobiology* (2020) 87:100–13. doi: 10.1159/000504394
45. Ohta M, Kitadai Y, Tanaka S, Yoshihara M, Yasui W, Mukaida N, et al. Monocyte Chemoattractant Protein-1 Expression Correlates With Macrophage Infiltration and Tumor Vascularity in Human Esophageal Squamous Cell Carcinomas. *Int J Cancer* (2002) 102:220–4. doi: 10.1002/ijc.10705
46. Koide N, Nishio A, Sato T, Sugiyama A, Miyagawa S. Significance of Macrophage Chemoattractant Protein-1 Expression and Macrophage Infiltration in Squamous Cell Carcinoma of the Esophagus. *Am J Gastroenterol* (2004) 99:1667–74. doi: 10.1111/j.1572-0241.2004.30733.x
47. Hu JM, Liu K, Liu JH, Jiang XL, Wang XL, Yang L, et al. The Increased Number of Tumor-Associated Macrophage is Associated With Overexpression of VEGF-C, Plays an Important Role in Kazakh ESCC Invasion and Metastasis. *Exp Mol Pathol* (2017) 102:15–21. doi: 10.1016/j.yexmp.2016.12.001
48. Hu JM, Liu K, Liu JH, Jiang XL, Wang XL, Chen YZ, et al. CD163 as a Marker of M2 Macrophage, Contribute to Predicte Aggressiveness and Prognosis of Kazakh Esophageal Squamous Cell Carcinoma. *Oncotarget* (2017) 8:21526–38. doi: 10.18632/oncotarget.15630
49. Liu J, Li C, Zhang L, Liu K, Jiang X, Wang X, et al. Association of Tumor-Associated Macrophages With Cancer Cell EMT, Invasion, and Metastasis of Kazakh Esophageal Squamous Cell Cancer. *Diagn Pathol* (2019) 14:55. doi: 10.1186/s13000-019-0834-0
50. Zhou J, Zheng S, Liu T, Liu Q, Chen Y, Tan D, et al. Il-1beta From M2 Macrophages Promotes Migration and Invasion of ESCC Cells Enhancing Epithelial-Mesenchymal Transition and Activating NF-kappaB Signaling Pathway. *J Cell Biochem* (2018) 119:7040–52. doi: 10.1002/jcb.26918
51. Li J, Xie Y, Wang X, Li F, Li S, Li M, et al. Prognostic Impact of Tumor-Associated Macrophage Infiltration in Esophageal Cancer: A Meta-Analysis. *Future Oncol* (2019) 15:2303–17. doi: 10.2217/fon-2018-0669
52. Yagi T, Baba Y, Okadome K, Kiyozumi Y, Hiyoshi Y, Ishimoto T, et al. Tumor-Associated Macrophages are Associated With Poor Prognosis and Programmed Death Ligand 1 Expression in Esophageal Cancer. *Eur J Cancer* (2019) 111:38–49. doi: 10.1016/j.ejca.2019.01.018
53. Guo SJ, Lin DM, Li J, Liu RZ, Zhou CX, Wang DM, et al. Tumor-Associated Macrophages and CD3-zeta Expression of Tumor-Infiltrating Lymphocytes in Human Esophageal Squamous-Cell Carcinoma. *Dis Esophagus* (2007) 20:107–16. doi: 10.1111/j.1442-2050.2007.00655.x
54. Cao W, Peters JH, Nieman D, Sharma M, Watson T, Yu J. Macrophage Subtype Predicts Lymph Node Metastasis in Esophageal Adenocarcinoma and Promotes Cancer Cell Invasion In Vitro. *Br J Cancer* (2015) 113:738–46. doi: 10.1038/bjc.2015.292
55. Balan S, Saxena M, Bhardwaj N. Dendritic Cell Subsets and Locations. *Int Rev Cell Mol Biol* (2019) 348:1–68. doi: 10.1016/bs.ircmb.2019.07.004
56. Yang W, Yu J. Immunologic Function of Dendritic Cells in Esophageal Cancer. *Dig Dis Sci* (2008) 53:1739–46. doi: 10.1007/s10620-007-0095-8
57. Liu J, Lu G, Li Z, Tang F, Liu Y, Cui G. Distinct Compartmental Distribution of Mature and Immature Dendritic Cells in Esophageal Squamous Cell Carcinoma. *Pathol Res Pract* (2010) 206:602–6. doi: 10.1016/j.prp.2010.03.011
58. Matsuda H, Mori M, Tsujitani S, Ohno S, Kuwano H, Sugimachi K. Immunohistochemical Evaluation of Squamous Cell Carcinoma Antigen and S-100 Protein-Positive Cells in Human Malignant Esophageal Tissues.

Cancer (1990) 65:2261–5. doi: 10.1002/1097-0142(19900515)65:10<2261::AID-CNCR2820651017>3.0.CO;2-8
59. Tran Janco JM, Lamichhane P, Karyampudi L, Knutson KL. Tumor-Infiltrating Dendritic Cells in Cancer Pathogenesis. *J Immunol* (2015) 194:2985–91. doi: 10.4049/jimmunol.1403134
60. Nishimura J, Tanaka H, Yamakoshi Y, Hiramatsu S, Tamura T, Toyokawa T, et al. Impact of Tumor-Infiltrating LAMP-3 Dendritic Cells on the Prognosis of Esophageal Squamous Cell Carcinoma. *Esophagus* (2019) 16:333–44. doi: 10.1007/s10388-019-00669-w
61. Suzuki Y, Mimura K, Yoshimoto Y, Watanabe M, Ohkubo Y, Izawa S, et al. Immunogenic Tumor Cell Death Induced by Chemoradiotherapy in Patients With Esophageal Squamous Cell Carcinoma. *Cancer Res* (2012) 72:3967–76. doi: 10.1158/0008-5472.CAN-12-0851
62. Chen SR, Luo YP, Zhang JK, Yang W, Zhen ZC, Chen LX, et al. Study on Immune Function of Dendritic Cells in Patients With Esophageal Carcinoma. *World J Gastroenterol* (2004) 10:934–9. doi: 10.3748/wjg.v10.i7.934
63. Ikeguchi M, Ikeda M, Tatebe S, Maeta M, Kaibara N. Clinical Significance of Dendritic Cell Infiltration in Esophageal Squamous Cell Carcinoma. *Oncol Rep* (1998) 5:1185–9. doi: 10.3892/or.5.5.1185
64. Bobryshev YV, Tran D, Killingsworth MC, Buckland M, Lord RV. Dendritic Cells in Barrett's Esophagus and Esophageal Adenocarcinoma. *J Gastrointest Surg* (2009) 13:44–53. doi: 10.1007/s11605-008-0613-9
65. Bobryshev YV, Lu J, Lord RV. Expression of C1q Complement Component in Barrett's Esophagus and Esophageal Adenocarcinoma. *J Gastrointest Surg* (2010) 14:1207–13. doi: 10.1007/s11605-010-1230-y
66. Liu J, Lu G, Tang F, Liu Y, Cui G. Localization of Indoleamine 2,3-Dioxygenase in Human Esophageal Squamous Cell Carcinomas. *Virchows Arch* (2009) 455:441–8. doi: 10.1007/s00428-009-0846-3
67. Kiyozumi Y, Baba Y, Okadome K, Yagi T, Ishimoto T, Iwatsuki M, et al. Ido1 Expression is Associated With Immune Tolerance and Poor Prognosis in Patients With Surgically Resected Esophageal Cancer. *Ann Surg* (2019) 269:1101–8. doi: 10.1097/SLA.0000000000002754
68. Kelly RJ. The Emerging Role of Immunotherapy for Esophageal Cancer. *Curr Opin Gastroenterol* (2019) 35:337–43. doi: 10.1097/MOG.0000000000000542
69. Tsujitani S, Saito H, Oka S, Sakamoto T, Kanaji S, Tatebe S, et al. Prognostic Significance of RCAS1 Expression in Relation to the Infiltration of Dendritic Cells and Lymphocytes in Patients With Esophageal Carcinoma. *Dig Dis Sci* (2007) 52:549–54. doi: 10.1007/s10620-006-9408-6
70. Tao H, Qian P, Wang F, Yu H, Guo Y. Targeting CD47 Enhances the Efficacy of Anti-PD-1 and CTLA-4 in an Esophageal Squamous Cell Cancer Preclinical Model. *Oncol Res* (2017) 25:1579–87. doi: 10.3727/096504017X14900505020895
71. Gabrilovich DI. Myeloid-Derived Suppressor Cells. *Cancer Immunol Res* (2017) 5:3–8. doi: 10.1158/2326-6066.CIR-16-0297
72. Kumar V, Patel S, Tcyganov E, Gabrilovich DI. The Nature of Myeloid-Derived Suppressor Cells in the Tumor Microenvironment. *Trends Immunol* (2016) 37:208–20. doi: 10.1016/j.it.2016.01.004
73. Gabrilovich DI, Nagaraj S. Myeloid-Derived Suppressor Cells as Regulators of the Immune System. *Nat Rev Immunol* (2009) 9:162–74. doi: 10.1038/nri2506
74. Huang H, Zhang G, Li G, Ma H, Zhang X. Circulating CD14(+)HLA-DR (-/low) Myeloid-Derived Suppressor Cell is an Indicator of Poor Prognosis in Patients With ESCC. *Tumor Biol* (2015) 36:7987–96. doi: 10.1007/s13277-015-3426-y
75. Chen X, Wang L, Li P, Song M, Qin G, Gao Q, et al. Dual TGF-beta and PD-1 Blockade Synergistically Enhances MAGE-A3-specific Cd8(+) T Cell Response in Esophageal Squamous Cell Carcinoma. *Int J Cancer* (2018) 143:2561–74. doi: 10.1002/ijc.31730
76. Chen MF, Kuan FC, Yen TC, Lu MS, Lin PY, Chung YH, et al. IL-6-Stimulated CD11b+ Cd14+ HLA-DR- Myeloid-Derived Suppressor Cells, are Associated With Progression and Poor Prognosis in Squamous Cell Carcinoma of the Esophagus. *Oncotarget* (2014) 5:8716–28. doi: 10.18632/oncotarget.2368
77. Chen MF, Chen PT, Kuan FC, Chen WC. The Predictive Value of Pretreatment Neutrophil-to-Lymphocyte Ratio in Esophageal Squamous Cell Carcinoma. *Ann Surg Oncol* (2019) 26:190–9. doi: 10.1245/s10434-018-6944-1
78. Stairs DB, Bayne LJ, Rhoades B, Vega ME, Waldron TJ, Kalabis J, et al. Deletion of p120-catenin Results in a Tumor Microenvironment With Inflammation and Cancer That Establishes it as a Tumor Suppressor Gene. *Cancer Cell* (2011) 19:470–83. doi: 10.1016/j.ccr.2011.02.007
79. Karakasheva TA, Waldron TJ, Eruslanov E, Kim SB, Lee JS, O'Brien S, et al. Cd38-Expressing Myeloid-Derived Suppressor Cells Promote Tumor Growth in a Murine Model of Esophageal Cancer. *Cancer Res* (2015) 75:4074–85. doi: 10.1158/0008-5472.CAN-14-3639
80. Hurt B, Schulick R, Edil B, El Kasmi KC, Barnett CJr. Cancer-Promoting Mechanisms of Tumor-Associated Neutrophils. *Am J Surg* (2017) 214:938–44. doi: 10.1016/j.amjsurg.2017.08.003
81. Fridlender ZG, Sun J, Kim S, Kapoor V, Cheng G, Ling L, et al. Polarization of Tumor-Associated Neutrophil Phenotype by TGF-beta: "N1" Versus "N2" Tan. *Cancer Cell* (2009) 16:183–94. doi: 10.1016/j.ccr.2009.06.017
82. Mishalian I, Bayuh R, Levy L, Zolotarov L, Michaeli J, Fridlender ZG. Tumor-Associated Neutrophils (TAN) Develop Pro-Tumorigenic Properties During Tumor Progression. *Cancer Immunol Immunother* (2013) 62:1745–56. doi: 10.1007/s00262-013-1476-9
83. Wu L, Saxena S, Awaji M, Singh RK. Tumor-Associated Neutrophils in Cancer: Going Pro. *Cancers (Basel)* (2019) 11:564. doi: 10.3390/cancers11040564
84. Guthrie GJ, Charles KA, Roxburgh CS, Horgan PG, McMillan DC, Clarke SJ. The Systemic Inflammation-Based Neutrophil-Lymphocyte Ratio: Experience in Patients With Cancer. *Crit Rev Oncol Hematol* (2013) 88:218–30. doi: 10.1016/j.critrevonc.2013.03.010
85. Han F, Liu Y, Cheng S, Sun Z, Sheng C, Sun X, et al. Diagnosis and Survival Values of Neutrophil-Lymphocyte Ratio (NLR) and Red Blood Cell Distribution Width (RDW) in Esophageal Cancer. *Clin Chim Acta* (2019) 488:150–8. doi: 10.1016/j.cca.2018.10.042
86. Yodying H, Matsuda A, Miyashita M, Matsumoto S, Sakurazawa N, Yamada M, et al. Prognostic Significance of Neutrophil-to-Lymphocyte Ratio and Platelet-to-Lymphocyte Ratio in Oncologic Outcomes of Esophageal Cancer: A Systematic Review and Meta-Analysis. *Ann Surg Oncol* (2016) 23:646–54. doi: 10.1245/s10434-015-4869-5
87. Sun Y, Zhang L. The Clinical Use of Pretreatment NLR, PLR, and LMR in Patients With Esophageal Squamous Cell Carcinoma: Evidence From a Meta-Analysis. *Cancer Manag Res* (2018) 10:6167–79. doi: 10.2147/CMAR.S171035
88. Nakamura K, Yoshida N, Baba Y, Kosumi K, Uchihara T, Kiyozumi Y, et al. Elevated Preoperative Neutrophil-to-Lymphocytes Ratio Predicts Poor Prognosis After Esophagectomy in T1 Esophageal Cancer. *Int J Clin Oncol* (2017) 22:469–75. doi: 10.1007/s10147-017-1090-5
89. Xie X, Luo KJ, Hu Y, Wang JY, Chen J. Prognostic Value of Preoperative Platelet-Lymphocyte and Neutrophil-Lymphocyte Ratio in Patients Undergoing Surgery for Esophageal Squamous Cell Cancer. *Dis Esophagus* (2016) 29:79–85. doi: 10.1111/dote.12296
90. Liew PX, Kubes P. The Neutrophil's Role During Health and Disease. *Physiol Rev* (2019) 99:1223–48. doi: 10.1152/physrev.00012.2018
91. Rayes RF, Mouhanna JG, Nicolau I, Bourdeau F, Giannias B, Rousseau S, et al. Primary Tumors Induce Neutrophil Extracellular Traps With Targetable Metastasis Promoting Effects. *JCI Insight* (2019) 4:e128008. doi: 10.1158/1538-7445.SABCS18-1508
92. Coffelt SB, Wellenstein MD, de Visser KE. Neutrophils in Cancer: Neutral No More. *Nat Rev Cancer* (2016) 16:431–46. doi: 10.1038/nrc.2016.52
93. Munch NS, Fang HY, Ingermann J, Maurer HC, Anand A, Kellner V, et al. High-Fat Diet Accelerates Carcinogenesis in a Mouse Model of Barrett's Esophagus Via Interleukin 8 and Alterations to the Gut Microbiome. *Gastroenterology* (2019) 157:492–506.e492. doi: 10.1053/j.gastro.2019.04.013
94. Chen CL, Wang Y, Huang CY, Zhou ZQ, Zhao JJ, Zhang XF, et al. Il-17 Induces Antitumor Immunity by Promoting Beneficial Neutrophil Recruitment and Activation in Esophageal Squamous Cell Carcinoma. *Oncoimmunology* (2017) 7:e1373234. doi: 10.1080/2162402X.2017.1373234
95. da Silva EZ, Jamur MC, Oliver C. Mast Cell Function: A New Vision of an Old Cell. *J Histochem Cytochem* (2014) 62:698–738. doi: 10.1369/0022155414545334

96. Ribatti D, Tamma R, Crivellato E. The Dual Role of Mast Cells in Tumor Fate. *Cancer Lett* (2018) 433:252–8. doi: 10.1016/j.canlet.2018.07.005
97. Elpek GO, Gelen T, Aksoy NH, Erdogan A, Dertsiz L, Demircan A, et al. The Prognostic Relevance of Angiogenesis and Mast Cells in Squamous Cell Carcinoma of the Esophagus. *J Clin Pathol* (2001) 54:940–4. doi: 10.1136/jcp.54.12.940
98. Tomita M, Matsuzaki Y, Edagawa M, Shimizu T, Hara M, Sekiya R, et al. Association of Mast Cells With Tumor Angiogenesis in Esophageal Squamous Cell Carcinoma. *Dis Esophagus* (2001) 14:135–8. doi: 10.1046/j.1442-2050.2001.00171.x
99. Fakhrjou A, Niroumand-Oscoei SM, Somi MH, Ghojazadeh M, Naghashi S, Samankan S. Prognostic Value of Tumor-Infiltrating Mast Cells in Outcome of Patients With Esophagus Squamous Cell Carcinoma. *J Gastrointest Cancer* (2014) 45:48–53. doi: 10.1007/s12029-013-9550-2
100. Dlamini Z, Bhoola KD. Upregulation of Tissue Kallikrein, Kinin B1 Receptor, and Kinin B2 Receptor in Mast and Giant Cells Infiltrating Esophageal Squamous Cell Carcinoma. *J Clin Pathol* (2005) 58:915–22. doi: 10.1136/jcp.2004.021444
101. Wang B, Li L, Liao Y, Li J, Yu X, Zhang Y, et al. Mast Cells Expressing Interleukin 17 in the Muscularis Propria Predict a Favorable Prognosis in Esophageal Squamous Cell Carcinoma. *Cancer Immunol Immunother* (2013) 62:1575–85. doi: 10.1007/s00262-013-1460-4
102. Robida PA, Puzzovio PG, Pahima H, Levi-Schaffer F, Bochner BS. Human Eosinophils and Mast Cells: Birds of a Feather Flock Together. *Immunol Rev* (2018) 282:151–67. doi: 10.1111/imr.12638
103. Weller PF, Spencer LA. Functions of Tissue-Resident Eosinophils. *Nat Rev Immunol* (2017) 17:746–60. doi: 10.1038/nri.2017.95
104. Davoine F, Lacy P. Eosinophil Cytokines, Chemokines, and Growth Factors: Emerging Roles in Immunity. *Front Immunol* (2014) 5:570. doi: 10.3389/fimmu.2014.00570
105. Travers J, Rothenberg ME. Eosinophils in Mucosal Immune Responses. *Mucosal Immunol* (2015) 8:464–75. doi: 10.1038/mi.2015.2
106. Doyle AD, Masuda MY, Kita H, Wright BL. Eosinophils in Eosinophilic Esophagitis: The Road to Fibrostenosis is Paved With Good Intentions. *Front Immunol* (2020) 11:603295. doi: 10.3389/fimmu.2020.603295
107. Fukuchi M, Sakurai S, Suzuki M, Naitoh H, Tabe Y, Fukasawa T, et al. Esophageal Squamous Cell Carcinoma With Marked Eosinophil Infiltration. *Case Rep Gastroenterol* (2011) 5:648–53. doi: 10.1159/000332441
108. Ohashi Y, Ishibashi S, Suzuki T, Shineha R, Moriya T, Satomi S, et al. Significance of Tumor Associated Tissue Eosinophilia and Other Inflammatory Cell Infiltrate in Early Esophageal Squamous Cell Carcinoma. *Anticancer Res* (2000) 20:3025–30.
109. Yang X, Wang L, Du H, Lin B, Yi J, Wen X, et al. Prognostic Impact of Eosinophils in Peripheral Blood and Tumor Site in Patients With Esophageal Squamous Cell Carcinoma Treated With Concurrent Chemoradiotherapy. *Med (Baltimore)* (2021) 100:e24328. doi: 10.1097/MD.0000000000024328
110. Lagisetty KH, McEwen DP, Nancarrow DJ, Schiebel JG, Ferrer-Torres D, Ray D, et al. Immune Determinants of Barrett's Progression to Esophageal Adenocarcinoma. *JCI Insight* (2021) 6:e143888. doi: 10.1172/jci.insight.143888
111. Warner K, Ohashi PS. ILC Regulation of T Cell Responses in Inflammatory Diseases and Cancer. *Semin Immunol* (2019) 41:101284. doi: 10.1016/j.smim.2019.101284
112. Mattiola I, Diefenbach A. Innate Lymphoid Cells and Cancer At Border Surfaces With the Environment. *Semin Immunol* (2019) 41:101278. doi: 10.1016/j.smim.2019.06.001

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.

frontiers in Immunology

CORRECTION
published: 27 May 2021
doi: 10.3389/fimmu.2021.708705

Corrigendum: Innate Immune Cells in the Esophageal Tumor Microenvironment

Kele Cui[1,3,4*†], *Shouxin Hu*[2,3,4†], *Xinyu Mei*[5] *and Min Cheng*[2,3,4*]

[1] Department of Clinical Laboratory, The First Affiliated Hospital of USTC, Division of Life Sciences and Medicine, University of Science and Technology of China, Hefei, China, [2] Department of Geriatrics, Gerontology Institute of Anhui Province, The First Affiliated Hospital of USTC, Division of Life Sciences and Medicine, University of Science and Technology of China, Hefei, China, [3] Anhui Provincial Key Laboratory of Tumor Immunotherapy and Nutrition Therapy, Hefei, China, [4] Cancer Immunotherapy Center, The First Affiliated Hospital of USTC, Division of Life Sciences and Medicine, University of Science and Technology of China, Hefei, China, [5] Department of Thoracic Surgery, The First Affiliated Hospital of USTC, Division of Life Sciences and Medicine, University of Science and Technology of China, Hefei, China

Keywords: innate immune cells, crosstalk, regulation, esophageal tumor microenvironment, immunotherapy strategy

A Corrigendum on

Innate Immune Cells in the Esophageal Tumor Microenvironment
by Cui K, Hu S, Mei X and Cheng M (2021). Front. Immunol. 12:654731. doi: 10.3389/fimmu.2021.654731

OPEN ACCESS

Approved by:
Frontiers Editorial Office, Frontiers Media SA, Switzerland

***Correspondence:**
Kele Cui
ckele@mail.ustc.edu.cn
Min Cheng
chengmin@ustc.edu.cn

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Molecular Innate Immunity, a section of the journal Frontiers in Immunology

Received: *12 May 2021*
Accepted: *14 May 2021*
Published: *27 May 2021*

Citation:
Cui K, Hu S, Mei X and Cheng M (2021) Corrigendum: Innate Immune Cells in the Esophageal Tumor Microenvironment. Front. Immunol. 12:708705. doi: 10.3389/fimmu.2021.708705

In the published article, there was an error in affiliations 1, 2, 4 and 5. Instead of "[1] Department of Clinical Laboratory, Division of Life Sciences and Medicine, The First Affiliated Hospital of USTC, University of Science and Technology of China, Hefei, China; [2] Department of Geriatrics, Gerontology Institute of Anhui Province, Division of Life Sciences and Medicine, The First Affiliated Hospital of USTC, University of Science and Technology of China, Hefei, China; [4] Cancer Immunotherapy Center, Division of Life Sciences and Medicine, The First Affiliated Hospital of USTC, University of Science and Technology of China, Hefei, China; [5] Department of Thoracic Surgery, Division of Life Sciences and Medicine, The First Affiliated Hospital of USTC, University of Science and Technology of China, Hefei, China", it should be "[1] Department of Clinical Laboratory, the First Affiliated Hospital of USTC, Division of Life Sciences and Medicine, University of Science and Technology of China, Hefei, China; [2] Department of Geriatrics, Gerontology Institute of Anhui Province, the First Affiliated Hospital of USTC, Division of Life Sciences and Medicine, University of Science and Technology of China, Hefei, China; [4] Cancer Immunotherapy Center, the First Affiliated Hospital of USTC, Division of Life Sciences and Medicine, University of Science and Technology of China, Hefei, China; [5] Department of Thoracic Surgery, the First Affiliated Hospital of USTC, Division of Life Sciences and Medicine, University of Science and Technology of China, Hefei, China".

The authors apologize for this error and state that this does not change the scientific conclusions of the article in any way. The original article has been updated.

frontiers in Immunology

ORIGINAL RESEARCH
published: 29 April 2021
doi: 10.3389/fimmu.2021.672729

Phenotypical Diversification of Early IFNα-Producing Human Plasmacytoid Dendritic Cells Using Droplet-Based Microfluidics

Laura C. Van Eyndhoven[1,2], *Eleni Chouri*[1,2], *Nikita Subedi*[1,2] *and Jurjen Tel*[1,2*]

[1] Laboratory of Immunoengineering, Department of Biomedical Engineering, Eindhoven University of Technology, Eindhoven, Netherlands, [2] Institute for Complex Molecular Systems (ICMS), Eindhoven University of Technology, Eindhoven, Netherlands

Plasmacytoid dendritic cells (pDCs) are a rare type of highly versatile immune cells that besides their specialized function of massive type I interferon (IFN-I) production are able to exert cytotoxic effector functions. However, diversification upon toll like receptor (TLR)-induced activation leads to highly heterogeneous responses that have not been fully characterized yet. Using droplet-based microfluidics, we showed that upon TLR7/8 and TLR9-induced single-cell activation only 1-3% secretes IFNα, and only small fractions upregulate cytotoxicity markers. Interestingly, this 1-3% of early IFN-producing pDCs, also known as first responders, express high levels of programmed death-ligand 1 (PD-L1) and TNF-related apoptosis-inducing ligand (TRAIL), which makes these hybrid cells similar to earlier described IFN-I producing killer pDCs (IKpDCs). IFN-I priming increases the numbers of IFNα producing cells up to 40%, but does not significantly upregulate the cytotoxicity markers. Besides, these so-called second responders do not show a cytotoxic phenotype as potent as observed for the first responders. Overall, our results indicate that the first responders are the key drivers orchestrating population wide IFN-I responses and possess high cytotoxic potential.

Keywords: plasmacytoid dendritic cells, droplet-based microfluidics, diversification, heterogeneity, interferons, cytotoxicity, stochasticity

OPEN ACCESS

Edited by:
Alexandre Corthay,
Oslo University Hospital, Norway

Reviewed by:
Stefanie Scheu,
Universitätsklinikum Düsseldorf, Germany
Evelien Smits,
University of Antwerp, Belgium
Ivan Hirsch,
Charles University, Czechia

***Correspondence:**
Jurjen Tel
j.tel@tue.nl

Specialty section:
This article was submitted to Molecular Innate Immunity, a section of the journal Frontiers in Immunology

Received: *26 February 2021*
Accepted: *19 April 2021*
Published: *29 April 2021*

Citation:
Van Eyndhoven LC, Chouri E, Subedi N and Tel J (2021) Phenotypical Diversification of Early IFNα-Producing Human Plasmacytoid Dendritic Cells Using Droplet-Based Microfluidics. Front. Immunol. 12:672729. doi: 10.3389/fimmu.2021.672729

INTRODUCTION

Plasmacytoid dendritic cells (pDCs) are a rare type of highly versatile immune cells constituting an extremely promising therapeutic target for infectious diseases, autoimmune diseases, and cancer. Besides their highly specialized function to produce massive amounts of type I interferons (IFN-I), previous studies highlighted their multifaceted biology that goes beyond the scopes of viral immunity (1, 2). In fact, it is widely appreciated that pDCs can also exert cytotoxic effector functions (3). However, in pathological conditions, pDCs are often in a hypofunctional or aberrant state, with tumor-infiltrating pDCs correlating with bad prognoses (4–6). Therefore, it is crucial to

Abbreviations: IFN-I, type I interferon; IKpDC, interferon producing killer plasmacytoid dendritic cell; PBMC, peripheral blood mononuclear cell; pDC, plasmacytoid dendritic cell; PD-L1, programmed death-ligand 1; TLR, toll-like receptor; TLRL, toll-like receptor ligand; TRAIL, TNF-related apoptosis-inducing ligand.

understand how pDCs can be properly activated, thereby enhancing their antitumoral potential and enabling them to restore local immune responses (7–9).

Over the past years, studies have revealed that pDCs are heterogeneous, showing a wide range of diversification upon stimulation. Subsequently, multiple different pDC subsets, specialized in multiple different functions, have been described (10–13). One of the most recent examples is the diversification observed in bulk-activated pDCs for the markers CD80 and Programmed death-ligand 1 (PD-L1), giving rise to three stable subpopulations of which one is specialized in IFN-I production (14). Although these studies have been very informative, an improved fundamental understanding of pDC diversification is needed to reveal their true intrinsic behaviors and capabilities. Therefore, experimental approaches require not only single-cell resolution, but also single-cell activation in which all types of cellular interactions that average out individual cellular responses are excluded. Droplet-based microfluidics is a high-throughput technique that allows highly controllable single-cell activation by encapsulating individual cells in picolitre-sized droplets that function as tiny bioreactors (15, 16). Inside these droplets, cells receive their input while being excluded from any kind of interaction with other cells, thereby revealing their intrinsic behaviors upon single-cell activation. This change in approach revealed that upon single-cell activation only fractions of 1-3% produce IFN-Is, which is a phenomenon that has been observed in multiple different settings, for different types of immune cells, both *in vitro* and *in vivo* (17–21). Additionally, IFN-I priming revealed the crucial role of paracrine signaling in amplifying the TLR-induced IFN-I response by increasing the fraction of IFN-producing pDCs up to 40%, which comes to a halt upon the administration of neutralizing antibodies (16).

In this study, we set out to investigate the cytotoxic diversification of pDCs to a similar extend as observed for IFN-I production. More specifically, we studied whether the small fractions of early IFN-I producers express cytotoxic markers, since so-called IFN-producing killer (p)DCs (IKDC/IKpDCs) have been described in the past (22–27). PD-L1 and TNF-related apoptosis-inducing ligand (TRAIL) have been known for their role in pDC cytotoxicity [reviewed in (3, 7)], and were therefore, together with a functional IFN-I readout, the main focus of this study. Additionally, intracellular granzyme B was assessed, since this cytolytic molecule has proven its anti-tumor effects in a variety of pDC and NK-cell dependent killing (3, 7, 28). Up to now, all that is known about pDC cytotoxicity and IKpDCs has been tested only in bulk-activated pDCs. Therefore, to the best of our knowledge, we are the first to study the diversification of IFN-I production combined with the expression of cytotoxic markers in single TLR-activated pDCs.

MATERIALS AND METHODS

Cell Isolation and Culture

Human primary pDCs were isolated from buffy coats of healthy donors (Sanquin), according to institutional guidelines and after informed consent per the Declaration of Helsinki. Peripheral blood mononuclear cells (PBMCs) were isolated *via* Lymphoprep density gradient centrifugation (Stemcell Technologies, 07861). The isolated PBMCs were washed thrice with phosphate-buffered saline (PBS, Thermo Fisher Scientific, 20012027) supplemented with 0.6 w/v% sodium citrate dehydrate tri-basic and 0.01 w/v% bovine serum albumin (Sigma Aldrich, C8532; A9418). To deplete monocytes, PBMCs were resuspended in RPMI cell culture medium (Thermo Fischer Scientific, 11875093) supplemented with 2% human serum (pooled; Sanquin), 1% antibiotics (penicillin-streptomycin, Thermo Fisher Scientific, 11548876), and incubated for 1 hour at 37 degrees Celsius in T150 cell culture flasks. Afterwards, non-adherent cells were collected while washing the cells thrice with PBS. Next, pDCs were isolated using magnet-activated cell sorting (MACS) by positive selection using the CD304 Microbeat Kit (Miltenyi Biotec, 130-090-532), according to manufacturer's instructions. For purity assessment, a small sample was washed with PBS supplemented with 0.5% bovine serum albumin (later referred to as PBA) and stained for 20 minutes at 4 degrees Celsius using FITC-labeled anti-CD123 and APC-labeled anti-CD303. The pDCs were identified as CD123+ CD303+ (average 92%, SD 4.4%, $n = 12$).

Soft Lithography and Microfluidic Setup

Microfluidic devices were fabricated with polydimethylsiloxane (PDMS) base and curing agent at a ratio of 10:1 (Sylgard 184; Sigma-Aldrich, 101697). After proper mixing, the PDMS mix was poured onto a master silicon wafer and cured at 65°C for 3 hours. The surface of the devices was OH-terminated by exposure to plasma (Emitech K1050X), and were sealed with plasma-treated glass slides to yield closed microchannels. Channels were treated with 2% silane in fluorinated HFE-7500 3M Novec (Fluorochem, 051243). Droplets were produced with a three-inlet microfluidic device. Liquids were dispensed from syringes driven by computer-controlled pumps (Nemesys, Cetoni GmbH). 2.5 v/v% Pico-Surf surfactant (Sphere Fluidics, C024) was used in fluorinated HFE-7500 3M Novec.

Bulk Activation Assay

Freshly isolated pDCs were incubated in 100 µL per 10^6 cells PBA containing the IFNα Cytokine Catch Reagent (Miltenyi Biotec, 130-092-605) at 4 degrees Celsius for 20 minutes. Next, cells were washed and resuspended in X-Vivo 15 cell culture medium (Lonza), supplemented with 2% human serum (pooled; Sanquin), 1% antibiotics (penicillin-streptomycin), at 25.000 cells per 100 µL in U-bottom microwell plates. For intracellular IFNα stainings, cells were not pre-incubated with Cytokine Catch Reagent, but directly transferred to the microwells upon isolation.

Single-Cell Activation Assay

Freshly isolated pDCs were incubated in 100 µL per 10^6 cells PBA containing the IFNα Cytokine Catch Reagent (Miltenyi Biotec) at 4 degrees Celsius for 20 minutes. For the primed conditions, pDCs were primed with 500 U/mL IFNβ, at a concentration of 10^5 cells per 100 µL in U-bottom microwell plates, prior to the incubation with IFNα Cytokine Catch Reagent. Next, cells were

washed and resuspended in X-Vivo 15 cell culture medium (Lonza, BE02-060Q), supplemented with 2% human serum (pooled; Sanquin), 1% antibiotics (penicillin-streptomycin), at $2.6*10^6$ cells/mL for single-cell encapsulation in 92 pL droplets. Stimulus was dissolved in medium at twice the desired concentration to account for 2x dilution in the microfluidic device. For a list of al utilized stimuli, see **Supplementary Table 1**. For droplet production, flow rates of 900 µL/h for the oil phase and 300 µL/h for the aqueous phases were used. Droplet production and encapsulation rates were carefully monitored using a microscope (Nikon) at 10x magnification and a high-speed camera. The droplet emulsion was collected and covered with culture medium to protect droplets from evaporation. The encapsulated cells were incubated in Eppendorf tubes with a few punched holes to allow gas exchange, at 37 degrees Celsius and 5% CO2. After 18 hours of incubation, the droplets were de-emulsified by adding 100 µL 20 v/v% 1H,1H,2H,2H-Perfluoro-1-octanol (Sigma Aldrich, 370533) in HFE-7500.

Antibody Staining

Cells were washed once with PBS and dead cells were stained with Zombie Green fixable viability dye (BioLegend, 423111), 1:10.000 in PBS, 100 µL) at 4 degrees Celsius for 20 minutes. Subsequently, cells were washed once with PBS and incubated with antibodies against surface proteins in 50 µL PBA at 4 degrees Celsius for 20 minutes. Next, cells were fixed and permeabilized with Cytofix/Cytoperm solution (BD Biosciences, 554714) at 4 degrees Celsius for 20 minutes. Cells were washed with Perm/Wash buffer (BD) and incubated with antibodies against intracellular proteins in 50 µL Perm/Wash buffer. For a full list of all utilized antibodies and reagents, see **Supplementary Table 2**.

Flow Cytometry

Acquisition was performed in PBA on FACS Aria (BD Biosciences). Flow cytometry data were analyzed using FlowJo X (Tree Star). FMO stainings served as controls for gating strategy. For the gating strategy, the readers are referred to **Supplementary Figure 1**.

Data Analysis and Statistics

Analysis and data visualization was performed using PRISM for windows version 9 (GraphPad). For statistical analysis, Mann-Whitney test, and Wilcoxon signed-rank test were performed. T-SNE multidimensional data analyses were performed using FlowJo X (Tree Star).

RESULTS

Droplet-Based Microfluidics Allows for Single-Cell Activation of pDCs

In vivo, pDCs operate in complex microenvironments that influence their cellular behavior. One way to reveal this complexity is by studying pDCs in highly controlled microenvironments, enabled by utilizing innovative techniques and approaches such as droplet-based microfluidics. Cleverly designed microfluidic chips allow for the generation of thousands of droplets with high precision and control over the content of each individual droplet.

In order to investigate the diversification of pDCs upon single-cell activation, freshly isolated human pDCs were either stimulated in bulk or individually in picolitre-sized droplets (**Figure 1A**). Briefly, pDCs were isolated from human buffy coats, primed with IFNβ or left untreated, and coated with cytokine Catch Reagent for the IFN-I readout. Therefore, during incubation, secreted IFNα was captured on the cells' surface (**Figure 1B**). The concentrations used for stimuli and cytokines for IFN-I priming have been extensively tested, as described elsewhere (16). Additionally, the different concentrations used for the droplet conditions compared to the bulk conditions were considerably chosen to correct for the total amount of volume per cell, to ensure similar absolute quantities of stimuli across experimental conditions. Besides, throughout this study, IL-3 stimulation served as a control to assess the TLR-induced up/down-regulation of the markers of interest, while this cytokine can be considered as a growth factor which barely induces phenotypical or functional differentiation in pDCs (14).

A microfluidic device was used to encapsulate individual cells in water-in-oil droplets (**Figures 1C, D**), as previously optimized and described (16). The Tip-Loading method was applied for proper cell encapsulation (29). The produced droplet volumes ranged between 70 and 110 pL (**Figure 1E**). Since the encapsulation of cells in droplets is a random process following the Poisson distribution, theoretical fractions of total droplets and cell-containing droplets were calculated (**Figure 1F**). Single-cell encapsulation was ensured by using relatively low cell seeding densities that resulted in over 95% of cell-containing droplets with only one cell. After incubation, pDCs were retrieved from droplets by breaking the emulsion for additional down-stream analysis. Although the de-emulsification of water-in-oil droplets can be toxic to the cells, cell viability remained high throughout all procedures, allowing reliable analysis (**Figure 1G**).

Taken together, our droplet-based microfluidics platform allows for studying diversification of single TLR-activated pDCs in highly controllable microenvironments.

Single-Cell Activation Limits the Production of IFN-Is and the Upregulation of Cytotoxicity Markers to Only Small Fractions of pDCs

Recent studies have demonstrated that bulk-activated pDCs analyzed at single-cell resolution show significant levels of cellular heterogeneity (10, 14). These results prompted us to explore the levels of diversification upon single-cell activation, focusing on IFNα production, the costimulatory molecule CD80, the cytotoxicity markers PD-L1 and TRAIL, the adhesion molecule CD2, and the secretion of granzyme B. Based on these markers, pDC subsets have been described in the past (1). In order to validate our single-cell findings, bulk-activated pDCs of matching donors served as an important control to literature and enabled us to assess the role of paracrine signaling

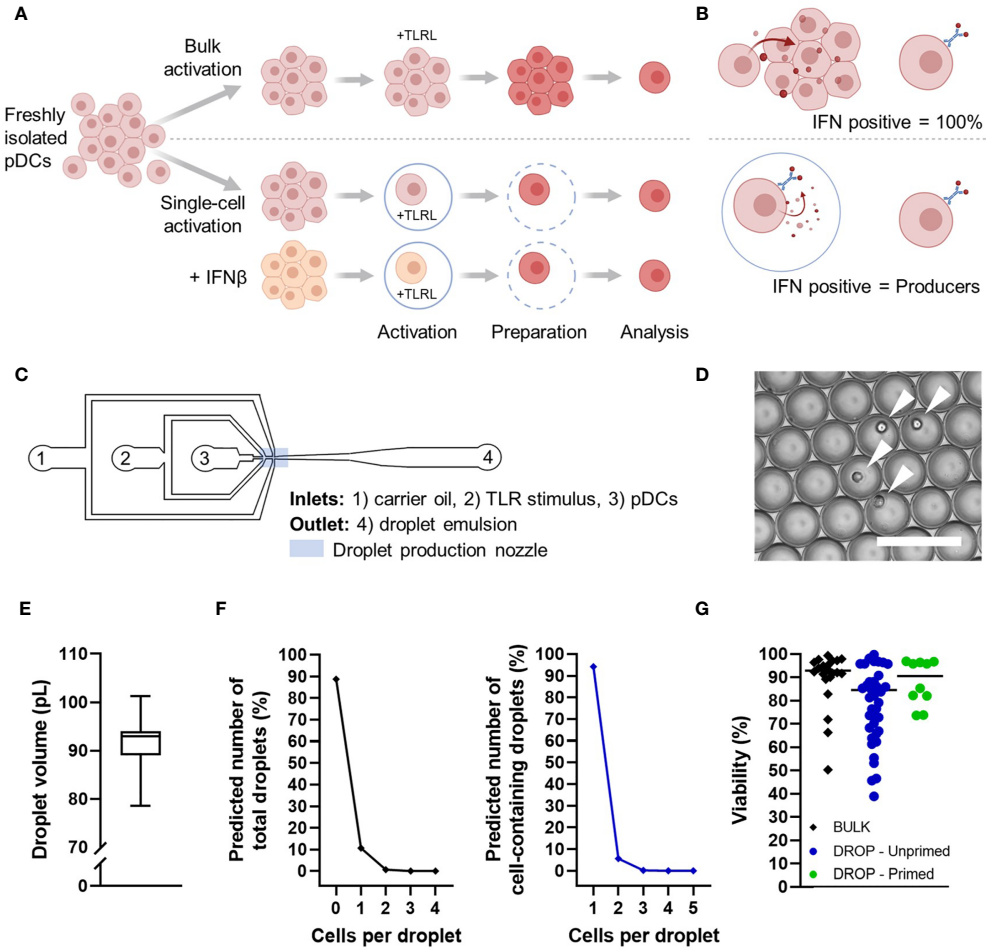


FIGURE 1 | Schematic overview of experimental conditions and microfluidic device. **(A)** Human pDCs were isolated from buffy coats, primed with 500 U/mL IFNβ or left untreated. Next, all pDCs were coated with IFNα Catch Reagent and incubated with TLR ligands (TLRL) in bulk or droplets for 18 hours. Single cell encapsulation was achieved with concentrations of 1,300,000 cells/mL in 92 pL droplets on average. After incubation, encapsulated pDCs were retrieved from droplets by de-emulsification. Next, pDCs were washed thoroughly, fixed and permeabilized, stained for viability, extracellular IFNα, surface marker expression, intracellular granzyme B, and analyzed by flow cytometry. **(B)** Schematic representation of IFNα production in bulk vs in droplets. **(C)** Schematic overview of microfluidic chip with inlets (1–3), outlet (4), and droplet production nozzle in blue rectangle. **(D)** Microscopic image of the droplet emulsion with single-encapsulated pDCs indicated with white arrows. Scale bar equals 100 µm. **(E)** Box plot depicting droplet volumes in pL, n = 100 droplets. **(F)** Theoretical fractions of total droplets (left) and cell-containing droplets (right) with x number of cells per droplet, according to the Poisson distribution, average droplet size of 92 pL, and cell concentration of 1,300,000 million cells/mL. **(G)** Viability of analyzed cells per experimental condition.

in the upregulation of the markers of interest. In turn, IFN-I priming mimics the effects of paracrine signaling, but since the activation of the cells in these conditions occurs in droplets, the intrinsic behaviors are still elicited rather than the averaged outcomes that would occur in bulk-activated pDCs.

In our experimental setup, bulk-activated pDCs reached up to 100% positivity for IFNα, as a result of the full saturation of IFNα binding to available Catch Reagents (**Figure 2A**). These IFNα molecules are produced by only a fraction of around 30% of IFNα-producing cells, according to numerous examples in literature (14). However, in droplets, the produced IFNα can only bind to the pDCs that have actually produced it. Therefore, the number of IFNα positive cells in the droplet conditions represent the actual IFNα producers. Notably, our results of unprimed single TLR-activated pDCs, representing the first responders, show that only a small fraction, of around 0.5-0.7%, of cells secrete IFNα (median for CpG-C; R848: 0.54; 0.76 respectively, n =10). Upon IFN-I priming, representing the second responders, percentages of IFN-producing pDCs increased up to 40%. This increase is the result of the IFN-I induced effects initiated during priming, which enhances the response rate upon activation compared to the numbers observed in naïve cells. This phenomenon is elicited upon both TLR7/8 (R848) and TLR9 (CpG-C) signaling, although the observed effect of IFN-I priming was higher for R848 activation ($p = 0.0005$) compared to CpG-C activation ($p = 0.006$).

In accordance with literature, bulk-activated pDCs show already levels of diversification upon TLR activation, with

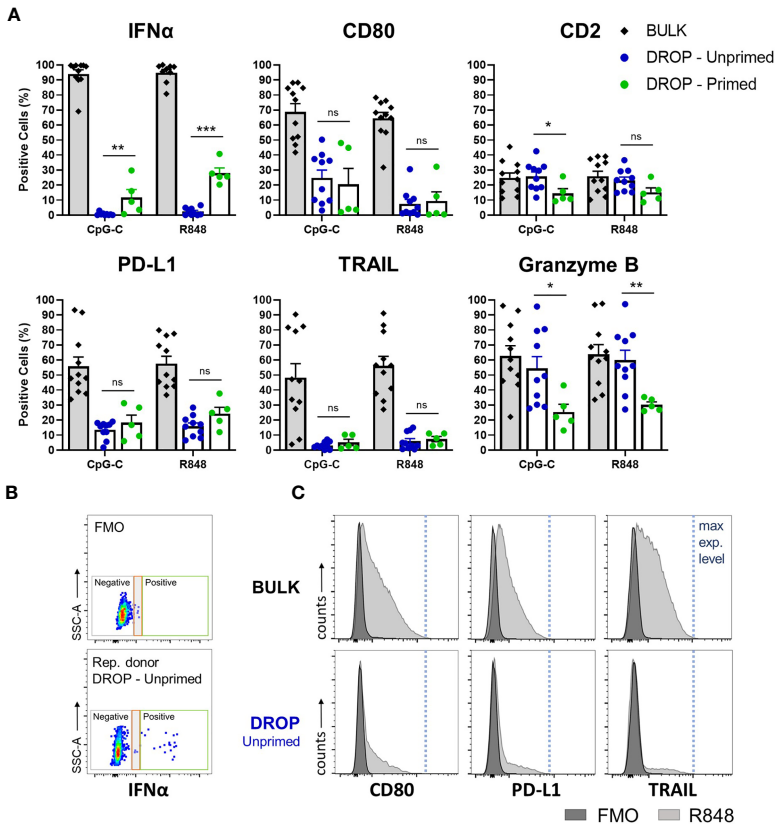


FIGURE 2 | Single-cell analysis reveals functional and phenotypical heterogeneity within single-activated pDCs. Freshly isolated pDCs were either primed or left unprimed, coated with IFNα Catch Reagent, and activated in bulk or droplets as described before. The pDCs were activated with 5 or 50 μg/mL CpG-C or R848, for bulk and droplet conditions respectively, for 18h. Next, pDCs were stained for viability, IFNα secretion, marker expression of CD80, PD-L1, TRAIL, CD2, granzyme B, and analyzed by flow cytometry. **(A)** Data represent mean percentages of IFNα-secreting and marker-expressing pDCs plotted against treatment condition; error bars indicate SEM; BULK n = 11; DROP-Unprimed n = 10; DROP-Primed n = 5 for both CpG-C and R848. **(B)** Dot plots depicting IFNα expression levels of FMO control and representative donor, DROP-Unprimed. Depicted in the colored boxes are negative events (gray) technological noise (red), and positive events (green). **(C)** pDCs were treated as described above. The histograms are representatives of expression levels of viable pDCs stimulated with R848 in either bulk or droplets from one donor, compared to FMO controls. Indicated with the blue dotted line are the maximum expression levels. Mann–Whitney test *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$. ns, non-significant.

varying expression levels of CD80, PD-L1 and TRAIL (**Figure 2A**). Upon single-cell activation, the percentages of positive cells for these markers were significantly lower ($p < 0.0001$ for all markers, compared to bulk percentages). Interestingly, the effects of priming did not significantly affect the upregulation of these markers, indicating an underlying mechanism other than induced by IFN-Is. In contrast and in accordance with literature, the expression of CD2 remained relatively stable over time, independent of experimental conditions or stimulus (**Supplementary Figure 2**). Only priming seemed to have affected CD2 expression, with a significant reduction upon CpG-C activation. The differences were not significant when activated with R848. The secretion of granzyme B, assessed by intracellular staining, was similar for both the bulk and droplet-unprimed conditions, while priming significantly enhanced the secretion. The interpretation of these results was based on the finding that all pDCs that were not activated with TLR ligands expressed granzyme B, reaching up to 100% positivity (**Supplementary Figure 2**). This finding is in accordance with literature (30, 31).

The relatively low percentages of positive cells among the single TLR-activated pDCs, compared to the bulk-activated pDCs, raised the question whether these results were possibly due to technical limitations of the experimental setup. Therefore, we evaluated the expression levels of both IFNα and the individual markers from the single-activated pDCs, and compared these with the expression levels obtained from the FMO controls and bulk-activated pDCs. Despite the very low numbers of IFN-positive events in the unprimed droplet conditions, these events could be properly distinguished from technical noise, indicating that these events represent true IFN producers (**Figure 2B**). Besides, the expression levels of the markers of interest of single-activated pDCs reach up to those found for bulk-activated pDCs, indicating that the pDCs

encapsulated in droplets can be properly activated and can induce their expression in a proper manner (**Figure 2C** and **Supplementary Figure 3**).

We conclude that single TLR-activated pDCs display a high degree of diversification, leading to heterogeneous outcomes, both functionally and phenotypically. Only small fractions secrete IFNα and induce the expression of surface markers CD80, PD-L1, and TRAIL. Moreover, IFN-I priming only increases the percentage of IFN producers and does not significantly affect the expression of cytotoxicity markers.

PD-L1+/CD80- Subsets Are Specialized IFNα Producers and Express High Levels of TRAIL

While still most of the subset characterization is based on phenotypes, functional readouts are often correlated as a proof of concept and to reveal the importance, relevance and potential of the different subsets. Over the years, much progress has been made in identifying IFN-secreting subsets in the pDC compartment. One of the most recent studies identified and characterized 3 populations, with distinct phenotypes and functionalities, that arise upon activation in bulk. Based on the expression of PD-L1 and CD80, populations could be defined (14). PD-L1+/CD80- pDCs, referred to as P1, were found to be specialized in the production of IFN-Is and prevalent in patients with IFN-I-mediated autoimmune diseases. PD-L1-/CD80+ pDCs, referred to as P3, were found to promote T cell activation and T_h2 differentiation. PD-L1+/CD80+ pDCs, referred to as P2, was not further addressed as being specialized in either IFN-I production or T cell activation.

In this study, we set out to investigate whether we could find similar IFNα secretory profiles for P1, as described by Alculumbre et al. Therefore, bulk activated pDCs were plotted for PD-L1 and CD80 expression accordingly. Based on the expression levels obtained upon IL-3 stimulation, the 3 populations were defined (**Figure 3A**). Only the data for the R848-actived pDCs are shown, while for all CpG-C-activated pDCs similar results were found, across all experimental conditions (data not shown). Although our plots were slightly different from the data presented in Alculumbre et al., which could be explained by the different stimuli and incubation times used, we were able to define the 3 populations. Unfortunately, our experimental setup did not allow an interpretation regarding the IFNα readout, as described in prior results sections. The expression of TRAIL and CD2 did not significantly differ between the 3 different populations.

Next, the 3 populations were defined for the unprimed single-activated pDCs (**Figure 3B**). Interestingly, P2 and P3 were barely present, suggesting that the diversification into these 2 populations is dependent on paracrine and/or juxtacrine

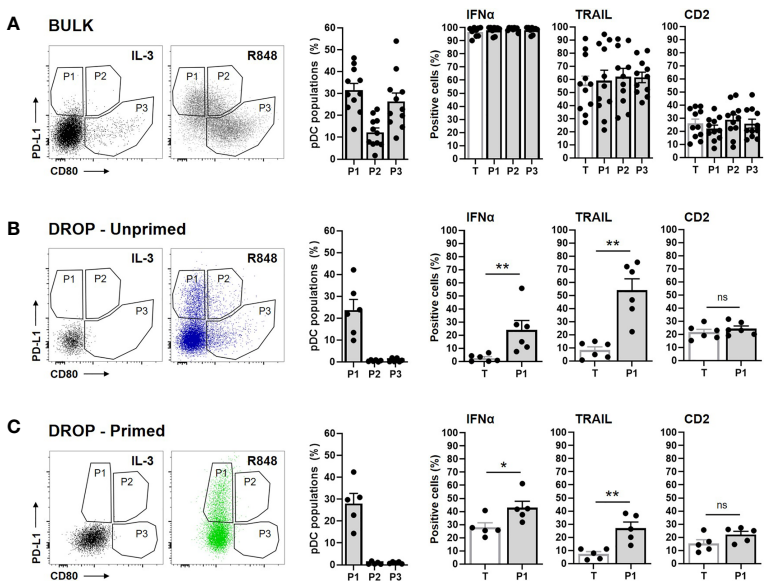


FIGURE 3 | PD-L1+/CD80- subsets are specialized IFNα producers and express high levels of TRAIL. Freshly isolated pDCs were either primed or left unprimed, coated with IFNα Catch Reagent, and activated in bulk or droplets as described before. The pDCs were activated with 5 or 50 μg/mL R848, for bulk and droplet conditions respectively, for 18h. Next, pDCs were stained for viability, IFNα secretion, surface marker expression of CD80, PD-L1, TRAIL, CD2, granzyme B and analyzed *via* flow cytometry. **(A)** Based on the expression of PD-L1 and CD80, 3 populations were defined for the bulk activated pDCs. IL-3 stimulated pDCs served as a control for gating. The first histogram represents the percentages of the 3 populations. Other histograms represent the percentages of positive cells for the indicated markers, for the total viable pDCs (T) and the 3 populations (P1, P2, P3). **(B)** Data as in panel **(A)**, for unprimed single-activated pDCs. Expression of markers is only plotted for T and P1, because of the low percentages of P2 and P3 present in total. **(C)** Data as in **(B)**, for primed single-activated pDCs. Bars represent mean percentages; error bars indicate SEM; BULK n = 11; DROP-Unprimed n = 6; DROP-Primed n = 5. Mann–Whitney test *$p < 0.05$, **$p < 0.01$. ns, non-significant.

signaling. Because of the low abundance, the expression of IFNα, TRAIL and CD2 could not be properly examined in these two populations and therefore left out of the analysis. Hence, only P1, characterized by PD-L1+/CD80- was further characterized. Interestingly, in agreement with Alculumbre et al., we found that P1 expressed significantly more IFNα, compared to the total pDCs (T), implying that this population is specialized in IFNα production. Besides, the expression of TRAIL was also significantly higher in P1 compared to the total pDCs.

Finally, the 3 populations were defined for the primed single-activated pDCs (**Figure 3C**). Similar to the unprimed single-activated pDCs, P2 and P3 were barely present. A similar increase in IFNα-expressing and TRAIL expressing pDCs was observed for P1 as for the unprimed single-activated pDCs, although less potent.

A similar approach was undertaken for the CD2+ and CD2- pDCs. Although literature characterized the co-expression of IFN-Is, TRAIL and granzyme B for CD2+ cells, we couldn't find any significant differences between the CD2+ and CD2- populations (data not shown) (10).

Together, in agreement with Alculumbre et al., our data demonstrate an increased IFNα signature for PD-L1+/CD80- pDCs, with high expression of TRAIL.

Early IFN-Producing pDCs Share Phenotypical Characteristics Assigned to Previously Described Killer pDCs

The combination of IFN-I production and strong cytotoxic or antitumor activities has been described for (p)DCs in numerous studies (22–27). Therefore, we next zoomed in on the expression of cytotoxic markers on the fraction of early IFN-producing pDCs. For further analysis we used stringent gating conditions to correct for any possible technological noise that could arise from the binding of excessive IFNα to non-IFN-producing pDCs during the droplet de-emulsification phase, as described earlier.

Our results show that early IFN-producing pDCs, those who produce IFN-Is in the unprimed conditions, express relatively high levels of PD-L1 and TRAIL compared to non-IFN-producing pDCs (**Figures 4A, B**). This finding is in line with the earlier described IFN-I-induced effect upon autocrine signaling that results in the upregulation of these markers, but it might as well be inherent to this specific fraction of pDCs (32, 33). Besides, considering the merged expression levels of all donors, the secretion of granzyme B seemed to be enhanced in IFN-producing pDCs compared to non-producing pDCs, though not significantly (**Figures 4A, B**).

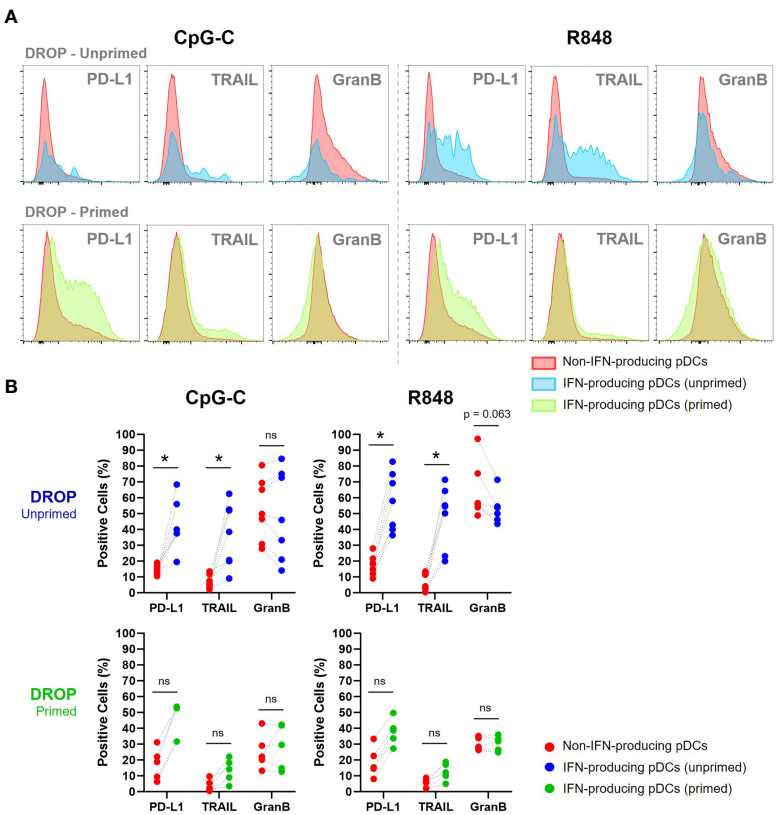


FIGURE 4 | IFNα producing pDCs are phenotypically similar to previously described killer pDCs. **(A)** Single-encapsulated pDCs were either primed or left unprimed, incubated with IFNα Catch Reagent, activated with 50 μg/mL CpG-C or R848 for 18h and gated for IFN-producing (blue; unprimed, and green; primed) and non-IFN-producing pDCs (red). Histograms represent relative expression levels of PD-L1, TRAIL and granzyme B for merged data of 7 and 5 representative donors for unprimed and primed conditions, respectively. **(B)** Scatter plots show paired data of non-IFN-producing pDCs, IFN-producing pDCs and their corresponding percentage of positive cells for PD-L1, TRAIL and granzyme B per donor. Wilcoxon signed-rank test. *p < 0.05. ns, non-significant.

For the populations of second responders, those who produce IFN-Is in the primed conditions, we found similar results, although less potent. This could be explained by the difference in secretion quantities between first and second responders, as literature suggests that first responders produce more cytokines than second responders due to multiple autocrine feedback loops (34, 35).

Altogether, first responders share phenotypical characteristics assigned to previously described IKpDCs, with relatively high expression of PD-L1 and TRAIL. The killer phenotype of the second responders, however, was not as evident as for first responders.

Multi-Dimensional Data Analysis Shows Clustering of Early IFN-Producing pDCs With Co-Expression of PD-L1 and TRAIL

Next, we performed multidimensional data analysis to explore the possible existence of different pDC subsets. t-Distributed Stochastic Neighbor Embedding (t-SNE) is a data-visualization technique for dimensionality reduction that is particularly well suited for high-dimensional datasets. Since our panel consisted of only 6 dimensions (IFNα, CD80, PD-L1, TRAIL, CD2, and granzyme B), this technique is, in case of our study, more useful for visualizing co-expression of different markers and the indication of potential pDC subsets, rather than proving them.

Data of all representative donors were combined for robustness of the analysis to generate the t-SNE plots. Notably, the unprimed single-TLR activated pDCs show that the first responders form a separate cluster, as visualized with the corresponding heatmaps, irrespective of the TLR ligand used for activation (**Figure 5**). These clusters of first responders showed relatively high levels of cytotoxic marker expression compared to the non-IFN producing cells. Interestingly, the t-SNE plots revealed high numbers of co-expression of the two cytotoxicity markers in this cluster (percentage PD-L1 positive events of TRAIL positive events: 88.9%). Overall, for both IFN producers and non-producers, the vast majority of TRAIL positive events co-expressed PD-L1 (co-expression numbers of total events; median ± median absolute deviation: 82.8% ± 6.5%, R848 activated, n = 7).

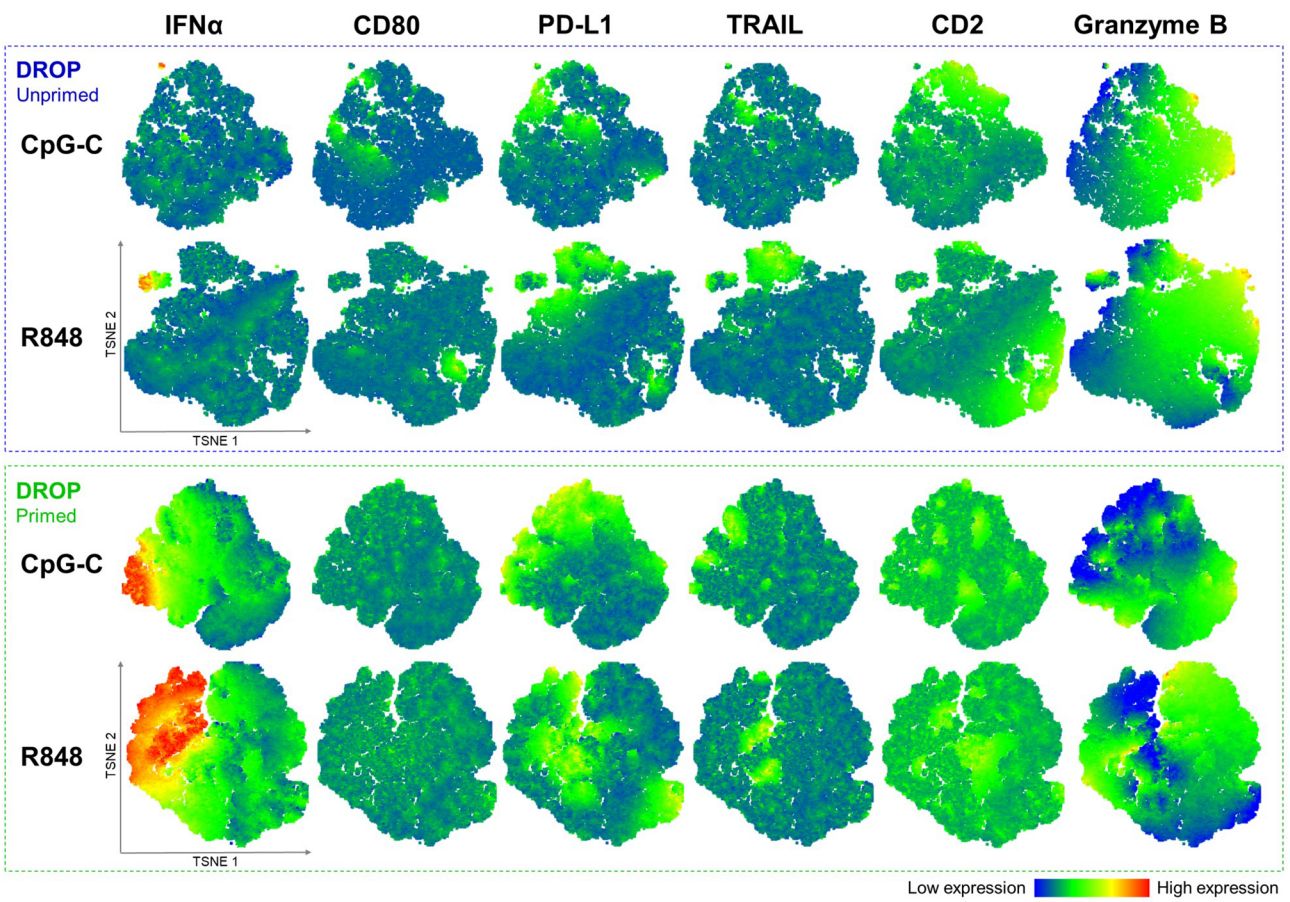


FIGURE 5 | Multi-dimensional data analysis shows clustering of early IFNα-producing pDCs with co-expression of PD-L1 and TRAIL. Single pDCs were either primed or left unprimed, incubated with IFNα Catch Reagent, activated with 50 μg/mL CpG-C or R848 for 18 hours, retrieved, stained for viability and markers of interest, and analyzed with flow cytometry as described earlier. Depicted are t-SNE plots based on the reanalysis of earlier presented data, based expression of IFNα, CD80, PD-L1, TRAIL, CD2, and granzyme B; DROP-Unprimed n = 10; DROP-Primed n = 5. Individual events are colored according to the expression of each indicated marker. Corresponding heatmap indicates the level of expression from low (blue) to high (red). Iterations 1000, perplexity 30.

Similarly, t-SNE plots were generated for all primed single-TLR activated pDCs. Results show that primed IFN-producing pDCs, referred to as second responders, do no longer segregate in a separate population as we observed for the first responders (**Figure 5**). As previously shown, we confirmed that priming does not significantly affect the expression of PD-L1 and TRAIL, with t-SNE plots showing similar patterns of expression and co-expression numbers as compared to the unprimed data (co-expression numbers of total events; median ± median absolute deviation: 85.4 ± 2.6%, R848 activated, n = 5).

Overall, the multi-dimensional data analysis visualized by t-SNE plots showed that the unprimed IFN-producing pDCs form a separated cluster. These results indicate that the 1-3% of early IFN-producing pDCs are potentially a distinct subset with relatively high expression of PD-L1 and TRAIL.

Early IFN-Producing pDCs Are Key Drivers for Orchestrating Population-Wide IFN-I Responses

Advancing single-cell technologies and refined experimental approaches have highlighted the multi-layered stochasticity present in the IFN-I system [reviewed in (36)]. This stochasticity turned out to be the underlying driver of the cellular heterogeneity that is observed today. Computational tools and mathematical models have helped interpret and understand the observed cellular heterogeneity and started to reveal the importance of first, second and non-responders involved in all kinds of biological processes. Therefore, we wanted to further characterize and investigate the role of first responders in eliciting population-wide effects.

To assess population-wide IFN dynamics without the functional presence of the first responders, freshly isolated pDCs were first activated in droplets for 18h to eradicate the first responders (**Figure 6A**). During this incubation, the first responders will produce IFN-Is, after which they will enter an IFN-desensitized state (37–39). Next, all remaining pDCs, those that did not produce IFN-Is during the first activation, will receive a second activation in bulk. As a control, freshly isolated pDCs were directly activated in bulk to assess the IFN-I dynamics for the first crucial hours, measured by intracellular IFNα. These results show that within the first 4 hours the peak of IFN positive cells has passed, reaching up to 40% positivity (**Figure 6B**). This number corresponds with the second responders fraction described in prior result sections. In the

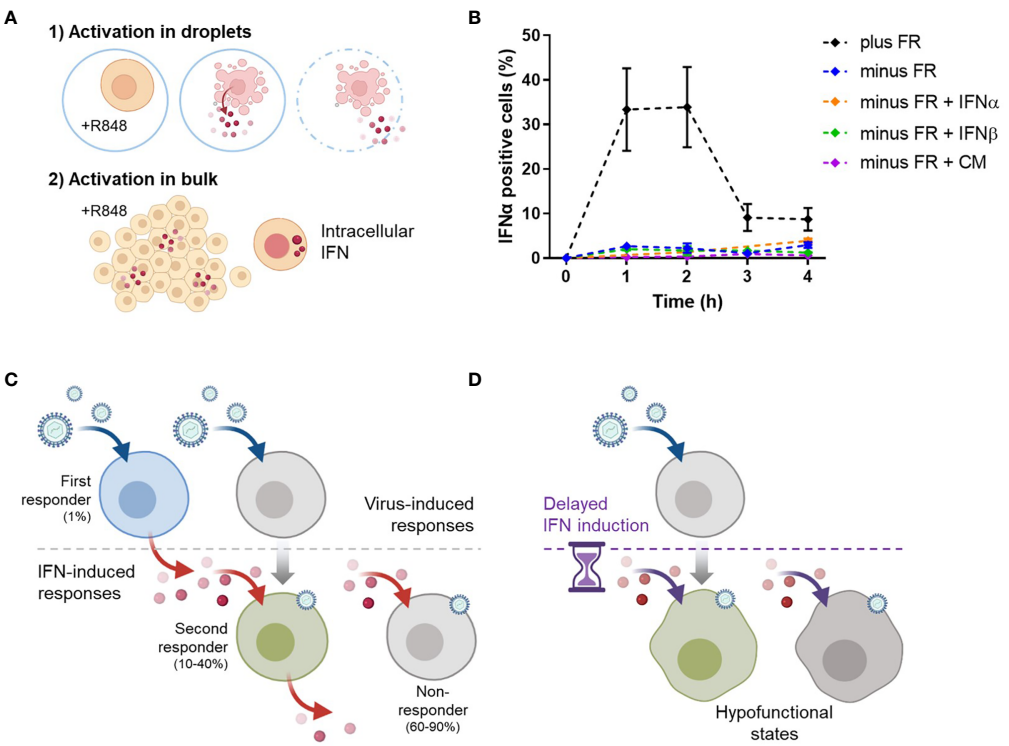


FIGURE 6 | Timing of IFN-I dynamics is key for orchestrating population-wide IFN-I responses. **(A)** Freshly isolated pDCs were either only activated in bulk with 5 μg/mL R848 (2) or first activated in droplets with 50 μg/mL R848 (1), followed by the activation in bulk (2). In theory, activation 1) leads to the eradication of first responders (minus FR), whereas activation 2), without prior activation 1), will still include the first responders (plus FR). Next, cells were fixed, permeabilized and stained for intracellular IFNα over the course of the first 4 hours after bulk activation (2). **(B)** Depicted are the dynamics of IFNα positive cells per condition. The additives IFNα (100pg), IFNβ (250pg), and conditioned medium (CM, 100%) were added upon activation 2). Dots represent mean percentages; error bars represent SEM, n = 6. **(C)** Schematic overview of proper IFN-I induction by a small fraction of first responders, induced upon viral detection. Subsequently, the secreted IFN-Is induce the larger fraction of second responders, which arise from the population that did not become a first responder in the earlier phase. The majority is left unresponsive. **(D)** Without the proper timing of the IFN-induced responses by the first responders, pDCs will enter a hypofunctional state, as observed in panel **(B)**.

absence of first responders, these levels of IFNα positive cells were not met. In fact, levels remained within the 1-3% range.

Following that, we assumed that the IFN-I response could be restored by supplying the approximate amount of IFN-Is that is usually produced by the first responders. Therefore, we added small amounts of IFNα or IFNβ to the conditions without the first responders present. Different amounts of IFN-Is were tested, ranging from 1 pg up to 1 μg (data not shown). However, none of it led to the increase in IFNα positive events. Finally, we added conditioned medium obtained from 1h bulk activation to the conditions without the first responders present, to ensure the right range of input for the second responders to get activated. Interestingly, also this approach did not led to the increase in IFNα positive events. This made us hypothesize that, under physiological conditions, population-wide IFN-I responses are driven by fractions of first responders, because of their fast and potent peak in IFNs that elicits the second responders (**Figure 6C**). When this trigger is delayed, the activated cells enter a hypofunctional state (**Figure 6D**).

To conclude, our results indicate that first responders can potentially serve as the key drivers of population wide IFN-I responses. In our experimental setup, IFN-I responses are impaired without their proper functioning.

DISCUSSION

The multifaceted biology of pDCs makes this versatile immune cell a promising therapeutic target for a wide range of diseases. The combination of their highly specialized function to produce massive amounts of IFNs, together with their ability to exert cytotoxic effector functions, makes pDCs the ideal treatment target to improve current immune therapies. Over the past decades, multiple different studies revealed heterogeneity present in the pDC compartment (10–14, 16). By changing the experimental approach from bulk activation to single-cell activation, we were able to further dive into the complex pDC biology by highlighting the important role for the first responders in eliciting population-wide IFN-I responses.

Similar to the phenomenon of a small fraction producing IFN-Is upon single TLR-activation, only small fractions of pDCs upregulate important cytotoxic markers, similar to the earlier described IKpDCs. It would be interesting to evaluate whether, besides their specialized function to secrete massive amounts of IFN-Is, IKpDCs share functional and phenotypical characteristics with NK cells, since several observations have suggested that these cells could represent a subset of NK cells (24). Additionally, our results are in agreement with previous studies that describe the effect of removing cell-to-cell communication resulting in a decreased expression of surface markers (16, 19). The phenomenon of a small fraction of cells amplifying population wide responses by activating surrounding cells *via* paracrine signaling has been extensively studied and proven for IFN-I signaling and multiple other immune reactions (17, 19, 40, 41). By making only a small fraction of cells respond to the initial stimulus, the response can be tightly regulated. For IFN signaling, this tight balance is crucial to avoid the harmful effects of excessive amounts of IFNs. For that matter, pDCs were classically described as being refractory to IFNα production upon repeated TLR activation, which could explain the non-responsiveness we observed in our study (42, 43). However, to the best of our knowledge, we are the first that have activated pDCs twice, of which the first activation happened in total isolation. Therefore, we can now conclude that this hypofunctional state is not initiated by soluble mediators released from the responding cells, but an intrinsic factor of pDC biology mediated by TLR-signaling.

Up to now, only HIV-stimulated pDCs allow for persistent IFNα production upon repeated stimulation (44). This persistent IFNα production was correlated with increased levels of IRF7 and was dependent upon the autocrine IFNα/β receptor feedback loop. Besides, multiple studies have proven the correlation between high basal expression levels of IRF7 with responsiveness, for both pDCs and other cell types (35, 45–48). IRF7 deficiency can even lead to recurrent influenza infection in humans, emphasizing its crucial role in eliciting proper IFN-I responses (49). This makes us hypothesize that the basal quantitative level of IRF7 present in single cells could potentially be one of the determining factors influencing cellular decision making, in particular whether to become a first or second responder, even though it has been stated that pDCs constitutively express background levels of IRF7 (50).

The multi-layered stochasticity underlying the IFN-I dynamics, thereby determining the first, second and non-responders, remains elusive. Though, the first responders seem to comprise a subset of high importance due to their dedication to the mass production of IFN-Is. Therefore, this subset is able to drive and orchestrate the population-wide IFN responses. In addition, possibly due to the autocrine signaling, the cytotoxic markers PD-L1 and TRAIL get upregulated, making them an ideal target for immune therapies. The question remains whether such hybrid cell will excel in performing both tasks *in vivo*. Although this subset might not excel in both, it could still be a crucial driver of all kinds of pDC functionalities. Therefore, a better understanding on the role of these highly specialized hybrid cells in orchestrating population-wide responses might push the field of pDC-based immune therapies to a new therapeutic level.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

Human primary pDCs were isolated from buffy coats of healthy donors (Sanquin), according to institutional guidelines and after informed consent per the Declaration of Helsinki. No additional ethics approval was required for their use, and written informed consent was provided by all donors.

AUTHOR CONTRIBUTIONS

LE, EC, NS, and JT designed the study. LE performed all experiments. LE analyzed the data. LE, EC and JT wrote the article. JT supervised the research. All authors contributed to the article and approved the submitted version.

ACKNOWLEDGMENTS

The authors want to thank Prof. Carlijn V.C. Bouten for all advice and suggestions on the manuscript, Bart Tiemeijer for all enthusiastic discussions, and Liesbeth Varion-Verhagen for advice on the flow cytometric analyses. This work is supported by an ERC Starting Grant ImmunoCode (802791), and furthermore we acknowledge generous support by the Eindhoven University of Technology.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2021.672729/full#supplementary-material

REFERENCES

1. Reizis B. Plasmacytoid Dendritic Cells: Development, Regulation, and Function. *Immunity* (2019) 50:37–50. doi: 10.1016/j.immuni.2018.12.027
2. Swiecki M, Colonna M. The Multifaceted Biology of Plasmacytoid Dendritic Cells. *Nat Rev Immunol* (2015) 15:471–85. doi: 10.1038/nri3865
3. Koucký V, Bouček J, Fialová A. Immunology of Plasmacytoid Dendritic Cells in Solid Tumors: A Brief Review. *Cancers (Basel)* (2019) 11(4):470. doi: 10.3390/cancers11040470
4. Conrad C, Gregorio J, Wang YH, Ito T, Meller S, Hanabuchi S, et al. Plasmacytoid Dendritic Cells Promote Immunosuppression in Ovarian Cancer Via ICOS Costimulation of Foxp3+ T-Regulatory Cells. *Cancer Res* (2012) 72:5240–9. doi: 10.1158/0008-5472.CAN-12-2271
5. Sisirak V, Faget J, Gobert M, Goutagny N, Vey N, Treilleux I, et al. Impaired IFN-α Production by Plasmacytoid Dendritic Cells Favors Regulatory T-cell Expansion That may Contribute to Breast Cancer Progression. *Cancer Res* (2012) 72:5188–97. doi: 10.1158/0008-5472.CAN-11-3468
6. Labidi-Galy SI, Sisirak V, Meeus P, Gobert M, Treilleux I, Bajard A, et al. Quantitative and Functional Alterations of Plasmacytoid Dendritic Cells Contribute to Immune Tolerance in Ovarian Cancer. *Cancer Res* (2011) 71:5423–34. doi: 10.1158/0008-5472.CAN-11-0367
7. Tel J, Anguille S, Waterborg CEJ, Smits EL, Figdor CG, de Vries IJM. Tumoricidal Activity of Human Dendritic Cells. *Trends Immunol* (2014) 35:38–46. doi: 10.1016/j.it.2013.10.007
8. Thomann S, Boscheinen JB, Vogel K, Knipe DM, DeLuca N, Gross S, et al. Combined Cytotoxic Activity of an Infectious, But non-Replicative Herpes Simplex Virus Type 1 and Plasmacytoid Dendritic Cells Against Tumour Cells. *Immunology* (2015) 146:327–38. doi: 10.1111/imm.12509
9. Boudewijns S, Bloemendal M, Gerritsen WR, de Vries IJM, Schreibelt G. Dendritic Cell Vaccination in Melanoma Patients: From Promising Results to Future Perspectives. *Hum Vaccines Immunother* (2016) 12:2523–8. doi: 10.1080/21645515.2016.1197453
10. Matsui T, Connolly JE, Michnevitz M, Chaussabel D, Yu C-I, Glaser C, et al. Cd2 Distinguishes Two Subsets of Human Plasmacytoid Dendritic Cells With Distinct Phenotype and Functions. *J Immunol* (2009) 182:6815–23. doi: 10.4049/jimmunol.0802008
11. Zhang H, Gregorio JD, Iwahori T, Zhang X, Choi O, Tolentino LL, et al. A Distinct Subset of Plasmacytoid Dendritic Cells Induces Activation and Differentiation of B and T Lymphocytes. *Proc Natl Acad Sci USA* (2017) 114:1988–93. doi: 10.1073/pnas.1610630114
12. Alcántara-Hernández M, Leylek R, Wagar LE, Engleman EG, Keler T, Marinkovich MP, et al. High-Dimensional Phenotypic Mapping of Human Dendritic Cells Reveals Interindividual Variation and Tissue Specialization. *Immunity* (2017) 47:1037–50.e6. doi: 10.1016/j.immuni.2017.11.001
13. Villani AC, Satija R, Reynolds G, Sarkizova S, Shekhar K, Fletcher J, et al. Single-Cell RNA-seq Reveals New Types of Human Blood Dendritic Cells, Monocytes, and Progenitors. *Science* (2017) 356(6335):eaah4573. doi: 10.1126/science.aah4573
14. Alculumbre SG, Saint-André V, Di Domizio J, Vargas P, Sirven P, Bost P, et al. Diversification of Human Plasmacytoid Predendritic Cells in Response to a Single Stimulus. *Nat Immunol* (2018) 19:63–75. doi: 10.1038/s41590-017-0012-z
15. Sinha N, Subedi N, Tel J. Integrating Immunology and Microfluidics for Single Immune Cell Analysis. *Front Immunol* (2018) 9:2373. doi: 10.3389/fimmu.2018.02373
16. Wimmers F, Subedi N, van Buuringen N, Heister D, Vivié J, Beeren-Reinieren I, et al. Single-Cell Analysis Reveals That Stochasticity and Paracrine Signaling Control Interferon-Alpha Production by Plasmacytoid Dendritic Cells. *Nat Commun* (2018) 9:3317. doi: 10.1038/s41467-018-05784-3
17. Bauer J, Dress RJ, Schulze A, Dresing P, Ali S, Deenen R, et al. Cutting Edge: Ifn-β Expression in the Spleen is Restricted to a Subpopulation of Plasmacytoid Dendritic Cells Exhibiting a Specific Immune Modulatory Transcriptome Signature. *J Immunol* (2016) 196:4447–51. doi: 10.4049/jimmunol.1500383
18. Scheu S, Dresing P, Locksley RM. Visualization of Ifnβ Production by Plasmacytoid Versus Conventional Dendritic Cells Under Specific Stimulation Conditions In Vivo. *Proc Natl Acad Sci USA* (2008) 105:20416–21. doi: 10.1073/pnas.0808537105
19. Shalek AK, Satija R, Shuga J, Trombetta JJ, Gennert D, Lu D, et al. Single-Cell RNA-seq Reveals Dynamic Paracrine Control of Cellular Variation. *Nature* (2014) 510:363–9. doi: 10.1038/nature13437
20. Berggren O, Hagberg N, Weber G, Alm GV, Rönnblom L, Eloranta ML. B Lymphocytes Enhance Interferon-α Production by Plasmacytoid Dendritic Cells. *Arthritis Rheum* (2012) 64:3409–19. doi: 10.1002/art.34599
21. Hjorton K, Hagberg N, Pucholt P, Eloranta ML, Rönnblom L. The Regulation and Pharmacological Modulation of Immune Complex Induced Type III IFN Production by Plasmacytoid Dendritic Cells. *Arthritis Res Ther* (2020) 22 (1):130. doi: 10.1186/s13075-020-02186-z
22. Welner RS, Pelayo R, Garrett KP, Chen X, Perry SS, Sun XH, et al. Interferon-Producing Killer Dendritic Cells (Ikdcs) Arise Via a Unique Differentiation Pathway From Primitive c-kitHiCD62L+ Lymphoid Progenitors. *Blood* (2007) 109:4825–31. doi: 10.1182/blood-2006-08-043810
23. Vremec D, O'Keeffe M, Hochrein H, Fuchsberger M, Caminschi I, Lahoud M, et al. Production of Interferons by Dendritic Cells, Plasmacytoid Cells, Natural Killer Cells, and Interferon-Producing Killer Dendritic Cells. *Blood* (2007) 109:1165–73. doi: 10.1182/blood-2006-05-015354
24. Terme M, Mignot G, Ullrich E, Bonmort M, Minard-Colin V, Jacquet A, et al. The Dendritic Cell-Like Functions of IFN-producing Killer Dendritic Cells Reside in the CD11b+ Subset and are Licensed by Tumor Cells. *Cancer Res* (2009) 69:6590–7. doi: 10.1158/0008-5472.CAN-08-4473
25. Huarte E, Rynda-Apple A, Riccardi C, Skyberg JA, Golden S, Rollins MCF, et al. Tolerogen-Induced Interferon-Producing Killer Dendritic Cells (Ikdcs) Protect Against EAE. *J Autoimmun* (2011) 37:328–41. doi: 10.1016/j.jaut.2011.09.005
26. Chan CW, Crafton E, Fan HN, Flook J, Yoshimura K, Skarica M, et al. Interferon-Producing Killer Dendritic Cells Provide a Link Between Innate and Adaptive Immunity. *Nat Med* (2006) 12:207–13. doi: 10.1038/nm1352
27. Taieb J, Chaput N, Ménard C, Apetoh L, Ullrich E, Bonmort M, et al. A Novel Dendritic Cell Subset Involved in Tumor Immunosurveillance. *Nat Med* (2006) 12:214–9. doi: 10.1038/nm1356

28. Drobits B, Holcmann M, Amberg N, Swiecki M, Grundtner R, Hammer M, et al. Imiquimod Clears Tumors in Mice Independent of Adaptive Immunity by Converting pDCs Into Tumor-Killing Effector Cells. *J Clin Invest* (2012) 122:575–85. doi: 10.1172/JCI61034
29. Sinha N, Subedi N, Wimmers F, Soennichsen M, Tel J. A Pipette-Tip Based Method for Seeding Cells to Droplet Microfluidic Platforms. *J Vis Exp* (2019) 144:57848. doi: 10.3791/57848
30. Kalb ML, Glaser A, Stary G, Koszik F, Stingl G. Trail + Human Plasmacytoid Dendritic Cells Kill Tumor Cells in Vitro: Mechanisms of Imiquimod- and IFN-α–Mediated Antitumor Reactivity. *J Immunol* (2012) 188:1583–91. doi: 10.4049/jimmunol.1102437
31. Tel J, Smits EL, Anguille S, Joshi RN, Figdor CG, De Vries IJM. Human Plasmacytoid Dendritic Cells are Equipped With Antigen-Presenting and Tumoricidal Capacities. *Blood* (2012) 120:3936–44. doi: 10.1182/blood-2012-06-435941
32. Gougeon ML, Herbeuval JP. Ifn-α and TRAIL: A Double Edge Sword in HIV-1 Disease? *Exp Cell Res* (2012) 318:1260–8. doi: 10.1016/j.yexcr.2012.03.012
33. Saïdi H, Bras M, Formaglio P, Melki MT, Charbit B, Herbeuval JP, et al. Hmgb1 Is Involved in IFN-α Production and TRAIL Expression by HIV-1-Exposed Plasmacytoid Dendritic Cells: Impact of the Crosstalk With NK Cells. *PloS Pathog* (2016) 12:e1005407. doi: 10.1371/journal.ppat.1005407
34. Kandhaya-Pillai R, Miro-Mur F, Alijotas-Reig J, Tchkonia T, Kirkland JL, Schwartz S. Tnfα-Senescence Initiates a STAT-dependent Positive Feedback Loop, Leading to a Sustained Interferon Signature, DNA Damage, and Cytokine Secretion. *Aging (Albany NY)* (2017) 9:2411–35. doi: 10.18632/aging.101328
35. Tomasello E, Naciri K, Chelbi R, Bessou G, Fries A, Gressier E, et al. Molecular Dissection of Plasmacytoid Dendritic Cell Activation In Vivo During a Viral Infection. *EMBO J* (2018) 37:e98836. doi: 10.15252/embj.201798836
36. Talemi SR, Höfer T. Antiviral Interferon Response At Single-Cell Resolution. *Immunol Rev* (2018) 285:72–80. doi: 10.1111/imr.12699
37. Ivashkiv LB, Donlin LT. Regulation of Type I Interferon Responses. *Nat Rev Immunol* (2014) 14:36–49. doi: 10.1038/nri3581
38. Czerkies M, Korwek Z, Prus W, Kochańczyk M, Jaruszewicz-Błońska J, Tudelska K, et al. Cell Fate in Antiviral Response Arises in the Crosstalk of IRF, Nf-κb and JAK/STAT Pathways. *Nat Commun* (2018) 9:1–14. doi: 10.1038/s41467-017-02640-8
39. Schneider WM, Chevillotte D, Rice CM. Interferon-Stimulated Genes: A Complex Web of Host Defenses. *Annu Rev Biochem* (2014) 32:513–45. doi: 10.1146/annurev-immunol-032713-120231
40. Doğanay S, Lee MY, Baum A, Peh J, Hwang SY, Yoo JY, et al. Single-Cell Analysis of Early Antiviral Gene Expression Reveals a Determinant of Stochastic: IFNB1 Expression. *Integr Biol (United Kingdom)* (2017) 9:857–67. doi: 10.1039/c7ib00146k
41. Drayman N, Patel P, Vistain L. HSV-1 Single-Cell Analysis Reveals the Activation of Anti-Viral and Developmental Programs in Distinct Sub-Populations. *Elife* (2019) 8:1–25. doi: 10.7554/eLife.46339
42. Ito T, Kanzler H, Duramad O, Cao W, Liu YJ. Specialization, Kinetics, and Repertoire of Type 1 Interferon Responses by Human Plasmacytoid Predendritic Cells. *Blood* (2006) 107:2423–31. doi: 10.1182/blood-2005-07-2709
43. Björck P. Dendritic Cells Exposed to Herpes Simplex Virus in Vivo do Not Produce Ifn-α After Rechallenge With Virus in Vitro and Exhibit Decreased T Cell Alloreactivity. *J Immunol* (2004) 172:5396–404. doi: 10.4049/jimmunol.172.9.5396
44. O'Brien M, Manches O, Sabado RL, Baranda SJ, Wang Y, Marie I, et al. Spatiotemporal Trafficking of HIV in Human Plasmacytoid Dendritic Cells Defines a Persistently IFN-α-Producing and Partially Matured Phenotype. *J Clin Invest* (2011) 121:1088–101. doi: 10.1172/JCI44960
45. Ma S, Wan X, Deng Z, Shi L, Hao C, Zhou Z, et al. Epigenetic Regulator CXXC5 Recruits DNA Demethylase Tet2 to Regulate TLR7/9-elicited IFN Response in Pdcs. *J Exp Med* (2017) 214:1471–91. doi: 10.1084/jem.20161149
46. Shalek AK, Satija R, Adiconis X, Gertner RS, Gaublomme JT, Raychowdhury R, et al. Single-Cell Transcriptomics Reveals Bimodality in Expression and Splicing in Immune Cells. *Nature* (2013) 498:236–40. doi: 10.1038/nature12172
47. Zhao M, Zhang J, Phatnani H, Scheu S, Maniatis T. Stochastic Expression of the Interferon-β Gene. *PloS Biol* (2012) 10:e1001249. doi: 10.1371/journal.pbio.1001249
48. Doğanay S, Lee MY, Baum A, Peh J, Hwang SY, Yoo JY, et al. Single-Cell Analysis of Early Antiviral Gene Expression Reveals a Determinant of Stochastic: IFNB1 Expression. *Integr Biol* (2017) 9:857–67. doi: 10.1039/c7ib00146k
49. Ciancanelli MJ, Huang SXL, Luthra P, Garner H, Itan Y, Volpi S, et al. Life-Threatening Influenza and Impaired Interferon Amplification in Human IRF7 Deficiency. *Science* (2015) 348:448–53. doi: 10.1126/science.aaa1578
50. Barchet W, Cella M, Odermatt B, Asselin-Paturel C, Colonna M, Kalinke U. Virus-Induced Interferon α Production by a Dendritic Cell Subset in the Absence of Feedback Signaling In Vivo. *J Exp Med* (2002) 195:507–16. doi: 10.1084/jem.20011666

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.

frontiers in Immunology

METHODS
published: 04 May 2021
doi: 10.3389/fimmu.2021.660037

Comparison of Two Strategies to Generate Antigen-Specific Human Monoclonal Antibodies: Which Method to Choose for Which Purpose?

Anna M. Ehlers[1,2]*, *Constance F. den Hartog Jager*[1,2], *Tineke Kardol-Hoefnagel*[1], *Miriam M.D. Katsburg*[1], *André C. Knulst*[1,2] *and Henny G. Otten*[1]

[1] Center for Translational Immunology, University Medical Center Utrecht, Utrecht University, Utrecht, Netherlands,
[2] Department of Dermatology/Allergology, University Medical Center Utrecht, Utrecht University, Utrecht, Netherlands

Human monoclonal antibodies (mAbs) are valuable tools to link genetic information with functional features and to provide a platform for conformational epitope mapping. Additionally, combined data on genetic and functional features provide a valuable mosaic for systems immunology approaches. Strategies to generate human mAbs from peripheral blood have been described and used in several studies including single cell sequencing of antigen-binding B cells and the establishment of antigen-specific monoclonal Epstein-Barr Virus (EBV) immortalized lymphoblastoid cell lines (LCLs). However, direct comparisons of these two strategies are scarce. Hence, we sought to set up these two strategies in our laboratory using peanut 2S albumins (allergens) and the autoantigen anti-Rho guanosine diphosphate dissociation inhibitor 2 (RhoGDI2, alternatively 'ARHGDIB') as antigen targets to directly compare these strategies regarding costs, time expenditure, recovery, throughput and complexity. Regarding single cell sequencing, up to 50% of corresponding V(D)J gene transcripts were successfully amplified of which 54% were successfully cloned into expression vectors used for heterologous expression. Seventy-five percent of heterologously expressed mAbs showed specific binding to peanut 2S albumins resulting in an overall recovery of 20.3%, which may be increased to around 29% by ordering gene sequences commercially for antibody cloning. In comparison, the establishment of monoclonal EBV-LCLs showed a lower overall recovery of around 17.6%. Heterologous expression of a mAb carrying the same variable region as its native counterpart showed comparable concentration-dependent binding abilities. By directly comparing those two strategies, single cell sequencing allows a broad examination of antigen-binding mAbs in a moderate-throughput manner, while the establishment of monoclonal EBV-LCLs is a powerful tool to select a small number of highly reactive mAbs restricted to certain B cell subpopulations. Overall, both strategies, initially set-up for peanut 2S albumins, are suitable to obtain human mAbs and they are easily transferrable to other target antigens as shown for ARHGDIB.

Keywords: monoclonal antibodies, single cell sequencing, immortalization, limiting dilution, antigen-specific B cells

OPEN ACCESS

Edited by:
Florian Wimmers,
Stanford University, United States

Reviewed by:
Zoltan Prohaszka,
Semmelweis University, Hungary
Katharina Roeltgen,
Stanford University, United States

***Correspondence:**
Anna M. Ehlers
a.m.ehlers@umcutrecht.nl

Specialty section:
This article was submitted to Molecular Innate Immunity, a section of the journal Frontiers in Immunology

Received: *28 January 2021*
Accepted: *12 April 2021*
Published: *04 May 2021*

Citation:
Ehlers AM, den Hartog Jager CF, Kardol-Hoefnagel T, Katsburg MMD, Knulst AC and Otten HG (2021) Comparison of Two Strategies to Generate Antigen-Specific Human Monoclonal Antibodies: Which Method to Choose for Which Purpose? Front. Immunol. 12:660037. doi: 10.3389/fimmu.2021.660037

INTRODUCTION

Antibody diversity enables the adaptive immune system to generate a humoral response against virtually any antigen. Gene recombination of variable (V), diversity (D) and joining (J) gene segments for the heavy chain and V and J gene segments for the corresponding light chain results in a wide variety of antibodies with distinct specificities. The diversity is even enlarged by imprecision during the V(D)J gene rearrangement process (1, 2). The introduction of somatic hypermutations is an additional tool to increase diversity but more importantly to strengthen antibody's affinity against the respective target (3).

Disease-related specific antibody repertoires are often studied by comparing specific B cell subpopulations between patients and healthy donors using next generation sequencing (4–7). This powerful approach, however, does not provide any information about antigen reactivity, affinity and functionality. For this reason, several studies included the generation of human monoclonal antibodies (mAbs) in order to assess their functionality and to map their characteristics related to their genetic features (8, 9). Besides identifying genetic features associated with health or disease, human mAbs can support the mapping of conformational epitopes formed by closely located amino acids upon folding (9). While linear epitopes, comprised of sequential amino acids, can easily be mapped by e.g. peptide microarrays, the mapping of conformational epitopes requires more sophisticated techniques like mass spectrometry, nuclear magnetic resonance spectroscopy and/or mutation libraries (10). Since these techniques can hardly be executed with patient serum containing polyclonal antibodies, human mAbs are powerful tools to overcome this obstacle. Moreover, data obtained with mAbs derived from humans are thought to be more easily translatable to clinical research compared to data obtained with e.g. mouse-derived mAbs (9, 11). Additionally, these genetic and functional data can provide a valuable mosaic for systems immunology approaches, which model the complexity of the immune system *in silico*.

The first human mAbs were obtained in the early 90's by phage display technology using single chain or Fab fragment libraries (12, 13). Nearly simultaneously, transgenic animals consisting of human immunoglobulin genes provided an additional tool (14). These approaches, however, are artificial and cannot represent a complete human antibody repertoire, emphasizing the advantage of mAbs generated from human peripheral blood or tissues. The first strategy to generate human mAbs from peripheral blood included the establishment of immortalized B cell lines by Epstein-Barr Virus (EBV) infection followed by limiting dilution cloning (15, 16). More recently, the immortalization by EBV was partly replaced by BCl-6 overexpression mimicking a germinal center status accompanied by constant antibody secretion (17). The most recent technique, however, is single cell sequencing with subsequent heterologous antibody expression (18).

This study provides a detailed description and comparison of two different strategies to generate human mAbs including single cell sequencing of antigen-binding B cells (Method 1) and the establishment of monoclonal EBV-immortalized B cell lines (Method 2). Both methods were used to generate human mAbs against peanut 2S albumins, major allergens in peanut allergy, and anti-Rho guanosine diphosphate dissociation inhibitor 2 (RhoGDI2, alternative and used 'ARHGDIB'), a non-HLA target potentially involved in graft failure upon kidney transplantation (19). The generation of peanut 2S albumin-specific mAbs enables the evaluation of differences between specific antibody repertoires of peanut allergic and tolerant patients sensitized to these allergens. mAbs directed against ARHGDIB, on the other hand, enables the assessment of the role of ARHGDIB-specific mAbs in the pathogenesis of non-HLA related graft failure upon kidney transplantation, which can potentially lead to new therapeutic concepts in this context. Method 1 is especially suitable for a broad examination of antigen-binding mAbs due to a less selective process compared to Method 2. Method 2, however, is a powerful tool to select a small number of highly reactive mAbs potentially applicable in the development of treatment strategies. A schematic overview of both methods is shown in **Figure 1**.

MATERIALS

Kits

EZ-Link™ Sulfo-NHS-Biotin, 21217, ThermoFisher Scientific, Waltham, Massachusetts, US

Pierce™ NHS-Activated Magnetic Beads, 88826, ThermoFisher Scientific, Waltham, Massachusetts, US

Transcriptor First Strand cDNA Synthesis Kit, 04379012001, Roche, Basel, Switzerland

Human B cell isolation kit II, 130-091-151, Miltenyi Biotec, Bergisch Gladbach, Germany

NucleoSpin gel and PCR clean up, REF 740609.250, Macherey Nagel, Düren, Germany

NucleoSpin Plasmid EasyPure, REF 740727.50, Macherey Nagel, Düren, Germany

Enzymes

RNAse inhibitor, N8080119, ThermoFisher Scientific, Waltham, Massachusetts, US

RNAsin, N2515, Promega, Madison, Wisconsin, US

SuperScript™ III Reverse Transcriptase, 18080044, ThermoFisher Scientific, Waltham, Massachusetts, US

AmpliTaq Gold™ DNA polymerase, N8080249, ThermoFisher Scientific, Waltham, Massachusetts, US

Exonuclease I (*E. coli*), M0293L, New England BioLabs, Ipswich, Massachusetts, US

FASTAP Thermosensitive Alkaline Phosphatase, EF0651, ThermoFisher Scientific, Waltham, Massachusetts, US

Phusion high-fidelity DNA polymerase, M0530L, New England BioLabs, Ipswich, Massachusetts, US

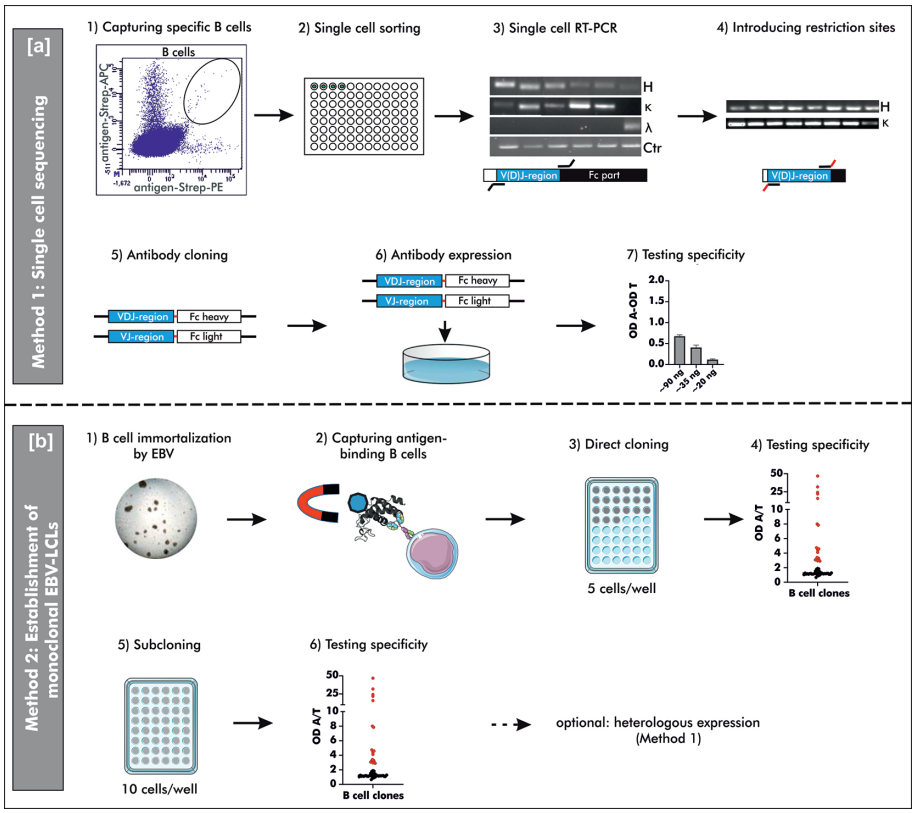


FIGURE 1 | Schematic overview on two different methods to generate human mAbs. **(A)** Method 1: "Single cell sequencing": B cells for single cell sorting were captured by double-positive antigen-tetramer staining (1 + 2). Upon transcription of total mRNA into cDNA, the V(D)J gene transcripts of the heavy and corresponding light chain were amplified by reverse transcriptase polymerase chain reaction (RT-PCR) using multiplex primers (3). The V(D)J gene usage was determined by Sanger sequencing and respective specific simplex primers were used to introduce restriction sites for subsequent cloning (4). Amplified V(D)J gene transcripts were cloned into commercially available pFUSEss-IgH vectors (Invivogen) carrying either the heavy or light chain constant backbone (Fc) (5). Vectors with correctly incorporated V(D)J gene transcripts were used for transient mammalian cell transfection (6) and the specificity was examined upon heterologous expression using a direct ELISA (7). OD A, optical density antigen; OD T, optical density transferrin (control). **(B)** Method 2: "Establishment of monoclonal EBV-LCLs": Antigen-binding B cells were captured from B cells immortalized by Epstein-Barr Virus (EBV) using antigen-coupled magnetic beads (1 + 2). Based on theoretical frequency of antigen-binding B cells, cells were directly cloned by seeding 5 cells/well on top of irradiated PBMCs as feeder cells (3). After 4 weeks of culturing, the supernatants were checked for antibodies binding specifically to the antigen of interest (4). B cells with supernatant containing specific antibodies were seeded for an additional round of cloning (5) and the resulting supernatant was screened for specific antibodies (6) after additional 4 weeks of culturing. Monoclonality was checked by Sanger sequencing and the mAbs can be heterologously expressed as described for Method 1. LCLs, lymphoblastoid cell lines; OD A, optical density antigen; OD T, optical density transferrin (control).

AmpliTaq polymerase, N8080161, ThermoFisher Scientific, Waltham, Massachusetts, US

EcoRI restriction enzyme, R0101L, New England BioLabs, Ipswich, Massachusetts, US

NheI restriction enzyme, R0131L, New England BioLabs, Ipswich, Massachusetts, US

BsiWI restriction enzyme, R0553L, New England BioLabs, Ipswich, Massachusetts, US

AvrII restriction enzyme, R0174L, New England BioLabs, Ipswich, Massachusetts, US

T4 Polynucleotide kinase, M0201L, New England BioLabs, Ipswich, Massachusetts, US

T4 DNA ligase, M0202L, New England BioLabs, Ipswich, Massachusetts, US

Antibodies

Antibodies for Single Cell Sorting

CD45-Pacific Orange, clone HI30, final dilution 1:40, MHCD4530, ThermoFisher Scientific, Waltham, Massachusetts, US

CD3-Pacific Blue, clone UCHT1, final dilution 1:160, 558117, BD Biosciences, East Rutherford, New Jersey, US

CD19-Fluorescein isothiocyanate, clone 4G7, final dilution 1:20, 345776, BD Biosciences, East Rutherford, New Jersey, US

CD14-R-phycoerythrin-cyanine dye Cy7, clone M5E2, final dilution 1: 800, 2109070, Sony Biotechnology, Inc., San Jose, California, US

CD16-R-phycoerythrin-cyanine dye Cy7, clone 3G8, final dilution 1:1000, 557744, BD Biosciences, East Rutherford, New Jersey, US

Streptavidin-Allophycocyanin, 17-4317-82, ThermoFisher Scientific, Waltham, Massachusetts, US

Streptavidin-R-phycoerythrin, 12-4317-87, ThermoFisher Scientific, Waltham, Massachusetts, US

Antibodies for ELISA

Goat anti-human IgE coupled with horse radish peroxidase (HRP), 5210-0158, KPL, Gaithersburg, Maryland, US:

Goat anti-human kappa antibody-HRP, 2060-05, SouthernBiotech, Homewood, Alabama, US

Goat anti-human lambda antibody-HRP, 2070-05, SouthernBiotech, Homewood, Alabama, US

Goat anti-human IgG FcY-HRP, 109-035-098, Jackson ImmunoResearch, Philadelphia, Pennsylvania, US

Goat anti-human IgG-PE antibody, 109-116-170, Jackson ImmunoResearch, Philadelphia, Pennsylvania, US

Donkey Anti-Goat IgG (H+L)-PE, 705-116-147, Jackson ImmunoResearch, Philadelphia, Pennsylvania, US

Vectors

pFUSEss-CHIg-hE (hEF1-HTLV promotor, IL-2 signal sequence for secretion, multiple cloning site upstream the constant region and zeocin resistance gene for selection), pfusess-hche2, Invivogen, San Diego, California, US

pFUSEss-CHIg-hG1 (hEF1-HTLV promotor, IL-2 signal sequence for secretion, multiple cloning site upstream the constant region and zeocin resistance gene for selection), pfusess-hchg1, Invivogen, San Diego, California, US

pFUSEss-CLIg-hk (hEF1-HTLV promotor, IL-2 signal sequence for secretion, multiple cloning site upstream the constant region and blasticidin resistance gene for selection), pfuse2ss-hclk, Invivogen, San Diego, California, US

pFUSEss-CHIg-hl (hEF1-HTLV promotor, IL-2 signal sequence for secretion, multiple cloning site upstream the constant region and blasticidin resistance gene for selection), pfuse2ss-hcll, Invivogen, San Diego, California, US

pAdVAntage™ Vector, E1711, Promega, Madison, Wisconsin, US

Cell Lines

Human embryonic kidney cells (HEK293-F (RRID : CVCL_6642), ThermoFisher Scientific, Waltham, Massachusetts, US

E. coli TOP10 cells, C404010, ThermoFisher Scientific, Waltham, Massachusetts, US

Irradiated (35 Gy) PBMCs, feeder cells for EBV-LCLs, freshly obtained from blood bank donors

Medium

Fast-Media Zeo Agar (Lysogency Broth (LB) medium containing agar and zeocin), fas-zn-s, Invivogen, San Diego, California, US

Fast-Media Zeo Liquid (LB medium containing zeocin), fas-zn-s, Invivogen, San Diego, California, US

Fast-Media Blas Agar (LB medium containing agar and blasticidin), fas-bl-s, Invivogen, San Diego, California, US

Fast-Media Blas Liquid (LB medium containing blasticidin), fas-bl-s, Invivogen, San Diego, California, US

FreeStyle 293 expression medium, ThermoFisher Scientific, Waltham, Massachusetts, US

RPMI-1640, 11554526, ThermoFisher Scientific, Waltham, Massachusetts, US supplemented with 20% fetal calf serum and 1% Penicillin-Streptomycin

Virus

Epstein-Barr virus was obtained from the supernatant of B95-8 cells, ACC 100, German Collection of Microorganisms and Cell Cultures GmbH, Braunschweig, Germany

Antigens

Native peanut 2S albumins: isolated from roasted peanuts as described previously by Ehlers and colleagues (20)

ARHGDIB, heterologously expressed in HEK 293F cells, as described previously by Kamburova and colleagues (21)

METHODS

For the evaluation of both methods, we obtained blood from 14 blood bank donors who gave broad written informed consent at the research blood bank of the University Medical Center Utrecht (Mini Donor Dienst). Additionally, we included data from patients who were recruited for a study on peanut allergy and tolerance including 6 peanut-allergic, 6 peanut-tolerant patients and 5 non-atopic controls (obtained from the research blood bank) for single cell sorting of 2S albumin-binding B cells. Peanut allergic and tolerant patients gave written informed consent at the time point of inclusion. The corresponding study was approved by the Medical Ethical Committee at the University Medical Center Utrecht under the number 17-945 (20). An overview on which sample was used for which evaluation is shown **Table 1**.

Method 1: "Single Cell Sequencing of Antigen-Binding B Cells"

Antigen-Tetramer Formation

Isolated peanut 2S albumins (20) or heterologously expressed ARHGDIB (21) were treated with an excess of biotin in accordance with manufacturer's instructions. The excess of biotin resulted in on average four biotin molecules per one molecule protein (EZ-Link™ Sulfo-NHS-Biotin). Antigen-tetramers were subsequently formed by separately adding streptavidin-R-phycoerythrin or streptavidin-Allophycocyanin to the biotinylated 2S albumin fraction or ARHGDIB in a molecular ratio of 1:1 (streptavidin: protein) (22, 23).

TABLE 1 | Patient samples used for each part of the protocol.

	Antigen	Patient samples
Method 1: "Single cell sequencing of antigen-binding B cells"		
FACS set-up	Peanut 2S albumins	4 blood bank donors†
Amplification		6 blood bank donors†
Ab expression		Study population
FACS set-up	ARHGDIB	3 blood bank donors‡
Ab expression		1 blood bank donor‡
Method 2: "Establishment of monoclonal EBV-LCLs"		
Optimal seeding density	Peanut 2S albumins	1 blood bank donor
Cloning round 1		2 blood bank donors§
Cloning round 2		2 blood bank donors§
Cloning round 1	ARHGDIB	2 blood bank donors¶
Cloning round 2		2 blood bank donors¶

†, ‡, §, ¶: indicating which samples were used for different steps

Single-Cell Sorting of Antigen Specific B Cells

Blood was drawn into heparin-coated tubes and freshly processed within 24 hours. Peripheral blood mononuclear cells (PBMCs) were obtained by density gradient centrifugation (Ficoll paque, SigmaAldrich) and they were finally resuspended in PBS containing 0.5% bovine serum albumin (BSA) and 2 mM EDTA. Subsequently, B cells were negatively enriched from the PBMCs fraction using antibody-coated magnetic beads (B cell isolation kit II). Enriched B cells were centrifuged (1000 x g, 20 °C, 5 min) and resuspended in PBS containing 0.5% BSA and 2 mM EDTA to obtain a concentration of 2*10^6 cells/40 µl. These B cells were stained for antigen specificity with 10 µl staining mix per 40 µl cell suspension using CD45-Pacific Orange (final dilution: 1:40), CD3-Pacific Blue (final dilution: 1:160), CD19-Fluorescein isothiocyanate (final dilution: 1:20), CD14-R-phycoerythrin-cyanine dye Cy7 (final dilution: 1:800), CD16-R-phycoerythrin-cyanine dye Cy7 (final dilution: 1:1000) and antigen-tetramers [75 pg[1] 2S albumins, 1.1 ng[2] ARHGDIB (24)]. Stained B cells were gated for CD45+, CD3-, CD19+, antigen tetramers (double positive) + and CD14/16- and single cell sorted into 96 wells fully-skirted PCR plates supplemented with 14 mM DTT and 11.2 U RNAse inhibitor in a total volume of 4 µl 0.5x PBS. Plates containing single-sorted B cells were immediately put on dry ice and stored at -80 °C until further processing. Before single cell sorting, the Aria II was calibrated to sort single cells into the middle of the respective well to minimize the probability that the sorted cell hits the wall.

For the selection of 2S albumin-binding B cells, a positive control comprising PBMCs from a healthy donor with detectable 2S albumin-binding B cells in previous experiments was included (healthy donor from the research blood bank but not belonging to the healthy donor pool used for analysis). B cells stained with biotin were used as negative control (Ctr 1). The gating was based on the respective negative control 1. An additional negative control (Ctr 2) was added to evaluate the optimal dilution of the used antigen-tetramers. This control consisted of a pre-incubation step with non-biotinylated antigen (75 pg 2S albumins, 1.1 ng ARHGDIB) followed by the normal staining protocol described above (**Figure 1A**, Step 1 and 2). The final gating strategy is shown in **Figure 2**.

[1] ≙ 30*10^6 antigen-tetramers sufficient for approximately 150-300 resting B cells.

[2] ≙ 112*10^6 antigen-tetramers sufficient for approximately 500-1000 resting B cells.

Gene Amplification of B Cell Receptors

V(D)J gene transcripts amplification was performed as previous described with minor modifications (18, 22, 25). While keeping the frozen B cells on dry ice, they were supplemented with 1.4% NP-40, 3 U RNAse inhibitor and 7 µM hexamer primers reaching a total volume of 7.5 µl. For primer annealing, this mixture was subsequently incubated for one minute at 68 °C and cDNA transcription was performed in accordance with manufacturer's instruction (SuperScript III). The resulting cDNA was subsequently used as template for the amplification of the heavy chain and its corresponding light chain gene transcript. Amplification was accomplished with 1.25 U hot-start DNA polymerase (AmpliTaq Gold), 2.5 mM $MgCl_2$, 1 mM dNTP mix and 40 nM multiplex primers (**Table 2**) for 50 cycles. The annealing temperature was set to 62, 60 and 58 °C for heavy, κ and λ light chain, respectively (**Figure 1A**, Step 3).

Note: Primers aliquots (10 µM) should be stored in 5 mM Tris-HCl buffer, pH 8.0 instead of Milli-Q water to prevent degradation. Before using the primers for amplification, the primers are diluted 1:10 (1 µM) using Milli-Q water.

Sequence Analysis of B Cell Receptors

Amplified V(D)J gene transcripts were purified by adding 0.5 µl Exo RI (0.01 U) and 1 µl FAST-AP (1 U) and incubating this mixture for 30 min at 37 °C followed by 20 min at 80 °C. Purified heavy chain gene transcripts were Sanger sequenced by Macrogen service using 200 nM multiplex reverse or framework (FR)1 forward primers. Light chain gene transcripts were sequenced using 200 nM of the respective reverse primers (**Table 2**). To check the quality of the sequences, they were evaluated using Chromas Lite, version 2.6.5. Double peaks, potentially resulting from errors in the beginning of the amplification reaction, were aligned to their germline and corrected if plausible. Quality-checked gene sequences were saved as FASTA files and subsequently used for automatic germline alignment using the IgBLAST web interface (reference: IMGT database) (26). The resulting output was written into a SQLite database using R 3.6.3.

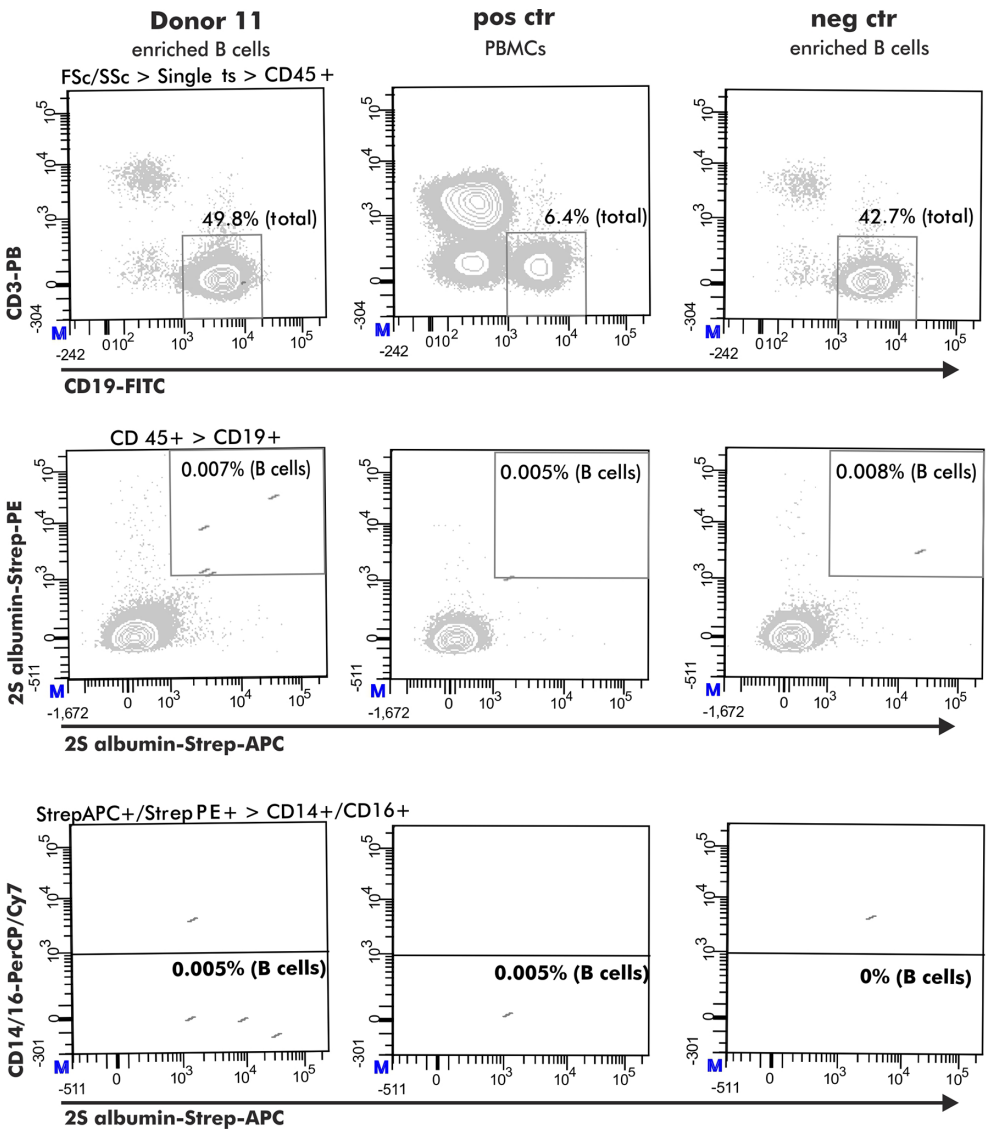


FIGURE 2 | Flow cytometry gating strategy to obtain antigen-specific B cells. The selection of antigen-specific B cells (derived from a peanut-tolerant patient) was based on expression of CD45+ (CD45-PO), CD3- (CD3-PB), CD19+ (CD19-FITC) and double-positive binding to biotinylated antigen coupled with either streptavidin-PE or streptavidin-APC. This population must be negative for CD14 (CD14-PerCP/Cy7) and CD16 (CD16-PerCP/Cy7) expression.

Cloning of V(D)J Gene Transcripts

To ensure the cloning of the entire V(D)J gene transcript without the introduction of additional amino acids, restriction sites were introduced with specific V gene forward and J gene reverse primers (**Table 2**). This introduction was achieved using 0.5 U Phusion high-fidelity DNA polymerase in presence of 1x reaction buffer, 2.5 mM $MgCl_2$, 1 mM dNTP mix and 400 nM respective forward and reverse primers. The annealing temperature was set to 62, 60 and 58 °C for heavy, κ and λ light chain, respectively and the amplification reaction was performed for 30 cycles. Gene products, purified accordingly to manufacturer's instructions (NucleoSpin gel and PCR clean up), were digested with 1 U of the respective restriction enzymes Eco RI and NhE I (heavy chain), BsWi (κ light chain) or Avr II (λ light chain). Before cloning the digested gene products into human IgH and IgL pFUSEss expression vectors, they were purified as already described and phosphorylated with 1 U T4-polynucleotide kinase for 60 min at 37 °C followed by 20 min at 65 °C. To prevent self-ligation, 1 µg of digested - 1 U respective restriction enzymes - and gel-purified parent vector was dephosphorylated with 4 U FAST-AP for 10 min at 37 °C followed by 10 min at 70 °C. Dephosphorylated vectors were mixed with digested gene products in a molecular ratio of 5:1 and incubated with 1 U T4 ligase for 1 hour at room temperature.

TABLE 2 | Primers used for single cell amplification, sequencing and cloning control; underlined: restriction sites; iltalic: additional nucleotide to ensure correct insertion.

Primer	Sequence
Heavy chain primer forward	
Single cell RT-PCR	
5' L-Vh1	ACAGGTGCCCACTCCCAGGTGCAG
5' L-Vh3	AAGGTGTCCAGTGTGARGTGCAG
5' L-Vh4/6	CCCAGATGGGTCCTGTCCCAGGTGCAG
5' L-Vh5	CAAGGAGTCTGTTCCGAGGTGCAG
Introduction restriction sites	
5' EcoRI_VH1	ATATTGAATTC*G*CAGGTGCAGCTGGTGCAG
5' EcoRI_VH1/5	ATATTGAATTC*G*GAGGTGCAGCTGGTGCAG
5' EcoRI_VH1-18	ATATTGAATTC*G*CAGGTTCAGCTGGTGCAG
5' EcoRI_VH1-24	ATATTGAATTC*G*CAGGTCCAGCTGGTACAG
5' EcoRI_VH3	ATATTGAATTC*G*GAGGTGCAGCTGGTGGAG
5' EcoRI_VH3-23	ATATTGAATTC*G*GAGGTGCAGCTGTTGGAG
5' EcoRI_VH3-33	ATATTGAATTC*G*CAGGTGCAGCTGGTGGAG
5' EcoRI_VH3-9	ATATTGAATTC*G*GAAGTGCAGCTGGTGGAG
5' EcoRI_VH4	ATATTGAATTC*G*CAGGTGCAGCTGCAGGAG
5' EcoRI_VH4-34	ATATTGAATTC*G*CAGGTGCAGCTACAGCAGTG
5' EcoRI_VH4-39	ATATTGAATTC*G*CAGCTGCAGCTGCAGGAG
5' EcoRI_VH6-1	ATATTGAATTC*G*CAGGTACAGCTGCAGCAG
Sanger Sequencing	
5' Vh1-FR1_(1-2)	GGCCTCAGTGAAGGTCTCCTGCAAG
5' Vh2-FR1_(2-5)	GTCTGGTCCTACGCTGGTGAAACCC
5' Vh3-FR1_(3-7)	CTGGGGGGTCCCTGAGACTCTCCTG
5' Vh4-FR1_(4-4)	CTTCGGAGACCCTGTCCCTCACCTG
5' Vh5-FR1_(5-51)	CGGGGAGTCTCTGAAGATCTCCTGT
5' Vh6-FR1_(6-1)	TCGCAGACCCTCTCACTCACCTGTG
Colony screening	
5' Vh1-FR2_(1-2)	CTGGGTGCGACAGGCCCCTGGACAA
5' Vh2-FR2_(2-5)	TGGATCCGTCAGCCCCCAGGGAAGG
5' Vh3-FR2_(3-7)	GGTCCGCCAGGCTCCAGGGAA
5' Vh4-FR2_(4-4)	TGGATCCGCCAGCCCCCAGGGAAGG
5' Vh5-FR2_(5-51)	GGGTGCGCCAGATGCCCGGGAAAGG
5' Vh6-FR2_(6-1)	TGGATCAGGCAGTCCCCATCGAGAG
5' Vh7-FR2_(7)	TTGGGTGCGACAGGCCCCTGGACAA
Heavy chain primer reverse	
Single cell RT-PCR	
3' CH1_IgA*	AGCCCTGGACCAGGCA
3' CH1_IgE*,**	GAAGACGGATGGGCTCTGT
3' CH1_IgG*,**	GGAAGGTGTGCACGCCGCTG
3' CH1_IgM*	GGGAATTCTCACAGGAGACG
Introduction restriction sites	
3' NheI_JH1/2/4/5	ATGCTAGCTGAGGAGACGGTGACCAG
3' NheI_JH3	ATGCTAGCTGAAGAGACGGTGACCATTG
3' NheI_JH6	ATGCTAGCTGAGGAGACGGTGACCGTG
Kappa light chain primer forward	
Single cell RT-PCR	
5' L-Vκ1/2	ATGAGGSTCCCYGCTCAGCTGGTGG
5' L-Vκ3	CTCTTCCTCCTGCTACTCTGGCTCCCAG
5' L-Vκ4	ATTTCTCTGTTGCTCTGGATCTCTG
Introduction restriction sites	
5' EcoRI_Vκ1-5	ATATTGAATTC*A*GACATCCAGATGACCCAGTC
5' EcoRI_Vκ1-9	ATATTGAATTC*A*GACATCCAGTTGACCCAGTCT
5' EcoRI_Vκ1D-43	ATATTGAATTC*A*GCCATCCGGATGACCCAGTC
5' EcoRI_Vκ2-24	ATATTGAATTC*A*GATATTGTGATGACCCAGAC
5' EcoRI_Vκ2-28	ATATTGAATTC*A*GATATTGTGATGACTCAGTC
5' EcoRI_Vκ2-30	ATATTGAATTC*A*GATGTTGTGATGACTCAGTC
5' EcoRI_Vκ3-11	ATATTGAATTC*A*GAAATTGTGTTGACACAGTC
5' EcoRI_Vκ3-15	ATATTGAATTC*A*GAAATAGTGATGACGCAGTC
5' EcoRI_Vκ3-20	ATATTGAATTC*A*GAAATTGTGTTGACGCAGTCT
5' EcoRI_Vκ4-1	ATATTGAATTC*A*GACATCGTGATGACCCAGTC
Colony screening	
5' Vk1f/6	TCAAGGTTCAGCGGCAGTGGATCTG
5' Vk2f	GGCCTCCATCTCCTGCAGGTCTAGTC
5' Vk3f	CCCAGGCTCCTCATCTATGATGCATCC
5' Vk4_int	CAACTGCAAGTCCAGCCAGAGTGTTTT
5' Vk5_int	CCTGCAAAGCCAGCCAAGACATTGAT
5' Vk7_int	GACCGATTTCACCCTCACAATTAATCC
Kappa light chain primer reverse	
Single cell RT-PCR	
3' Cκ 494*,**	GTGCTGTCCTTGCTGTCCTGCT
Introduction restriction sites	
3' BsiWI_Jκ1/4	ATCGTACGTTTGATYTCCACCTTGGTC
3' BsiWI_Jκ2	ATCGTACGTTTGATCTCCAGCTTGGTC
3' BsiWI_Jκ3	ATCGTACGTTTGATATCCACTTTGGTC
3' BsiWI_Jκ5	ATCGTACGTTTAATCTCCAGTCGTGTC
Lambda light chain primer forward	
Single cell RT-PCR	
5' L-Vλ1	GGTCCTGGGCCCAGTCTGTGCTG
5' L-Vλ2	GGTCCTGGGCCCAGTCTGCCCTG
5' L-Vλ3	GCTCTGTGACCTCCTATGAGCTG
5' L-Vλ4/5	GGTCTCTCTCSCAGCYGTTGCTG
5' L-Vλ6	GTTCTTGGGCCAATTTTATGCTG
5' L-Vλ7	GGTCCAATTCYCAGGCTGTGGTG
5' L-Vλ8	GAGTGGATTCTCAGACTGTGGTG
Introduction restriction sites	
5' EcoRI_Vλ1	ATATTGAATTC*G*CAGTCTGTGCTGACKCAG
5' EcoRI_Vλ2	ATATTGAATTC*G*CAGTCTGCCCTGACTCAG
5' EcoRI_Vλ3	ATATTGAATTC*G*TCCTATGAGCTGACWCAG
5' EcoRI_Vλ4/5	ATATTGAATTC*G*CAGCYTGTGCTGACTCA
5' EcoRI_Vλ6	ATATTGAATTC*G*AATTTTATGCTGACTCAG
5' EcoRI_Vλ7/8	ATATTGAATTC*G*CAGRCTGTGGTGACYCAG
Colony screening	
5' Vλ1/2_int	ATTCTCTGGCTCCAAGTCTGGC
5' Vλ3_int	GGATCCCTGAGCGATTCTCTGG
Lambda light chain primer reverse	
Single cell RT-PCR	
3' Cλ*,**	CACCAGTGTGGCCTTGTTGGCTTG
Introduction restriction sites	
3' AvrII_Jλ1	ATTCCTAGGACGGTGACCTTGGT
3' AvrII_Jλ2/3	ATTCCTAGGACGGTCAGCTTGGT
3' AvrII_Jλ6	ATTCCTAGGACGGTCACCTTGGT
3' AvrII_Jλ7-1	ATTCCTAGGACGGTCAGCTGGGT
3' AvrII_Jλ7-2	ATTCCTAGGGCGGTCAGCTGGGT

(Continued)

TABLE 2 | Continued

**also used for Sanger Sequencing; **also used for colony screening.*

The ligated vector was transformed into competent *E. coli* Top10 cells by allowing them to rest on ice for 30 min and subsequently, by incubating the mixture for 45 seconds at 42 °C (heat shock). Overnight grown colonies (LB Agar containing Zeocin or Blasticidin) were screened for incorporated V(D)J gene transcripts by PCR using 400 nM forward primers binding to the respective FR2 region and 400 nM reverse primers suitable for the constant part of the respective vector (**Table 2**) in presence of 2.5 U AmpliTaq polymerase, 2.5 mM $MgCl_2$ and 1 mM dNTP mix. The annealing temperature was set to 63, 62 and 61 °C for heavy, κ and λ light chain, respectively and the amplification cycle was repeated for 30 times. Positive clones were grown overnight in 3 ml LB medium containing either zeocin or blasticidin. Purified vectors (NucleoSpin Plasmid EasyPure) were Sanger sequenced using the Macrogen service and the correctness was verified by aligning the vector sequence to the first sequence result (**Figure 1A**, Step 4 and 5).

Heterologous Expression of Monoclonal Antibodies

For heterologous expression of human mAbs, human embryonic kidney (HEK) 293F cells were cultured in FreeStyle 293 expression medium using 125 ml shaking culture flasks. Exponentially growing cells at a confluence of 80% and viability of 90% were transiently transfected with VH and VL expression vectors in a ratio of 2:3 (total 0.5 µg plasmid DNA per 1*10^6 cells) using 150 mm culture plates and 293fectin (2 µl/µg plasmid DNA, ThermoFisher Scientific). To ensure a sufficient intake, the expression vectors were supplemented with 0.5 µg pAdVAntage plasmid. The supernatant was harvested three days upon transfection and stored at -20 °C for further analyses (**Figure 1A**, Step 6).

Method 2: "Establishment of Monoclonal EBV-LCLs"

Immortalization of Enriched B Cells by Epstein-Barr Virus

B cells, isolated from heparin blood as described for Method 1, were immortalized with EBV in presence of the TLR9 agonist CpG 2006 and the immunosuppressive Cyclosporine A to establish lymphoblastoid cell lines (LCLs) (15). In detail, 1 ml of EBV-containing supernatant, obtained from growing B-95.8 cells, was added to 5*10^6 pelleted B cells and incubated for 1 hour at 37 °C and 5% CO_2. Upon washing with 1 ml PBS, infected B cells were resuspended in 3 ml RPMI-1640 supplemented with 20% fetal calf serum (FCS), 2.5 µg/ml CpG 2006, 1 µg/ml Cyclosporine A and 1% Penicillin-Streptomycin (Pen/Strep). The suspension was cultured at 37 °C and 5% CO_2 until visible clusters were formed and the medium changed its color from red to yellow due to acidification (**Figure 1B**, Step 1).

Isolation of Antigen-Binding B Cells

Antigen-binding B cells were isolated from LCLs with an excess of antigen-coupled Pierce™ NHS-activated magnetic beads. The antigen was coupled to the beads in accordance with manufacturer's instruction. For the isolation, LCLs (10*10^6 cells) were pelleted and cooled on ice for 60 min. The pellet was subsequently resuspended in 100 µl PBS supplemented with 0.5% BSA and 2 mM EDTA accompanied by 30 µl of antigen-coupled beads and cooled on ice for another 10 min. This suspension was applied on a magnetic column (MACS Cell Separation Columns, Miltenyi Biotec B.V.) and the separation was achieved in accordance with manufacturer's instruction. The elution fraction was collected as antigen-binding B cells and was used to generate antigen-binding monoclonal LCLs (**Figure 1B**, Step 2).

Direct Cloning

The number of isolated antigen-binding LCLs was estimated based on the frequency of antigen-binding B cells determined by flow cytometry analysis (~ 0.01% of the B cells - Method 1). The elution fraction of the magnetic separation was mixed with 1*10^6/ml irradiated PBMCs (35 Gy) suspended in RPMI-1640 containing 20% FCS and 1% Pen/Strep to achieve a concentration of 25 antigen-binding LCLs/ml. 200 µl of this mixture was transferred to one well of a flat-bottom 96 wells plate (= 5 cells/well). The cells were allowed to grow for 4 weeks without re-feeding and the plate was tilted after the first week of culturing to keep the cells in close contact after the feeder cells died off. The plate was straightened again after an additional week of culturing to avoid too close contact between steadily proliferating cells (16) (**Figure 1B**, Step 3).

Note: It is crucial to tilt the 96 well plate after one week of culturing and straightening it again after 2 weeks of culturing.

Generating Monoclonal LCLs by Limiting Dilution Cloning

LCLs with supernatant containing antibodies specific to the antigen of interest were used for a second round of cloning to generate monoclonal LCLs. To this end, positive LCLs from the direct cloning step were counted and diluted to 50 cells/ml in RPMI-1640 supplemented with 20% FCS, 1% Pen/Strep and 1*10^6/ml irradiated PBMCs (35 Gy). This mixture was seeded in a volume of 200 µl/well in a flat-bottom 96 wells plate (10 cells/well) and incubated for 4 weeks as described above. For defining the optimal seeding density, a range from 0.3 to 10 cells/well was used.

Grown LCLs secreting antibodies specific for the antigen of interest were transferred to a 5 mL round bottom polystyrene test tube and cultured until a visible pellet was observed and the medium color changed from red to yellow. For further expansion, the LCLs were first transferred to a 25 cm^2 culture flask and subsequently to a 75 cm^2 culture flask. Expanded LCLs secreting specific mAbs were frozen at -80 °C and the supernatant containing mAbs was stored at -20 °C. Monoclonality was checked by Sanger Sequencing as described for Method 1 (**Figure 1B**, Step 4 to 6).

RNA Extraction

Total RNA was isolated accordingly to manufacturer's instruction (RNA-Bee, BioConnect). Briefly, pelleted LCLs were homogenized in 1 ml RNA-Bee and the RNA was separated from the genomic DNA by adding 200 µl of chloroform and spinning for 15 min at 12.000 x g and 4 °C. The colorless phase was transferred to 500 µl ice-cold isopropanol and incubated for 10 min on ice to precipitate RNA. The precipitate was washed with 75% ethanol and the resulted pellet was resolved in RNAse free water for 15 min at 55 °C. Extracted RNA was either stored at -80 °C or used immediately.

cDNA Transcription and Gene Transcript Amplification

RNA was transcripted into cDNA in accordance with manufacturer's instructions (Transcriptor First Strand cDNA Synthesis Kit). Briefly, 2 µl of random hexamer nucleotides (5 µM) were mixed with 8 µl of RNA template (1 µg) and the mixture was incubated for 3 min at 85 °C. Upon cooling down, 10 µl cDNA reaction mixture containing 1X RT-buffer, 5 mM $MgCl_2$, 1 mM dNTP mix, 5U RNA inhibitor, 1.5 U avian myeloblastosis virus (AMV)-RT and 10 mM gelatin was added and incubated for 90 min at 42 °C. The reaction was stopped by inactivating the AMV-RT for 3 min at 85 °C. cDNA was either stored at -20 °C or immediately used for V(D)J gene transcript amplification as

described for Method 1. Contrary to Method 1, the primers concentrations were adjusted to 400 nM instead of 40 nM.

Examination of Specific Antigen-Binding

Specific Binding to Peanut 2S Albumins

Specificity of native and heterologously produced mAbs to peanut 2S albumins was tested using a direct ELISA. Briefly, plates were coated by applying either 0.3 µg/well Ara h 2 and 6 (27) or transferrin (negative control) overnight at room temperature. On the following day, the plate was blocked with PBS supplemented with 1% BSA and 0.1% Tween-20 (blocking buffer) for one hour at room temperature. Subsequently, supernatants were applied upon 1:2 dilution in blocking buffer (EBV LCLs) or in serial dilution (1 to 10 µg/ml) (heterologously expressed mAbs) for one hour at room temperature under continuous shaking. Bound antibodies from EBV-LCLs were detected with anti-human kappa (final dilution: 1:10.000) and anti-human lambda antibodies coupled with horse radish peroxidase (HRP) (final dilution: 1:5000) for one hour at room temperature under continuous shaking. Bound heterologously expressed mAbs, on the other hand, were detected by either goat anti-human IgE (final dilution: 1:5000) or goat anti-human IgG1 (final dilution: 1:2000) antibodies coupled with HRP under the same conditions. Visualization was provided by adding tetramethylbenzidine (TMB) for 15 minutes in the dark and the optical density (OD) was measured at 450 nm. Native Abs were considered for further cloning or analysis when the sample OD was at least two times higher than the OD of the negative control for round 1 and 2.5 times higher for round 2. Heterologously expressed mAbs were defined as specific if the sample OD was, upon subtraction of the negative control OD value, at least 1.5 times greater than the OD obtained with culture medium containing no mAbs. This evaluation was chosen, in comparison to the evaluation of native Abs, because the use of anti-IgE and anti-IgG-HRP as secondary antibodies resulted in higher background OD values when measuring binding to transferrin (coated mock antigen) compared with OD values to peanut 2S albumins as coated antigen. mAbs with OD values above 1 at a concentration of 10 µg/ml were considered as strong binders, mAbs with OD values between 0.07 and 1 at all concentration steps were considered as moderate binders and mAbs with increased OD values at the highest concentration of 10 µg/ml, but no detectable OD at the lowest concentration of 1 µg/ml were considered as weak binders.

Specific Binding to ARHGDIB

Antibody specificity to ARHGDIB was evaluated using ARHGDIB-coupled microspheres diluted in PBS supplemented with 0.1% BSA (wash buffer) (21). IgG-coupled microspheres served as positive control whilst empty and transferrin coupled microspheres served as negative controls. All incubation steps were performed in the dark, at room temperature and with continuous shaking.

For each antibody to be tested, 1500 microspheres consisting of 4 colors, each individually coated, were incubated with 50 µl undiluted HEK293 (Method 1) or EBV-LCLs supernatant (Method 2) containing the respective mAb and incubated overnight. Upon washing with a Bio-Plex Pro Wash station (Bio-Rad), bound mAbs were detected with either an 1-step (Method 1) or a 2-step procedure (Method 2). For the 1-step procedure, 50 µl of 1:50 diluted PE-conjugated goat-anti human IgG antibody was added and incubated for 30 min. For the 2-step procedure, 50 µl of goat anti-human kappa (final dilution: 1:100) and goat anti-human lambda antibody (final dilution: 1:32) was added and incubated for 30 min. For the second step, 50 µl of 1:100 diluted PE-conjugated donkey anti-goat IgG antibody was added and incubated for additional 30 min. For the readout of both procedures, 50 µl of washing buffer was added and the median fluorescence intensities (MFI) were measured on a Luminex 200 flow analyzer (Luminex Corp) (50 counts, 75 µl sample volume, 90s time out).

Recovery Rate and Cost Calculation

The recovery rate of specific human mAbs was calculated for both methods as followed:

Method 1:

$$\begin{aligned} Recovery\ \ rate &= 100\,\% * amlification\ \ efficiency * cloning \\ &\quad efficiency * portion\ \ specific\ \ mAbs \\ &= 100\,\% * 0.5 * 0.54 * 0.75 \\ &= 20.3\,\% \end{aligned}$$

Method 2:

$$\begin{aligned} Recovery\ \ rate &= \frac{n\left({}^{s}_{\mathrm{dc}}wells\right) * efficiency_{dc} * n\left(\frac{mAbs}{{}^{s}_{\mathrm{dc}}wells}\right) * efficiency_{mc}}{n(antigen{-}binding\ \ B\ \ cells_{theo})} * 100\,\% \\ &= \frac{200 * 0.8 * 2 * 0.55}{1000} * 100\,\% \\ &= 17.6\,\% \end{aligned}$$

$({}^{s}_{\mathrm{dc}})wells$ = seeded wells for direct cloning based on theoretically antigen-binding B cells

$efficiency_{dc}$ = efficiency of direct cloning (approximate 2S albumins and ARHGDIB)

$n(\frac{mAbs}{{}^{s}_{\mathrm{dc}}wells})$ = approximate number of mAbs obtained from one seeded well for direct cloning

$efficiency_{mc}$ = efficiency to obtain monoclonal cell lines upon 2nd round of cloning

Costs for the generation of one single human mAb was calculated by adding up the expenses and correcting it for the recovery rate of the respective strategy. Personal costs were included based on an average salary of a research technician and the approximate working hours needed to produce one single mAb.

RESULTS

Method 1: "Single Cell Sequencing"

Double Tetramer-Staining Reduced the Selection of Non-specific CD19+ B Cells

Antigen-binding B cells for subsequent single cell V(D)J gene transcript amplification were detected by flow cytometry using antigen-tetramers formed with fluorophore-labeled streptavidin. For the development of an optimal staining protocol, enriched

B cells, derived from the same blood bank donor, were stained with 2S albumin-tetramers accompanied by either a single (APC) or two distinct fluorophores (APC and PE). As shown in **Figure 3A**, B cells single-positive for 2S albumin-tetramer binding accounted for 0.1% of the total CD19+ B cell fraction. Both controls - staining with biotin without the antigen (Ctr 1) and blocking with unlabeled antigen (Ctr 2) - showed, however, a comparable percentage of 2S albumin-binding B cells (Ctr 1: 0.092%; Ctr 2: 0.097%). The subtraction of Ctr 1 resulted in a final percentage of 0.008% 2S albumin-binding B cells. In comparison, the fraction of double-positive 2S albumin-binding B cells was already reduced to a percentage of 0.004% without any background staining, indicating the potential of double antigen-tetramer staining for identifying bona fide antigen-binding B cells. For validation purposes, double-tetramer staining was used for the detection of 2S albumin (**Figure 3B**) or ARHGDIB-binding B cells (**Figure 3C**) in 3 independent blood bank donors, respectively (n=6, **Table 1**). 2S albumin-binding B cells ranged from 0.002 to 0.007% whilst ARHGDIB-binding B cells ranged from 0.005 to 0.015%, indicating a good reproducibility of the developed staining protocol.

V(D)J Gene Transcript Amplification Efficiency Is Donor-Dependent

Amplification efficiency upon single cell sorting of 2S albumin-binding B cells was evaluated in 6 independent blood bank donors. Amplification efficiency of the heavy chain VDJ gene transcript from 24 to 50 individual B cells ranged from 29 to 63% whilst the percentage of successfully amplified V(D)J gene transcripts from the heavy and corresponding light chain was reduced and ranged from 17 to 50%. Although a low amplification efficiency was shown for two donors, most of the donors (4/6) showed sufficient amplification efficiencies of approximately 50%[3]. Cloning of successfully amplified heavy and corresponding light chain gene transcripts from peanut allergic, peanut tolerant and non-atopic controls resulted into an overall cloning efficiency of 54% which may be increased by ordering the not successfully cloned gene sequences commercially. Individual cloning efficiencies were estimated to 76% for the heavy chain gene transcripts and 71% for the light chain gene transcripts (78% kappa and 31% lambda). In the present study, 4 gene sequences, whose cloning originally failed, were commercially obtained and the cloning of those sequences were subsequently successful. Hence, the overall cloning efficiency might be increased to 76% when obtaining gene sequences partly commercially.

Antibodies Derived From Double-Positive Tetramer-Binding B Cells Are Mostly Specific

The specificity of human mAbs, generated by single cell sequencing, can only be examined upon heterologous expression in the end of the workflow. Overall, 33 heterologously expressed mAbs from 10 different donors carried the variable region of B cells selected for their putative binding to peanut 2S albumins and 1 mAb carried the variable region of one B cell selected for putative binding to ARHGDIB. These mAbs were randomly expressed as IgE or IgG1 mAbs as this enabled the assessment of their potency to induce degranulation (IgE) and their ability to block serum IgE binding (IgG1) in the study of Ehlers and colleagues (20). Their concentrations varied between 0.1 and 10 µg/ml.

Binding to 2S albumins was observed in 75% (25/33) of all heterologously expressed mAbs. While mAbs with the ability to specifically bind to peanut 2S albumins originated from 11 IgM+, 9 IgG+ and 5 IgA+ B cells, all (n=8) mAbs without specific binding to peanut 2S albumins originated from IgM+ B cells. Based on their binding abilities, examined by comparing their achieved OD values at different concentrations, they were roughly categorized into weak (n=13), moderate (n=8) and strong (n=2) binders (**Figures 4A, B**). This variability in target binding indicates that our selection strategy was not restricted to only strongly binding B cells and implicates no selection bias regarding distinct affinities. Our staining protocol can also be adjusted to different antigen targets as the mAb generated from an ARHGDIB-binding B cell (IgA+) showed strong binding to ARHGDIB-coupled microspheres with an ARHGDIB/transferrin ratio greater than 2 (**Figure 4C**).

Method 2: "Establishment of Monoclonal EBV-LCLs"

LCLs Isolated With Antigen-Coupled Beads Are Mostly Antigen-Specific

Native Abs can be produced by *in-vitro* culturing of primary B cells secreting polyclonal Abs upon activation. Since *in-vitro* culturing of primary B cells is challenging (28), especially in the absence of a particular antigen, we chose for the establishment of LCLs by EBV immortalization. Immortalized LCLs were selected for their binding to either peanut 2S albumins or ARHGDIB. As shown in **Figure 5A**, almost all polyclonal EBV-LCLs (98.8%) obtained from the first round of cloning secreted antibodies specifically binding to 2S albumins (2S albumin/transferrin ratio ≥ 2) compared to the supernatant of irradiated PBMCs (Ctr). This large number of positive EBV-LCLs (100%) was confirmed by selecting for 2S albumin-binding EBV-LCLs using a second blood bank donor. Although the number of positive clones was reduced for ARHGDIB-binding B cells (Donor 1: 67%, Donor 2: 56%), a representative number of EBV-LCLs was selected for a second round of cloning to achieve monoclonality (**Figure 5B**).

Optimal Seeding Density to Achieve Monoclonal LCL Clones

LCLs positive for antigen-binding were used for an additional round of cloning to achieve LCLs secreting monoclonal instead of polyclonal Abs. EBV-LCLs from the first blood bank donor were used to determine the optimal seeding density. Even though a seeding density of 0.3 cells/well results theoretically in the highest probability of obtaining monoclonal LCLs, only a small number of wells contained LCLs showing proliferation.

[3] Note: We observed higher amplification efficiencies when primer aliquots were stored in 50 mM Tris-HCl buffer supplemented with 2 mM EDTA instead of RNAse free water.

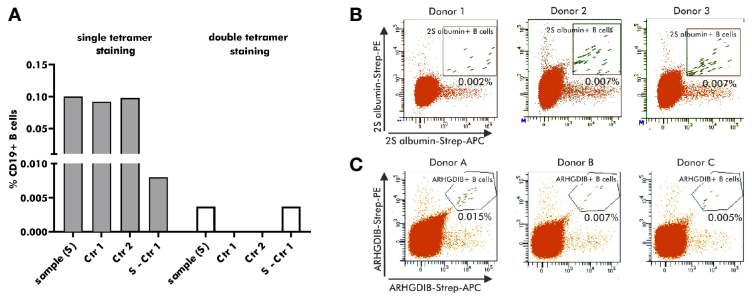


FIGURE 3 | Selection of antigen-binding B cells using antigen-tetramers. **(A)** Percentage of 2S albumin-binding B cells (blood bank donors) in relation to the total CD19+ B cell fraction using single or double-tetramer staining. Control 1 (Ctr 1) represents the staining with only biotin and Control 2 (Ctr 2) is executed by pre-blocking with unlabeled 2S albumin prior normal staining procedure. The final percentage is calculated by subtracting Ctr 2 from the sample. **(B, C)** FACS plots representing antigen-binding B cells in relation to the total B cell population of 6 independent blood bank donors. 2S albumin-binding B cells are shown in **(B)** and ARHGDIB-binding B cells in **(C)**.

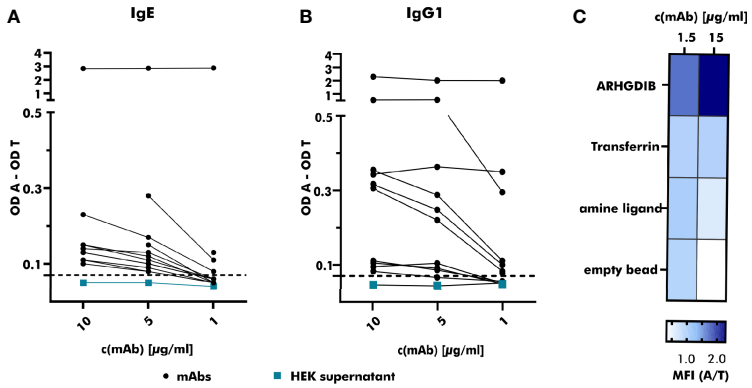


FIGURE 4 | Specific binding of produced monoclonal antibodies towards the respective antigen. **(A)** The binding of serial-diluted (1 to 10 µg/ml) human monoclonal antibodies (mAbs) (expressed as IgE; derived from peanut-allergic and peanut-tolerant patients) towards peanut 2S albumins was measured using anti-human IgE-horse radish peroxidase (HRP) as detection antibody. Measured optical density (OD) values were corrected by subtracting the OD value measured for the control antigen (transferrin) and compared to the OD value obtained with HEK supernatant not containing any antibodies (turquoise). The threshold was set to an OD value 1.5 times greater than the OD value obtained with HEK supernatant containing no mAb. OD A, optical density antigen; OD T, optical density transferrin (control). **(B)** The binding of serial-diluted (1 to 10 µg/ml) human mAbs (expressed as IgG1, derived from peanut-allergic and peanut-tolerant patients) towards peanut 2S albumins was measured using anti-human IgG-HRP as detection antibody. The measured OD value was corrected as described for **(A)**. **(C)** The specificity of the mAb derived from an ARHGDIB-binding B cell, derived from a blood bank donor, was examined using ARHGDIB-coupled microspheres. The median fluorescence intensity (MFI) towards ARHGDIB was evaluated in relation to the MFI towards transferrin as control antigen. A, antigen (ARHGDIB); T, transferrin.

By increasing the seeding density, the number of LCLs secreting Abs binding to 2S albumins rose in accordance with the number of LCLs seeded per well. Despite a high seeding density of 10 cells/well, 55% of sequenced EBV-LCLs (6/11) showed monoclonality, leading to a compromise between a high number of proliferating EBV-LCLs and a reasonable rate of achieved monoclonal EBV-LCLs (**Figure 6A**). Seeding 2S albumin-binding EBV-LCLs from the second blood bank donor at a density of 10 cells/well resulted in an approximate recovery of 2 EBV-LCLs with specific binding to 2S albumins per seeded plate (**Figure 6B**) when setting the threshold to a ratio (OD antigen/OD transferrin) of 2.5. This increased threshold was set to enlarge the probability to select truly specific LCLs combined with a certain amount of secreted antibodies. A comparable recovery was achieved for ARHGDIB-binding EBV-LCLs from another two independent blood bank donors (**Figure 6C**).

Heterologously Expressed mAbs Showed Comparable Binding to 2S Albumins

To validate heterologous expression of mAbs secreted by established EBV-LCLs, one pair of heavy and corresponding light chain variable regions was selected from the pool of 2S albumin-specific monoclonal EBV-LCLs. Upon heterologous expression with a

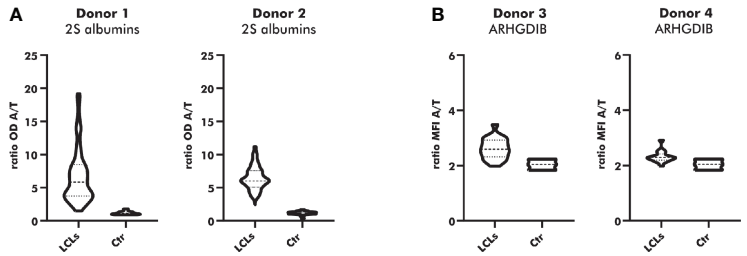


FIGURE 5 | Direct cloning of peanut 2S albumin and ARHGDIB-binding EBV-LCLs from blood bank donors. **(A)** Antibodies secreted by directly cloned EBV-LCLs were screened for their binding to peanut 2S albumins using a direct ELISA. EBV-LCLs secreting antibodies with an OD ratio (2S albumin/transferrin) ≥ 2 were selected for an additional round of cloning. The data are presented as violin blot with the median and the inter-quartile range. Control (Ctr): irradiated PBMCs without EBV-LCLs. LCLs, lymphoblastoid cell lines; OD, optical density; A, antigen; T, transferrin (mock antigen). **(B)** Antibodies secreted by directly cloned EBV-LCLs were screened for their binding to ARHGDIB using antigen-coupled microspheres. EBV-LCLs secreting antibodies with a MFI ratio (ARHGDIB/transferrin) ≥ 2 were selected for an additional round of cloning. The data are presented as violin blot with the median and the inter-quartile range. Control (Ctr): irradiated PBMCs without EBV-LCLs. LCLs, lymphoblastoid cell lines; MFI, median fluorescent intensity; A, antigen; T, transferrin (mock antigen).

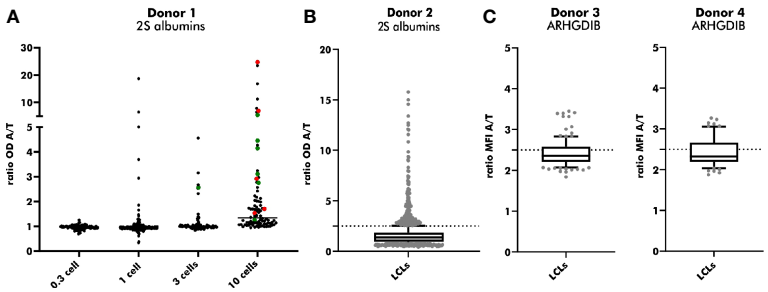


FIGURE 6 | Subcloning of 2S albumin and ARHGDIB-binding EBV-LCLs from blood bank donors. **(A)** Determination of the optimal seeding density for the second round of EBV-LCLs cloning. EBV-LCLs with specific binding to peanut 2S albumins (ratio 2S albumins/transferrin ≥ 2) were considered for expansion. The V(D)J gene transcript of these EBV-LCLs were Sanger sequenced to check for monoclonality. Green: monoclonal EBV-LCLs, red: polyclonal EBV-LCLs, black: not sequenced. LCLs, lymphoblastoid cell lines; OD, optical density; A, antigen; T, transferrin (mock antigen). **(B)** Antibodies secreted by subcloned EB V-LCLs were screened for their binding to peanut 2S albumins. EBV-LCLs with a ratio (2S albumin/transferrin) ≥ 2.5 were considered for expansion. The data are presented as box blot with the median and the inter-quartile range. The lowest and highest 10% are presented as grey, filled dots. LCLs, lymphoblastoid cell lines; OD, optical density; A, antigen; T, transferrin (mock antigen); dash line: set threshold to ratio ≥ 2.5. **(C)** Antibodies secreted by subcloned EBV-LCLs were screened for their binding to ARHGDIB. EBV-LCLs with a ratio (ARHGDIB/transferrin) ≥ 2.5 were considered for expansion. The lowest and highest 10% are presented as grey, filled dots. LCLs, lymphoblastoid cell lines; MFI, median fluorescent intensity; A, antigen; T, transferrin (mock antigen); dash line: set threshold to ratio ≥ 2.5.

vector containing the IgG1 backbone, native and heterologously expressed counterparts were applied on a direct ELISA in a serial dilution as shown in **Figure 7**. The measured OD values were comparable between the native IgM mAb and the heterologously expressed IgG1 mAb (light chain detection), indicating that mAbs from subcloned EBV-LCLs can also be heterologously expressed with comparable binding abilities.

Single Cell Sequencing Allows Higher Throughput

Both methods described within this chapter are characterized by several up- and downsides influencing the choice which method to use. A comparison regarding costs, expenditure of time, recovery, throughput and complexity is shown in **Table 3**. Without considering any further characterization, the generation of one single human mAb can add up to 750€. The establishment of monoclonal EBV-LCLs, however, is generally less expensive (400 - 500€) compared with the generation by single sequencing (650 - 750€), resulting from expensive reagents required for single cell amplification. On the other hand, the workflow of single cell sequencing shortens the required time from 12 weeks including long culturing periods to only 3 weeks. Even though both methods showed comparable recovery rates in our experiments, the establishment of monoclonal EBV-LCLs reaches easily a capacity threshold limiting the overall throughput and leading to random selection of clones to proceed with. Hence, the establishment of monoclonal EBV-LCLs is a suitable tool to generate a limited number of human mAbs, especially due to the ability to screen for binding, functionality and neutralization capacity throughout the workflow. Single cell sequencing is, however, a more suitable

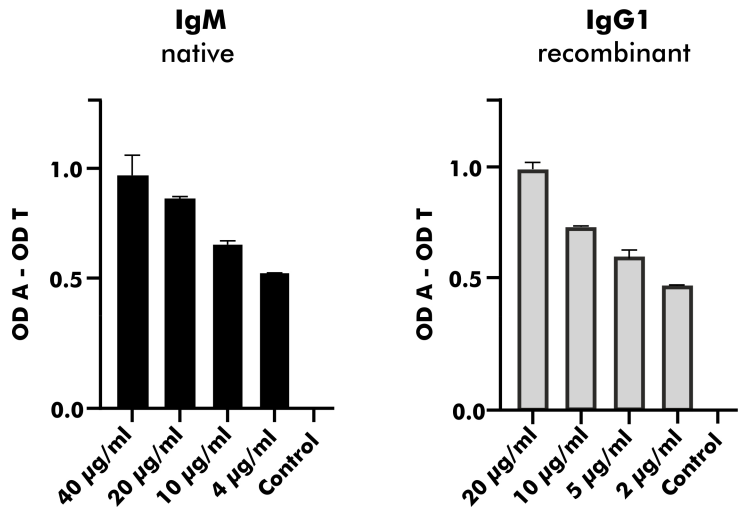


FIGURE 7 | Heterologous expression of a human mAb with specific binding to peanut 2S albumins derived from a blood bank donor. A mAb carrying the same variable region as native antibodies obtained from a monoclonal EBV-LCL was heterologously expressed with a vector containing the IgG1 backbone. Its binding ability expressed as OD values (corrected for binding to transferrin) was compared to the binding ability of its native counterpart (IgM) using anti-kappa-HRP as secondary antibody. The data are presented as the mean with the standard deviation of two replicates. OD, optical density; A, antigen; T, transferrin (mock antigen).

TABLE 3 | Comparison of Method 1 and 2 to generate human mAbs.

	Method 1: "Single cell"	Method 2: "EBV-LCLs"
*Costs	*~650 - 750€	*~450€
Time	3 weeks	12 weeks
Antigen adaptation	sorting strategy: approx. 3 weeks	No special adaptation needed (only antigen coupling to beads – 1 day)
Recovery	20.3 - 29%**	17.6%***
Throughput	+++	+
Complexity	+++	+++

*costs per one single human mAb.
**Cloning efficiency might be increased from 54% to 76% by purchase gene sequences commercially.
***The number of subcloned B cells is reduced by reaching a certain capacity threshold. This has been reflected with the limited throughput.
+: low, ++: moderate, +++: high.

tool for a broader examination of antigen-specific B cells, since a much higher throughput can be achieved. Unfortunately, specificity and functionality can only be examined in the end of the workflow. Both methods are laborious and highly complex.

DISCUSSION

Studying human-derived mAbs offers the possibility to link genetic information to functional features, making them indispensable in modern molecular biology research. In this study, we directly compared two distinct strategies to generate (highly) specific mAbs from peripheral blood of human donors, i.e.: single cell sequencing and the establishment of monoclonal EBV-LCLs. While single cell sequencing is a suitable tool to generate a large panel of mAbs with distinct binding features, the establishment of monoclonal EBV-LCLs provides the possibility to screen for specificity and functionality throughout the workflow. Both strategies, initially set-up for mAbs specifically directed against peanut 2S albumins, were easily adaptable to other antigen targets as shown for ARHGDIB.

Corresponding V(D)J gene transcripts were successfully amplified from up to 50% of all single cell sorted 2S albumin-binding B cells. The overall recovery of 2S albumin-specific human mAbs, however, was reduced to around 20% by taking the cloning efficiency (54%) and proportion of mAbs with proven specificity (75%) into consideration. The cloning efficiency may be increased by ordering gene sequences without cloning success commercially. Overall, our amplification efficiency (~50%) corresponds to the work of Tiller and colleagues who described an amplification efficiency of up to 60% (18). Moreover, a comparable overall recovery of 27% has been shown for single cell sequencing of antigen-specific B cells from guinea pigs (29). Increased amplification efficiencies of 90 to even 100% have been described for performing a comparable amplification protocol in triplicates (30). However, such an approach will simultaneously increase the probability of amplification errors, potentially hampering gene analysis and antibody cloning.

Regarding the establishment of antigen-specific monoclonal EBV-LCLs, the overall recovery was estimated to around 18%. This is in accordance with cloning efficiencies of around 15% observed for limiting dilution approaches upon EBV immortalization (16). However, we lack the information what proportion of antigen-specific B cells were initially immortalized

by EBV. To overcome this limitation, Fraussen and colleagues described in their protocol the immortalization of 50 cells/well upon sorting the desired B cell subpopulation (31). However, we were not able to immortalize such small numbers of B cells successfully in our laboratory.

Both strategies are characterized by their own strengths and limitations. Single cell sequencing, on the one hand, enables the execution in a moderate-throughput manner and the selection of antigen-specific B cells is not restricted to certain subpopulations as EBV immortalization is restricted to CpG-activated memory B cells (15). These advantages make the single cell sequencing platform a suitable tool for a broad examination of antigen-binding B cells and their corresponding mAbs. The establishment of monoclonal EBV-LCLs, on the other hand, provides continuous screening for specific binding and functionality throughout the workflow, resulting in an easy selection of mAbs with high affinity towards their targets and making this strategy a powerful tool in therapeutic research. Moreover, this strategy enables the comparison of heterologously expressed mAbs with their natural counterparts, allowing the identification of potential structure alterations by post-translational modifications during heterologous expression (32, 33). An additional advantage of generating human mAbs by establishing monoclonal EBV-LCLs is their limited need for expensive reagents.

In conclusion, both strategies - single cell sequencing and establishment of monoclonal EBV-LCLs - are able to generate (highly) antigen-specific human mAbs and they are easily adaptable to other target antigens. The recommended method to choose is dependent on the research question to explore as both strategies have their own strengths and limitations.

DATA AVAILABILITY STATEMENT

All data are presented in this article. Genetic information about the generated antibodies can be found under https://www.ncbi.nlm.nih.gov/genbank/, MW271045 - MW271525.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Medical Ethical Committee of the University Medical Center Utrecht. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

AE, CHJ, and HO: experimental design. AE, CHJ, TK-H, and MK: experimental performance. AE, CHJ, TK-H, and MK: data collection and analyses. AE: drafting the manuscript. AE, CHJ, TK-H, MK, AK, and HO: contribution to data interpretation. CHJ, TK-H, MK, AK, and HO: critical revision of the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

For financial support, we want to acknowledge EUROIMMUN AG, Lübeck, Germany.

REFERENCES

1. VanDyk L, Meek K. Assembly of IgH CDR3: Mechanism, Regulation, and Influence on Antibody Diversity. *Int Rev Immunol* (1992) 8(2-3):123–33. doi: 10.3109/08830189209055568
2. Gough NM, Bernard O. Sequences of the Joining Region Genes for Immunoglobulin Heavy Chains and Their Role in Generation of Antibody Diversity. *Proc Natl Acad Sci* (1981) 78(1):509–13. doi: 10.1073/pnas.78.1.509
3. Mishra A, Mariuzza R. Insights Into the Structural Basis of Antibody Affinity Maturation From Next-Generation Sequencing. *Front Immunol* (2018) 9:117. doi: 10.3389/fimmu.2018.00117
4. O'Connell A, Volpi S, Dobbs K, Fiorini C, Tsitsikov E, de Boer H, et al. Next Generation Sequencing Reveals Skewing of the T and B Cell Receptor Repertoires in Patients With wiskott-Aldrich Syndrome. *Front Immunol* (2014) 5:340. doi: 10.3389/fimmu.2014.00340
5. Hansen A, Dörner T, Lipsky P. Use of Immunoglobulin Variable-Region Genes by Normal Subjects and Patients With Systemic Lupus Erythematosus. *Int Arch Allergy Immunol* (2000) 123(1):36–45. doi: 10.1159/000024422
6. Cameron E, Spencer S, Lazarini J, Harp C, Ward E, Burgoon M, et al. Potential of a Unique Antibody Gene Signature to Predict Conversion to Clinically Definite Multiple Sclerosis. *J Neuroimmunol* (2009) 213(1-2):123–30. doi: 10.1016/j.jneuroim.2009.05.014
7. Pantazes R, Reifert J, Bozekowski J, Ibsen K, Murray J, Daugherty P. Identification of Disease-Specific Motifs in the Antibody Specificity Repertoire Via Next-Generation Sequencing. *Sci Rep* (2016) 6:30312. doi: 10.1038/srep30312
8. Titcombe P, Wigerblad G, Sippl N, Zhang N, Shmagel A, Sahlström P, et al. Pathogenic Citrulline-Multispecific B Cell Receptor Clades in Rheumatoid Arthritis. *Athritis Rheumatol* (2018) 70(12):1933–45. doi: 10.1002/art.40590
9. Wilson P, Andrews S. Tools to Therapeutically Harness the Human Antibody Response. *Nat Rev Immunol* (2012) 12(10):709–19. doi: 10.1038/nri3285
10. Abbott W, Damschroder M, Lowe D. Current Approaches to Fine Mapping of Antigen–Antibody Interactions. *Immunology* (2014) 142(4):526–35. doi: 10.1111/imm.12284
11. Rajan S, Kierny M, Mercer A, Wu J, Tovchigrechko A, Wu H, et al. Recombinant Human B Cell Repertoires Enable Screening for Rare, Specific, and Natively Paired Antibodies. *Commun Biol* (2018) 1:5. doi: 10.1038/s42003-017-0006-2
12. McCafferty J, Griffiths A, Winter G, Chiswell D. Phage Antibodies: Filamentous Phage Displaying Antibody Variable Domains. *Nature* (1990) 348:552–4. doi: 10.1038/348552a0
13. Rondot S, Koch J, Breitling F, Dübel S. A Helper Phage to Improve Single-Chain Antibody Presentation in Phage Display. *Nat Biotechnol* (2001) 19:75–8. doi: 10.1038/83567
14. Lonberg N. Human Monoclonal Antibodies From Transgenic Mice. *Ther Antibodies* (2008) 181:69–97. doi: 10.1007/978-3-540-73259-4_4
15. Traggiai E, Becker S, Subbarao K, Kolesnikova L, Uematsu Y, Gismondo M, et al. An Efficient Method to Make Human Monoclonal Antibodies From Memory B Cells: Potent Neutralization of SARS Coronavirus. *Nat Med* (2004) 10(8):871–5. doi: 10.1038/nm1080
16. Steinitz M. Production of Human Monoclonal Antibodies by Epstein-Barr Virus Method. *Methods Mol Biol* (2014) 1060:111–22. doi: 10.1007/978-1-62703-586-6_6
17. Kwakkenbos M, Diehl S, Yasuda E, Bakker A, van Geelen C, Lukens M, et al. Generation of Stable Monoclonal Antibody-Producing BCR+ Human Memory B Cells by Genetic Programming. *Nat Med* (2010) 16(1):123–8. doi: 10.1038/nm.2071

18. Tiller T, Meffre E, Yurasov S, Tsuiji M. Nussenzweig M and Wardemann H, Efficient Generation of Monoclonal Antibodies From Single Human B Cells by Single Cell RT-PCR and Expression Vector Cloning. *J Immunol Methods* (2008) 329(1-2):112–24. doi: 10.1016/j.jim.2007.09.017
19. Kamburova EG, Gruijters ML, Kardol-Hoefnagel T, Wisse BW, Joosten I, Allebes WA, et al. Antibodies Against ARHGDIB are Associated With Long-Term Kidney Graft Loss. *Am J Transplant* (2019) 19(12):3335–44. doi: 10.1111/ajt.15493
20. Ehlers AM, den Hartog Jager CF, Knulst AC, Otten HG. Distinction Between Peanut Allergy and Tolerance by Characterization of B Cell Receptor Repertoires. *Authorea Preprints* (2020). doi: 10.22541/au.160588804.43147170/v1
21. Kamburova EG, Kardol-Hoefnagel T, Wisse BW, Joosten I, Allebes WA, van der Meer A, et al. Development and Validation of a Multiplex non-Hla Antibody Assay for the Screening of Kidney Transplant Recipients. *Front Immunol* (2018) 9:3002. doi: 10.3389/fimmu.2018.03002
22. Patil S, Ogunniyi A, Calatroni A, Tadigotla V, Ruiter B, Ma A, et al. Peanut Oral Immunotherapy Transiently Expands Circulating Ara H 2-Specific B Cells With a Homologous Repertoire in Unrelated Individuals. *J Allergy Clin Immunol* (2015) 136(1):125–34. doi: 10.1016/j.jaci.2015.03.026
23. Franz B, May K. Dranoff G and Wucherpfennig K, Ex Vivo Characterization and Isolation of Rare Memory B Cells With Antigen Tetramers. *Blood* (2011) 118(2):348–57. doi: 10.1182/blood-2011-03-341917
24. Yang J, Reth M. Oligomeric Organization of the B-cell Antigen Receptor on Resting Cells. *Nature* (2010) 467(7314):465–9. doi: 10.1038/nature09357
25. Hoh R, Joshi S, Liu Y, Wang C, Roskin K, Lee J, et al. Single B Cell Deconvolution of Peanut-specific Antibody Responses in Allergic Patients. *J Allergy Clin Immunol* (2016) 137(1):157–67. doi: 10.1016/j.jaci.2015.05.029
26. Ye J, Ma N, Madden T, Ostell J. Igblast: An Immunoglobulin Variable Domain Sequence Analysis Tool. *Nuc Acid Res* (2013) 41:W34–40. doi: 10.1093/nar/gkt382
27. Koppelman SJ, De Jong GAH, Laaper-Ertmann M, Peeters KABM, Knulst AC, Hefle SL, et al. Purification and Immunoglobulin E-binding Properties of Peanut Allergen Ara H 6: Evidence for Cross-Reactivity With Ara H 2. *Clin Exp Allergy* (2005) 35(4):490–7. doi: 10.1111/j.1365-2222.2005.02204.x
28. Kumar A, Guido E, Liu R-S, Saedi M. Long-Term Culture of Primary B Cells and in Vitro Expression of an Exogenous Gene. *Immunol Lett* (1995) 47 (3):193–7. doi: 10.1016/0165-2478(95)00092-7
29. Lei L, Tran K, Wang Y, Steinhardt J, Xiao Y, Chiang C-I, et al. Antigen-Specific Single B Cell Sorting and Monoclonal Antibody Cloning in Guinea Pigs. *Front Microbiol* (2019) 10:672. doi: 10.3389/fmicb.2019.00672
30. Guselnikov S, Belovezhets T, Kulemzin S, Gorchakov A, Taranin A. A Simple Way to Increase Recovery of the Expressed VH and VL Genes in Single-Sorted Human B Cells. *Biotechniques* (2019) 67(4):184–7. doi: 10.2144/btn-2019-0079
31. Fraussen J, Vrolix K, Martinez-Martinez P, Losen M, Meulemans E, De Baets M, et al. A Novel Method for Making Human Monoclonal Antibodies. *J Autoimmun* (2010) 35(2):130–4. doi: 10.1016/j.jaut.2010.05.001
32. Jefferis R. Posttranslational Modifications and the Immunogenicity of Biotherapeutics. *J Immunol Res* (2016) 2016:5358272. doi: 10.1155/2016/5358272
33. Tokmakov A, Kurotani A, Takagi T, Toyama M, Shirouzu M, Fukami Y, et al. Multiple Post-translational Modifications Affect Heterologous Protein Synthesis. *J Biol Chem* (2012) 287(32):27106–16. doi: 10.1074/jbc.M112.366351

Conflict of Interest: The research position of AE was partially funded by EUROIMMUN AG, Lübeck, Germany.

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.

frontiers in Immunology

ORIGINAL RESEARCH
published: 24 May 2021
doi: 10.3389/fimmu.2021.669400

Integration of *Immunome* With Disease-Gene Network Reveals Common Cellular Mechanisms Between IMIDs and Drug Repurposing Strategies

Abhinandan Devaprasad[1,2], *Timothy R. D. J. Radstake*[1,2] *and Aridaman Pandit*[1,2*]

[1] Division Internal Medicine and Dermatology, University Medical Center Utrecht, Utrecht, Netherlands, [2] Center for Translational Immunology, University Medical Center Utrecht, Utrecht, Netherlands

OPEN ACCESS

Edited by:
Moncef Zouali, Institut National de la Santé et de la Recherche Médicale (INSERM), France

Reviewed by:
Mohan Maddur, Pfizer, United States
Giuseppe Murdaca, University of Genoa, Italy

***Correspondence:**
Aridaman Pandit
a.pandit@umcutrecht.nl

Specialty section:
This article was submitted to Molecular Innate Immunity, a section of the journal Frontiers in Immunology

Received: *18 February 2021*
Accepted: *04 May 2021*
Published: *24 May 2021*

Citation:
Devaprasad A, Radstake TRDJ and Pandit A (2021) Integration of Immunome With Disease-Gene Network Reveals Common Cellular Mechanisms Between IMIDs and Drug Repurposing Strategies. Front. Immunol. 12:669400. doi: 10.3389/fimmu.2021.669400

Objective: Development and progression of immune-mediated inflammatory diseases (IMIDs) involve intricate dysregulation of the disease-associated genes (DAGs) and their expressing immune cells. Identifying the crucial disease-associated cells (DACs) in IMIDs has been challenging due to the underlying complex molecular mechanism.

Methods: Using transcriptome profiles of 40 different immune cells, unsupervised machine learning, and disease-gene networks, we constructed the Disease-gene IMmune cell Expression (DIME) network and identified top DACs and DAGs of 12 phenotypically different IMIDs. We compared the DIME networks of IMIDs to identify common pathways between them. We used the common pathways and publicly available drug-gene network to identify promising drug repurposing targets.

Results: We found $CD4^+$Treg, $CD4^+$Th1, and NK cells as top DACs in inflammatory arthritis such as ankylosing spondylitis (AS), psoriatic arthritis, and rheumatoid arthritis (RA); neutrophils, granulocytes, and $BDCA1^+CD14^+$ cells in systemic lupus erythematosus and systemic scleroderma; ILC2, $CD4^+$Th1, $CD4^+$Treg, and NK cells in the inflammatory bowel diseases (IBDs). We identified lymphoid cells ($CD4^+$Th1, $CD4^+$Treg, and NK) and their associated pathways to be important in HLA-B27 type diseases (psoriasis, AS, and IBDs) and in primary-joint-inflammation-based inflammatory arthritis (AS and RA). Based on the common cellular mechanisms, we identified lifitegrast as a potential drug repurposing candidate for Crohn's disease and other IMIDs.

Conclusions: Existing methods are inadequate in capturing the intricate involvement of the crucial genes and cell types essential to IMIDs. Our approach identified the key DACs, DAGs, common mechanisms between IMIDs, and proposed potential drug repurposing targets using the DIME network. To extend our method to other diseases, we built the DIME tool (https://bitbucket.org/systemsimmunology/dime/) to help scientists uncover

the etiology of complex and rare diseases to further drug development by better-determining drug targets, thereby mitigating the risk of failure in late clinical development.

Keywords: IMID (immune-mediated inflammatory diseases), disease-associated cells, disease-associated genes, drug repurposing, machine learning, *immunome*

INTRODUCTION

Genetic and epigenetic heterogeneity plays a significant role in the development and progression of complex diseases. The past two decades have seen a major surge in studies that characterize genes and loci associated with diseases (1). The use of high-throughput omics technology and functional screenings have boosted our knowledge about genetic, epigenetic, and metabolic factors underlying complex diseases (1). As a result of these genetic and epigenetic screenings, we now know that most complex diseases and genes/loci have a many-to-many relationship. Meaning that complex diseases are linked to many different genes, and a gene/loci might be associated with several diseases (2). Thus, it is essential to identify and characterize these disease-associated genes (DAGs) to understand diseases better and develop therapy accordingly.

Extensive high-throughput screening studies and multi-omics have helped in the identification of DAGs. However, in most studies, DAGs were identified using bulk tissue or whole blood, a caveat since each gene's expression is known to vary between tissues and cell types (3, 4). Thus, bulk tissue- or blood-based studies on DAGs do not consider the role of different cells and tissues in disease biology. To improve the understanding and molecular basis of complex diseases, a large number of research groups and consortiums have started to functionally identify disease-associated cells (DACs) or tissue types (3–7). The Genotype-Tissue Expression (GTEx) is one such valuable project, which maps gene expression profiles of 54 different human tissue types and the corresponding expression quantitative trait loci (eQTLs) (5–7). Furthermore, the growth of single-cell technologies propelled our understanding of diseases and helped in identifying DACs for complex conditions, including cancer (8), Alzheimer's (9), rheumatoid arthritis (10), among others. Among these studies, the role of immune cells has been central to disease etiology and progression.

The immune system plays a vital role in developing and progressing immune-mediated and non-immune mediated chronic diseases. Many association and functional studies have shown that immune cells express multiple DAGs, and perturbing these DAGs can modulate immune cell functions (11). However, very few studies have explored the impact of DAGs on specific cell types and even fewer on immune cells, many of which focus on a limited number of cell subsets (12–16). Recently, Schmiedel et al. studied the effect of genetic variants on gene expression in 13 different immune cell types (17). However, this study primarily focused on analyzing genetic variants and their impact on a total of 13 immune cell types: monocytes (classical and non-classical), NK cells, naïve B-cells, and nine sub-populations of T-cells. The study identified several genetic variants to have a role in specific immune cell subsets in autoimmune disorders. For example, the modulatory effects of the variant rs12936231 in asthma and other autoimmune diseases are seen in lymphoid rather than myeloid subsets as previously described (17). Such new insights into specific immune cells' role led us to believe that specific immune cells and their DAGs remain poorly understood even in immune disorders.

The immune-mediated inflammatory diseases (IMIDs) are complex among the immune disorders, involving several immune cells. For example, in rheumatoid arthritis, the immune cells such as B-cells, T-cells, macrophages, mast cells, dendritic cells, and NK cells play a significant role (18). However, the exact mechanism of these cell types remains unknown. Insights on the precise mechanism of action are crucial for developing successful therapies, which becomes particularly challenging for IMIDs due to several cell types involved. The massive undertaking of GWAS has enabled the mapping of some of the molecular mechanisms of the IMIDs (19–22). However, further research is required to understand the etiology of IMIDs taking into account the several different immune cells at play and the contributing DAGs for each immune cell type. By identifying the critical immune cells and their mechanism, we would set a robust rationale for identifying any mechanistic overlap between diseases and exploiting them to develop therapeutic strategies.

This study mapped the largest available and expert-curated disease-gene network (from the DisGeNet curated from 16 different databases) (23) on the most extensive *immunome* data comprising gene expression profiles of 40 different immune cell types, curated by us. We then used an unsupervised machine learning algorithm, the disease-gene network, and the *immunome* to create the Disease-gene IMmune cell Expression (DIME) network. Using this approach, we built a tool called DIME. Using DIME, we quantified the effects of 3957 DAGs on the *immunome* to identify DACs for 12 phenotypically different IMIDs. We used the DIME to (1) study the underlying cell-specific mechanisms (2); identify common DACs and their top-weighted DAGs between disease pairs (referred to as the common cell-gene network); and (3) identify drug repurposing targets using the common cell-gene network. The DIME is available as a user-friendly R tool (https://bitbucket.org/systemsimmunology/dime), to determine the top genes and cells associated with the disease of interest for (1): diseases from the DisGeNet (2), diseases from the EBI genome-wide association study (GWAS) catalog, or (3) custom set of genes defined by the user.

METHODS

Transcriptome Data - *Immunome*

The transcriptome data consists of RNA-sequencing datasets of 40 different immune cell types curated using 316 samples from a total of 27 publicly available datasets (see **Supplementary Table S1** for list of GEO datasets and samples used). The 40 different immune cells cover the entire hematopoietic stem cell differentiation tree comprising nine progenitors, 19 lymphoid, and 12 myeloid cell types. The samples used here were manually curated considering only the unstimulated (except for monocyte-derived macrophages) immune cells that were sorted using Fluorescence-activated cell sorting (FACS) and were isolated from either blood, bone marrow, or cord blood from healthy donors.

All the selected datasets (**Supplementary Table S1**) were downloaded as FASTQ files using the fastq-dump tool from sratoolkit. The "split-files" option was given if the library type was paired-end sequencing. FASTQ files were then aligned to the reference genome (GRCH.Hg38.79) using the STAR aligner (24). The result is a SAM file, which was then converted into a sorted BAM file using the samtools program (25). These were then used to calculate the count of aligned reads using the HTSeq program (26) with the "intersection non-empty" option. HTSeq was run for all possible stranded mode options, the count file with the maximum counts was chosen as the respective count file for the sample.

The data was then filtered by removing all genes that had less than 20 read counts in 95 percent of the samples using R programming. The filtered data was then lane normalized using the "betweenLaneNormalization" function from the RUVSeq package (27). The RUVr method from RUVSeq was used to identify residual factors contributing to the batch effect. The resulting filtered, batch corrected, and normalized data had expression for 34,906 genes that were void of any observable batch effect. We calculated counts per million (CPM) for the filtered genes and used log2(CPM +1) as the gene expression measure. We then used the median gene expression for each cell type for the rest of the analysis. This processed, batch corrected, and normalized data of the 40 immune cells is referred to here as the *immunome*.

Disease-Gene Network From DisGeNet

The disease-gene network from DisGeNet (23) was downloaded from the DisGeNet database (www.disgenet.org/downloads). All HLA associated genes were removed from the network; this was done to ensure that bias towards myeloid cells and B cells are removed since the HLA genes are primarily expressed by these cells. The resulting network was further filtered to include only those genes that were present in the *immunome*.

IMID Disease-Gene Network

To study and identify the DACs of the IMIDs, we extracted the DAGs of 12 IMIDs extracted from the above DisGeNet. The IMID gene network for the 12 diseases comprised of 3579 DAGs. The 12 diseases that broadly represent the IMIDs in this study include: ankylosing spondylitis (CUI: C0038013), arthritis (CUI: C0003864), Crohn's disease (CUI: C0010346), diabetes mellitus - non-insulin-dependent (CUI: C0011860), systemic lupus erythematosus (CUI: C0024141), multiple sclerosis (CUI: C0026769), psoriasis (CUI: C0033860), psoriatic arthritis (CUI: C0003872), rheumatoid arthritis (CUI: C0003873), Sjogren's syndrome (CUI: C1527336), systemic scleroderma (CUI: C0036421), and ulcerative colitis (CUI: C0009324). CUI, used in DisGeNet, is the concept-unique-identifier for the disease term defined by the unified medical language system (28). The disease term arthritis (CUI: C0003864) comprises DAGs that pan over several arthropathies such as spondyloarthropathy, osteoarthritis, gout, allergic arthritis, etc., that fall under the broad arthritis MeSH term.

Identification of Top DAC and DAG Using Machine Learning

Briefly, we used an unsupervised machine learning algorithm called non-negative matrix factorization (NMF) to map the disease-gene network to the *immunome* and identify the top DACs and DAGs of the 12 IMIDs. The NMF algorithm clusters the input gene expression data into 'k' clusters, such that the DAGs of a cluster are expressed by the DACs of the same cluster, thus forming DAC-DAG pairs in each cluster (29). We used the coefficients and weights identified by the NMF algorithm as the DAC and DAG scores, respectively. The scores were scaled between 0 and 1, with 1 being the highest score. Those in the top 25 percentile of the scores were regarded as the top DACs and DAGs, respectively. We calculated the Frobenius norm for each cluster to weigh and rank the clusters; the rank 1 cluster is the top cluster having the highest Frobenius norm value. The top cluster comprises the DAC-DAG pair that maximally captures/represents the input gene expression matrix. Using the top DAC-DAG pairs of all clusters, we constructed the Disease-gene IMmune cell Expression (DIME) network for the 12 IMIDs. Detailed description of the DIME method is as follows.

Mapping Disease-Gene Network to *Immunome* Data

For a given disease D and its DAG, we first extracted the corresponding *immunome* expression matrix (X_D). X_D comprised the gene expression of the DAG across the 40 cell types. X_D was used as an input matrix for the NMF algorithm.

Using NMF to Cluster X_D Into k Classes

We used the NMF package (30) in R and applied Brunet's NMF algorithm (29) on X_D to factor it into two matrices, namely W_D and H_D such that.

$$X_D \approx W_D H_D \tag{1}$$

$$W_D H_D = \begin{bmatrix} | & | & | & | \\ w_D^1 & w_D^2 & \dots & w_D^k \\ | & | & | & | \end{bmatrix} \begin{bmatrix} - & h_D^1 & - \\ - & h_D^2 & - \\ - & \vdots & - \\ - & h_D^k & - \end{bmatrix} \tag{2}$$

$$W_D H_D = \Sigma_{i=1}^{k} \; w_D^i \; h_D^i; \; i \in \{1, \dots, k\} \tag{3}$$

Where W_D and H_D are the basis and coefficient matrices computed by NMF. Here, k is the number of classes/clusters that splits the data such that it satisfies the above NMF equations. The W_D matrix comprises the weights of the DAGs across the k clusters (in each column), and the H_D matrix comprises the weights of the cells in the corresponding k clusters (in each row). We used Brunet's method to identify the ideal k value using the cophenetic correlation coefficient (29).

Identifying the Top DAG and DAC From W_D and H_D

The NMF algorithm clusters the data into k clusters (as shown in Equations 2 and 3) such that, in each cluster 'i', where $i \in (1, \ldots, k)$, the genes that have high values in w_D^i are constitutively expressed by the cells that have high values in h_D^i. Where, w_D^i is the i^{th} column of W_D and h_D^i is the i^{th} row of H_D. We used the scaled (between 0 and 1) values of h_D^i and w_D^i as the DAC and DAG scores, respectively. For each cluster i, we chose the DACs and DAGs that were in the top 25$^{\text{th}}$ percentile range of their DAC and DAG scores, respectively. These filtered DACs and DAGs are regarded as the top DACs and DAGs, respectively. The top DACs and DAGs were extracted for all clusters of $i \in (1, \ldots, k)$. The DIME network was constructed using the top DAC-DAG pairs from all clusters.

Identifying the Top Cluster

We then identified the largest weighted cluster (referred to as the top cluster) among the k clusters identified by the NMF. That is, the subset of DACs and DAGs of X_D that can capture most of its expression pattern. We did this by calculating the Frobenius norm of each $w_D^i h_D^i$ for all values of $i \in (1, \ldots, k)$ from Equation 3. We then identified the top cluster for which $||w_D^i h_D^i||_F$ is the maximum. This can be represented as:

$$top\ \ cluster\ \ = \text{argmax}\left(||w_D^i h_D^i||_F\right);\ \ i \in \{1, \ldots k\} \qquad (4)$$

Where the top cluster represents that which maximally captures/represents the expression matrix X_D. Thus, the top cluster is the rank 1 cluster of DIME. Subsequent ranks are the next highest weighted clusters.

Evaluating the Consistency of Top DACs and DAGs Identified by DIME

To check the consistency of the results from DIME, we performed 1000 jackknife simulations for each of the 12 IMIDs. For each simulation of each disease, we ran DIME with 70% random subsampling of either the DACs or the DAGs. And in each simulation, we identified the top cluster and the top DACs when DAGs were subsampled and vice versa. We compared the consistency of the top DACs (**Supplementary Figure S2**) and the top 10 DAGs (**Supplementary Figure S3**) identified by the original DIME run (100% of the sample) against the 1000 simulations. We computed the Pearson correlation coefficient between the DAC/DAG score of the top cluster of the original run to the number of times the DAC/DAG was found as the top DAC/DAG in the top cluster of the 1000 simulations. We used the p-value from the Pearson correlation test to state the significance of correlation and thus the statistical significance of the top DAC/DAG of the top cluster.

The Common Cell-Gene Network Between Diseases

To identify the common cell-gene network between two diseases, we looked at their overlapping DAC-DAG pairs in their corresponding DIME networks. We refer to the overlapping DAC-DAG pairs as the common cell-gene network between the two diseases. We then used the Jaccard index (JI) to measure the overlap between the two diseases and Fisher's exact test (FET) to obtain a confidence p-value for the given overlap.

Integrating Drug-Gene Network

We extracted the drug-gene target network from (1) DGIdb with the filter set to contain CHEMBL interactions of the drugs approved by the food and drug administration (FDA) of USA (31) (2); all drug-gene of CLUE database (32) and (3); all drug-gene of hPDI (33). The genes with drugs associated with them are labeled in the common cell-gene networks to highlight drugability (**Figures 6C–E**).

Statistical Analysis

We performed 1000 jackknife simulations to assess the consistency of the results from the DIME (Supplementary Methods and **Supplementary Figures S2–S4**). We used the Pearson correlation coefficient to measure the significance of the jackknife simulations compared to the original run (**Supplementary Figure S4**).

RESULTS

The Disease-Gene Network of the 12 IMIDs Reveal Several Common DAGs

In this study, we analyzed the DAGs of 12 different types of IMIDs that broadly include inflammatory arthropathies, spondyloarthropathies, rheumatic diseases, systemic IMIDs, and inflammatory bowel diseases (IBDs). And specifically, the 12 different IMIDs studied here are ankylosing spondylitis (298 DAGs), arthritis (567 DAGs), Crohn's disease (786 DAGs), diabetes mellitus - non-insulin-dependent (1415 DAGs), systemic lupus erythematosus (963 DAGs), multiple sclerosis (961 DAGs), psoriasis (689 DAGs), psoriatic arthritis (177 DAGs), rheumatoid arthritis (1612 DAGs), Sjogren's syndrome (229 DAGs), systemic scleroderma (494 DAGs), and ulcerative colitis (796 DAGs) (**Figures 1A, B**). The 12 IMIDs had a total of 3957 DAGs. Among these, several genes were linked to several IMIDs; for example, 74 DAGs were linked to only Crohn's disease (CD) and ulcerative colitis (UC), both IBDs. Calculating the Jaccard index and Fisher's exact test (FET) on all the overlapping DAGs between all IMIDs revealed that CD and UC had the highest significant overlap (**Figure 1C**). Interestingly, genes associated with CD had significant overlap (FET p-value ≤ 0.05) with all diseases except psoriatic arthritis and non-insulin-dependent diabetes mellitus. Rheumatoid arthritis (RA) had a significant overlap of DAGs with all IMIDs except non-insulin-dependent diabetes mellitus. However, non-insulin-dependent

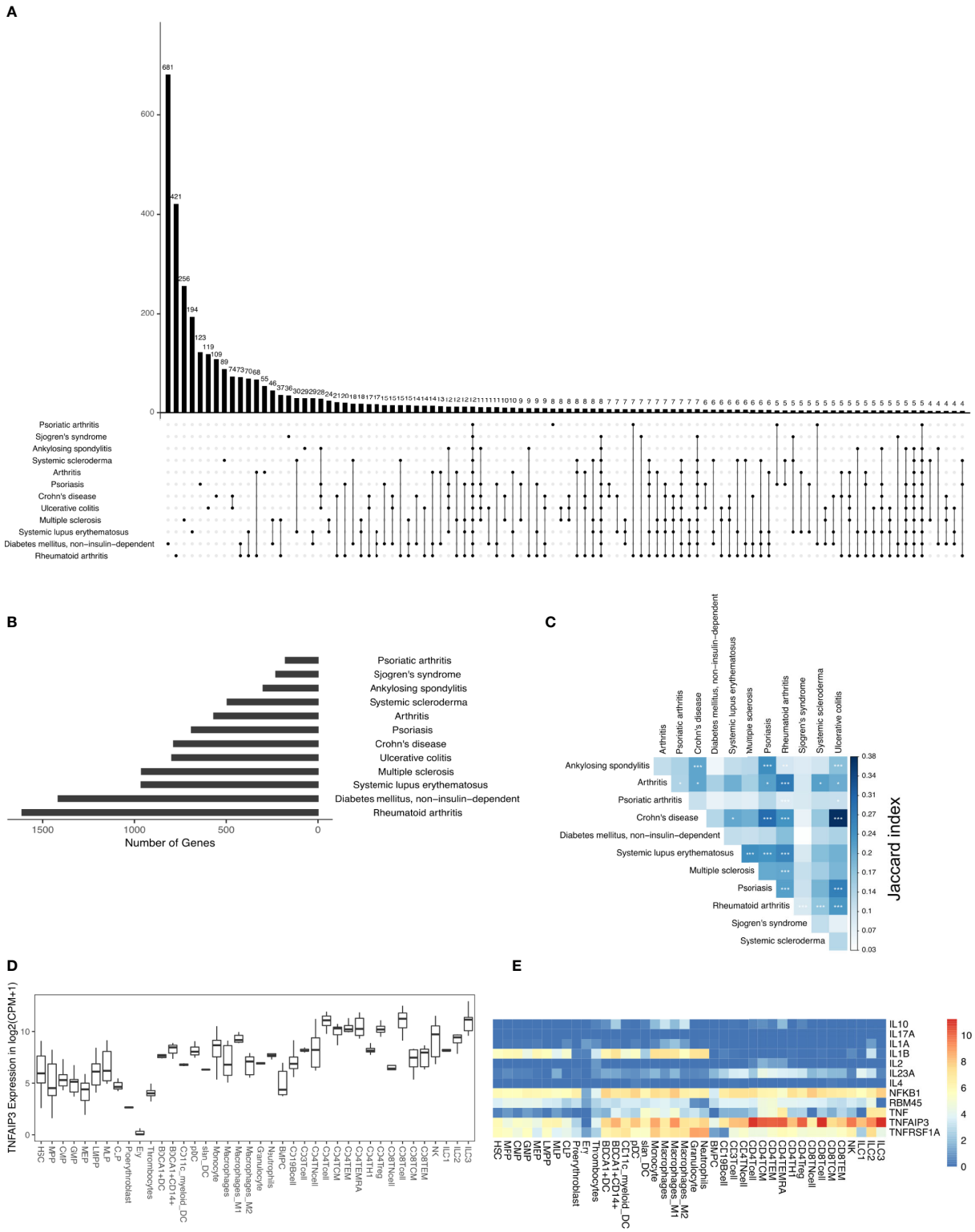


FIGURE 1 | DAGs of IMIDs: **(A)** intersection of DAGs for all comparisons of IMIDs. Comparisons are shown only for those diseases that have at least one intersecting DAG between them. **(B)** Barplot represents the number of DAGs in each IMID. **(C)** Heatmap depicting Jaccard index and Fisher exact test (FET) p-value calculated for each IMID comparison. Fisher exact test (FET) p-value denoted by * (***≤ 0.001, **≤ 0.01 and *≤ 0.05). **(D)** Gene expression of TNFAIP3. **(E)** Heatmap depicting gene expression of the 12 genes common to all 12 IMIDs. Gene expression values in log2(CPM+1). CPM denotes counts per million.

diabetes mellitus did not have a significant overlap of DAGs with any of the IMIDs. Arthritis, psoriasis, CD, and RA had a significant overlap of DAGs between each other. We found 12 DAGs that were associated with all the 12 IMIDs (**Figures 1A, E**). These DAGs were related to processes typically associated with inflammation, such as cytokine signaling (GO:0001817; GO:0019221), regulation of inflammatory response (GO:0050727), and regulation of interleukin-6 (GO:0032675; GO:0032635). We further explored these DAGs in the *immunome* and found the expression of TNFAIP3 to be the highest in $CD8^+$ T-cells, ILC3 and $CD4^+$ T-cells (**Figures 1D, E**). Likewise, IL1B was expressed by myeloid and progenitor cells; TNF was expressed by lymphoid and myeloid cells. Overall, specific myeloid and lymphoid cells specifically expressed some of the 12 genes linked to all the 12 IMIDs. Such cell-specific expression of the DAGs led us to question the immune cell types and genes essential for the 12 IMIDs. Hence, we used the DIME on the 12 IMIDs to identify their top DACs and DAGs. Briefly, DIME uses the *immunome*, input disease-gene network, and an unsupervised machine learning algorithm (NMF) to determine the clusters of top DACs and DAGs.

Inflammatory Arthritis Is Driven by $CD4^+$ Treg, $CD4^+$ Th1, and NK Cells

Inflammatory arthritis is an autoimmune disorder characterized by joint inflammation. And joint inflammation is the primary clinical feature observed in inflammatory arthritis types such as ankylosing spondylitis (AS) and RA. However, in other inflammatory arthritis types, such as psoriatic arthritis, inflammation is present in both the skin and joints. Interestingly, AS and psoriatic arthritis are both seronegative spondyloarthropathies (negative for rheumatoid factor and auto nuclear antibodies) characterized by enthesitis and predominant HLA-B27 genotype (34, 35). Such shared clinical features led us to question if the inflammatory arthritis types shared molecular mechanisms. So, we performed DIME on the different types of inflammatory arthritis to identify the critical DACs and DAGs and compare the molecular mechanism shared between them. As a reference, we used the broader arthritis disease term that encompassed several different kinds of arthropathies (including inflammatory arthritis).

In the DIME analysis, the clusters of the DIME network (**Figures 2–5**) are ordered based on the Frobenius norm. Cluster with the highest Frobenius norm represents the most crucial cluster and is designated as the top cluster comprising the most crucial DACs and DAGs. In each cluster, the DACs and DAGs are ordered based on the DIME score, high scoring nodes signify higher importance. The DIME analysis of AS revealed lymphoid cells such as NK cells, ILC3, $CD4^+$ T-cells (Th1, Treg, TEMRA) as the top DACs in the top cluster (**Figure 2A**). The top cluster's top DAGs comprised ETS1, HSPA5, TNFAIP3, IL2RG, WNK1, etc. that were associated with pathways such as interleukin signaling, antigen presentation, regulation of RUNX3, and BCR signaling (**Figure 2C**). We find RUNX3 expression highest in the NK cells, followed by $CD8^+$ T-cells and Th1 cells (**Supplementary Figure S1A**). The exact role of RUNX3 in AS is unclear and possibly involves regulation, differentiation, and activation of Th1 and T-bet cells (36). Further research is required to establish the exact role of RUNX3 in AS and the above-identified lymphoid cell subsets. In the second cluster, the top DACs included myeloid cells, and the top DAGs were associated with pathways such as interleukin (IL-4, IL-10, IL-13) signaling, MAPK3 activation, and MyD88 (**Figures 2A, C**). Thus, the key DACs of AS are diverse, as reported in the literature (37). However, according to DIME, the top DACs were NK cells, ILC3, $CD4^+$ T-cells (Th1, Treg, TEMRA).

The DIME analysis of psoriatic arthritis revealed lymphoid cells such as NK cells, ILC3, T-cells ($CD8^+$, TEMRA, Th1), ILC2 and myeloid cells like the macrophages and $BDCA1^+$ DC as the top DACs in the top cluster (**Figure 2B**). Likewise, T-cells, NK cells, and antigen-presenting cells have played a crucial role in psoriatic arthritis pathology (38). The top cluster's top DAGs were associated with interleukin (IL-4, IL-10, IL-13) signaling, PI3K, and NF-KB activation. (**Figure 2D**). Furthermore, the downstream genes of TNF-alpha such as TNFAIP3, TRAF5, NFKB1, and ICAM1 were top DAGs in the top cluster. Interestingly, TNF-alpha is a therapeutic target for psoriatic arthritis (39, 40), perhaps the downstream genes identified by DIME could also be explored as a therapeutic target for psoriatic arthritis. S100 calcium-binding proteins like S100A8 and S100A9 play a role in regulating inflammation in psoriatic arthritis (41). In the second cluster, we found the top DAGs included the S100 calcium-binding proteins, such as S100A9 and S100A8, which were highly expressed by neutrophils, granulocytes, monocytes, and dendritic cells (**Figure 2B** and **Supplementary Figure S1B**).

The crucial immune cells involved in RA are T-cells, B-cells, and APCs (42). While activation of $CD4^+$ Th1 and impairment of $CD4^+$ Tregs are essential for the pathology of rheumatoid arthritis (43), the DIME analysis of RA revealed several lymphoid cells such as $CD4^+$ Tregs, $CD4^+$ Th1, NK cells, etc., as the top DACs in the top cluster (**Figure 3A**). The top cluster's top DAGs were associated with pathways such as interleukin, TCR, FCERI, and BCR signaling (**Figure 3C**). In the second cluster, the top DACs included myeloid cells, and the top DAGs were associated with pathways such as interleukin (IL-10, IL-13) signaling, neutrophil degranulation, and ECM organization (**Figures 3A, C**). Activation, recruitment, and apoptosis of neutrophils are altered in RA, and under chronic inflammatory conditions, they release protease-rich granules (44).

The DIME analysis of the broader arthritis disease term revealed macrophages as the top DAC in the top cluster (**Figure 3B**). Macrophages play a central role in arthropathies, where they release cytokines and activate several immune cells such as T-cells, monocytes, neutrophils, and synovial fibroblasts. Besides, they are also the most abundant cells at the site of inflammation (45). The top DAGs of the top cluster were associated with pathways such as interleukin (IL-4, IL-13) signaling, extracellular matrix (ECM) related pathways, neutrophil degranulation, and toll-like receptor (TLR) cascades (**Figure 3D**). In the second cluster, the top DACs comprise neutrophils, granulocytes, and the top DAGs associated with pathways similar to the top cluster and included inflammasomes-related pathways (**Figures 3B, D**).

In conclusion, using DIME, we found that in addition to the shared clinical features, the three inflammatory arthritis types

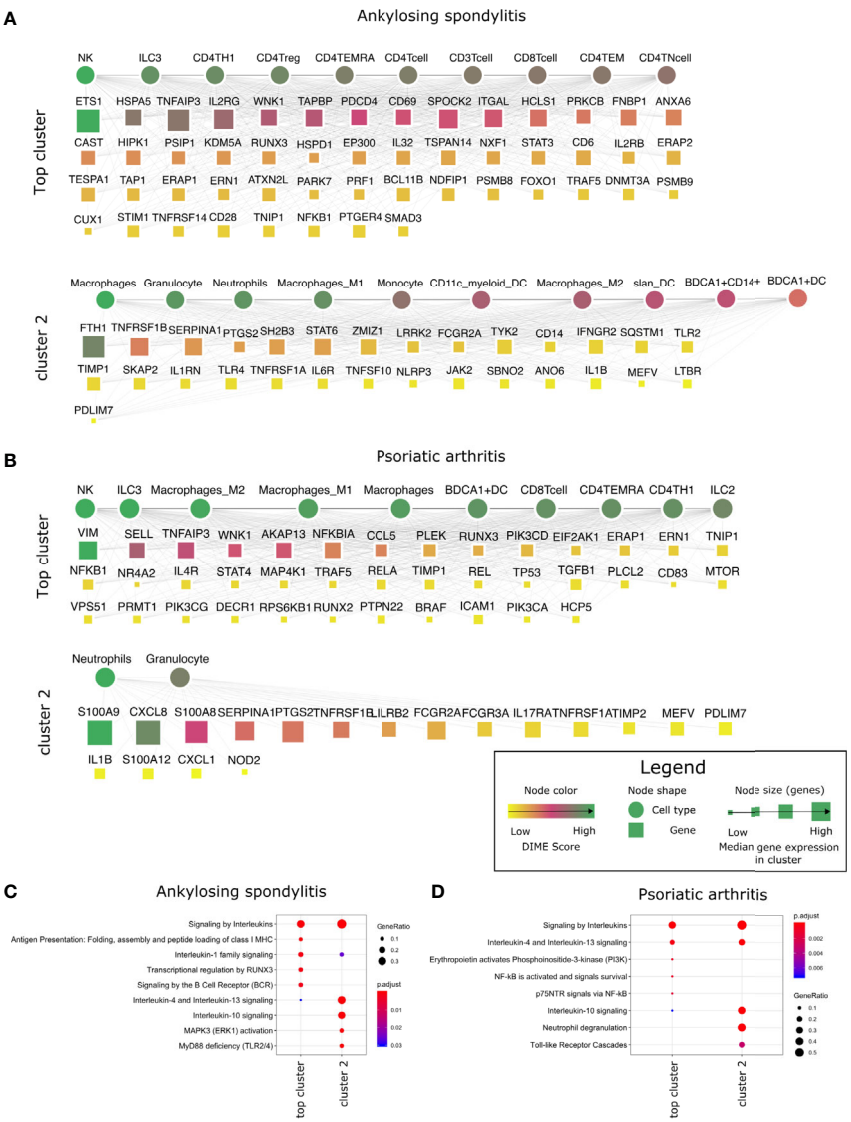


FIGURE 2 | Top DACs and DAGs of inflammatory arthritis: DIME network of **(A)** ankylosing spondylitis and **(B)** psoriatic arthritis. The DIME network shows top 25 percentile DACs and DAGs. Square nodes represent genes and circular nodes represent cell types, the color scheme signify the DIME score (higher score signify importance of the node in the cluster) and the node size represent the median gene expression of the gene in the given cluster. Edges in each cluster signify the relationship between the cell types and genes as identified by DIME, where the cell types in the cluster express the genes of the same cluster. To aid visualization, the DAGs in the network is pruned based on the DIME score (top 50 DAGs if present) and gene expression (> 5 median gene expression in the corresponding cluster's cell types). Pathway enrichment analysis of the top 25 percentile DAGs of **(C)** ankylosing spondylitis and **(D)** psoriatic arthritis.

also had a similar DAC profile comprising $CD4^+$ Treg, $CD4^+$ Th1, and NK cells as the top DACs, distinguishing them from the broader arthritis disease that showed macrophages as its top DAC. Perhaps these lymphoid cells contribute to inflammation in these arthropathies, and targeting them to reduce inflammation could be explored as a therapeutic strategy (46).

Myeloid Cells Are Essential to Systemic IMIDs

We performed the DIME analysis on the systemic IMIDs such as systemic lupus erythematosus (SLE) and systemic scleroderma (SSc) (**Figure 4**). SLE and SSc are type I interferon-mediated systemic autoimmune diseases, that unlike RA, primarily affect not just the joints but also the skin, kidney, heart, and other organs (47). In SLE, the continuous IFN production by pDC and neutrophils leads to activation of monocytes, T-cells, and B-cells (48). The DIME analysis of SLE revealed the myeloid cells (granulocytes, macrophages, $BDCA1^+$ $CD14^+$, monocytes) as the top DACs in the top cluster (**Figure 4A**). The top DAGs in the top cluster were associated with interleukin signaling pathways (IL-4, IL-13), neutrophil degranulation, cell-surface interactions at the vascular wall, and the TLR cascades

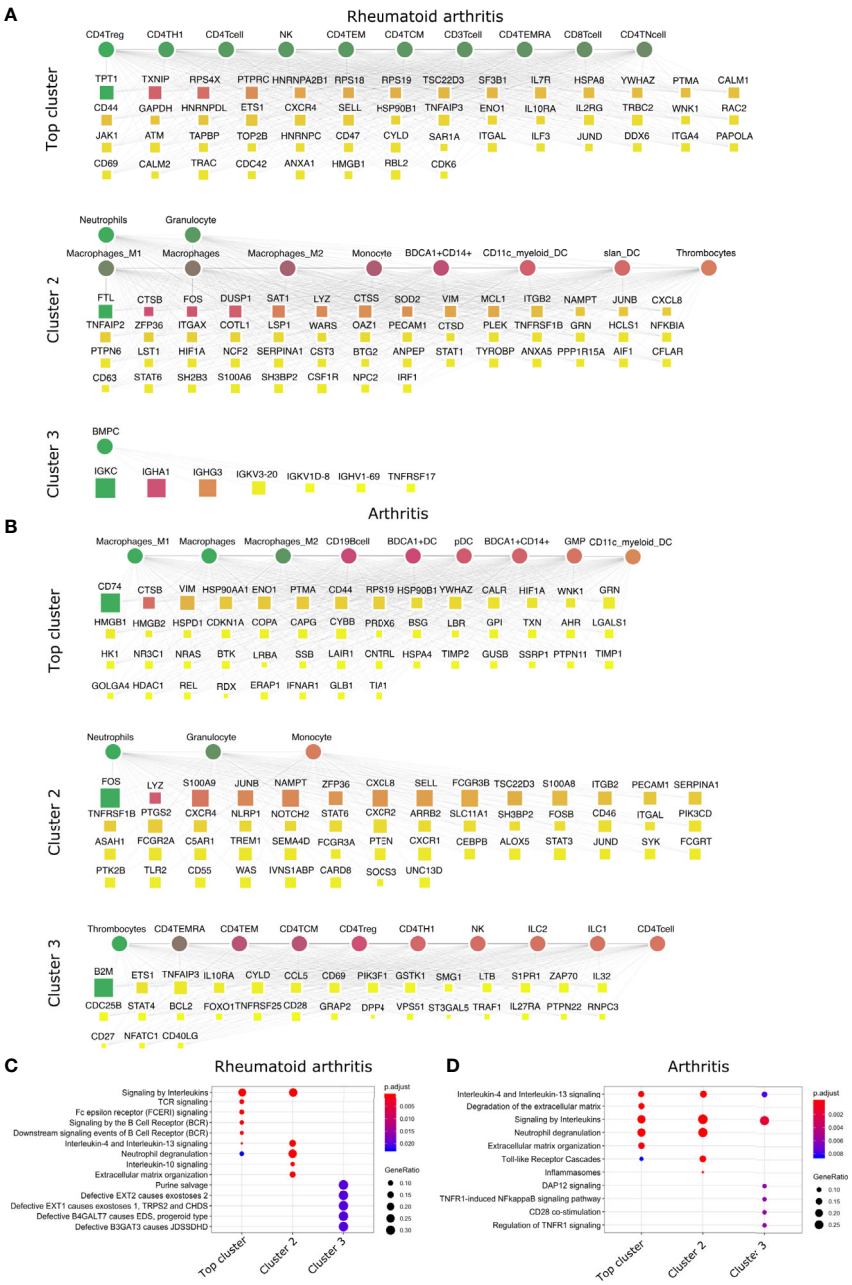


FIGURE 3 | Top DACs and DAGs of inflammatory arthritis: DIME network of **(A)** rheumatoid arthritis and **(B)** arthritis. Pathway enrichment analysis of the top 25 percentile DAGs of **(C)** rheumatoid arthritis and **(D)** arthritis. See **Figure 2** legend for network description.

(**Figure 4C**). Incidentally, the neutrophils in SLE undergo spontaneous NETosis (a form of suicidal cell death), and this process is dependent on TLR signaling (48). Additionally, T-cells in SLE have altered cytokine production with higher IL6, IL7, and IL10 secretions (48). In the second cluster, we found the top DACs included $CD4^+$ T-cells (TEMRA, TEM, TCM), and the top DACs were associated with pathways such as immunoregulatory interactions, Nef-associated factors (TNIP1, TNFAIP3), ZAP-70, VAV1 pathway (**Figures 4A, C**). Nef-associated factors (TNIP1, TNFAIP3) have played a role in T-cell activation *via* TCR signaling in SLE (49).

The DIME analysis of SSc revealed myeloid cells (neutrophils, granulocytes, $BDCA1^+$ $CD14^+$ cells) and lymphoid cells (NK cells, $CD4^+$ Treg, ILC2, and ILC3) as the top DACs in the top cluster (**Figure 4B**). The top DAGs in the top cluster were associated with interleukin signaling pathways (IL-4, IL-13), TGF beta signaling, NLR signaling, etc. (**Figure 4D**). In the second cluster, the top DACs included macrophages and the top

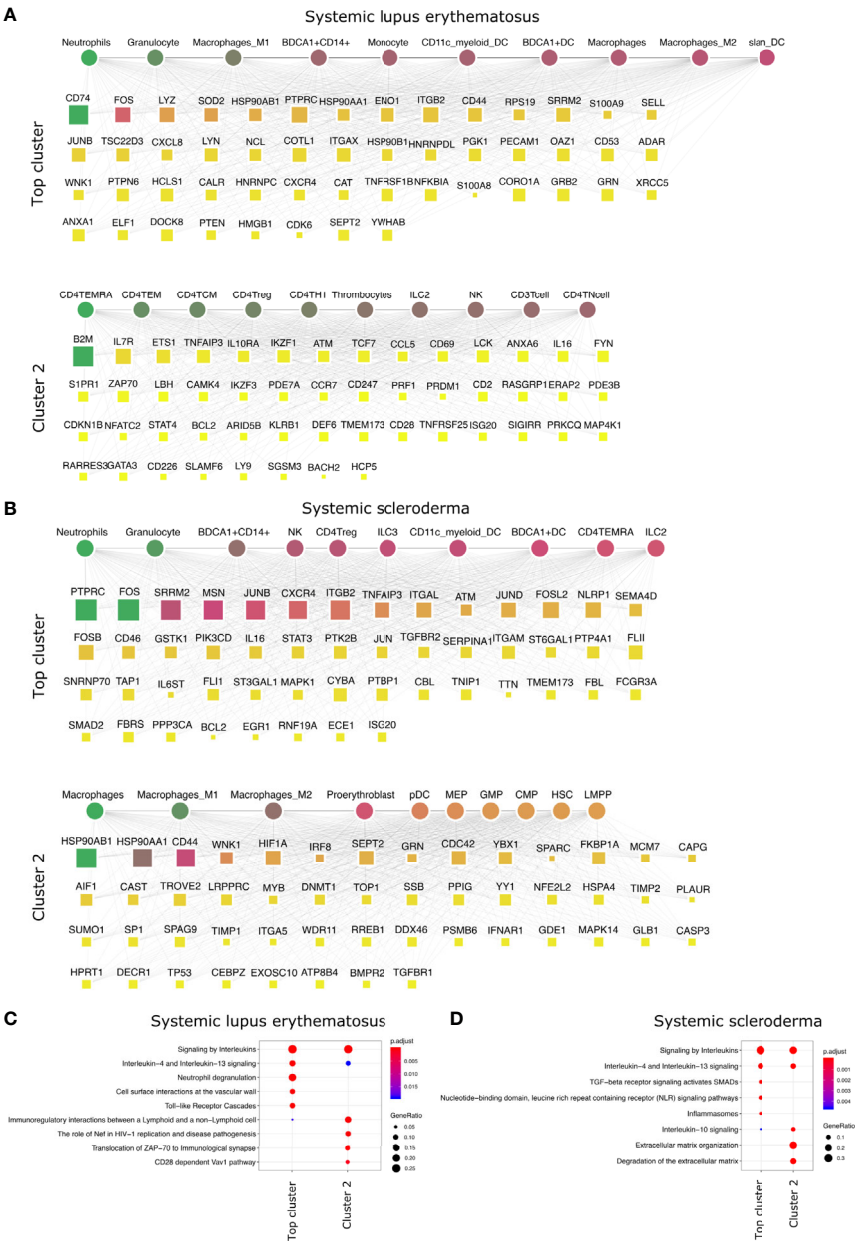


FIGURE 4 | Top DACs and DAGs of systemic diseases: DIME network of **(A)** SLE and **(B)** arthritis. Pathway enrichment analysis of the top 25 percentile DAGs of **(C)** SLE and **(D)** SSc. See **Figure 2** legend for network description.

DAGs were associated with pathways that included IL-10 signaling and degradation of ECM (**Figures 4B, D**). As described in the review by Caam et al., several studies have shown neutrophils, macrophages, NK cells, and Tregs to play a role in the profibrotic events in SSc by the production of profibrotic cytokines such as TGF beta, IL-4, IL-10, IL-13, etc., thus corroborating our findings (50).

Thus, the top DACs of the systemic IMIDs comprised myeloid cells such as neutrophils, granulocytes, BDCA1$^+$ CD14$^+$ cells, CD11c$^+$ myeloid cells, and BDCA1$^+$ DC. Exploring the role of these cells, their corresponding DAGs and pathways in the systemic IMIDs may be useful for gaining mechanistic insights into disease and for successful therapeutic strategy. Exploring the role of neutrophils and their degranulation genes may serve as stronger targets in SLE, as neutrophil degranulation precedes type 1 interferon signature (often observed in SLE patients) (51–54). While the DACs of the SSc are diverse populations of myeloid and lymphoid cells, neutrophils are still an interesting candidate and exploring the role of TGF beta may shed light on the profibrotic events in SSc.

Thus, based on the DIME analysis, attenuation of neutrophilic inflammation may be the key therapeutic strategy for the systemic IMIDs.

IBDs Are Primarily Lymphoid Driven

We then looked at IMIDs that involve chronic inflammation of the digestive system, categorized as IBDs. The two primary forms of IBDs are CD and UC. CD is known to be driven by $CD4^+$ Th1 cells, with a dominant Th1 cytokine profile leading to a pro-inflammatory effect (55). The DIME analysis of CD revealed lymphoid cells ($CD4^+$ Treg, ILC2, $CD4^+$ TEMRA, $CD4^+$ Th1) as the top DACs in the top cluster (**Figure 5A**). The top DAGs of the top cluster were associated with pathways such as interleukin (IL-4, IL-10, IL-13) signaling, TLR (TLR-5, TLR-10) signaling, MyD88, and neutrophil degranulation (**Figure 5C**). In the second cluster, the top DACs included granulocytes, neutrophils, monocytes, macrophages, etc., and the top DAGs were associated with pathways such as

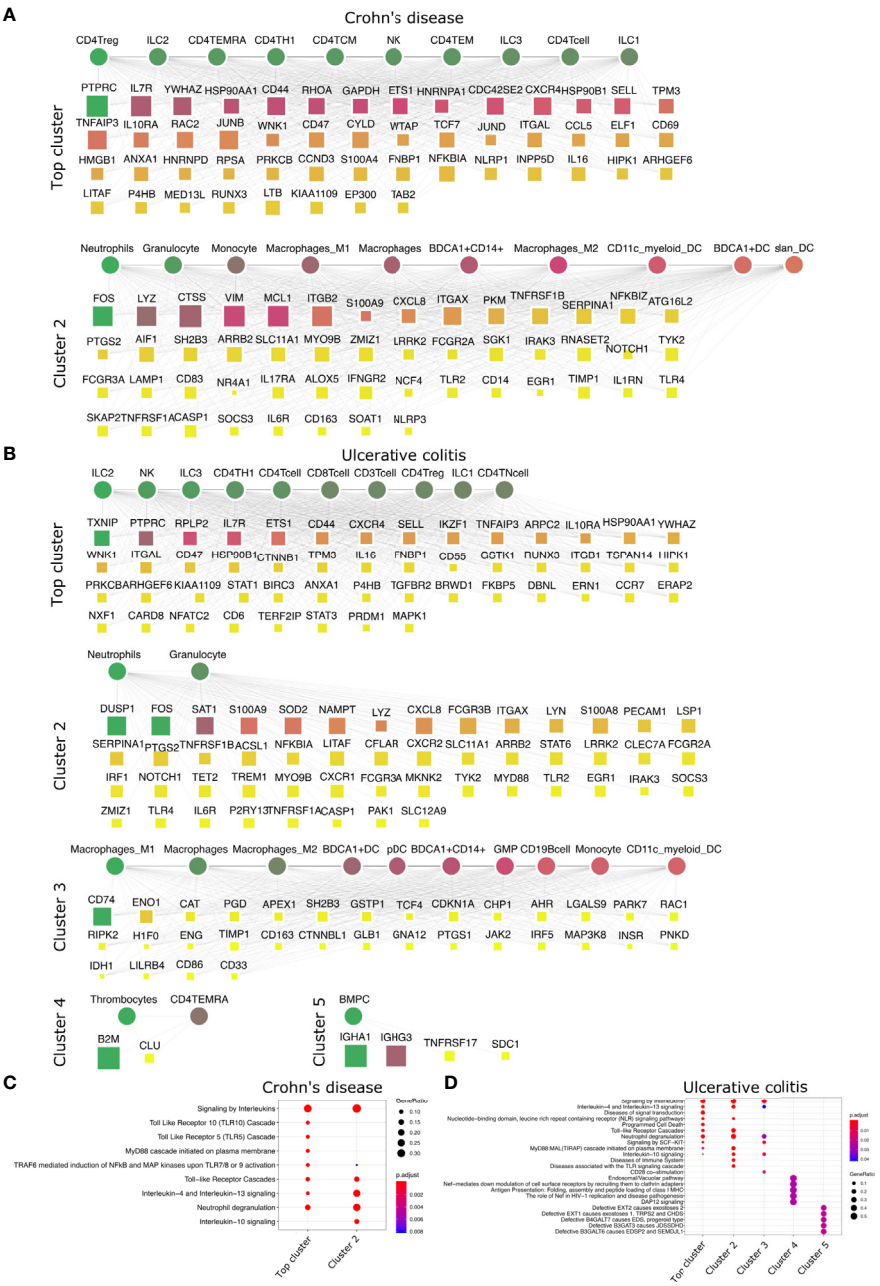


FIGURE 5 | Top DACs and DAGs of IBDs: DIME network of **(A)** Crohn's disease and **(B)** ulcerative colitis. Pathway enrichment analysis of the top 25 percentile DAGs of **(C)** Crohn's disease and **(D)** ulcerative colitis. See **Figure 2** legend for network description.

interleukin signaling, neutrophil degranulation, and TLR cascades (**Figures 5A, C**).

The T-cell profile of UC has been difficult to categorize due to discrepancies in its response among patients. However, there is evidence of Th2 cells, NK cells, macrophages, and neutrophils involved in the pathogenesis of UC (55). The DIME analysis of UC revealed lymphoid cells (ILC2, NK, ILC3, $CD4^+$ Th1, etc.) as the DACs in the top cluster (**Figure 5B**). The top DAGs of the top cluster were associated with pathways such as interleukin (IL-4, IL-13) signaling, TLR cascades, NLR signaling, neutrophil degranulation, etc. (**Figure 5D**). In the second cluster, the top DACs included granulocytes, $BDCA1^+$ $CD14^+$ cells, etc. The top DAGs were associated with interleukin signaling pathways (IL-4, IL-10, IL-13), neutrophil degranulation, and TLR cascades (**Figures 5B, D**). The DIME analysis of UC comprised 5 clusters in total, with myeloid cells in the third cluster, thrombocytes and TEMRA in the fourth cluster, and bone marrow plasma cells in the fifth cluster. This shows that the disease-gene network of UC involves genes that participate in many different cell types and are more complex to uncover their etiology. However, the highest weighted cluster signifies that the lymphoid cells may be the primary candidates for further analysis.

Thus, the top DACs of the IBDs include the ILC2, $CD4^+$ Th1, $CD4^+$ Treg, and the NK cells. The T-cells, innate lymphoid cells, and NK cells play an important role in the pathogenesis of IBDs (56–58), thus corroborating our findings. We hypothesize that Crohn's disease is driven primarily by lymphoid inflammatory response and downregulating TLR signaling pathways could be a potential therapeutic strategy. The DAC profile of ulcerative colitis was found to be diverse, and would require a deeper analysis into the participation of the different cell types involved in their etiology.

Statistically Significance of DIME Results

To evaluate DIME's consistency, we performed 1000 Jackknife simulations with random subsampling of DAC/DAG. We re-identified the top DAC/DAG for all IMIDs (see **Supplementary Methods**). The jackknife simulations revealed that the top DACs identified across all clusters in the simulations (**Supplementary Figure S2A**) showed a similar pattern compared to top DACs identified in the original run (**Supplementary Figure S2C**). For the top DACs of the top cluster, the pattern from the simulations (**Supplementary Figure S2B**) was comparable to the original run's DAC score (**Supplementary Figure S2D**). We used Pearson correlation to compare the pattern observed between the simulations and the original run. The Pearson correlation between the pattern observed in the simulated run (**Supplementary Figure S2B**) and the DAC scores of the original run for the top cluster revealed that the top DACs in the top cluster were significantly correlated (p-value ≤ 0.05) for all the IMIDs except ulcerative colitis (**Supplementary Figure S4**). This correlation shows that the top DACs of the top cluster identified by DIME is statistically significant for all IMIDs, except UC.

Likewise, we evaluated the consistency of the top DAGs. In all simulations, the top 10 DAGs of the original run's top cluster were present as the top DAG in any of the simulated run clusters (**Supplementary Figure S3**). The Pearson correlation between the pattern observed in the simulated run and the DAG scores of the original run for the top cluster was significantly correlated for all the IMIDs (**Supplementary Figure S4B**). Thus, the top DAGs of the top cluster identified by DIME is statistically significant for all IMIDs.

Why Are the top DACs of UC Insignificant?

In the case of UC, the top DACs were statistically insignificant from our 1000 jackknife simulations; the top DAGs, however, were significant (**Supplementary Figures S2–S4**). We found from 1000 simulations that the lymphoid cells identified by the original run (**Figure 5B**) were indeed present in the simulations. The myeloid cells were also part of the top DACs of the top cluster in the simulations (**Supplementary Figure S2B**). Furthermore, we found that the top cluster's top DAGs included genes associated with neutrophil degranulation pathways and other myeloid cell-related pathways (**Figures 5B, D**). Thus, owing to the non-convergence of NMF in accurately predicting the top DACs of the top cluster in UC. The top DACs of the top cluster of UC were ambiguous, as reported in the literature (55). From our simulations, we propose the inclusion of the myeloid cells in the top DACs of the top cluster in addition to the lymphoid cells previously identified (**Figure 5B**).

Common Cell-Gene Networks Reveal Common Mechanisms Between IMIDs and Potential Drug Targets

The DIME analysis revealed that several top DAGs and their corresponding DACs were present in many IMIDs. For example, in many IMIDs, the gene FOS was present as top DAG in the cluster typically containing myeloid cells (granulocytes, neutrophils, and dendritic cells) as the top DACs. We found several genes, like FOS, that were present as the top DAG in the same top DAC cluster between different pairs of diseases. We refer to these top DACs and DAGs present between the two diseases as the common cell-gene network (represented schematically in **Figure 6A**). Using the common cell-gene network, we suggest that these diseases may have a similar mechanism of action. We could exploit such common mechanisms to gain mechanistic insights between diseases and identify drug repurposing targets. Hence, we integrated the publicly available drug-gene networks to identify and reinforce drug repurposing targets based on the common mechanisms (cell-gene networks) determined from the DIME analysis (**Figure 6A**).

To identify the common mechanisms across the 12 IMIDs, we identified the common cell-gene networks between all disease comparisons (**Figure 6B**). We then used the Jaccard index and FET to measure the extent and significance of the overlap in the common cell-gene networks between the disease pairs. Compared to the analysis that looked at all DAGs, which showed several diseases to be statistically significant in the overlap, the common cell-gene network overlap was restricted to fewer diseases (**Figure 1C** and **Figure 6B**).

The comparative analysis revealed that CD had statistically significant common cell-gene networks with several diseases such as, psoriasis, RA, and UC (**Figure 6B**). CD and UC's common cell-gene network had the highest Jaccard index among all the IMIDs, both being IBDs with an aggressive T-cell response (55). CD and UC's common cell-gene network revealed that the top DACs included the lymphoid cells such as $CD4^+$ T-cell, $CD4^+$ Th1, $CD4^+$ Treg, ILC1, ILC2, ILC3, and NK cells in one cluster (**Figure 6C**). $CD4^+$ Th1 and NK cells are known to be implicated in both CD and UC (55). The top DAGs (represented by green border) such as CD44, CXCR4, SELL, HSP90AA1 etc., were highly expressed by cells of the lymphoid cluster and were also drug targets (genes that are druggable or have drug targeting them in our drug-gene network). Thus, making them potential drug repurposing candidates for CD and UC. Other potentially interesting DAGs (that were not drug targets) included PTPRC, ETS1, IL7R, TNFAIP3, etc. Collectively, the DAGs in the lymphoid cluster were enriched in interleukin signaling pathways (IL-4 and IL-13), NLR signaling, etc. (**Supplementary Figure S5A**). The other clusters consisted of myeloid cells such as the granulocytes, dendritic cells, monocytes, and macrophages, among which dendritic cells have been crucial for regulating the T-cell responses in IBDs. The top DAGs, such as SH2B3,

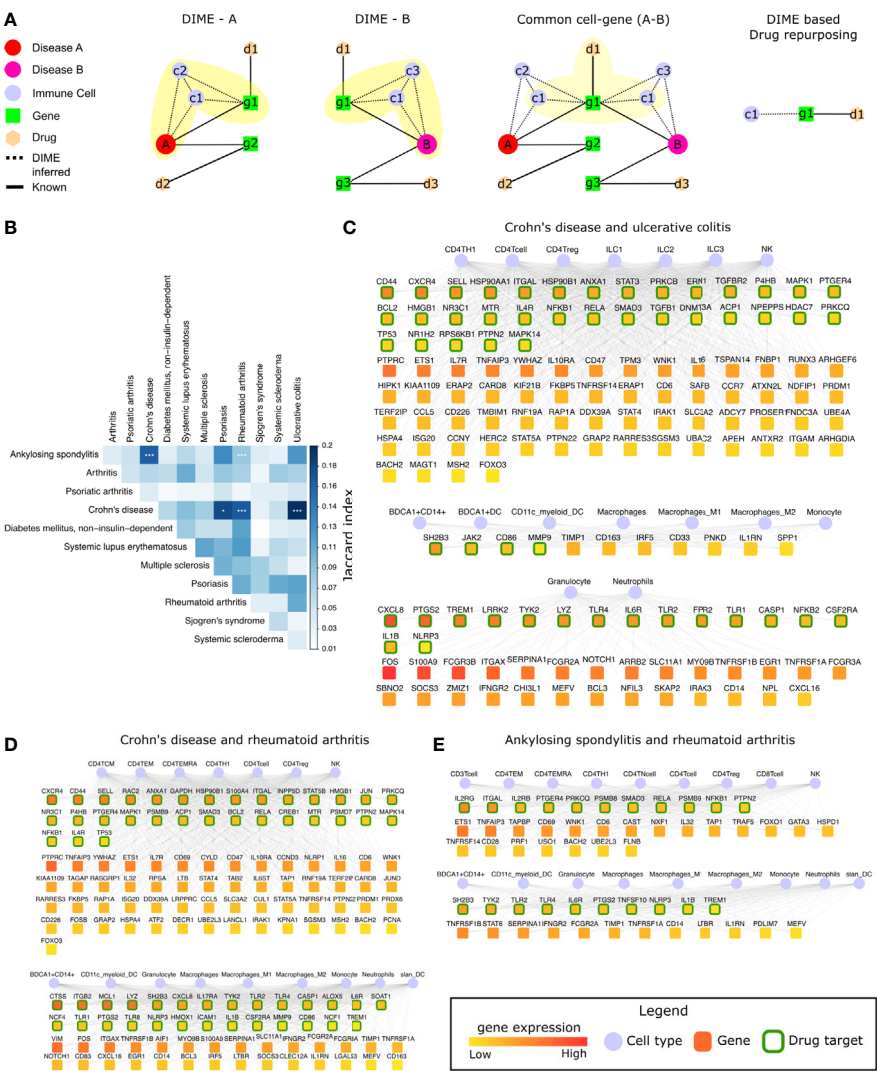


FIGURE 6 | Common mechanisms between IMIDs. **(A)** Steps involved in DIME based drug repurposing using the common cell-gene network. **(B)** Jaccard index and FET calculated for the common cell-gene between two diseases for all disease comparisons. Fisher exact test (FET) p-value denoted by *(***≤ 0.001 and *≤ 0.05). The common cell-gene network of **(C)** Crohn's disease and ulcerative colitis, **(D)** Crohn's disease and rheumatoid arthritis, and **(E)** ankylosing spondylitis and rheumatoid arthritis. The DAG's color is based on the median gene expression of the DAG in the corresponding DACs. DAGs that are drug targets have a green border, and the cells are shown in blue color.

JAK2, CD86, MMP9, CXCL8, PTGS2, TREM1, LRRK2, TYK2, etc., were highly expressed by the myeloid cluster cells and were potential drug repurposing candidates. These DAGs were enriched in interleukin signaling pathways (IL-10), TLR signaling, ECM degradation, etc. (**Supplementary Figure S5A**).

We next explored the common cell-gene network of the two distinct IMIDs that belonged to different pathophysiology, namely CD and RA. The common cell-gene network of CD and RA revealed that the top DACs comprised of the lymphoid cells that included all $CD4^+$ T-cells and NK cells in one cluster (**Figure 6D**). The top DAGs such as CXCR4, CD44, SELL, RAC2, ANXA1, etc., were highly expressed by this cluster's cells and were potential drug repurposing candidates. These DAGs were enriched for pathways associated with interleukin, TLR, MyD88 signaling, etc. (**Supplementary Figure S5B**). The other cluster comprised myeloid cells such as granulocytes, dendritic cells, monocytes, and macrophages. The top DAGs such as CTSS, ITGB2, MCL1, LYZ, SH2B3, etc., were highly expressed by this cluster's cells and were potential drug repurposing candidates. These DAGs were enriched for pathways associated with interleukin (IL-4, IL-13) signaling, neutrophil degranulation, etc. (**Supplementary Figure S5B**).

In addition to CD's common cell-gene networks, we found a statistically significant common cell-gene network between the two inflammatory arthropathies that have joint pain as the primary feature, namely AS and RA. The common cell-gene network of AS and RA revealed that the top DACs comprised of the lymphoid cells that included all the T-cells and NK cells in one cluster (**Figure 6E**). The top DAGs such as IL2RG, ITGAL, IL2Rb, PTGER4, PRKCQ, etc., were highly expressed by this cluster's cells and were potential drug repurposing candidates. These DAGs were enriched for interleukin (IL-1) signaling pathways, FCERI mediated NF–kB activation, TCR signaling, etc. The other clusters comprised myeloid cells such as granulocytes, dendritic cells, monocytes, and macrophages. The top DAGs, such as SH2B3, TYK2, TLR2, TLR4, IL6R, were highly expressed by this cluster's cells and were potential drug repurposing candidates. These DAGs were enriched for interleukin signaling pathways (IL-4, IL-10, IL-13) (**Supplementary Figure S5C**).

Thus, using the common cell-gene networks, we could uncover the common mechanisms between the IMID pairs (**Figure 6** and **Supplementary Figure S5**) and use them to identify potential drug targets. This novel method of computational drug repurposing is a combination of target-based and mechanism-based drug repurposing strategies (59). We found several DAGs such as IL1B, IL6R, ITGAL, PTGS2, TYK2, NFKB1, NLRP3, PRKCQ, PTGER4, PTPN2, RELA, SH2B3, SMAD3, TLR2, TLR4, and TREM1, that were drug targets and present in all the common cell-gene networks (**Figures 6C, E**). Among these DAGs, ITGAL was the only DAG that was a drug target and present as the top DAG of the top cluster (lymphoid cell cluster) in the DIME networks of CD, UC, AS, and RA. This is interesting as the lymphoid cells were identified as the top DAC of the top cluster for all of the above diseases. Using the drugs associated with these drug targets specifically for these diseases (CD, UC, AS, and RA) in therapy would require extensive experimental validation and clinical trials. Therefore, we explored (in the next section) the possibility of using some of these drug targets for repurposing based on existing studies and drugs that are already approved by the FDA. Thus, reinforcing and strengthening these targets and also the validity of our approach in identifying them.

Common Cell-Gene Networks Reveal Drug Repurposing Targets

To explore and validate the drug targets for repurposing, we focused on the top DAGs of the statistically significant (FET p-value ≤ 0.05) common cell-gene networks of all IMIDs (**Figure 6B**). To identify drug targets that were targets of FDA-approved drugs, we used the drug-gene network of CHEMBL. We found several drug targets such as IL1B, IL6R, ITGAL, and TYK2 to be present in all the statistically significant common cell-gene networks (**Table 1**). Here, we explore the possibility of using these drug targets for repurposing based on their current use as a therapeutic target in certain IMIDs and suggest where they can be repurposed based on their presence in the common cell-gene networks of those IMIDs and others.

Drug targets identified by us that are already in use for the different IMIDs include drugs that target IL1B, IL6, TYK2, and JAK2.IL1B was present in the common cell-gene network of AS, CD, psoriasis, RA, and UC, indicating a similar mechanism between these diseases (**Table 1**). Anti-IL1 therapy is used for psoriasis and RA (60–62), suggesting that it could be potentially used in IMIDs like AS, CD, and UC. Incidentally, preliminary studies indicate that anti-IL1 treatment has shown promising clinical response for treating AS, CD, and UC (63, 64). Anti-IL6 therapy (tocilizumab) showed a positive clinical response in a small group of patients in AS, CD, and RA, suggesting its application in other IMIDs like psoriasis and UC (65–67). However, anti-IL6 therapy had side effects in smaller studies on psoriasis and UC and must be explored carefully (68, 69). Tofacitinib, a TYK2 and JAK2 inhibitor developed for RA, is now making way to treatment options in other diseases such as, CD, UC, and psoriasis (70–73). Plerixafor (drug target: CXCR4) is a drug now used in cancer (lymphoma and multiple myeloma) after stem cell transplantation to initiate migration of stem cells in the bloodstream (74). This drug is now in clinical trials (NCT01413100) to be evaluated for use after autologous transplant in patients with SSc. We suggest extending such trials based on exploiting the CXCR4 mediated dysregulation of the immune system to other IMIDs like psoriasis, CD, RA, and UC.

Integrin based therapies (such as natalizumab and vedolizumab that targets ITGB2) are already used for CD (75). Exploring other integrin based therapies (such as Lifitegrast that targets ITGAL and ITGB2) for CD may be beneficial since ITGAL and ITGB2 are top DAGs in the DIME network and are also implicated in CD (76, 77). Lifitegrast could be a promising drug repurposing candidate for CD and perhaps for

TABLE 1 | The top DAGs and their FDA approved drug candidates identified from the DIME-based common cell-gene networks of the different IMIDs.

DAG/ Drug target	Diseases	Drugs
IL1B	AS, CD, Psoriasis, RA, UC	canakinumab, rilonacept, anakinra
IL6R	AS, CD, Psoriasis, RA, UC	tocilizumab
ITGAL	AS, CD, Psoriasis, RA, UC	lifitegrast
TYK2	AS, CD, Psoriasis, RA, UC	tofacitinib citrate
PSMB9	AS, CD, Psoriasis, RA	bortezomib, carfilzomib, ixazomib citrate
DNMT3A	AS, CD, Psoriasis, UC	azacitidine, decitabine
HDAC7	AS, CD, Psoriasis, UC	belinostat, panobinostat lactate, romidepsin
JAK2	AS, CD, Psoriasis, UC	baricitinib, ruxolitinib phosphate, tofacitinib citrate
PTGS2	AS, CD, RA, UC	acetaminophen, aminosalicylate potassium, aminosalicylate sodium, aspirin, balsalazide disodium, bismuth subsalicylate, bromfenac sodium, carprofen, diclofenac, diclofenac epolamine, diclofenac potassium, diclofenac sodium, diflunisal, etodolac, etoricoxib, fenoprofen calcium, flurbiprofen, flurbiprofen sodium, ibuprofen, ibuprofen lysine, ibuprofen sodium, indomethacin, indomethacin sodium, ketoprofen, ketorolac tromethamine, meclofenamate sodium, meloxicam, mesalamine, nabumetone, naproxen, naproxen etemesil, naproxen sodium, nepafenac, olsalazine sodium, oxaprozin, oxaprozin potassium, piroxicam, sulfasalazine, sulindac, tolmetin sodium
BCL2	CD, Psoriasis, RA, UC	venetoclax
CXCR4	CD, Psoriasis, RA, UC	plerixafor
IL4R	CD, Psoriasis, RA, UC	dupilumab
IL17RA	CD, Psoriasis, RA	brodalumab
ITGB2	CD, Psoriasis, RA	lifitegrast
PSMD7	CD, Psoriasis, RA	bortezomib, carfilzomib, ixazomib citrate
CD86	CD, RA, UC	abatacept, belatacept
CSF2RA	CD, RA, UC	sargramostim
NR3C1	CD, RA, UC	alclometasone dipropionate, amcinonide, beclomethasone dipropionate, betamethasone, betamethasone acetate, betamethasone benzoate, betamethasone dipropionate, betamethasone sodium phosphate, betamethasone valerate, budesonide, ciclesonide, clobetasol propionate, clocortolone pivalate, cortisone acetate, deflazacort, desonide, desoximetasone, dexamethasone, dexamethasone acetate, dexamethasone sodium phosphate, diflorasone diacetate, difluprednate, flumethasone pivalate, flunisolide, fluocinonide, fluorometholone, fluorometholone acetate, fluprednisolone, flurandrenolide, fluticasone furoate, fluticasone propionate, halcinonide, hydrocortamate hydrochloride, hydrocortisone, hydrocortisone acetate, hydrocortisone butyrate, hydrocortisone cypionate, hydrocortisone probutate, hydrocortisone sodium phosphate, hydrocortisone sodium succinate, hydrocortisone valerate, loteprednol etabonate, medrysone, meprednisone, methylprednisolone, methylprednisolone acetate, methylprednisolone sodium succinate, mifepristone, mometasone furoate, paramethasone acetate, prednicarbate, prednisolone, prednisolone acetate, prednisolone sodium phosphate, prednisolone tebutate, prednisone, rimexolone, triamcinolone, triamcinolone acetonide, triamcinolone diacetate, triamcinolone hexacetonide

(Continued)

TABLE 1 | Continued

DAG/ Drug target	Diseases	Drugs
P4HB	CD, RA, UC	lomitapide mesylate
IL2RB, IL2RG	AS, RA	basiliximab, daclizumab
PSMB8	AS, RA	bortezomib, carfilzomib, ixazomib citrate
ALOX5	CD, RA	balsalazide disodium, meclofenamate sodium, mesalamine, olsalazine sodium, sulfasalazine, zileuton

UC, AS, and RA. Since its target gene ITGAL, was the only top DAG of the top cluster (lymphoid cell cluster) that was also a drug target in the DIME networks of these diseases (**Figures 2**, **3**, **6C–E**). Thus, we propose lifitegrast as a novel drug repurposing candidate to be tested for CD, UC, AS, and RA.

DISCUSSION

Despite decades of experimental data, the knowledge of important cell types involved in the disease's pathogenesis remains limited. To address this gap, we used the *immunome* comprising 40 immune cells, the disease-gene network, and computational methods to identify the important DACs and DAGs of the disease. The integration of these parts resulted in the novel mechanisms being captured by our method, using which we built a tool called the DIME. Here, we highlight the important DACs, DAGs, and common mechanisms captured using DIME for 12 phenotypically different IMIDs. Using DIME, the top DACs were found to be $CD4^+$ Treg, $CD4^+$ Th1, and NK cells in inflammatory arthritis (AS, PsA, and RA); neutrophils, granulocytes, and $BDCA1^+$ $CD14^+$ cells in SLE and SSc; ILC2, NK, $CD4^+$ Th1, and $CD4^+$ Treg in the IBDs.

Lymphoid cells such as $CD4^+$ Th1, $CD4^+$ Treg, and NK cells were the key players in inflammatory arthritis (AS, PsA, and RA) and IBD (CD and UC). These diseases have been reported to have an intricate cross-play of the above lymphoid cells, where the NK cells influence the differentiation of $CD4^+$ Th cells into $CD4^+$ Th1 and $CD4^+$ Tregs; $CD4^+$ Th1 plays a crucial role in the initiation of inflammation by cytokine production; the $CD4^+$ Tregs are crucial for immune response modulation (78). Interestingly, the top DAGs of these diseases show pathways associated with the signaling of IL-4 and IL-13 that are crucial in this cross-play, thus corroborating DIME results.

Although we excluded HLA genes to prevent myeloid and B cell bias, the IMIDs associated with the HLA-B27, such as psoriasis, AS, and IBDs were found to have statistically significant common-cell gene networks. However, PsA (also associated with HLA-B27) is omitted here since it did not have a statistically significant common-cell gene network with any IMIDs (**Figure 6B**). Additionally, AS and RA, the two inflammatory arthritis with joint inflammation as the primary feature, also had a statistically significant common-cell gene network. Thus, the diseases with these shared clinical features also had common mechanisms as identified by DIME. The common mechanisms from these networks revealed several lymphoid and myeloid cells and their expressing DAGs. The lymphoid cells such as $CD4^+$ Th1, $CD4^+$ Treg, and NK were predominant in all the statistically significant common-cell gene networks, showing that these diseases were indeed mainly driven by the aggressive T-cell response (36–38, 55). Pathways such as interleukin (IL-4 and IL-13), TLR, TCR signaling, etc., was found to be enriched in the top DAGs of the common cell-gene networks of these IMIDs. Thus, the common cell-gene network revealed several common mechanisms between the diseases in accordance with the top DACs, DAGs, and their associated pathways.

We used the common mechanism from the common cell-gene network and the drug-gene networks to propose potential drug targets for repurposing. This novel computational drug repurposing strategy, a combination of target-based (literature drug-gene network) and mechanism-based (inferred from DIME), revealed several potential drug targets such as IL1B, IL6R, ITGAL, PTGS2, TYK2, NFKB1, NLRP3, PRKCQ, PTGER4, PTPN2, RELA, SH2B3, SMAD3, TLR2, TLR4, and TREM1. Further, we used these mechanism-based drug targets from DIME, and the FDA approved drug-gene network to propose several drug targets and their drugs that could expedite the drug repurposing process (**Table 1**). Thus, we were able to capture drug targets and their drugs currently being targeted or being explored for use in therapy for the IMIDs. We also found a few novel targets, such as the drug lifitegrast (used for dry eyes) for CD, UC, AS, and RA as an alternative to other integrin-based therapies already in use for CD. Lifitegrast is particularly interesting because it targets ITGAL, which was found to be important in the lymphoid cell cluster of CD, UC, AS, and RA. Thus, effectively targeting the exact mechanism. Perhaps the effect of lifitegrast could be used for down-regulating lymphoid cell-mediated inflammation in these diseases (79). Although Lifitegrast is currently available as an eye drop application and used to treat eye complications only, different formulations of this drug can be explored to treat CD, UC, AS, and RA. To the best of our knowledge, lifitegrast in the axis of ITGAL has not been investigated to treat CD, UC, AS, and RA. Thus, using DIME, we were able to propose a novel drug repurposing strategy from the analysis of the 12 IMIDs.

In conclusion, DIME helped identify (1) top DACs, DAGs of the IMIDs (2), Common mechanisms between the IMIDs, and (3) drug targets for repurposing. To enable DIME analysis

for other diseases from the DisGeNet, the GWAS network, or a user-defined set of genes, we built the DIME tool as a user-friendly shinyapp. We believe that this tool will help scientists uncover the etiology of complex and rare diseases and facilitate drug development by better-determining drug targets, thereby mitigating the risk of failure in late clinical development.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included inthe article/**Supplementary Material**. The DIME tool is available on https://bitbucket.org/systemsimmunology/dime. Further inquiries can bedirected to the corresponding author.

AUTHOR CONTRIBUTIONS

AD and AP were involved in the conception of the study. AD was involved in the data curation, visualization, and R shiny package development. AD and AP were involved in the data analysis and interpretation. AD and AP drafted the manuscript. TR helped in writing and revising the manuscript and discussions about clinical perspective. All authors contributed to the article and approved the submitted version.

FUNDING

This project was supported by the Netherlands Organization for Scientific Research (NWO; Grant number 016.Veni.178.027).

ACKNOWLEDGMENTS

The authors thank Ajinkya Kadu for sharing his expertise in linear algebra and multivariate analysis. AP would like to acknowledge the Netherlands Organization for Scientific Research (NWO; Grant number 016.Veni.178.027) for financial support.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2021.669400/full#supplementary-material

REFERENCES

1. Sun YV, Hu Y-J. Integrative Analysis of Multi-omics Data for Discovery and Functional Studies of Complex Human Diseases. *Adv Genet* (2016) 93:147–90. doi: 10.1016/bs.adgen.2015.11.004
2. Li Y, Agarwal P. A Pathway-Based View of Human Diseases and Disease Relationships. *PloS One* (2009) 4:e4346. doi: 10.1371/journal.pone.0004346
3. Camp JG, Platt R, Treutlein B. Mapping Human Cell Phenotypes to Genotypes With Single-Cell Genomics. *Sci (80-)* (2019) 365:1401–5. doi: 10.1126/science.aax6648
4. Rozenblatt-Rosen O, Stubbington MJT, Regev A, Teichmann SA. The Human Cell Atlas: From Vision to Reality. *Nature* (2017) 550:451–3. doi: 10.1038/550451a
5. Lonsdale J, Thomas J, Salvatore M, Phillips R, Lo E, Shad S, et al. The Genotype-Tissue Expression (Gtex) Project. *Nat Genet* (2013) 45:580–5. doi: 10.1038/ng.2653
6. Ardlie KG, Deluca DS, Segre AV, Sullivan TJ, Young TR, Gelfand ET, et al. The Genotype-Tissue Expression (Gtex) Pilot Analysis: Multitissue Gene Regulation in Humans. *Sci (80-)* (2015) 348:648–60. doi: 10.1126/science.1262110
7. GTEx Consortium, Laboratory, Data Analysis &Coordinating Center (LDACC)—Analysis Working Group, Statistical Methods groups—Analysis Working Group, Enhancing GTEx (eGTEx) groups, NIH Common Fund and NIH/NCI, et al. Genetic Effects on Gene Expression Across Human Tissues. *Nature* (2017) 550:204–13. doi: 10.1038/nature24277
8. Tirosh I, Suvà ML. Deciphering Human Tumor Biology by Single-Cell Expression Profiling. *Annu Rev Cancer Biol* (2019) 3:151–66. doi: 10.1146/annurev-cancerbio-030518-055609
9. Keren-Shaul H, Spinrad A, Weiner A, Matcovitch-Natan O, Dvir-Szternfeld R, Ulland TK, et al. A Unique Microglia Type Associated With Restricting Development of Alzheimer's Disease. *Cell* (2017) 169:1276–90.e17. doi: 10.1016/j.cell.2017.05.018
10. Kuo D, Ding J, Cohn IS, Zhang F, Wei K, Rao DA, et al. HBEGF+ Macrophages in Rheumatoid Arthritis Induce Fibroblast Invasiveness. *Sci Transl Med* (2019) 11:eaau8587. doi: 10.1126/scitranslmed.aau8587
11. Knight JC. Genomic Modulators of the Immune Response. *Trends Genet* (2013) 29:74–83. doi: 10.1016/j.tig.2012.10.006
12. Raj T, Rothamel K, Mostafavi S, Ye C, Lee MN, Replogle JM, et al. Polarization of the Effects of Autoimmune and Neurodegenerative Risk Alleles in Leukocytes. *Sci (80-)* (2014) 344:519–23. doi: 10.1126/science.1249547
13. Quach H, Rotival M, Pothlichet J, Loh Y-HE, Dannemann M, Zidane N, et al. Genetic Adaptation and Neandertal Admixture Shaped the Immune System of Human Populations. *Cell* (2016) 167:643–56.e17. doi: 10.1016/j.cell.2016.09.024
14. Lee MN, Ye C, Villani A-C, Raj T, Li W, Eisenhaure TM, et al. Common Genetic Variants Modulate Pathogen-Sensing Responses in Human Dendritic Cells. *Sci (80-)* (2014) 343:1246980–1246980. doi: 10.1126/science.1246980
15. Fairfax BP, Humburg P, Makino S, Naranbhai V, Wong D, Lau E, et al. Innate Immune Activity Conditions the Effect of Regulatory Variants Upon Monocyte Gene Expression. *Sci (80-)* (2014) 343:1246949–1246949. doi: 10.1126/science.1246949
16. Chen L, Ge B, Casale FP, Vasquez L, Kwan T, Garrido-Martín D, et al. Genetic Drivers of Epigenetic and Transcriptional Variation in Human Immune Cells. *Cell* (2016) 167:1398–414.e24. doi: 10.1016/j.cell.2016.10.026
17. Schmiedel BJ, Singh D, Madrigal A, Valdovino-Gonzalez AG, White BM, Zapardiel-Gonzalo J, et al. Impact of Genetic Polymorphisms on Human Immune Cell Gene Expression. *Cell* (2018) 175:1701–15.e16. doi: 10.1016/J.CELL.2018.10.022
18. Yap H-Y, Tee S, Wong M, Chow S-K, Peh S-C, Teow S-Y. Pathogenic Role of Immune Cells in Rheumatoid Arthritis: Implications in Clinical Treatment and Biomarker Development. *Cells* (2018) 7:161. doi: 10.3390/cells7100161
19. Lim J, Kim K. Genetic Variants Differentially Associated With Rheumatoid Arthritis and Systemic Lupus Erythematosus Reveal the Disease-Specific Biology. *Sci Rep* (2019) 9:1–7. doi: 10.1038/s41598-019-39132-2
20. Verstockt B, Smith KG, Lee JC. Genome-Wide Association Studies in Crohn's Disease: Past, Present and Future: Past. *Clin Transl Immunol* (2018) 7:e1001. doi: 10.1002/cti2.1001
21. Bentham J, Morris DL, Cunninghame Graham DS, Pinder CL, Tombleson P, Behrens TW, et al. Genetic Association Analyses Implicate Aberrant Regulation of Innate and Adaptive Immunity Genes in the Pathogenesis of Systemic Lupus Erythematosus. *Nat Genet* (2015) 47:1457–64. doi: 10.1038/ng.3434
22. Okada Y, Wu D, Trynka G, Raj T, Terao C, Ikari K, et al. Genetics of Rheumatoid Arthritis Contributes to Biology and Drug Discovery. *Nature* (2014) 506:376–81. doi: 10.1038/nature12873

23. Piñero J, Bravo À, Queralt-Rosinach N, Gutiérrez-Sacristán A, Deu-Pons J, Centeno E, et al. DisGeNET: A Comprehensive Platform Integrating Information on Human Disease-Associated Genes and Variants. *Nucleic Acids Res* (2017) 45:D833–9. doi: 10.1093/nar/gkw943
24. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: Ultrafast Universal RNA-seq Aligner. *Bioinformatics* (2013) 29:15–21. doi: 10.1093/bioinformatics/bts635
25. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map Format and Samtools. *Bioinformatics* (2009) 25:2078–9. doi: 10.1093/bioinformatics/btp352
26. Anders S, Pyl PT, Huber W. Htseq–a Python Framework to Work With High-Throughput Sequencing Data. *Bioinformatics* (2015) 31:166–9. doi: 10.1093/bioinformatics/btu638
27. Risso D, Ngai J, Speed TP, Dudoit S. Normalization of RNA-seq Data Using Factor Analysis of Control Genes or Samples. *Nat Biotechnol* (2014) 32:896–902. doi: 10.1038/nbt.2931
28. Bodenreider O. The Unified Medical Language System (Umls): Integrating Biomedical Terminology. *Nucleic Acids Res* (2004) 32:D267. doi: 10.1093/nar/gkh061
29. Brunet J-P, Tamayo P, Golub TR, Mesirov JP. Metagenes and Molecular Pattern Discovery Using Matrix Factorization. *Proc Natl Acad Sci U S A* (2004) 101:4164–9. doi: 10.1073/pnas.0308531101
30. Gaujoux R, Seoighe C. A Flexible R Package for Nonnegative Matrix Factorization. *BMC Bioinf* (2010) 11:367. doi: 10.1186/1471-2105-11-367
31. Cotto KC, Wagner AH, Feng Y-Y, Kiwala S, Coffman AC, Spies G, et al. Dgidb 3.0: A Redesign and Expansion of the Drug–Gene Interaction Database. *Nucleic Acids Res* (2018) 46:D1068–73. doi: 10.1093/nar/gkx1143
32. Subramanian A, Narayan R, Corsello SM, Peck DD, Natoli TE, Lu X, et al. A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles. *Cell* (2017) 171:1437–52.e17. doi: 10.1016/j.cell.2017.10.049
33. Xie Z, Hu S, Blackshaw S, Zhu H, Qian J. Hpdi: A Database of Experimental Human protein-DNA Interactions. *Bioinformatics* (2010) 26:287–9. doi: 10.1093/bioinformatics/btp631
34. Armstrong RD, Panayi GS, Welsh KI. Histocompatibility Antigens in Psoriasis, Psoriatic Arthropathy, and Ankylosing Spondylitis. *Ann Rheum Dis* (1983) 42:142–6. doi: 10.1136/ard.42.2.142
35. McGonagle D, Gibbon W, Emery P. Classification of Inflammatory Arthritis by Enthesitis. *Lancet* (1998) 352:1137–40. doi: 10.1016/S0140-6736(97)12004-9
36. Vecellio M, Cohen CJ, Roberts AR, Wordsworth PB, Kenna TJ. RUNX3 and T-bet in Immunopathogenesis of Ankylosing Spondylitis - Novel Targets for Therapy? *Front Immunol* (2019) 10:3132. doi: 10.3389/fimmu.2018.03132
37. Rezaiemanesh A, Abdolmaleki M, Abdolmohammadi K, Aghaei H, Pakdel FD, Fatahi Y, et al. Immune Cells Involved in the Pathogenesis of Ankylosing Spondylitis. *BioMed Pharmacother* (2018) 100:198–204. doi: 10.1016/j.biopha.2018.01.108
38. Veale DJ, Ritchlin C, FitzGerald O. Immunopathology of Psoriasis and Psoriatic Arthritis. In: *Annals of the Rheumatic Diseases*. Annals of the Rheumatic Diseases: BMJ Publishing Group Ltd (2005). p. ii26–9. doi: 10.1136/ard.2004.031740
39. Murdaca G, Gulli R, Spanò F, Lantieri F, Burlando M, Parodi A, et al. Tnf-α Gene Polymorphisms: Association With Disease Susceptibility and Response to Anti-TNF-α Treatment in Psoriatic Arthritis. *J Invest Dermatol* (2014) 134: P2503–9. doi: 10.1038/jid.2014.123
40. Murdaca G, Colombo BM, Cagnati P, Gulli R, Span F, Puppo F. Update Upon Efficacy and Safety of TNF-α Inhibitors. *Expert Opin Drug Saf* (2012) 11:1–5. doi: 10.1517/14740338.2012.630388
41. Fröhling M, Vogl T, Loser K, Paruzel P, Blackshear P, Stumpo D, et al. A1.30 A Key Role of S100A9 in the Pathogenesis of Psoriatic Arthritis in TTP/S100 Deficient Mice. *Ann Rheum Dis* (2016) 75:A13.1–A13. doi: 10.1136/annrheumdis-2016-209124.30
42. Hu XX, Wu YJ, Zhang J, Wei W. T-Cells Interact With B Cells, Dendritic Cells, and Fibroblast-Like Synoviocytes as Hub-Like Key Cells in Rheumatoid Arthritis. *Int Immunopharmacol* (2019) 70:428–34. doi: 10.1016/j.intimp.2019.03.008
43. Cope AP, Schulze-Koops H, Aringer M. The Central Role of T Cells in Rheumatoid Arthritis. *Clin Exp Rheumatol* (2007) 25:S4–11.
44. Cascão R, Rosário HS, Souto-Carneiro MM, Fonseca JE. Neutrophils in Rheumatoid Arthritis: More Than Simple Final Effectors. *Autoimmun Rev* (2010) 9:531–5. doi: 10.1016/j.autrev.2009.12.013
45. Udalova IA, Mantovani A, Feldmann M. Macrophage Heterogeneity in the Context of Rheumatoid Arthritis. *Nat Rev Rheumatol* (2016) 12:472–85. doi: 10.1038/nrrheum.2016.91
46. Louis C, Souza-Fonseca-Guimaraes F, Yang Y, D'Silva D, Kratina T, Dagley L, et al. NK Cell–Derived GM-CSF Potentiates Inflammatory Arthritis and is Negatively Regulated by CIS. *J Exp Med* (2020) 217:e20191421. doi: 10.1084/jem.20191421
47. Assassi S, Mayes MD, Arnett FC, Gourh P, Agarwal SK, McNearney TA, et al. Systemic Sclerosis and Lupus: Points in an Interferon-Mediated Continuum. *Arthritis Rheum* (2010) 62:589–98. doi: 10.1002/art.27224
48. Matta B, Barnes BJ. Coordination Between Innate Immune Cells, Type I Ifns and IRF5 Drives SLE Pathogenesis. *Cytokine* (2020) 132:154731. doi: 10.1016/j.cyto.2019.05.018
49. Vaughn SE, Kottyan LC, Munroe ME, Harley JB. Genetic Susceptibility to Lupus: The Biological Basis of Genetic Risk Found in B Cell Signaling Pathways. *J Leukoc Biol* (2012) 92:577–91. doi: 10.1189/jlb.0212095
50. Van Caam A, Vonk M, Van Den Hoogen F, Van Lent P, Van Der Kraan P. Unraveling SSc Pathophysiology; The Myofibroblast. *Front Immunol* (2018) 9:2452. doi: 10.3389/fimmu.2018.02452
51. Denny MF, Yalavarthi S, Zhao W, Thacker SG, Anderson M, Sandy AR, et al. A Distinct Subset of Proinflammatory Neutrophils Isolated From Patients With Systemic Lupus Erythematosus Induces Vascular Damage and Synthesizes Type I Ifns. *J Immunol* (2010) 184:3284–97. doi: 10.4049/jimmunol.0902199
52. Garcia-Romo GS, Caielli S, Vega B, Connolly J, Allantaz F, Xu Z, et al. Netting Neutrophils are Major Inducers of Type I Ifn Production in Pediatric Systemic Lupus Erythematosus. *Sci Transl Med* (2011) 3:73ra20–0. doi: 10.1126/SCITRANSLMED.3001201
53. Darrah E, Andrade F. Nets: The Missing Link Between Cell Death and Systemic Autoimmune Diseases? *Front Immunol* (2012) 3:428. doi: 10.3389/fimmu.2012.00428
54. Lindau D, Mussard J, Rabsteyn A, Ribon M, Kötter I, Igney A, et al. TLR9 Independent Interferon α Production by Neutrophils on NETosis in Response to Circulating Chromatin, a Key Lupus Autoantigen. *Ann Rheum Dis* (2014) 73:2199–207. doi: 10.1136/annrheumdis-2012-203041
55. Sartor RB. Mechanisms of Disease: Pathogenesis of Crohn's Disease and Ulcerative Colitis. *Nat Clin Pract Gastroenterol Hepatol* (2006) 3:390–407. doi: 10.1038/ncpgasthep0528
56. Giuffrida P, Di Sabatino A. Targeting T Cells in Inflammatory Bowel Disease. *Pharmacol Res* (2020) 159:105040. doi: 10.1016/j.phrs.2020.105040
57. Geremia A, Arancibia-Cárcamo CV. Innate Lymphoid Cells in Intestinal Inflammation. *Front Immunol* (2017) 8:1296. doi: 10.3389/fimmu.2017.01296
58. Poggi A, Benelli R, Venè R, Costa D, Ferrari N, Tosetti F, et al. Human Gut-Associated Natural Killer Cells in Health and Disease. *Front Immunol* (2019) 10:961. doi: 10.3389/fimmu.2019.00961
59. Park K. A Review of Computational Drug Repurposing. *Transl Clin Pharmacol* (2019) 27:59. doi: 10.12793/tcp.2019.27.2.59
60. Ito H, Takazoe M, Fukuda Y, Hibi T, Kusugami K, Andoh A, et al. A Pilot Randomized Trial of a Human Anti-Interleukin-6 Receptor Monoclonal Antibody in Active Crohn's Disease. *Gastroenterology* (2004) 126:989–96. doi: 10.1053/j.gastro.2004.01.012
61. Tsai Y-C, Tsai T-F. Anti-Interleukin and Interleukin Therapies for Psoriasis: Current Evidence and Clinical Usefulness. *Ther Adv Musculoskelet Dis* (2017) 9:277–94. doi: 10.1177/1759720X17735756
62. Alten R, Gomez-Reino J, Durez P, Beaulieu A, Sebba A, Krammer G, et al. Efficacy and Safety of the Human anti-IL-1beta Monoclonal Antibody Canakinumab in Rheumatoid Arthritis: Results of a 12-Week, Phase II, Dose-Finding Study. *BMC Musculoskelet Disord* (2011) 12:153. doi: 10.1186/1471-2474-12-153
63. Haibel H, Rudwaleit M, Listing J, Sieper J. Open Label Trial of Anakinra in Active Ankylosing Spondylitis Over 24 Weeks. *Ann Rheum Dis* (2005) 64:296–8. doi: 10.1136/ard.2004.023176
64. Hügle B, Speth F, Haas J-P. Inflammatory Bowel Disease Following anti-interleukin-1-treatment in Systemic Juvenile Idiopathic Arthritis. *Pediatr Rheumatol* (2017) 15:16. doi: 10.1186/s12969-017-0147-3
65. Merashli M, De Marco G, Podgorski M, McGonagle D, Marzo-Ortega H. Evidence of Response to IL-6 Inhibition in Some Cases of Refractory Spondyloarthritis Associated Peripheral Synovitis. *Ann Rheum Dis* (2016) 75:1418–20. doi: 10.1136/annrheumdis-2016-209275

66. Danese S, Vermeire S, Hellstern P, Panaccione R, Rogler G, Fraser G, et al. Randomised Trial and Open-Label Extension Study of an Anti-Interleukin-6 Antibody in Crohn's Disease (ANDANTE I and II). *Gut* (2019) 68:40–8. doi: 10.1136/gutjnl-2017-314562
67. Nishimoto N, Hashimoto J, Miyasaka N, Yamamoto K, Kawai S, Takeuchi T, et al. Study of Active Controlled Monotherapy Used for Rheumatoid Arthritis, an IL-6 Inhibitor (SAMURAI): Evidence of Clinical and Radiographic Benefit From an X Ray Reader-Blinded Randomised Controlled Trial of Tocilizumab. *Ann Rheum Dis* (2007) 66:1162–7. doi: 10.1136/ard.2006.068064
68. Blauvelt A. Il-6 Differs From TNF-α: Unpredicted Clinical Effects Caused by IL-6 Blockade in Psoriasis. *J Invest Dermatol* (2017) 137:541–2. doi: 10.1016/j.jid.2016.11.022
69. Atreya R, Billmeier U, Rath T, Mudter J, Vieth M, Neumann H, et al. First Case Report of Exacerbated Ulcerative Colitis After anti-interleukin-6R Salvage Therapy. *World J Gastroenterol* (2015) 21:12963–9. doi: 10.3748/wjg.v21.i45.12963
70. Van Der Heijde D, Deodhar A, Wei JC, Drescher E, Fleishaker D, Hendrikx T, et al. Tofacitinib in Patients With Ankylosing Spondylitis: A Phase II, 16-Week, Randomised, Placebo-Controlled, Dose-Ranging Study. *Ann Rheum Dis* (2017) 76:1340–7. doi: 10.1136/annrheumdis-2016-210322
71. Rogler G. JAK Efficacy in Crohn's Disease. *J Crohn's Colitis* (2019) 14:S746–54. doi: 10.1093/ecco-jcc/jjz186
72. Dhillon S. Tofacitinib: A Review in Rheumatoid Arthritis. *Drugs* (2017) 77:1987–2001. doi: 10.1007/s40265-017-0835-9
73. Tian F, Chen Z, Xu T. Efficacy and Safety of Tofacitinib for the Treatment of Chronic Plaque Psoriasis: A Systematic Review and Meta-Analysis. *J Int Med Res* (2019) 47:2342–50. doi: 10.1177/0300060519847414
74. Hübel K, Fresen MM, Salwender H, Basara N, Beier R, Theurich S, et al. Plerixafor With and Without Chemotherapy in Poor Mobilizers: Results From the German Compassionate Use Program. *Bone Marrow Transplant* (2011) 46:1045–52. doi: 10.1038/bmt.2010.249
75. Ley K, Rivera-Nieves J, Sandborn WJ, Shattil S. Integrin-Based Therapeutics: Biological Basis, Clinical Use and New Drugs. *Nat Rev Drug Discovery* (2016) 15:173–83. doi: 10.1038/nrd.2015.10
76. Ellinghaus D, Jostins L, Spain SL, Cortes A, Bethune J, Han B, et al. Analysis of Five Chronic Inflammatory Diseases Identifies 27 New Associations and Highlights Disease-Specific Patterns At Shared Loci. *Nat Genet* (2016) 48:510–8. doi: 10.1038/ng.3528
77. Bernstein CN, Sargent M, Rector E. Alteration in Expression of Beta 2 Integrins on Lamina Propria Lymphocytes in Ulcerative Colitis and Crohn's Disease. *Clin Immunol* (2002) 104:67–72. doi: 10.1006/clim.2002.5223
78. Gianchecchi E, Delfino DV, Fierabracci A. NK Cells in Autoimmune Diseases: Linking Innate and Adaptive Immune Responses. *Autoimmun Rev* (2018) 17:142–54. doi: 10.1016/j.autrev.2017.11.018
79. Perez VL, Pflugfelder SC, Zhang S, Shojaei A, Haque R. Lifitegrast, a Novel Integrin Antagonist for Treatment of Dry Eye Disease. *Ocul Surf* (2016) 14:207–15. doi: 10.1016/j.jtos.2016.01.001

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.

frontiers in Immunology

ORIGINAL RESEARCH
published: 10 June 2021
doi: 10.3389/fimmu.2021.642842

Transcriptional and Microenvironmental Landscape of Macrophage Transition in Cancer: A Boolean Analysis

Ugo Avila-Ponce de León[1,2], *Aarón Vázquez-Jiménez*[2], *Meztli Matadamas-Guzman*[2,3], *Rosana Pelayo*[4] *and Osbaldo Resendis-Antonio*[2,5]*

[1] Programa de Doctorado en Ciencias Biológicas, Universidad Nacional Autónoma de México, Ciudad de México, Mexico, [2] Human Systems Biology Laboratory, Instituto Nacional de Medicina Genómica (INMEGEN), Ciudad de México, Mexico, [3] Doctorado en Ciencias Biomédicas, Universidad Nacional Autónoma de México, Ciudad de México, Mexico, [4] Oncoimmunology Laboratory, Centro de Investigación Biomédica de Oriente, Instituto Mexicano del Seguro Social, Puebla, Mexico, [5] Coordinación de la Investigación Científica - Red de Apoyo a la Investigación, UNAM, Ciudad de México, Mexico

OPEN ACCESS

Edited by:
Federica Eduati,
Eindhoven University of Technology, Netherlands

Reviewed by:
Keehoon Jung,
Seoul National University, South Korea
Sang-Jun Ha,
Yonsei University, South Korea
Tomáš Helikar,
University of Nebraska-Lincoln, United States

***Correspondence:**
Osbaldo Resendis-Antonio
oresendis@inmegen.gob.mx

Specialty section:
This article was submitted to Molecular Innate Immunity, a section of the journal Frontiers in Immunology

Received: 16 December 2020
Accepted: 17 May 2021
Published: 10 June 2021

Citation:
Avila-Ponce de León U, Vázquez-Jiménez A, Matadamas-Guzman M, Pelayo R and Resendis-Antonio O (2021) Transcriptional and Microenvironmental Landscape of Macrophage Transition in Cancer: A Boolean Analysis. Front. Immunol. 12:642842. doi: 10.3389/fimmu.2021.642842

The balance between pro- and anti-inflammatory immune system responses is crucial to face and counteract complex diseases such as cancer. Macrophages are an essential population that contributes to this balance in collusion with the local tumor microenvironment. Cancer cells evade the attack of macrophages by liberating cytokines and enhancing the transition to the M2 phenotype with pro-tumoral functions. Despite this pernicious effect on immune systems, the M1 phenotype still exists in the environment and can eliminate tumor cells by liberating cytokines that recruit and activate the cytotoxic actions of TH1 effector cells. Here, we used a Boolean modeling approach to understand how the tumor microenvironment shapes macrophage behavior to enhance pro-tumoral functions. Our network reconstruction integrates experimental data and public information that let us study the polarization from monocytes to M1, M2a, M2b, M2c, and M2d subphenotypes. To analyze the dynamics of our model, we modeled macrophage polarization in different conditions and perturbations. Notably, our study identified new hybrid cell populations, undescribed before. Based on the *in vivo* macrophage behavior, we explained the hybrid macrophages' role in the tumor microenvironment. The *in silico* model allowed us to postulate transcriptional factors that maintain the balance between macrophages with anti- and pro-tumoral functions. In our pursuit to maintain the balance of macrophage phenotypes to eliminate malignant tumor cells, we emulated a theoretical genetically modified macrophage by modifying the activation of NFκB and a loss of function in HIF1-α and discussed their phenotype implications. Overall, our theoretical approach is as a guide to design new experiments for unraveling the principles of the dual host-protective or -harmful antagonistic roles of transitional macrophages in tumor immunoediting and cancer cell fate decisions.

Keywords: macrophage, phenotype, boolean models, systems immunology, gene regulatory network, cancer immunology

INTRODUCTION

Macrophages are essential cells in inflammatory responses and immune regulation. However, these cells have heterogeneous functions depending on their polarization. Despite the importance of the process, only the M1 and M2 phenotypes have been experimentally characterized (1, 2). Nevertheless, there are macrophages with intermediate functional phenotypes or hybrid stages that have not been fully described yet. M1, M2, and some hybrid states have specific gene expression and secretion profiles, depending on the microenvironment signals (3). For instance, M1 macrophages express NFκB or STAT1 and secrete TNF-α and IL-12, correlated with a pro-inflammatory response. On the contrary, the M2 macrophages are correlated with an anti-inflammatory response, but their expression profiles are subject to the hybrid stages. As far as we know M2 macrophages have 4 hybrid stages, called M2a, M2b, M2c, and M2d (4). M2a macrophages express STAT6 and secrete IL-10, TGF-β, and IL-1RA, associated with pro-fibrotic functions and inhibition of Th1, and a Th2 response (5). M2b macrophages express Erk and AP-1, and secrete IL-10, TNF- α, and IL-1, correlated with immune regulation (6). M2c macrophages express STAT3 and secrete IL-10 and TGF-β, involved in tissue repair, matrix remodeling, and immunosuppressive behavior (7). Finally, M2d macrophages express HIF1-α or a defective NFκB, enhance STAT1, and induce the secretion of IL-10, TGF-β, and VEGF. M2d is associated with tumors, enhancing the angiogenic process, metastasis, tumor growth, and regulating the immune system (8–10). In principle, hybrid stages in macrophage polarization depend on the microenvironment signals, such as cytokines, membrane receptors, and transcription factors, to determine their specific function and cell fate. However, macrophage hybrid states can transit from one to another depending on the stimuli. M1 macrophages can polarize reversibly into M2a, M2c, and M2d, while M2c can polarize reversibly into M2a (8–11). Moreover, the microenvironment signals can be secreted by other cells within the tissue or by the macrophages themselves (12). Despite the previous characterization, theoretically it has been proposed a continuum transition between the mentioned phenotypes (13). Therefore, integrating the signaling responses and molecular mechanisms involved in the macrophage polarization is required to untangle the complexity of the transitions between the different phenotypes.

As stated above, macrophages respond to tumor microenvironment stimuli in a coordinated and regulated manner. At the molecular level, transcription factors regulate the expression of other transcription factors or mRNA molecules inducing a change in the behavior based on the stimuli. These interactions can be summarized in a gene regulatory network where we can analyze their dynamic properties through mathematical models (14). Given that gene regulatory networks contemplate a large number of variables and the interactions among them, modeling their behavior using a continuous approach would be difficult due to the lack of kinetic parameters. However, gene regulatory networks represent variables as nodes and the interactions between them as edges, making a discrete approach suitable because it requires few parameters. A common approach is the Boolean modeling. This mathematical approach is, where we numerically represent the state of the nodes as 0 (OFF) or 1 (ON). The OFF state means that a transcription factor or molecule is below a certain threshold, meaning it is inactive, while ON indicates an active molecule. The activation and inactivation rules for each node are called Boolean functions. These rules, indicate how each node is activated or inhibited are obtained from experimental evidence and mathematically represented through logical operators: AND (stated as: &), OR (stated as: |), and NOT (stated as: ¬), obtained based on the experimental evidence on how each node is activated or inhibited. The system is solved *via* a synchronous or asynchronous update where we transform a static gene regulatory network into a dynamic one. By applying a synchronous all Boolean functions are updated simultaneously, and they will eventually arrive at a steady state, called an attractor. An attractor is a point where a group of states will converge to that point and they cannot leave this state until a perturbation occurs (15, 16). The biological interpretation of an attractor is a certain cell type or function defined by a gene pattern, it can be related to one or more subtypes of phenotypes and a phenotype can be associated with one or more attractors. This approach has been used to successfully model biological systems (17, 18), and study the cell fate decisions of the adaptive immune response by their specific signals (19, 20). Notably, this approach validated the continuum hypothesis of polarization states by integrating a network of external cytokine signals (21, 22). Although valuable endeavors have recovered and model some of the macrophage phenotypes (M1, M2a, M2b, and M2c) (21, 22). However, there is a lack of information about other phenotypes and their functionalities. We propose a new model capable of recreating the macrophage phenotype diversity considering relevant experimental evidence to fulfill this gap.

In this paper, we present a Boolean dynamical analysis over an updated signaling regulatory network of macrophages to assess the polarization of intermediate phenotypes immersed in the tumor microenvironment. Our Boolean model predicted known macrophage phenotypes, plus hybrid intermediate phenotypes with mixed physiological functions associated with tumor eradication, tissue repair, and progression of tumor cells. Then, we proceeded with an in-depth analysis of the functional properties of the hybrid states. Analyzing the stability of the new hybrid states. As a result, we acknowledge the importance of STAT1, NFκB for the M1 phenotype and HIF1-α for the M2 phenotype. We also proposed a theoretical approach of macrophage immunotherapy in an *in silico* breast cancer microenvironment, which showed promising results in a specific microenvironment towards a better prognosis for cancer eradication. Overall, this work represents a systems biology framework capable of characterizing known macrophage phenotypes and suggesting new hybrid cell populations. Also, it explores the typified macrophages functionality in a tumor microenvironment as well as their robustness and properties. Moreover, our model serves as a computational platform to explore how the tumor microenvironment can potentially modulate macrophages-cancer interaction.

MATERIAL AND METHODS

Mathematical Modeling of Network Phenotypes Through Discrete Variables

To explore the feasible space of phenotypes associated with our signaling network, we applied a Boolean approach. Briefly, Boolean approaches assume that each node in the network can be in one of two states (0 or 1), and their dynamic behavior is entirely governed by a Boolean function, which is defined by their regulation. Under this situation, the dynamic state of the i-esime node at time t is given by

$$x_i(t+1) = f_i(x_{i1}(t), x_{i2}(t), \ \ldots \ , \ x_{iki}(t))$$

Where x_i can take the values 0 or 1, and f_i represents the Boolean function of the i-esime node in the network. Note that every node has a Boolean function which determine how that specific node respond to the change of the neighbor regulators (here indicated by the arguments of f_i). Also, f_i maps a Boolean state including multiple Boolean variables at time t into a new Boolean state. After one unit of time, the dynamic behavior of the network is obtained when we simultaneously apply the transition function f_i over all the nodes. Starting from the 2^N initial states, the Boolean transition functions will allow the network to reach a finite set of states that could cycle between them in a fixed state. These recurrent states are called attractors: if the dynamics reaches a state and stays there, they are called simple attractors, but if they move irregularly in a set of states, they are called complex or loose attractors. All the set of initial states whose transition function guide to the same attractors will form the basin of attraction. The attractors are of great importance because they represent the long-term behavior of the Boolean model and are potentially associated with phenotypes in the system of study (20, 23).

Analysis and Simulation of the Attractors of Our Boolean Model

A gene regulatory network is a dynamic system. We obtained all of the attractors using the exhaustive search algorithm evaluating the space of 2^n (where n is the number of nodes) initial conditions. So we evaluated 536,870,912 initial conditions using BoolNet Library from R software (24). The obtained attractors carried important biological implications and were associated with macrophage phenotypes. Each attractor was labeled based on experimental evidence. The labeling was done with a function of R package BoolNetPerturb, located at https://github.com/mar-esther23/boolnet-perturb-0.1. To compare the differences between the obtained attractors we reduced the dimension of our variables to plot them in a 2D-space using the t-SNE function in R. Once we had our plot, we clustered them to find similarities between the attractors. To determine the optimal number of clusters, we develop rules based on each possible group value's associated error, varying from 10 to 30. Finally, the clusters were named based on the functions carried out by the phenotypes of macrophages in each cluster. The model is available in a txt file and an SBML file for it could be easily reproduced. Our Code, data, and implemented analysis can be accessed at: https://github.com/resendislab/M1-M2-Macrophage-Polarization.

Mutation Analysis

Mutation analysis was divided into two sections: gene deletion and activation. The number of initial conditions for each perturbation is 2^{n-1} where n–1 is the number of nodes minus the perturbed, so for every perturbation we evaluated 268,435456 initial conditions. We permanently set the state of the node as 0 for gene deletion and simulated the dynamics until it reached an attractor. Conversely, gene activation was accomplished by permanently fixing as 1 the state of one node in the network along with all units of time. These considerations were used to mimic the experimental condition of knock-outs or overexpression, respectively. By making this approach, we evaluated the robustness of a given macrophage subtype to resist a gene deletion or activation and how it affects its molecular activation pattern.

In the case of the cell fate map of macrophage polarization, we randomly changed the values of each attractor's nodes. by a single bit flip. Then we evaluated which nodes changed a given macrophage phenotype into another given phenotype; this perturbation will determine the one-state neighbors.

Robustness Analysis of Our Model

To validate the robustness of our model, we calculated the Derrida curve and developed the sensitivity analysis of the update rules. A Derrida curve allowed us to evaluate if a gene regulatory network is chaotic, ordered or critical. If the differences between the initial states diverge rapidly from the degrees diagonal, the network is chaotic. To obtain the Derrida curve we proceeded as follows. First, we sampled a random pair of initial states *X1(t)* and *X2(t)* at the same known time *t*. Once we have these states, we calculated the normalized Hamming distance *h(t)* which is the number of bits that differ from the two initial states divided by the number of nodes N. Then, we initialize in time the Boolean network and let both initial states evolve to their next states *X1(t+1)* and *X2(t+1)*, we calculated the normalized hamming distance denoted as *h(t+1)* run the dynamic. We repeated the previous steps for 10 000 random pairs and calculated the average behavior over the realizations. We plotted the value of *h(t)* in the x-axis and the y-axis the value of *h(t+1)*, as well we plotted the diagonal where *h(t)=h(t+1)* which means that the successor states and the initial states are the same. The Derrida curve was calculated with the R package BoolNetPerturb, located at https://github.com/mar-esther23/boolnet-perturb-0.1.

Complementary, the sensitivity analysis showed the differences on the results depending on alterations on each component, either turning on or off respectively. First, we created 50 000 randomly chosen states for each Boolean function define as *f(ki)* where *ki* is the different variables and *i* iterates from 1 to 29. Then, for each of these 50,000 initial states, we created a mutation randomly with only one-bit flip (Hamming distance of 1) labeled as *f(ki')*. Finally, we applied the update rule for *ki* and *ki'*, and we calculated the sensitivity of the update rule as the fraction of initial states where $f(ki) \neq f(ki')$.

The sensitivity analysis was carried out using the package of BoolNet in R.

RESULTS

Network Reconstruction and Boolean Analysis: The Molecular Basis of Macrophage Polarization in a Tumor Microenvironment

The tumor microenvironment contains several cell types, such as cancer cells, immune cells, and mesenchymal cells. These cells produce cytokines that attract other monocytes to the tumor as a natural immune response in the host. The interplay between tumor microenvironment interactions and macrophage functions must be taken into account to understand part of the inflammatory process involved in the interaction between macrophages and cancer. To this end, we reconstructed a signaling network that comprises a variety of regulatory elements associated with macrophage differentiation. We set up our interaction network of the macrophage polarization by a literature search of interactions between external components and how this affects the activation or inhibition of transcriptional factors associated with macrophage polarization. This approach is called a bottom-up methodology. The original reconstruction included 40 nodes. Then, we excluded all those nodes with less than two interactions (in and out) to analyze the dynamic behavior of the network. We did it because these nodes do not contribute information to the dynamic analysis. Therefore, we did the dynamic Boolean analysis onto a network of 29 nodes and 60 interactions. Overall, our transcriptional regulatory network (TRN), integrates extracellular-intracellular components, signal transduction cascades, and transcriptional regulatory mechanisms. This regulatory network of macrophage polarization had two parts, the extracellular components (green nodes) and the master transcriptional factors (blue nodes) that regulate macrophage polarization (**Figure 1** and **Table S1**).

Among the signaling cascades included in our reconstruction, we included the following interactions: Cancer dying cells produce HMGB1 activating Toll-like receptor 4 (TLR4), consequently activating NFκB (25). Briefly, these signals initiate the anti-tumor response in macrophages. However, the macrophage colony-stimulating factor (M-CSF) also triggers NFκB through JAK2 and STAT5. Besides, interferon triggers STAT1, so the cell releases cytokines to inhibit tumor growth (26). On the other side, we also described different activation processes for the M2 sub phenotypes (M2a, M2b, M2c, and M2d). First, M2a activation depends on IL-4 and IL-13 and triggers STAT6 (27). At the same time, STAT6 inhibits STAT1 and NFκB through SOCS1 and KLF4, avoiding the M1 phenotype. Then, M2b activation depends on the signals of AP1 or ERK. AP1 is activated by the presence of IL-1β, while ERK needs the binding to Fc receptors (6). Next, M2c activation relies on the presence of STAT3. IL-10, glucocorticoids, TGF-β, and adenosines trigger STAT3 (28). This cascade induces the secretion of IL-10 and the inactivation of NFκB. Finally, M2d activation hinges on IL-6 or HIF1, related to a hypoxic environment (29). This phenotype is present in solid tumors. This cascade induces the secretion of IL-10, VEGF, and other cytokines. Conversely, the activation of any M2 sub phenotypes culminates in the liberation of cytokines, such as IL-10, that enhance tumor growth. We also included interactions that described M2d phenotype, a tumor-associated-macrophage, and relevant inflammatory cytokines, such as TGF-β and TNF-α. We used this network to dynamically analyze the master regulators that drive the transition between several types of macrophage phenotypes. Hence, to unveil the effect of the tumor microenvironment in macrophage polarization, we performed a Boolean analysis over a reconstructed transcriptional regulatory

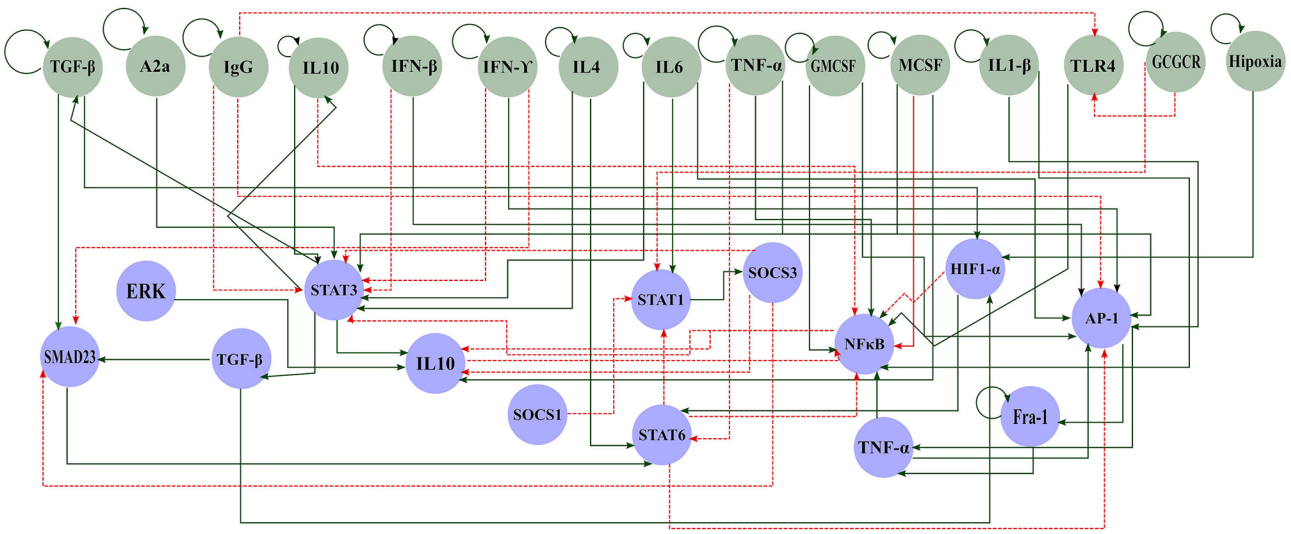


FIGURE 1 | Gene regulatory network of macrophage polarization in a tumor microenvironment. Green circles represent components of the extracellular space and blue circles represent the components of the internal machinery. Solid green arrows represent activation and dashed red arrows inhibition.

network for macrophages (**Table S1**). Our network contains a monocyte-derived macrophage regulatory mechanism, as depicted (**Figure 1**). We assumed that all the receptors for the external stimuli were constitutive. This assumption allowed us to eliminate linear interactions and to reduce the size of our network without affecting the dynamic and topology.

Boolean Modeling of the TRN and Its Spectrum of Macrophage Phenotypes

Based on our reconstructed signaling network, we determined the possible macrophage phenotypes existing in a tumoral microenvironment. Our Boolean model had interconnected signaling pathways, and their dynamics with the external inputs eventually converged to stable states, called single attractors. An attractor is a set of activated or inactivated genes that are time-invariant and potentially represent a macrophage phenotype. We named each attractor based on *in vitro* experimental evidence; we considered six possible steady states associated with a specific macrophage phenotype (**Table 1**). Using a synchronous update, our TRN converged to an entire landscape of 10,430 attractors, where 56 were cyclic, and the remaining were simple attractors. The simple attractors represented 13 phenotypes. Among them, four are already experimentally described (M0, M1, M2b, and M2d). The nine remaining were hybrid phenotypes with two or three macrophage phenotypes. To identify similarities among attractors, we made a bidimensional map using a distributed stochastic neighbor embedding (t-SNE) and a clustering analysis (**Figure S1**). Employing a 30 metrics consensus, we determined that the optimal cluster value was 10 (**Figure S2**). We obtained 10 clusters for 13 phenotypes (**Figure 2**), each one embedding more than one phenotype (**Figure S3**).

Among all the phenotypes, seven correlated with a pro-tumoral response and five with an anti-tumoral response. The tumor microenvironment enhances macrophages polarization to

TABLE 1 | Macrophage phenotypes used to name each attractor.

Phenotype	Transcriptional Factors Activated/External components	Literature Revised
M0	None	
M1	Activation of NFκB, or STAT1, or TNFα and AP1	(26, 30)
M2a	Activation of STAT6	(28, 31)
M2b	Activation of AP1 or ERK	(12)
M2c	Activation of STAT3	(29)
M2d	Activation of TLR4 and A2a or HIF1A	(32)

The attractors obtained after the simulation were labeled based on the macrophage phenotypes reported in the literature.

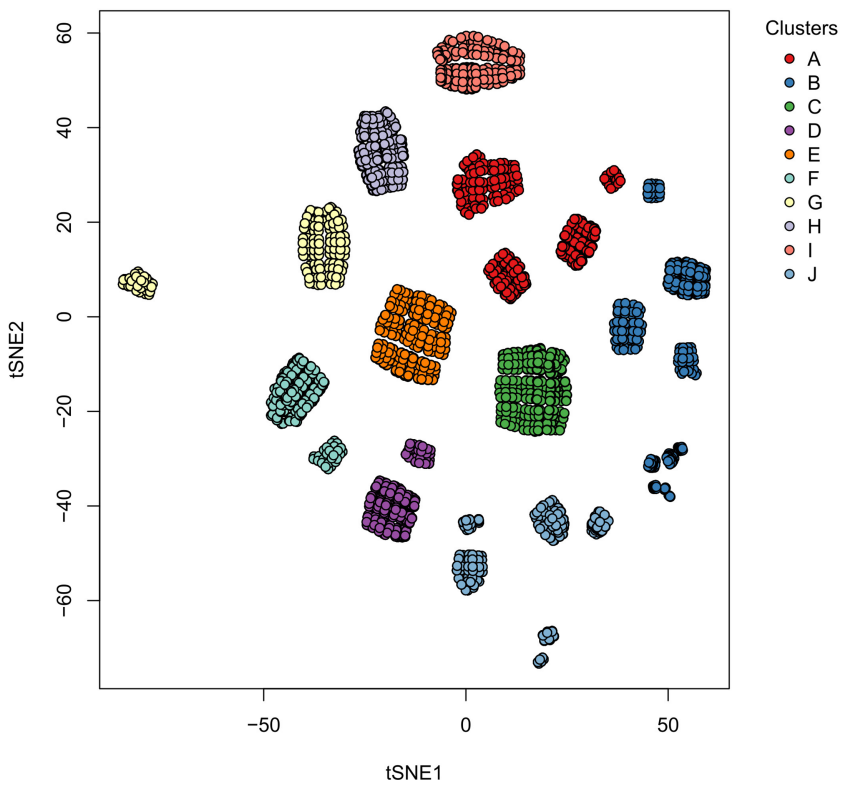


FIGURE 2 | Clusters of the single size attractors obtained in the Boolean simulations. Each circle represents a two-dimensional projection of one attractor, and the color represents a collection of attractors with similar phenotypes between them. The numbers associated with the clusters are explained in **Table 1**.

pro-tumoral phenotypes with the liberation of interleukins and chemokines. In return, macrophages release growth and immune regulatory factors allowing proliferation and progression of cancer cells. Notwithstanding, in a tumor microenvironment, there are macrophage phenotypes with anti-tumor activity, balancing the elimination of tumor cells without affecting the surrounding tissues.

Due to the massive amounts of stimuli in a tumor microenvironment, we hypothesized that many hybrid activation states exist. We handled macrophage hybrid phenotypes as a single cell identity rather than a heterogeneous population, as in previous works (32). We suggested that a hybrid will be phenotypically distinguished by a combination of expression markers associated with pure M1 and M2 phenotypes. For instance, cluster A contained two-hybrid (mixed macrophage behavior) phenotypes M2bM2d and M2aM2d, macrophages associated with wound healing and regulatory behavior. Based on their composition, we concluded that this cluster was beneficial for tumor growth due to the liberation of growth factors maintaining tumor proliferation and regulatory cytokines to maintain the immune system. On the other hand, clusters B and I consisted of macrophages that would eliminate tumor cells due to the presence of M1, which liberate cytotoxic cytokines to the microenvironment (30). Clusters C and D behavior was complex due to the heterogeneous phenotype composition. Cluster F was a wound-healing and regulatory macrophage due to a hybrid phenotype labeled as M2aM2b (**Table 2**). According to our results, a good scenario against cancer is given by a macrophage state the increases tumor clearance role, by releasing tumor cytotoxic factors, and diminishing the levels of components associated with pro-tumoral behavior.

Loss and Gain Function Alteration Dictates the Macrophages Subtypes

In this section, we computationally assessed the importance of each node in the stability of the phenotypes obtained. For this purpose, we performed a permanent deletion (set the value of the node to zero) or activation (set the value of the node to one) of each node in the dynamic analysis of the TRN. To quantify the effect produced by these perturbations, we applied the following equation:

$$log_2\left(\frac{M \pm i}{WT_i}\right)$$

where WT_i denotes the size of the basin of attraction for the "i" attractor without alterations, and $M \pm i$ denotes the basin of attraction for the "i" attractor for the deletion(-) and activating (+) of the node. As expected, the previous equation is constrained to those cases where $WT_i \neq 0$ and both variables are positively defined. The size of the basin of attraction for WT is depicted in **Figure S4**. **Figure 3** shows the log-fold of the size of the basin of attraction when we compared perturbed and unperturbed (WT) genes stated by columns. Gray areas indicate attractors that were in the wild-type but not in the perturbed state. Attractors with an M1 macrophage will be an anti-tumoral phenotype, and with the M2 macrophage, a pro-tumoral phenotype (rows). Each column is an activation or inactivation of a transcription factor.

In terms of the results, we observed that permanent activation of AP-1, STAT1, or NFκB was very similar because they induce hybrids with tumoricidal capacity (M1). AP-1 and STAT1 activation decreased at 100% pro-tumoral phenotypes and 20% anti-tumoral phenotypes. These transcriptional factors may recover the balance towards a tumor eliminating scenario in the tumor microenvironment (gray areas **Figures 3A** and **S5, S6** respectively). NFκB permanent activation maintained the same behavior as previous transcription factors because it diminished at a 100% pro-tumoral phenotype and 20% anti-tumoral phenotypes, creating a perfect balance of tumor elimination and recovery of the tissue due to the secretion of cytotoxic interleukins and interferons (gray areas in **Figures 3A** and **S5, S6** respectively). On the other hand, constitutive expressions of HIF1-α or TGF-β eliminated pure M1 macrophages (**Figure 3A**) and decreased the available profiles that converged to hybrids with M1 as a component. HIF1-α and TGF-β promoted the development of malignant behavior in macrophages. Once activated, HIF1-α decreased 22% anti-tumoral phenotypes (**Figure S6**) and 40% pro-tumoral phenotypes (**Figure S5**). We concluded cells should avoid the

TABLE 2 | Macrophage phenotypes associated with each cluster.

Cluster	Phenotypes	Behavior
A	M2bM2d and M2aM2d	Regulatory/wound healing macrophage
B	M1 and M1M2d	Classical activated macrophages/Pro-tumoral macrophages
C	M1M2bM2d, M1M2bM2cM2d, M1M2b, M2d and M2M2d	Complex behavior. Mixed phenotypes associated with different functions
D	M0, M1M2bM2d, M2b, M2d and M2bM2d	Complex behavior. Mixed phenotypes associated with different functions
E	M1M2bM2d and M2aM2d	Classical activated macrophage/Pro-tumor
F	M2aM2b	Wound-healing macrophages
G	M2aM2cM2d and M2aM2bM2d	Regulatory/wound-healing macrophages
H	M2bM2d and M1M2bM2d	Regulatory/Pro-tumoral macrophages
I	M1 and M1M2d	Classical activated macrophages/Pro-tumoral macrophages
J	M2aM2d and M2aM2bM2d	Regulatory/Pro-tumoral macrophages

The phenotype behavior of each cluster was approximated by the criteria of classification used by (33).
Description of the results obtained from the k-means clustering over the t-SNE 2D space.

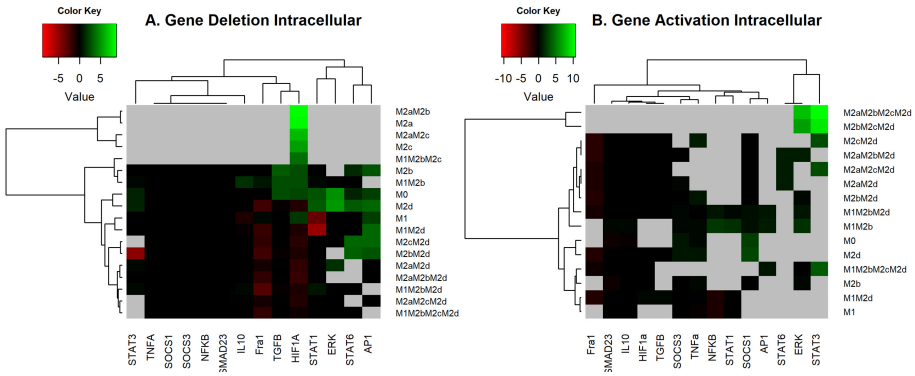


FIGURE 3 | Heat Map of the overexpression and knock-out of transcriptional factors. **(A)** Heat Map of the overexpression of transcriptional factors of macrophage polarization. We maintained the expression of the node as 1, simulated, and reviewed the attractors obtained. All overexpression was compared with the wild-type (original network) with log2 fold change. **(B)** Heat Map of Knock-outs of transcriptional factors of macrophage polarization. The nodes were permanently fixed with a value of 0; the perturbations were realized one by one until the attractors were reached. We compared the phenotypes of the wild-type with the perturbations by a log-fold change. Green and red regions indicate those attractors whose size of the basin of attraction increased or decreased after the perturbation. In black, we denote those attractors with few effects in the basin attraction size versus WT and after perturbation. Gray areas indicate those attractors that exist in the wild-type but do not remain perturbed.

presence of HIF1-α in a tumor microenvironment. Also, TGF-β had the same behavior as HIF1-α (**Figures S5, S6**, respectively), so cells should avert its secretion to revert cancer proliferation. Nevertheless, we hypothesize that the modulation of the balance to anti-tumoral phenotypes could be sufficient to diminish TGF-β.

Conversely, constitutive expressions of NFκB or STAT1 kept M1M2b and M1M2d phenotypes, respectively. Both hybrid phenotypes with well-known tumoricidal capacity. Interestingly, the overexpression of NFκB and STAT1 could be a potential therapeutic strategy because they promote phenotypes that could balance the behavior of macrophages in a tumor microenvironment, favoring tumoricidal capacity (**Figure 3A**).

On the other hand, we found an imbalance in the size of the basin attractors when we knocked out some transcriptional factor expressions separately (**Figure 3B**). Our *in silico* analysis allowed us to conclude that if we knocked out the expression of Fra-1, the M1 and the hybrid M1M2b produced a slight augment in their size of the basin of attraction, implementing a tumoral eradication capacity. Also, we observed that the lack of Fra-1 did not influence the development of the M2d phenotype. Turning off HIF1-α increased phenotypes associated with tumoricidal capacity (M1 and M1M2b), denoting the importance of not expressing these transcriptional factors for eliminating tumor cells (**Figure S7**). However, only affecting HIF1-α activity is insufficient because it enhances pro-tumoral phenotypes at 36% for the wild type macrophage (**Figures 3B** and **S8**). Thus, we had a typical host defense by activating the gamma receptors and liberating IFN-γ to the microenvironment. Notably, unlike gene activation analysis, we observed the emergence of some pure phenotypes, M2a, and M2c, tending to favor tumor progression when we knocked out HIF1-α.

Furthermore, we noted that knock-out of STAT3 decrease the proportion of phenotypes specific for tumor eradication and hybrid with anti-tumoral behavior at 33% (**Figure S7**). Inactivating STAT3 diminished 63% of pro-tumoral phenotypes involved in the regulation of the immune system (M2c) by activating NFκB (**Figure S8**). STAT6 knock-out does not affect anti-tumoral phenotypes (**Figure S7**), but it decreases the attraction basin of phenotypes. M2c phenotype releases IL-10 into the microenvironment creating a scenario for tumor evasion and enhancing tumor metastasis. In agreement with these findings, biological evidence suggests that the inactivation of STAT3 and STAT6 transcriptional factors are associated with reducing tumor growth and metastasis in a model of breast and lung cancer (31, 34).

Macrophage Polarization Develops Feedback With Microenvironments

Tumor microenvironments shape the polarization of macrophages. To evaluate how the microenvironmental signals alter this process, we analyzed six different signaling environments associated with macrophage phenotypes: M0, M1, M2a, M2b, M2c, and M2d states. We simulated signaling microenvironments activating permanently known profiles of cytokines and signaling metabolites associated with each macrophage phenotype (see **Table 3**) (33, 35).

Figure 4 depicts the logarithm of the basin of attraction of the attractors obtained in said microenvironments. Gray areas indicate that the phenotype was not present in the specific microenvironment. Notwithstanding that the Boolean model is a simplified analysis, we highlight a global behavior over how the macrophage phenotype is shaped. For instance, the monocyte microenvironment (Pro-M0) tends to induce three phenotypes at a low rate: M0, M1, and M2d. Fra1 may give a possible explanation underlying the increment in the size of the basin of attraction for M2d. The upregulation of Fra1 ultimately activates NFκB, which in turn will induce the M1 macrophage phenotype.

TABLE 3 | Microenvironments associated with specific macrophage subtypes.

Macrophage Phenotype	Signals
M0	No external stimuli
M1	IFN-γ and IFN-β
M2a	IL4 and TGF-β
M2b	IgG and glucocorticoids
M2c	IL10 and IL6 and MCSF
M2d	Adenosines and hypoxia and glucocorticoids

On the right side of the table, we have the phenotypes and the nodes that were kept on during the simulation until an attractor was reached.
Simulation of microenvironmental cues for the six macrophage phenotypes modeled in our gene regulatory network.

Furthermore, our computational analysis suggests that the Pro-M1 microenvironment diminishes regulatory and wound-healing phenotypes and enhances tumoricidal capacity (**Figure 4**). In counterpart, Pro-M2b microenvironment decreases tumoricidal capacity, favoring macrophages phenotype to tumor progression and immune regulation. This microenvironment only induces phenotypes either hybrid or simple associated with M2b. On the other hand, M2c microenvironments maintain hybrid phenotypes associated with tumor proliferation and wound-healing components, like angiogenesis. Besides, we noted that the regulatory cytokines like IL-10 and IL-6 in the microenvironment Pro-M2c diminished all possible cytotoxic hybrid activity in macrophages. This finding can explain cytotoxic activity in a pro-tumor microenvironment (36). However, the M2 population was preferred over M1 because they had specific microenvironment components (Pro-M2c) to maintain this imbalance (**Figure 4**). Furthermore, the M2d microenvironment does not generate any macrophage associated with tumoricidal capacity; it only activates macrophages implicated in wound-healing (angiogenesis and tissue recovery) and phenotypes regulating the actions of the immune system in eliminating tumors.

Lastly, macrophage phenotypes were susceptible to local microenvironments inside a tumor. These microenvironments can induce heterogeneity over the composition enhancing specific subtypes of macrophages. We believe that these phenotypes will increase their stability due to the continuum expression of their external signal. However, anti-tumoral behavior occurs even in the activation of pro-tumoral behavior by external signals but with lower stability. This type of modeling scheme may help develop therapeutic strategies to ensure stability despite the microenvironment acting against the treatment of a specific cancer type (37, 38).

The Importance of Molecular Components in Determining Cell Fate Macrophages

We analyzed each attractor's stability previously obtained through a gene perturbation analysis to evaluate the plasticity, closeness, and possible transition between the macrophage subtypes. As shown previously, these sets of analyses are beneficial to describe and uncover the network's global properties. We obtained a global landscape of the possible transitions among them (**Figure 5**). To accomplish this, we altered each node's state in the network. If the value of the node was 1, we changed to 0 and vice versa. Then we evaluated if this modification affected the stability of the phenotype in the network. Notably, we observed that only some transitions were allowed among attractors; besides, these transitions depended on different genetic alterations. As expected, phenotypes with the

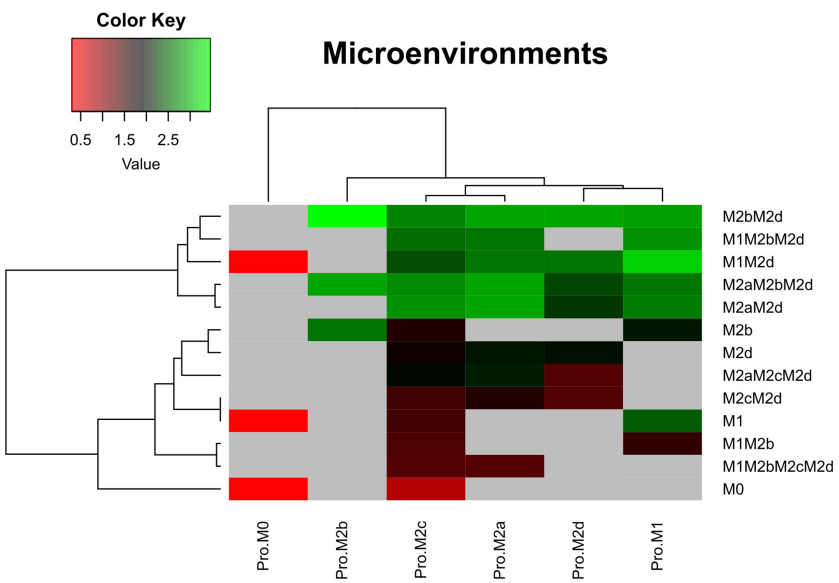


FIGURE 4 | Heat maps of the microenvironments of macrophage polarization. Microenvironments associated with the six phenotypes evaluated in this work. For these simulations we used the criteria of **Table 2**. Once the attractors were obtained, we applied a logarithmic transformation on the size of the basin of attraction for each phenotype. Red stands for a low basin of attraction and green for a high basin of attraction. Pro means we modelled the polarization in a microenvironment adjuvant for each macrophage subtype. M0, monocytes.

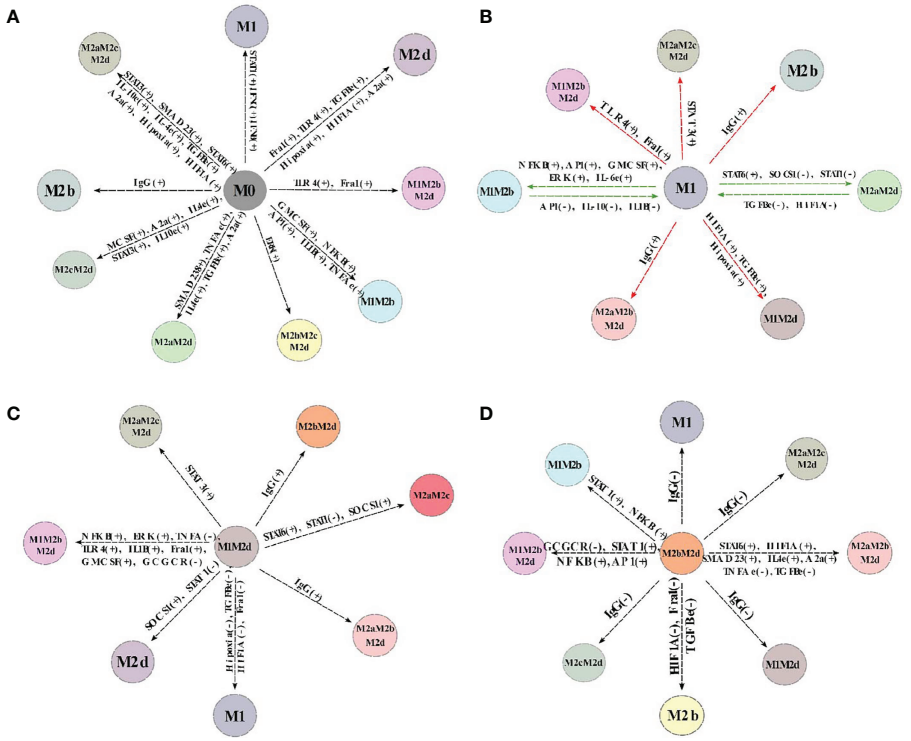


FIGURE 5 | Cell fate map of the macrophage polarization. Of all the attractors obtained, we changed the node's value and maintained this perturbation until an attractor was reached. If the attractor's transition to another phenotype, we represent it with a line and the new phenotype obtained by the perturbation. Plus sign (+) means the node was turned off, and we turn it on, while minus sign (-) means the perturbation was on, and we turned off. **(A)** Cell fate map of monocyte (M0). **(B)** Cell fate map of M1 macrophage. **(C)** Cell fate map of M1M2d macrophage. **(D)** Cell fate map of M2bM2d macrophage. Colors represent different states of macrophage polarization.

largest basin of attraction were more stable to perturbations (**Figure S4**). This study allowed us to conclude that M0 (monocyte) had a low stability, and it was prone to become other macrophage phenotypes under perturbations (**Figure 5A**). For example, M1 can differentiate from M0 through the activation of STAT1, IFN-γ, or IFN-β, but M1 can re-polarize to M0 (monocyte) activating the inhibitory function of SOCS1. These polarizations were irreversible, which means once they shift to the new phenotype, they cannot shift back to monocytes (**Figure S9**).

On the other hand, M2b transitioned from M0 by activating immunoglobulin G (IgG). B cells liberate IgG and can be cancer-derived (39, 40), so the M2b phenotype can be induced directly from monocytes and not necessarily has to pass first from a TLR4 activated macrophage and then shift to M2b (7). The transition from M0 to M2d was obtained by activating HIF1α or TGF-β or adenosines leading to the inactivation of any transcriptional factor of cytotoxic behavior; this transition was reversible (**Figure S9**). Notably, our simulations allowed us to postulate that the more complex phenotypes, the hybrids, can be obtained from M0. For instance, the tumoricidal/regulatory macrophage given by M1M2b obtained by activating AP1, NFκB, and IL1-β, had an, irreversible transition.

Another interesting observation is that the cytotoxic macrophage named M1 was stable for most perturbations (**Figure 5B**). For example, when we turn on AP1, the M1 phenotype becomes a hybrid phenotype (M1M2b) that maintains its cytotoxic capacity combined with a regulatory behavior. This shift is reversible macrophages can return to an M1 phenotype by inactivating AP1 (**Figure S9**). Furthermore, M2b attained from M1 by activating IgG, stating the importance of the immune complexes for selecting this trajectory. It does not need to be first a TLR4 activated macrophage to promote the transition to the M2b phenotype (7). Hybrid phenotype M1M2d (**Figure 5C**) can favor tumor clearance and recuperation of tissue. Based on our model, we observed this phenotype constrain the macrophage into a hypoxic condition, permanently activating TGF-β. The M1 phenotype can be obtained from M1M2d, turning off hypoxia or TGF-β, but both extracellular conditions are present most of the time in a tumor microenvironment. In this context, pursuing a hybrid phenotype with theoretical benefits such as M1M2d seems to be a better strategy than the induction of the M1 phenotype in a tumor microenvironment. We also observed that M1M2d, a tumoricidal/regulatory state, can shift to the M2d macrophage stage. This polarization is reached by turning on SOCS1

(inhibiting STAT1) or turning off STAT1 (M1 transcriptional factor). Notably, perturbation analysis suggests the emergence of a new phenotype from M1M2d; a hybrid labeled M2aM2c with an irreversible phenotype (**Figure S9**). M2aM2c has a regulatory/pro-tumoral behavior favoring tumor growth and tumor evasion, and it is promoted by turning on STAT6 from M1M2d. M2bM2d has more stability compared with other phenotypes (**Figure 5D**). IgG is a crucial factor that dictates this phenotype's behavior, so turning off this node, the hybrid phenotype can shift irreversibly to various tumoricidal macrophage stages, like M1 (**Figure S9**). M2b phenotype derived from M2bM2d by turning off HIF1α and the expression of ERK, which permits the secretion of IL-6, a pleiotropic cytokine and immune regulator of IL-10. Finally, turning on STAT1 or NFκB was sufficient to transit to a tumoricidal/regulatory phenotype from M2bM2d, labeled as M1M2b.

Lastly, there is growing evidence that plasticity is a property in the immune system's response altogether, here we supplied evidence that the response of macrophages is not the exception. Even though macrophages polarize depending on external factors, they can be manipulated to influence the outcome of their external signals. This analysis allowed us to identify potential genetic control points like NFκB and HIF1α, which could serve as potential molecular targets against cancer by modifying the macrophage phenotypes.

HIF1-α and NFκB as Potential Transcriptional Factors for a Theoretical Treatment Based Approach

Despite tumor microenvironment complexity and variability, their modulation is an appealing strategy to reduce the cancer cell phenotype (41). For instance, in our model, the most relevant activated transcription factors for the polarization to M1 phenotype and M1 hybrid are STAT1 and NFκB; these findings agree with a previous report (42). On the contrary, activation of HIF1-α is associated with the reduction of anti-tumoral macrophages by enhancing pro-tumoral macrophages. We concluded that the inactivation of HIF1-α and activation of NFκB or STAT1 might be critical to shift the balance to an anti-tumoral microenvironment rather than a pro-tumoral.

To verify these hypotheses, we designed two theoretical genetically modified macrophages (TGEM): HIF1-α =0 & NFκB =1 and HIF1-α =0 & STAT1 = 1. Dynamic analysis constrained by the inactivation of HIF1-α and activation of NFκB showed a significant reduction in the number of attractors, 4096, compared to the 10430 obtained initially. The HIF1-α =0 & STAT1 = 1 (STGEM) diminished the number of attractors (3840) distributed in four phenotypes (**Figure S10**). On the contrary, HIF1-α =0 & NFκB =1 had only two phenotypes (**Figure 6A**).

Interestingly, all the attractors belonged to two phenotypes, M1 and M1M2d hybrid. The M1M2d hybrid has a cytotoxic action, but the M2d counterpart diminishes the damage caused its cytotoxic function. M1 was the more stable due to its basin of attraction, while M1M2d was the least stable. We developed a cell fate map for our TGEM to acknowledge which genes can transit from one phenotype to another. Remarkably, we identified a reversible transition between M1 and M1M2d, mainly driven by Fra-1 and toll-like receptor 4 (TLR4) (**Figure 6B**). This means that our TGEM will cycle between these two phenotypes without developing new phenotypes; both are resistant to perturbations.

Our genetically modified macrophage with STAT1 activated (STGEM) developed four phenotypes. All of them had a hybrid state where the cytotoxic action was always present

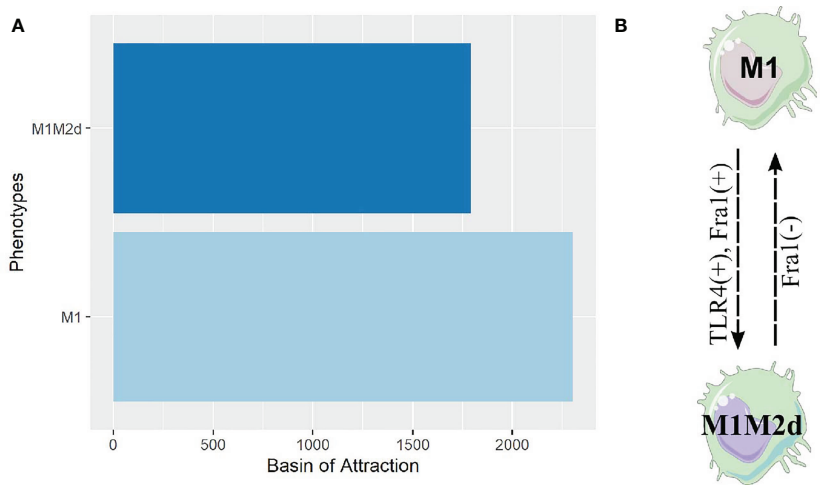


FIGURE 6 | Attractors and cell fate map of our TGEM. **(A)** Bar plot of the attractors obtained from our theoretical genetically modified macrophage. For this analysis we set the value of NFκB to 1 and HIF1-α to 0, and simulated until we obtained these attractors. **(B)** Cell fate map of our theoretical genetically modified macrophage. By analyzing the plasticity of phenotypes through single gene perturbation of the genetically modified macrophage, we obtained the rules of genetic perturbation that contribute transition between macrophages phenotypes. Here (-) means the node was turned off and (+) means the node was turned on.

(**Figure S10A**). M1M2b was the most stable phenotype, while M1M2d as well as the least stable. The cell fate map of STGEM is more complicated than TGEM. The transition from M1 to M1M2b marked by STAT3 and ERK activation. STAT3 and ERK are activated by common interleukins in a tumor microenvironment like IL-10, TNF-α, and IL1β. The rest of the map is depicted in **Figure S10B**, it showed that the four phenotypes cycle between them without any new phenotype. Based on this information, STGEM may be used as a strategy for eliminating tumor cells. Due to the nature and simplicity of the phenotypes obtained in TGEM, it can be an optimal option for an immunotherapeutic strategy to modulate the tumor microenvironment to eliminate tumor cells.

Finally, we concluded that our *in silico* TGEM (conceptually defined by the permanent activation and inactivation of NFκB and HIF1-α, respectively) induced phenotypes against cancer cells in the tumor microenvironment. We predicted balanced M1 characteristics that liberate all the components involved in the elimination phase of cancer immunosurveillance and the substances to resolve inflammation and tissue damage caused by cancer. Therefore, generating an M1/M2 ratio with a good prognosis. Furthermore, the phenotypes coming from our TGEM were robust, only certain transcription factors could redirect the polarization to another phenotype (**Figure 6B**).

Theoretical Genetically Modified Macrophage Resisted Perturbations in a Breast Cancer Microenvironment

Given the experimental evidence and previous results of the intricate and determining relationship of the microenvironments in macrophage polarization, we developed an *in-silico* microenvironment based on experimental evidence. We decided to simulate the alterations in the microenvironment of our previous TGEM. Microenvironment alterations are associated with adverse prognosis in other studies (43). We defined the state (active or inactive) of all the extracellular nodes present at the cancer microenvironment using cancer multi-omics data (44). We calculated the differential expressed genes between breast cancer cells and healthy tissue to estimate which extracellular nodes are highly present and then fixed the state of all extracellular nodes in our network. In the particular case of components such as immunoglobulin G (39), adenosines (45), glucocorticoids (46), and under hypoxia (47). All these are activated (node state:one) due to previous reports supporting their activity in samples of patients with breast cancer. Overall, we simulated our theoretical genetically modified macrophage (TGEM) in four different simplified breast cancer microenvironments to evaluate its behavior and determine if this pharmaceutical approach could be suitable. Under this microenvironment, we analyzed the effect of IgG and A2a; IL10 and TGF-β; IL-1β and IL 6; Hypoxia and glucocorticoids.

First, we evaluated the behavior modification of our TGEM in specific breast cancer microenvironments according to previous conditions. The phenotypes obtained changed globally, we found two new phenotypes: M1M2b and M1M2bM2d. Both phenotypes have the M1 function creating a cytotoxic/regulatory microenvironment (**Figure 7A**). The most complicated microenvironment (this microenvironment is implicated with metastasis) was the one with the presence of A2a and Ig G. We observed only two phenotypes: M1M2b and M1M2bM2d, while M1 and M1M2d were absent. This analysis suggested that IgG & A2a could inhibit functions of the M1 macrophage through the inactivation of TLR4 and consequently the activation of NFκB. A2a can inhibit NFκB as well; both components created a complicated microenvironment for the development of more M1 type macrophages.

Nevertheless, the phenotypes generated in this microenvironment can help eliminate tumor cells and balance the microenvironment to a more suitable one to create anti-tumoral macrophages. The remaining microenvironments develop four macrophage phenotypes (M1, M1M2b, M1M2bM2d, and M1M2d) with different proportions each. Therefore, we found a high proportion of pure M1 and M1 hybrid to shift the balance towards a tumor eliminating microenvironment, and having the regulatory components not cause more damage from the cytotoxic activity (**Figure 7A**).

In **Figures 7B–E**, we depict the cell fate map of our TGEM under the four breast cancer microenvironments evaluated. Overall, once STAT6, is activated, induce an irresistible into the phenotypes by adding the M2a components. Even though in the IgG & A2a environment, we did not have the M1 phenotypes, they can be irreversibly obtained by inhibiting the function of Erk in the M1M2b phenotypes. The M1M2b and M1M2bM2d phenotypes create feedback between them by activating Fra1 or TLR4 to transit from M1M2b to M1M2bM2d; this transition can be reversible by inhibiting the expression of Fra1. This complicated microenvironment in our TGEM may be an obstacle. However, from the phenotypes obtained and the one (M1) obtained, our TGEM may still eliminate tumor cells beside the microenvironment (**Figure 7B**).

In IL1β & IL6 microenvironment, M1 and M12b developed feedback between them where the transitions are reversible. This feedback may create a balance in the environment where the immune system can eliminate tumor cells and recuperate the damaged tissue caused by the cytotoxic functions (**Figure 7C**). Under hypoxic conditions (**Figure 7D**), our TGEM still develops macrophages with the capacity to eliminate tumor cells. The lack of oxygen does not affect the feedback between M1-M1M2b and M1M2d-M1M2bM2d. In this hypoxic microenvironment, the activation of STAT6 may serve as an adaptation agent. STAT6 is associated with the regulation of the lipid metabolism, our macrophages can obtain energy from lipids to liberate cytotoxic interleukins and eliminate tumor cells. From a theoretical perspective, STAT6 activation in our TGEM in a hypoxic scenario is not one we want to avoid (48, 49). Finally, in **Figure 7E**, a microenvironment with regulatory activity in macrophage cytotoxic functions does not affect creating phenotypes to eliminate tumor cells. M1 and M1M2b develop the same feedback as the previous microenvironment. M1M2bM2d and M1M2d create feedback between them; therefore, the transitions are reversible.

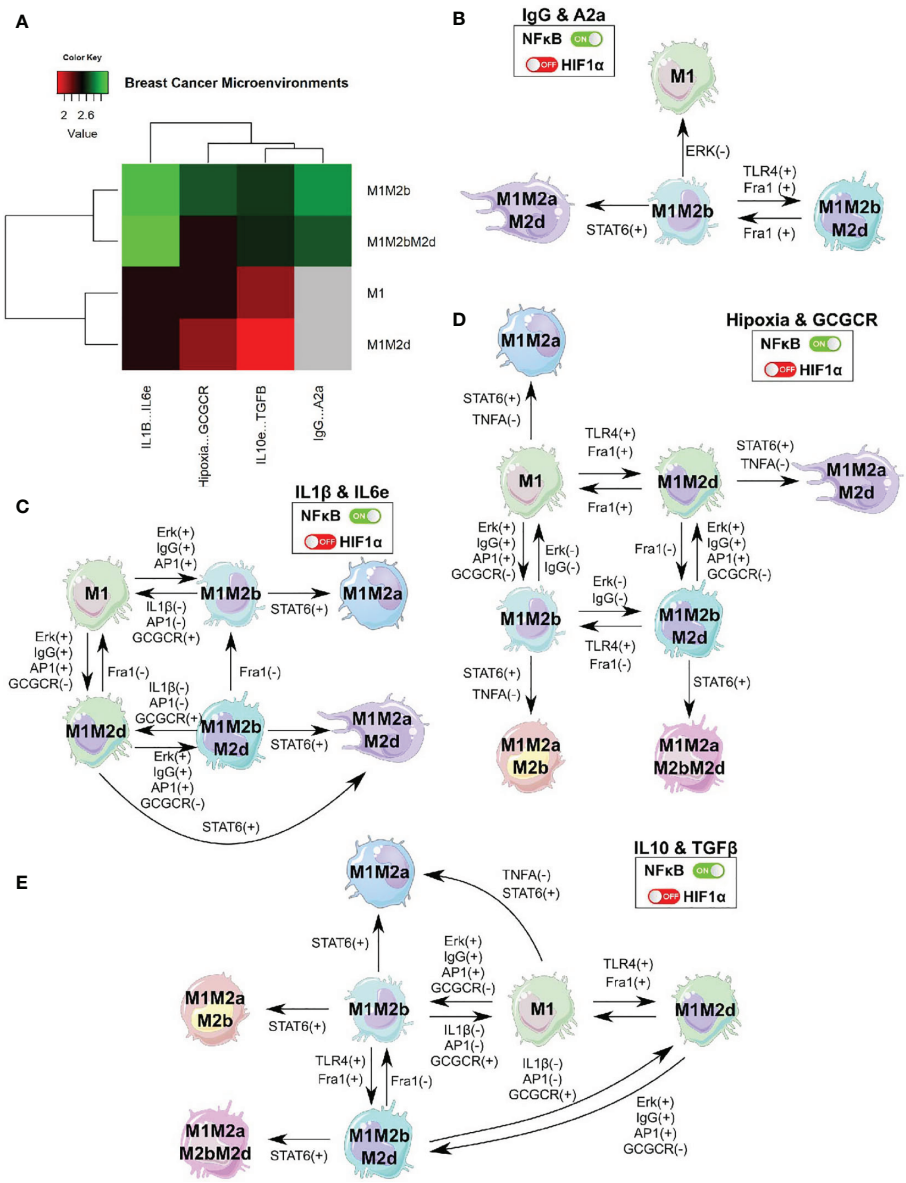


FIGURE 7 | TGEM in a breast cancer microenvironment. **(A)** We engineered different microenvironments associated with breast cancer and evaluate how our TGEM behaved. **(B)** Cell fate map of the theoretical genetically modified macrophage in breast cancer microenvironment for the expression of IgG and adenosines. **(C)** Cell fate map of the theoretical genetically modified macrophage in breast cancer microenvironment for the expression of IL1-β and IL-6. **(D)** Cell fate map of the theoretical genetically modified macrophage in breast cancer microenvironment for the expression of Hypoxia and glucocorticoids (GCGCR). **(E)** Cell fate map of the theoretical genetically modified macrophage in breast cancer microenvironment for the expression of IL-10 and TGF-β. This analysis was to evaluate the stability of our pharmaceutical approach in a breast cancer scenario.

In summary, the Boolean modeling of the TRN mimicking the feasible space of macrophage phenotypes serves as a platform to create hypotheses of the control mechanism that promotes cancer phenotype. We postulate that TGEM associated with pro-inflammatory is a promising pharmaceutical approach because it is robust to permanent perturbation in a breast cancer microenvironment. This hypothesis has to be experimentally proven, an aim that constitutes a perspective of this paper.

Robustness and Sensitivity Analysis of Our Transcriptional Regulatory Network of Macrophage Polarization in a Tumor Microenvironment

A robust gene regulatory network is resistant to perturbations in the network. We defined robustness as a characteristic of behavior invariance, according to changes in their inner variables. Nevertheless, gene regulatory networks or any

biological network representing the phenotypic behavior of cells have to gain an ability to respond appropriately to external stimuli and filter molecular perturbations.

A mathematical representation of a gene regulatory network can be dynamically classified as ordered, critical, or chaotic regime. Ordered networks are not affected by filter external stimuli. Critical networks are intermediate from an ordered and chaotic behaviors, they are resilient to a specific subset of perturbations. Finally, chaotic networks are severely affected by perturbations, inducing a plethora of responses (50). To evaluate the dynamical behavior of macrophage polarization in our TRN we obtained the Derrida curve, see **Figure 8A**. To determine in what extend the slope of the curve around x=0 differ from identity line and thus classify the dynamic behavior of our network, we applied a chi-squared Goodness of fit test with the coordinates of the average Hamming values before the intersection point between both curves. We consider the observed values as those average Hamming distances depicted in the y-axis of the plot. Simultaneously, we considered the expected values as the average Hamming distances outlined in the identity line (y=x). Despite our models seeming to draw a curve slightly above the identity line around x=0, our statistical test allowed us to conclude that there is no significant difference in the Hamming distance between the observed and the expected Hamming distances (exact multinomial test of goodness-of-fit, p-value=0.5804, statistical significance= 0.05). Given that there was not a statistical difference between expected and observed values, we supplied evidence that the dynamic behavior of our network falls into the critical region. There is evidence that criticality is a trademark for living systems for simultaneously co-exist in robustness and evolvability capacities (51, 52). This finding suggests that TRN for macrophage polarization in the tumor microenvironment seems to be robust to perturbations and permits evolve according to the environmental cues.

To evaluate the sensitivity in our TRN of macrophage polarization, we modeled multiple perturbations of each update rule. The sensitivity analysis of a gene or node evaluates the influence of the other genes of our network. **Figure 8B** outlines the average sensitivity of every node and the entire network (golden line). TGF-β and IFN-γ are the nodes with higher sensitivity, followed by glucocorticoid receptor (GCGCR), interleukin 4 (IL-4), and suppressor of cytokine signaling 1 (SOCS1). Consequently, the extracellular cytokines more susceptible to stimuli or molecular noise. For example, IFN-γ is associated with an anti-tumoral microenvironment and a cytotoxic M1 macrophage function. Contrary, TGF-β is associated with pro-tumoral microenvironment and a regulatory wound healing M2 phenotypes. TGF-β is associated with inhibiting M1 functions and enhance tumor proliferation. GCGCR and IL-4 are associated with regulatory and wound healing behavior as well. SOCS1 can inhibit STAT1 and diminish the expression of IFN-γ in the microenvironment. NFκB, STAT1, and STAT3 have lower average sensitivity, thus, they are robust to transient changes. Therefore, we acknowledged NFκB and STAT1 as a therapeutic strategy for immunotherapy in tumor eradication. The network's average sensitivity is lower than one, hence our network behaves orderly (**Figure 8A**).

DISCUSSION

The interaction between the immune system, microenvironment, and cancer is one appealing topic to design effective treatments. Moreover, the development of computational approaches that contribute to clarify their mechanisms is a needed task. By and extensive Boolean analysis, we presented a high quality curated signaling regulatory network between cancer-derived factors and macrophages. Unlike previous reports, we present a model with

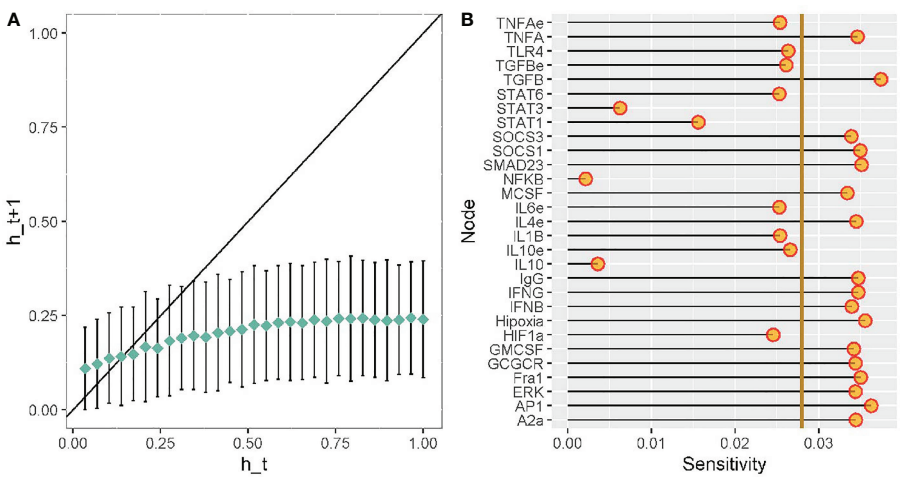


FIGURE 8 | Robustness analysis of our transcriptomic regulatory network of macrophage polarization. Sensitivity analysis of each node from our transcriptional regulatory network of macrophage polarization. **(A)** Derrida curves of our transcriptional regulatory network of macrophage polarization in a tumor microenvironment. **(B)** Sensitivity analysis of the variables in our transcriptional regulatory network of macrophage polarization in a tumor microenvironment. Golden line is the mean of the value of the sensitivity analysis.

additional interactions experimentally validated and an in-depth analysis of the macrophages' hybrid phenotypes. As Palma et al. previously concluded, these intermediary steady-states support the hypothesis that macrophage polarization is a continuum process and not binary, as usually suggested. Furthermore, our extended analysis allowed us to identify phenotypes that suggested new physiological interpretations associated with hybrid macrophage phenotypes. Based on our results, we concluded that the hybrid phenotypes evolve constrained by a permanent interaction with the environment. We postulate that some of them have therapeutic implications by enhancing tumoricidal capacity (M1M2d and M1M2b) or promoting regulatory mechanisms (M2bM2d, M2aM2d, and M2cM2d) against cancer.

Notably, our analysis allowed us to build hypotheses towards the hybrid states that favor or contrast cancer phenotype. We concluded that M1M2d, a tumoricidal-regulatory macrophage, could potentially eliminate tumor cells due to the secretion of cytotoxic cytokines and IL-12, the latest helps to differentiate CD4+ T cells. As we know, M1 is an essential player in host defense, but if it is not regulated, it could cause tissue damage. Nonetheless, the M2d part of the hybrid phenotype will secrete IL-10 and TGF-β. These regulatory cytokines would eventually maintain at bay the action caused by the cytotoxic cytokines. The M1M2d phenotype would induce a cytotoxic/regulatory cytokine ratio that could eradicate tumor cells and avoid tissue damage. Furthermore, M1M2b has a similar function as the M1M2d. It would regulate with IL-10 the damage caused by the cytotoxic actions. However, it would eventually eliminate tumor cells through the action of M1 and heal the damage caused by the elimination of tumor cells aside from the action of M2b.

Contrastingly, M2bM2d, M2aM2d, and M2cM2d are hybrid states with a malignant phenotype that express HIF1α. When a tumor grows and increases its diameter, the oxygen supply becomes insufficient in inner regions, creating hypoxic or necrotic areas (47). Hypoxia response *via* HIF1α not only affects macrophage polarization (53) also recruits macrophages and mesenchymal stromal cells to these regions (54, 55). All three macrophages phenotypes have different actions on tumor progression. For example, M2bM2d would favor the angiogenic process with the contribution of vascular endothelial growth factor (VEGF-A).

Meanwhile, the M2aM2d hybrid would be the most dangerous phenotype because it could heal the "wound" caused by tumor growth. Moreover, M2aM2d would favor tumor angiogenesis by secreting PDGF, TGF-β, IL-8, CXCL12, and VEGF-A, thus contributing to tumor angiogenesis metastasis. A shred of additional experimental evidence describes this hypoxia adaptation is given by M2a, which inhibits T cell expansion, reducing tumor clearance (56). Finally, M2cM2d and other malignant hybrids would be spatially constrained into a hypoxic area, behaving as regulators secreting IL-10 and TGF-β, creating a tumor proliferation scenario progression. Hypothetically, the inhibition of IL-10 or the coactivation of CD40, IL-12, IL-8, and TNF-α, could repolarize these macrophages to the M1 phenotype, where reverse tumor development has been demonstrated (57). Overall, we postulate that the behavior of these macrophages hybrids is malignant, given their characteristics they could promote carcinogenesis and regulate the immune system in favor of tumor phenotype (58, 59).

Recent advances have shed the different cells that constitute specialized niches and the mechanisms that promote cell-to-cell interaction. The knowledge of the cellular content and diversity of the tumor microenvironment in malignant transformation and other metastatic diseases is relevant. Most human tissues, including the breast, sustain their continuous replenishment from primary stem cells. The microenvironment that maintains homeostasis promotes cell differentiation according to functional demands and suppresses aberrant cells' potential emergence. The interaction can be direct or bystander between malignant and non-malignant macrophages presumably powerfully influence the disease outcome, and cytokines secretion by macrophages is a fundamental factor for epithelial to mesenchymal transition (EMT) in breast cancer. Whether they promote retention of primitive cells within their niches to avoid mobilization into bloodstream or external tissues, as occurs in bone marrow, stills a matter in question (57, 60). Even though it is beyond this paper's scope, it is becoming clearer the critical role of the M1/M2 macrophage polarization in the malignant progression of TNBC. These malignant macrophage hybrids are mostly going to express type II cytokines, promoting cancer stems cells (CSC) growth. M2 macrophages protect CSC from the immune surveillance mechanisms and induce anti-inflammatory microenvironments that plays as onco-promoters. Additionally, M2 macrophage contribute to drug resistance, particularly at the late stages of tumorigenesis (57).

In our TGEM approach, we mutated NFκB and HIF1α. NFκB is a master regulator with high activation and translocation efficiency. Based on our results, it is a promising key player to eliminate tumor cells. Having vital roles in the macrophage function. The nanoparticle approach develops a shift from an M2 macrophage type to M1 (61), very similar to our TGEM approach; nevertheless, only targeting NFκB is not sufficient. Hence, we knocked out HIF1α, because this factor is affected by hypoxia, a constitutive condition in a tumor microenvironment, that triggers the M2 polarizing effect (62). In our TGEM, we obtained three hybrid phenotypes all with M1 combined with a regulatory behavior macrophage. Contrastingly under the knockout of HIF1α, the macrophages still going to migrate to hypoxic regions but not for the same reason described earlier. This time, the spatial accumulation of macrophages in the hypoxic areas is due to their ability to scavenge apoptotic cells. Additionally, the macrophages activate the secretion of IL-12 and IFN-γ generating positive feedback with CD4+ T cells that will terminate theoretically in tumor clearance. By this matter, hypoxia can no longer influence our TGEM. It will not be immobilized in hypoxia regions, contributing to an adverse prognosis in breast cancer.

Even though this approach shows good macrophage behavior in a tumor microenvironment, we must take care of the Boolean model's limitations. There is no grading scale in the associated factors, because Boolean models only consider 0 or 1 values. So, we cannot evaluate the amount of a transcription factor is

sufficient to obtain a specific macrophage subtype. As we had proven, this reconstruction includes the necessary information to explore and supply evidence of the macrophage polarization transformation. However, we should update the components and their Boolean rules (**Figure 1**), as more experimental evidence will be appearing, and draw a more accurate description of this phenomenon. Finally, we considered the M0 phenotype as a monocyte, which is a precursor of macrophage activation. However, this type of cell has heterogeneity that could be manipulated differently in cancer cells (63). We believe that the monocyte continuum or reprogramming by cancer cells can be an essential variable in determining the macrophage continuum's behavior. This behavior should be studied for understanding completely macrophage polarization and the complexity in between. Alternatively considering the heterogeneity of the myeloid (monocytes and neutrophils) population set an appealing future perspective.

Here, we aimed to understand the molecular and external mechanisms that orchestrate macrophage polarization and develop potential therapeutic strategies. Several strategies focus on diminishing the recruitment of macrophages to the tumor site, showing the limitation of tumor vascularization and metastasis in mouse models (64). However, these treatments only reduce tumor growth by reducing the M2 cells (65), losing the beneficial properties of M1 type macrophages. We showed that our designed macrophage diminished malignant hybrid phenotypes and adapted to the perturbations caused by the tumor microenvironment. Our theoretical approach follows a combined strategy, which we believe will be the key to design potential therapies with macrophages. Our theoretical model also suggested that the best approach to defeat cancer is not shifting M2 to M1, as most pharmaceutical methods do. This shift is easily lost because it is affected by the tumor microenvironment (66). Instead, an optimal therapeutic strategy could be highlighting the best characteristics in the pro and anti-inflammatory scenarios. W concluded that the simultaneous action of HIF1α and NFκB allows us achieving "control" over the influence of hypoxia and the cytotoxic behavior of macrophages. Finally, our computational approaches may contribute to set the foundations of the macrophage population dynamics under the phases of the cancer immunoediting (elimination, equilibrium and escape), in the hallmarks of breast cancer. Future directions and remaining challenges in investigating transitional biology from immunosurveillance to suppressor macrophages will include high throughput genomic or cytometric analyses of M1 and M2 populations in breast cancer-associated tumor microenvironment. The understanding of the dynamic process from immunosurveillance to malignant progression may unravel the principles of the dual host-protective or -harmful roles of M1 and M2 macrophages in tumors immunoediting. In recap, computational modeling has become a crucial tool for paving novel avenues in immunotherapies capable to implement optimal strategies for fighting against highly invasive breast cancer.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**. All code used for analysis is available at https://github.com/resendislab/M1-M2-Macrophage-Polarization.

AUTHOR CONTRIBUTIONS

UA-P and OR-A conceived and designed the mathematical model. UA-P performed all computational analysis and analyzed the data. MM-G and AV-J helped with the analyses of the data. RP with OR-A supervised the mathematical model and the *in-silico* analysis. All authors contributed to the article and approved the submitted version

FUNDING

The authors thank the financial support from PAPIIT-UNAM (IA202720). UA-P received a doctoral fellowship from Consejo Nacional de Ciencia y Tecnología (CONACYT) (CVU: 774988).

ACKNOWLEDGMENTS

UA-P is a doctoral student from Programa de Doctorado en Ciencias Biológicas of the Universidad Nacional Autónoma de México (UNAM). This paper is part of the requirements to obtain the degree of doctor in said program. UA-P also received a fellowship (CVU: 774988) from Consejo Nacional de Ciencia y Tecnología (CONACYT). UA-P wants to thank Dr. Eduardo García Zepeda for the valuable contributions to this research work. Without him, this original research could not be possible; rest in peace.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2021.642842/full#supplementary-material

REFERENCES

1. Mills CD, Kincaid K, Alt JM, Heilman MJ, Hill AM. M-1/M-2 Macrophages and the Th1/Th2 Paradigm. *J Immunol* (2000) 164:6166–73. doi: 10.4049/jimmunol.164.12.6166
2. Mills CD. Anatomy of a Discovery: M1 and M2 Macrophages. *Front Immunol* (2015) 6:212. doi: 10.3389/fimmu.2015.00212
3. Rhee I. Diverse Macrophages Polarization in Tumor Microenvironment. *Arch Pharm Res* (2016) 39:1588–96. doi: 10.1007/s12272-016-0820-y
4. Biswas SK, Sica A, Lewis CE. Plasticity of Macrophage Function During Tumor Progression: Regulation by Distinct Molecular Mechanisms. *J Immunol* (2008) 180:2011–7. doi: 10.4049/jimmunol.180.4.2011
5. Van Dyken SJ, Locksley RM. Interleukin-4- and interleukin-13-mediated Alternatively Activated Macrophages: Roles in Homeostasis and Disease.

Annu Rev Immunol (2013) 31:317–43. doi: 10.1146/annurev-immunol-032712-095906
6. Wang L-X, Zhang S-X, Wu H-J, Rong X-L, Guo J. M2b Macrophage Polarization and its Roles in Diseases. *J Leukoc Biol* (2019) 106:345–58. doi: 10.1002/JLB.3RU1018-378RR
7. Martinez FO, Gordon S. The M1 and M2 Paradigm of Macrophage Activation: Time for Reassessment. *F1000Prime Rep* (2014) 6:13. doi: 10.12703/P6-13
8. Biswas SK, Gangi L, Paul S, Schioppa T, Saccani A, Sironi M, et al. A Distinct and Unique Transcriptional Program Expressed by Tumor-Associated Macrophages (Defective NF-kappaB and Enhanced IRF-3/STAT1 Activation). *Blood* (2006) 107:2112–22. doi: 10.1182/blood-2005-01-0428
9. Wang Q, Ni H, Lan L, Wei X, Xiang R, Wang Y. Fra-1 Protooncogene Regulates IL-6 Expression in Macrophages and Promotes the Generation of M2d Macrophages. *Cell Res* (2010) 20:701–12. doi: 10.1038/cr.2010.52
10. Duluc D, Delneste Y, Tan F, Moles M-P, Grimaud L, Lenoir J, et al. Tumor-Associated Leukemia Inhibitory Factor and IL-6 Skew Monocyte Differentiation Into Tumor-Associated Macrophage-Like Cells. *Blood* (2007) 110:4319–30. doi: 10.1182/blood-2007-02-072587
11. Hagemann T, Wilson J, Burke F, Kulbe H, Li NF, Plüddemann A, et al. Ovarian Cancer Cells Polarize Macrophages Toward A Tumor-Associated Phenotype. *J Immunol* (2006) 176:5023–32. doi: 10.4049/jimmunol.176.8.5023
12. Seager RJ, Hajal C, Spill F, Kamm RD, Zaman MH. Dynamic Interplay Between Tumour, Stroma and Immune System Can Drive or Prevent Tumour Progression. *Convergent Sci Phys Oncol* (2017) 3:034002. doi: 10.1088/2057-1739/aa7e86
13. Mosser DM, Edwards JP. Exploring the Full Spectrum of Macrophage Activation. *Nat Rev Immunol* (2008) 8:958–69. doi: 10.1038/nri2448
14. Kitano H. Systems Biology: A Brief Overview. *Science* (2002) 295:1662–4. doi: 10.1126/science.1069492
15. Wang R-S, Saadatpour A, Albert R. Boolean Modeling in Systems Biology: An Overview of Methodology and Applications. *Phys Biol* (2012) 9:055001. doi: 10.1088/1478-3975/9/5/055001
16. Xiao Y. A Tutorial on Analysis and Simulation of Boolean Gene Regulatory Network Models. *Curr Genomics* (2009) 10:511–25. doi: 10.2174/138920209789208237
17. Albert R, Othmer HG. The Topology of the Regulatory Interactions Predicts the Expression Pattern of the Segment Polarity Genes in Drosophila Melanogaster. *J Theor Biol* (2003) 223:1–18. doi: 10.1016/S0022-5193(03)00035-3
18. Mendoza L, Thieffry D, Alvarez-Buylla ER. Genetic Control of Flower Morphogenesis in Arabidopsis Thaliana: A Logical Analysis. *Bioinformatics* (1999) 15:593–606. doi: 10.1093/bioinformatics/15.7.593
19. Martinez-Sanchez ME, Mendoza L, Villarreal C, Alvarez-Buylla ER. A Minimal Regulatory Network of Extrinsic and Intrinsic Factors Recovers Observed Patterns of CD4+ T Cell Differentiation and Plasticity. *PloS Comput Biol* (2015) 11:e1004324. doi: 10.1371/journal.pcbi.1004324
20. Méndez A, Mendoza L. A Network Model to Describe the Terminal Differentiation of B Cells. *PloS Comput Biol* (2016) 12:e1004696. doi: 10.1371/journal.pcbi.1004696
21. Rex J, Albrecht U, Ehlting C, Thomas M, Zanger UM, Sawodny O, et al. Model-Based Characterization of Inflammatory Gene Expression Patterns of Activated Macrophages. *PLoS Comput Biol* (2016) 12:e1005018. doi: 10.1371/journal.pcbi.1005018
22. Palma A, Jarrah AS, Tieri P, Cesareni G, Castiglione F. Gene Regulatory Network Modeling of Macrophage Differentiation Corroborates the Continuum Hypothesis of Polarization States. *Front Physiol* (2018) 9:1659. doi: 10.3389/fphys.2018.01659
23. Schwab JD, Kestler HA. Automatic Screening for Perturbations in Boolean Networks. *Front Physiol* (2018) 9:431. doi: 10.3389/fphys.2018.00431
24. Müssel C, Hopfensitz M, Kestler HA. Boolnet—an R Package for Generation, Reconstruction and Analysis of Boolean Networks. *Bioinformatics* (2010) 26:1378–80. doi: 10.1093/bioinformatics/btq124
25. Kawasaki T, Kawai T. Toll-Like Receptor Signaling Pathways. *Front Immunol* (2014) 5:461. doi: 10.3389/fimmu.2014.00461
26. Malyshev I, Malyshev Y. Current Concept and Update of the Macrophage Plasticity Concept: Intracellular Mechanisms of Reprogramming and M3 Macrophage "Switch" Phenotype. *BioMed Res Int* (2015) 2015:341308. doi: 10.1155/2015/341308
27. Kratochvill F, Neale G, Haverkamp JM, Van de Velde L-A, Smith AM, Kawauchi D, et al. Tnf Counterbalances the Emergence of M2 Tumor Macrophages. *Cell Rep* (2015) 12:1902–14. doi: 10.1016/j.celrep.2015.08.033
28. Hutchins AP, Diez D, Miranda-Saavedra D. The IL-10/STAT3-mediated Anti-Inflammatory Response: Recent Developments and Future Challenges. *Brief Funct Genomics* (2013) 12:489–98. doi: 10.1093/bfgp/elt028
29. Duluc D, Corvaisier M, Blanchard S, Catala L, Descamps P, Gamelin E, et al. Interferon-γ Reverses the Immunosuppressive and Protumoral Properties and Prevents the Generation of Human Tumor-Associated Macrophages. *Int J Cancer* (2009) 125:367–73. doi: 10.1002/ijc.24401
30. Solinas G, Germano G, Mantovani A, Allavena P. Tumor-Associated Macrophages (TAM) as Major Players of the Cancer-Related Inflammation. *J Leukoc Biol* (2009) 86:1065–73. doi: 10.1189/jlb.0609385
31. Binnemars-Postma K, Bansal R, Storm G, Prakash J. Targeting the Stat6 Pathway in Tumor-Associated Macrophages Reduces Tumor Growth and Metastatic Niche Formation in Breast Cancer. *FASEB J* (2018) 32:969–78. doi: 10.1096/fj.201700629R
32. O'Carroll C, Fagan A, Shanahan F, Carmody RJ. Identification of a Unique Hybrid Macrophage-Polarization State Following Recovery From Lipopolysaccharide Tolerance. *J Immunol* (2014) 192:427–36. doi: 10.4049/jimmunol.1301722
33. Long KB, Collier AI, Beatty GL. Macrophages: Key Orchestrators of a Tumor Microenvironment Defined by Therapeutic Resistance. *Mol Immunol* (2019) 110:3–12. doi: 10.1016/j.molimm.2017.12.003
34. Iriki T, Ohnishi K, Fujiwara Y, Horlad H, Saito Y, Pan C, et al. The Cell-Cell Interaction Between Tumor-Associated Macrophages and Small Cell Lung Cancer Cells Is Involved in Tumor Progression Via STAT3 Activation. *Lung Cancer* (2017) 106:22–32. doi: 10.1016/j.lungcan.2017.01.003
35. Belli C, Trapani D, Viale G, D'Amico P, Duso BA, Della Vigna P, et al. Targeting the Microenvironment in Solid Tumors. *Cancer Treat Rev* (2018) 65:22–32. doi: 10.1016/j.ctrv.2018.02.004
36. Jackute J, Zemaitis M, Pranys D, Sitkauskiene B, Miliauskas S, Vaitkiene S, et al. Distribution of M1 and M2 Macrophages in Tumor Islets and Stroma in Relation to Prognosis of Non-Small Cell Lung Cancer. *BMC Immunol* (2018) 19:3. doi: 10.1186/s12865-018-0241-4
37. Jin M-Z, Jin W-L. The Updated Landscape of Tumor Microenvironment and Drug Repurposing. *Signal Transduct Target Ther* (2020) 5:166. doi: 10.1038/s41392-020-00280-x
38. Baghban R, Roshangar L, Jahanban-Esfahlan R, Seidi K, Ebrahimi-Kalan A, Jaymand M, et al. Tumor Microenvironment Complexity and Therapeutic Implications At A Glance. *Cell Commun Signaling* (2020) 59:1–19. doi: 10.1186/s12964-020-0530-4
39. Wang J, Lin D, Peng H, Huang Y, Huang J, Gu J. Cancer-Derived Immunoglobulin G Promotes Tumor Cell Growth and Proliferation Through Inducing Production of Reactive Oxygen Species. *Cell Death Dis* (2013) 4:e945. doi: 10.1038/cddis.2013.474
40. Hu D, Zheng H, Liu H, Li M, Ren W, Liao W, et al. Immunoglobulin Expression and Its Biological Significance in Cancer Cells. *Cell Mol Immunol* (2008) 5:319–24. doi: 10.1038/cmi.2008.39
41. Singh SR, Rameshwar P, Siegel P. Targeting Tumor Microenvironment in Cancer Therapy. *Cancer Lett* (2016) 380:203–4. doi: 10.1016/j.canlet.2016.04.009
42. Liu H, Shi B, Huang C-C, Eksarko P, Pope RM. Transcriptional Diversity During Monocyte to Macrophage Differentiation. *Immunol Lett* (2008) 117:70–80. doi: 10.1016/j.imlet.2007.12.012
43. Choi J, Gyamfi J, Jang H, Koo JS. The Role of Tumor-Associated Macrophage in Breast Cancer Biology. *Histol Histopathol* (2018) 33:133–45. doi: 10.14670/HH-11-916
44. Fernandez-Banet J, Esposito A, Coffin S, Horvath IB, Estrella H, Schefzick S, et al. OASIS: Web-Based Platform for Exploring Cancer Multi-Omics Data. *Nat Methods* (2016) 13:9–10. doi: 10.1038/nmeth.3692
45. Kumar V. Adenosine as an Endogenous Immunoregulator in Cancer Pathogenesis: Where to Go? *Purinergic Signal* (2013) 9:145–65. doi: 10.1007/s11302-012-9349-9
46. Obradović MMS, Hamelin B, Manevski N, Couto JP, Sethi A, Coissieux M-M, et al. Glucocorticoids Promote Breast Cancer Metastasis. *Nature* (2019) 567:540–4. doi: 10.1038/s41586-019-1019-4

47. Michiels C, Tellier C, Feron O. Cycling Hypoxia: A Key Feature of the Tumor Microenvironment. *Biochim Biophys Acta* (2016) 1866:76–86. doi: 10.1016/j.bbcan.2016.06.004
48. Szanto A, Balint BL, Nagy ZS, Barta E, Dezso B, Pap A, et al. STAT6 Transcription Factor Is a Facilitator of the Nuclear Receptor Pparγ-Regulated Gene Expression in Macrophages and Dendritic Cells. *Immunity* (2010) 33:699–712. doi: 10.1016/j.immuni.2010.11.009
49. Varga T, Czimmerer Z, Nagy L. Ppars are a Unique Set of Fatty Acid Regulated Transcription Factors Controlling Both Lipid Metabolism and Inflammation. *Biochim Biophys Acta* (2011) 1812:1007–22. doi: 10.1016/j.bbadis.2011.02.014
50. Lloyd-Price J, Gupta A, Ribeiro AS. Robustness and Information Propagation in Attractors of Random Boolean Networks. *PloS One* (2012) 7:e42018. doi: 10.1371/journal.pone.0042018
51. Balleza E, Alvarez-Buylla ER, Chaos A, Kauffman S, Shmulevich I, Aldana M. Critical Dynamics in Genetic Regulatory Networks: Examples From Four Kingdoms. *PLoS One* (2008) 3:e2456. doi: 10.1371/journal.pone.0002456
52. Aldana M, Balleza E, Kauffman S, Resendiz O. Robustness and Evolvability in Genetic Regulatory Networks. *J Theor Biol* (2007) 245:433–48. doi: 10.1016/j.jtbi.2006.10.027
53. Henze A-T, Mazzone M. The Impact of Hypoxia on Tumor-Associated Macrophages. *J Clin Invest* (2016) 126:3672–9. doi: 10.1172/JCI84427
54. Leek RD, Landers RJ, Harris AL, Lewis CE. Necrosis Correlates With High Vascular Density and Focal Macrophage Infiltration in Invasive Carcinoma of the Breast. *Br J Cancer* (1999) 79:991–5. doi: 10.1038/sj.bjc.6690158
55. Murdoch C, Giannoudis A, Lewis CE. Mechanisms Regulating the Recruitment of Macrophages Into Hypoxic Areas of Tumors and Other Ischemic Tissues. *Blood* (2004) 104:2224–34. doi: 10.1182/blood-2004-03-1109
56. Cordeiro-da-Silva A, Tavares J, Araújo N, Cerqueira F, Tomás A, Kong Thoo Lin P, et al. Immunological Alterations Induced by Polyamine Derivatives on Murine Splenocytes and Human Mononuclear Cells. *Int Immunopharmacol* (2004) 4:547–56. doi: 10.1016/j.intimp.2004.02.009
57. Vahidian F, Duijf PHG, Safarzadeh E, Derakhshani A, Baghbanzadeh A, Baradaran B. Interactions Between Cancer Stem Cells, Immune System and Some Environmental Components: Friends or Foes? *Immunol Lett* (2019) 208:19–29. doi: 10.1016/j.imlet.2019.03.004
58. Cacho-Díaz B, García-Botello DR, Wegman-Ostrosky T, Reyes-Soto G, Ortiz-Sánchez E, Herrera-Montalvo LA. Tumor Microenvironment Differences Between Primary Tumor and Brain Metastases. *J Transl Med* (2020) 18:1. doi: 10.1186/s12967-019-02189-8
59. Rubio C, Munera-Maravilla E, Lodewijk I, Suarez-Cabrera C, Karaivanova V, Ruiz-Palomares R, et al. Macrophage Polarization as a Novel Weapon in Conditioning Tumor Microenvironment for Bladder Cancer: can We Turn Demons Into Gods? *Clin Transl Oncol* (2019) 21:391–403. doi: 10.1007/s12094-018-1952-y
60. Szade K, Gulati GS, Chan CKF, Kao KS, Miyanishi M, Marjon KD, et al. Where Hematopoietic Stem Cells Live: The Bone Marrow Niche. *Antioxid Redox Signal* (2018) 29:191–204. doi: 10.1089/ars.2017.7419
61. Li K, Lu L, Xue C, Liu J, He Y, Zhou J, et al. Polarization of Tumor-Associated Macrophage Phenotype Via Porous Hollow Iron Nanoparticles for Tumor Immunotherapy In Vivo. *Nanoscale* (2020) 12:130–44. doi: 10.1039/C9NR06505A
62. Raggi F, Pelassa S, Pierobon D, Penco F, Gattorno M, Novelli F, et al. Regulation of Human Macrophage M1-M2 Polarization Balance by Hypoxia and the Triggering Receptor Expressed on Myeloid Cells-1. *Front Immunol* (2017) 8:1097. doi: 10.3389/fimmu.2017.01097
63. Jeong J, Suh Y, Jung K. Context Drives Diversification of Monocytes and Neutrophils in Orchestrating the Tumor Microenvironment. *Front Immunol* (2019) 10:1817. doi: 10.3389/fimmu.2019.01817
64. Qian B, Deng Y, Im JH, Muschel RJ, Zou Y, Li J, et al. A Distinct Macrophage Population Mediates Metastatic Breast Cancer Cell Extravasation, Establishment and Growth. *PLoS One* (2009) 4:e6562. doi: 10.1371/journal.pone.0006562
65. Fritz JM, Tennis MA, Orlicky DJ, Lin H, Ju C, Redente EF, et al. Depletion of Tumor-Associated Macrophages Slows the Growth of Chemically Induced Mouse Lung Adenocarcinomas. *Front Immunol* (2014) 5:587. doi: 10.3389/fimmu.2014.00587
66. Andón FT, Digifico E, Maeda A, Erreni M, Mantovani A, Alonso MJ, et al. Targeting Tumor Associated Macrophages: The New Challenge for Nanomedicine. *Semin Immunol* (2017) 34:103–13. doi: 10.1016/j.smim.2017.09.004

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.

frontiers in Immunology

ORIGINAL RESEARCH
published: 23 August 2021
doi: 10.3389/fimmu.2021.688257

A Stochastic Intracellular Model of Anthrax Infection With Spore Germination Heterogeneity

Bevelynn Williams[1], Martín López-García[1], Joseph J. Gillard[2], Thomas R. Laws[2], Grant Lythe[1], Jonathan Carruthers[3], Thomas Finnie[3] and Carmen Molina-París[1,4]**

[1] Department of Applied Mathematics, School of Mathematics, University of Leeds, Leeds, United Kingdom, [2] CBR Division, Defence Science and Technology Laboratory, Salisbury, United Kingdom, [3] Emergency Response Department, Public Health England, Salisbury, United Kingdom, [4] T-6, Theoretical Biology and Biophysics, Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM, United States

OPEN ACCESS

Edited by:
Rob J. De Boer, Utrecht University, Netherlands

Reviewed by:
Tom Chou, University of California, Los Angeles, United States
Joel Bozue, United States Army Medical Research Institute of Infectious Diseases (USAMRIID), United States

***Correspondence:**
Martín López-García
m.lopezgarcia@leeds.ac.uk
Carmen Molina-París
molina-paris@lanl.gov

Specialty section:
This article was submitted to Molecular Innate Immunity, a section of the journal Frontiers in Immunology

Received: *30 March 2021*
Accepted: *01 July 2021*
Published: *23 August 2021*

Citation:
Williams B, López-García M, Gillard JJ, Laws TR, Lythe G, Carruthers J, Finnie T and Molina-París C (2021) A Stochastic Intracellular Model of Anthrax Infection With Spore Germination Heterogeneity. Front. Immunol. 12:688257. doi: 10.3389/fimmu.2021.688257

We present a stochastic mathematical model of the intracellular infection dynamics of *Bacillus anthracis* in macrophages. Following inhalation of *B. anthracis* spores, these are ingested by alveolar phagocytes. Ingested spores then begin to germinate and divide intracellularly. This can lead to the eventual death of the host cell and the extracellular release of bacterial progeny. Some macrophages successfully eliminate the intracellular bacteria and will recover. Here, a stochastic birth-and-death process with catastrophe is proposed, which includes the mechanism of spore germination and maturation of *B. anthracis*. The resulting model is used to explore the potential for heterogeneity in the spore germination rate, with the consideration of two extreme cases for the rate distribution: continuous Gaussian and discrete Bernoulli. We make use of approximate Bayesian computation to calibrate our model using experimental measurements from *in vitro* infection of murine peritoneal macrophages with spores of the Sterne 34F2 strain of *B. anthracis*. The calibrated stochastic model allows us to compute the probability of rupture, mean time to rupture, and rupture size distribution, of a macrophage that has been infected with one spore. We also obtain the mean spore and bacterial loads over time for a population of cells, each assumed to be initially infected with a single spore. Our results support the existence of significant heterogeneity in the germination rate, with a subset of spores expected to germinate much later than the majority. Furthermore, in agreement with experimental evidence, our results suggest that most of the spores taken up by macrophages are likely to be eliminated by the host cell, but a few germinated spores may survive phagocytosis and lead to the death of the infected cell. Finally, we discuss how this stochastic modelling approach, together with dose-response data, allows us to quantify and predict individual infection risk following exposure.

Keywords: *Bacillus anthracis*, Markov process, intracellular model, spore germination, rupture size distribution, approximate Bayesian computation, time to macrophage rupture, dose-response

INTRODUCTION

Anthrax is an infectious disease, caused by the bacterium *Bacillus anthracis*. Under adverse conditions, *B. anthracis* forms a dormant spore, unable to replicate. These spores monitor their environment, and when favourable conditions are detected, such as the nutrient content of a host, the spores begin to germinate into vegetative bacteria, which can replicate inside the host (1). *B. anthracis* spores in the surrounding air inhaled into the lungs can lead to inhalational anthrax, which is usually fatal if not rapidly detected and treated (2). Even with treatment, fatality rates for inhalational anthrax can be rather high if the treatment is not started early enough after exposure and symptoms onset. *B. anthracis* spores can be produced and preserved, making inhalational anthrax a potential bio-terror threat (3).

Inhalational anthrax is initiated by ungerminated, dormant *B. anthracis* spores, inhaled by a host. The spores travel through the air passages and eventually reach the alveoli of the lungs. There is some evidence that spores may be able to germinate extracellularly in the lungs (4). However the generally accepted model of inhalational anthrax infection is the Trojan horse model, which assumes that ungerminated spores must be engulfed by alveolar phagocytes before they begin to germinate (5). Once the spores have been phagocytosed, the infected phagocytes migrate into the nearby lymph nodes in the mediastinum. Macrophages play a key role in the early infection stages of anthrax, since they can induce microbicidal defences against intracellular pathogens and help to clear the infection (6). It has also been shown that 1-2 hours after phagocytosis, newly germinated bacteria are able to escape from macrophage phagosomes and begin to replicate in the cytosol, before being released from the macrophage into the extracellular environment when the host cell ruptures and dies (7). Dendritic cells are also thought to play a role in the trafficking of *B. anthracis* to the lymph nodes during the early stages of infection, since they have been found to readily engulf *B. anthracis* spores (8, 9), and transport them to the lymph nodes in a mouse model of inhalational infection (10). Once an infected host cell ruptures, the extracellular bacteria continue to multiply, leading to oedema and haemorrhage of the mediastinal lymph nodes, and large amounts of fluid in the pleural cavity, which can severely affect breathing (11). The bacteria can also spread into the bloodstream and other organs to establish a systemic infection (12). One of the characteristic virulence factors of *B. anthracis* is the production of toxins. The two anthrax toxins, oedema toxin and lethal toxin, cause different cellular responses and are essential factors for the survival of bacteria in the infected host. Lethal toxin disrupts cell signalling pathways of macrophages and some other cells, leading to cell death, whereas oedema toxin inhibits the phagocytosis of bacteria by neutrophils (13). In some cell types, oedema toxin also increases the levels of cyclic adenosine monophosphate, which is a chemical messenger that plays a major role in controlling many intracellular processes. Together, the anthrax toxins cause suppression of the host's immune system, often leading to death of the host. Another important factor for the survival of *B. anthracis* bacteria in the host is the antiphagocytic capsule, which allows extracellular, vegetative bacteria to avoid eradication by the immune system by preventing the bacteria being phagocytosed and destroyed (14).

In this paper, we propose a stochastic model for the intracellular infection dynamics of inhalational anthrax, which adapts and extends the deterministic one of Pantha et al. (15). The model by Pantha et al. is a system of ordinary differential equations (ODEs) representing the interaction between macrophages and *B. anthracis* spores, and considers two intracellular bacterial populations: newly germinated bacteria, which are susceptible to macrophage killing but unable to replicate, and vegetative bacteria, which are susceptible to macrophage killing and capable of replicating. Spores germinate into newly germinated bacteria, and the newly germinated bacteria must mature into vegetative bacteria before they can begin to replicate. Similarly to the model in (15), our model considers the germination of spores, replication of bacteria, and killing of bacteria by the host cell. Still, we make use of a stochastic approach, instead of a deterministic one, to describe the population dynamics of spores and bacteria. We follow the methods recently developed by Carruthers et al. (16) for *Francisella tularensis* infection, extended here to include spores and spore germination, since *B. anthracis* is a spore-forming bacteria and *F. tularensis* is not.

An important addition in our stochastic model is the consideration of macrophage rupture, not explicitly considered in (15). The rupture of host cells and the release of bacteria into the extracellular environment is an important mechanism in the pathogenesis of anthrax. Thus, incorporating this event into the model allows one to better understand both the timescales of macrophage rupture, and the rupture size distribution (*i.e.*, the number of vegetative bacteria released upon rupture). These summary statistics can then play an important role when considering within-host infection dynamics, such as in the model by Day et al. (17), or when linking to dose-response data (18), as considered in the *Discussion* section. In the same way as Carruthers et al. (16), we assume that an infected macrophage's rupture probability per unit time is proportional to its bacterial load. Thus, cells with a high bacterial load at a given time are more likely to rupture than those with a lower one. This hypothesis is supported by Ruthel et al. (19), who suggest that the intracellular bacterial load may be a contributing factor to whether a macrophage will survive an infection.

A second improvement in our model is the consideration of spore germination heterogeneity. Motivation for this comes from the work by Setlow (1, 20, 21), where it is shown that germination rates are highly heterogeneous for the *Bacillus* species spores, with germination times ranging from a few minutes to longer than 24 hours. It is thought that this spore germination heterogeneity is primarily due to variation in the germinant receptor levels between individual spores. Setlow mentions that spores with very low germinant receptor levels germinate extremely slowly and are termed superdormant (1, 20, 21). Hence, in our model we explore two hypotheses for this heterogeneity. The first hypothesis is that the germination rate is continuously distributed in a population of spores and follows a truncated normal distribution. The second hypothesis is that the

population of spores can be roughly split into two discrete groups, with different germination rates, where one group corresponds to the spores with "average" germinant receptor levels, and the other corresponds to the superdormant spores.

Our resulting stochastic model is a linear birth-and-death process with catastrophe, extended to account for spore germination heterogeneity. For this model, we show how to compute the probability of either rupture or recovery of the infected cell and the conditional mean times taken to reach these fates. Furthermore, we adapt some of the results from (16) in order to compute the probability distribution of rupture times, which is shown to be proportional to the mean number of vegetative bacteria in the cell over time. We are also able to compute the probability distribution of the rupture size, which is the number of bacteria that are eventually released into the extracellular environment from one single infected cell. We carry out parameter calibration by means of Approximate Bayesian Computation Sequential Monte Carlo (ABC-SMC) (22), and by making use of the spore and bacterial counts experimentally measured by Kang et al. (23). We then present numerical results to quantify the implications of the calibrated model. Finally, we discuss the possible application of our intracellular model to explain dose-response data for anthrax infection and explore the relationship between mechanistic approaches and existing dose-response models.

MATERIALS AND METHODS

Stochastic Model for the Dynamics of Spores and Bacteria in a Single Infected Cell

In this section we introduce a stochastic model for the dynamics of spores and bacteria in a single infected phagocyte, starting at the time when the cell phagocytoses a spore, and ending either with rupture and death of the cell and the release of bacteria into the extracellular environment, or with recovery of the cell and the elimination of any intracellular spores or bacteria. When considering low dose exposures, for which the multiplicity of infection (MOI) will be low, it is reasonable to assume that each phagocyte will only engulf at most one spore. Therefore, in what follows we only consider infection of a cell that has phagocytosed a single spore. The model presented here includes germination of the spore into a newly germinated bacterium, maturation of the newly germinated bacterium into a vegetative bacterium, replication of vegetative bacteria, death of bacteria, and rupturing of the host cell to release the intracellular bacteria (see **Figure 1**).

Our intracellular infection model, depicted in **Figure 1**, corresponds to a continuous-time Markov chain (CTMC), $\chi = \{X(t), t \geq 0\}$, taking values in the state-space $\mathcal{S} = \{1_S, 1_{NGB}, 0,1,2, \ldots\} \cup \{R\}$, where:

- 1_S corresponds to the state where the host cell contains a spore and no bacteria,
- 1_{NGB} corresponds to the state where the host cell contains a single newly germinated bacterium,
- {0,1,2, ... } represent the number of vegetative bacteria inside the host cell, so that 0 corresponds to the cell's recovery, and
- R corresponds to the state of the host cell having ruptured.

Time $t = 0$ is the time at which the spore is phagocytosed, so the initial state of the process is assumed to be $X(0) = 1_S$. Experimental work using murine macrophages has indicated that the ungerminated spore form of anthrax is not easily eliminated by host cells (23), so we assume that the initial spore will not die or be killed by the cell. We represent the germination-maturation time as an Erlang(2, g) distribution, as depicted in **Figure 1**, where the rate g has units $hours^{-1}$. This

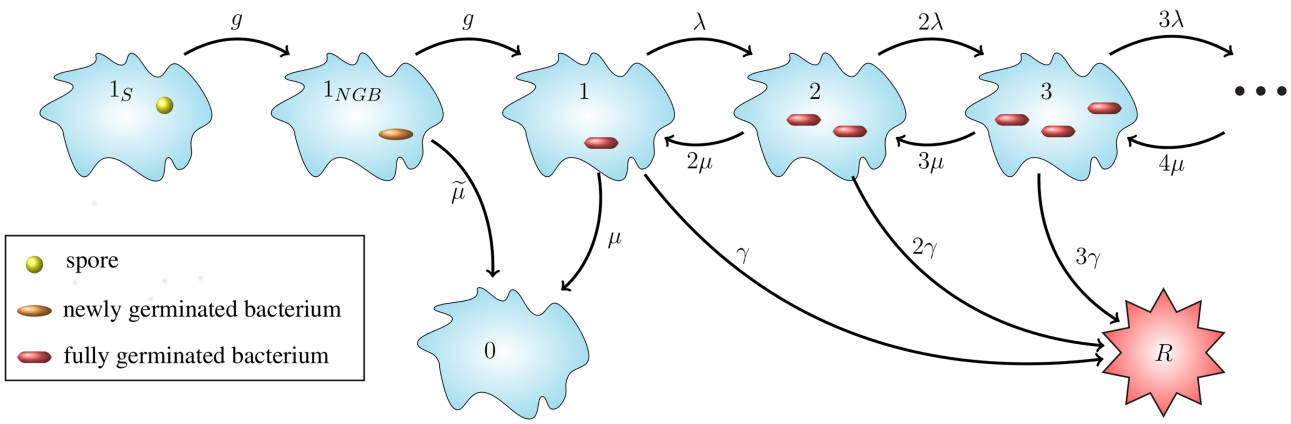


FIGURE 1 | Intracellular infection model. State 1_S represents a phagocytosed spore and state 1_{NGB} a newly germinated bacterium (NGB). The germination rate from spore to NGB, and also the maturation rate from NGB to vegetative bacterium for a given spore is denoted by g $hours^{-1}$, which leads to an Erlang(2, g) distribution for the germination-maturation time. The rate g is assumed to vary between spores. The death rate of the newly germinated bacterium is given by $\tilde{\mu}$ $hours^{-1}$. States $i \in \mathbb{N} \cup \{0\}$ represent i intracellular bacteria. State 0 represents recovery and state R the rupture of the cell. 0 and R are absorbing states for the stochastic process. Transitions between states $i \in \mathbb{N}$ represent three types of events: transition to state $i + 1$ (division of a bacterium), to state $i - 1$ (death of a bacterium), and to state R (rupture of the host cell with release of i bacteria). The per bacterium division, death, and rupture rates are $\lambda > 0$, $\mu > 0$ and $\gamma > 0$, respectively, all with units $(bacteria \cdot hours)^{-1}$, leading to a linear birth-and-death process with catastrophe. The infected cell survives for as long as it does not reach state R.

Erlang(2, g) representation is the simplest approach to consider a non-exponential distribution for the time that it takes the spore to become a vegetative bacterium, while including an intermediate, susceptible state as done in (15), and keeping the process Markovian. During the germination process, there is an increase in the hydration of the spore core and the spore cortex is broken down (1). These changes mean that the spore loses some of its resistance against the anti-microbial environment within the host cell and may be killed by the reactive oxygen and nitrogen species, and anti-microbial peptides within the phagolysosome (13). As a result, modelling the germination-maturation process with two stages allows one to incorporate this loss of resistance into the intermediate state 1_{NGB}. We refer the reader to the *Discussion* section for some additional considerations on the Erlang(2, g) choice. The newly germinated bacterium can be killed by the cell, with rate $\tilde{\mu}$ $hours^{-1}$, but cannot replicate unless it matures into a vegetative bacterium, represented by state 1. If the stochastic process reaches state 1, the subsequent replication of bacteria, death of bacteria, and rupture of the host cell is modelled as a birth-and-death process with catastrophe (24), with state-space {0,1,2, ... } ∪ {R} (see **Figure 1**). We introduce the per bacterium division, death and rupture rates λ, μ and γ, respectively, all with units $(bacteria \cdot hours)^{-1}$. The stochastic process has two absorbing states: the recovery state, 0, representing elimination of any intracellular spores and bacteria, and the rupture state, R, representing rupture of the infected cell and release of its entire content of bacteria.

For a CTMC, χ, the transition probability from state $i \in \mathcal{S}$ to state $j \in \mathcal{S}$ in time t is defined by,

$$p_{i,j}(t) = \mathbb{P}(X(t) = j | X(0) = i)\,. \tag{1}$$

We are interested in the dynamics of the process when the initial state is $X(0) = 1_S$, representing an intracellular spore which was just phagocytosed. Therefore, when the initial state is $i = 1_S$, we will omit the first index in the notation, so that the probability that the process is in state j at time t, given that it started with one spore, is denoted by

$$p_j(t) = \mathbb{P}(X(t) = j | X(0) = 1_S)\,. \tag{2}$$

Spore Germination

To describe the dynamics of the germination-maturation process, one can consider the transient state probabilities, $p_j(t)$, for each of the states, $j \in \{1_S, 1_{NGB}\}$. These probabilities obey the following differential equations, with initial conditions $p_{1_S}(0) = 1$ and $p_{1_{NGB}}(0) = 0$,

$$\frac{dp_{1_S}}{dt} = -gp_{1_S}, \tag{3}$$

$$\frac{dp_{1_{NGB}}}{dt} = gp_{1_S} - (g + \tilde{\mu})p_{1_{NGB}}. \tag{4}$$

If a phagocyte engulfs a spore with germination rate g at time $t = 0$, then the cell will contain one spore at time t if $X(t) = 1_S$, and zero spores if $X(t) \neq 1_S$. Hence, the mean number of intracellular spores at time t is equal to $p_{1_S}(t)$. Similarly, the mean number of intracellular newly germinated bacteria at time t is equal to $p_{1_{NGB}}(t)$. These are given by the solutions to Eqs. (3) and (4),

$$p_{1_S}(t) = e^{-gt}, \tag{5}$$

$$p_{1_{NGB}}(t) = \frac{g}{\tilde{\mu}}\left(e^{-gt} - e^{-(g+\tilde{\mu})t}\right), \tag{6}$$

for any $t \geq 0$. For each spore, we assume that its germination rate is equal to its maturation rate, and denote it by g. However, in order to reflect the heterogeneity in spore germination times as reported in the literature (1, 20, 21), we assume that the germination rate varies between spores. We consider here two different distributions for the germination rate and will explore and compare these two hypotheses in the *Parameter Calibration* section.

Continuous Gaussian Distribution

First, we assume that the germination rate varies continuously among spores, according to a normal distribution on the positive axis. In this case, the germination rate g for a given spore is a realisation of the continuous random variable $G \sim N_{(0,+\infty)}(\mu_g, \sigma_g^2)$, which is normally distributed with mean μ_g $hours^{-1}$ and standard deviation σ_g $hours^{-1}$, and is truncated to the interval $(0,+\infty)$. Therefore, the germination rate has probability density function given by

$$f_G(g) = \frac{1}{Z}\frac{1}{\sigma_g\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{g-\mu_g}{\sigma_g}\right)^2}, \text{ for } \quad g \in (0, +\infty), \tag{7}$$

where $Z = \Phi\left(\frac{\mu_g}{\sigma_g}\right)$ is a normalisation factor, and Φ is the cumulative distribution function of the standard normal distribution. Given this truncated normal distribution for the germination rate, the mean number of spores inside the infected cell at time $t \geq 0$ is given by

$$\begin{aligned} S(t) &= \int_0^{+\infty} f_G(g)p_{1_S}(t)dg \\ &= \int_0^{+\infty} f_G(g)e^{-gt}dg = \frac{1}{Z}\exp\left(\frac{\sigma_g^2 t^2}{2} - \mu_g t\right)\Phi\left(\frac{\mu_g - \sigma_g^2 t}{\sigma_g}\right). \end{aligned} \tag{8}$$

Similarly, the mean number of newly germinated bacteria inside the infected cell at time $t \geq 0$ is given by

$$\begin{aligned} B_{NGB}(t) &= \int_0^{+\infty} f_G(g)p_{1_{NGB}}(t)dg = \int_0^{+\infty} f_G(g)\frac{g}{\tilde{\mu}}\left(e^{-gt} - e^{-(g+\tilde{\mu})t}\right)dg \\ &= \frac{1}{\tilde{\mu}Z}\left(1 - e^{-\tilde{\mu}t}\right)\left[\frac{\sigma_g}{\sqrt{2\pi}}e^{-\frac{\mu_g^2}{2\sigma_g^2}} + (\mu_g - \sigma_g^2 t)\exp\left(\frac{\sigma_g^2 t^2}{2} - \mu_g t\right)\Phi\left(\frac{\mu_g - \sigma_g^2 t}{\sigma_g}\right)\right]. \end{aligned} \tag{9}$$

Discrete Bernoulli Distribution

Setlow in Refs. (20, 21) explains that some spores can be described as superdormant because they have very low germinant receptor levels and germinate extremely slowly,

taking many hours or even days. Therefore, we propose a second choice for the germination rate, which assumes that the population of spores can be roughly split into two discrete groups, type A and type B, with type A having a faster germination rate than type B. Here, type A corresponds to the spores with "typical" germinant receptor levels, and type B corresponds to spores with significantly lower levels. We represent this as a discrete Bernoulli distribution with probability mass function, as follows

$$f_B(g) = \begin{cases} \varepsilon & \text{if } g = g_A, \\ 1-\varepsilon & \text{if } g = g_B, \end{cases} \tag{10}$$

for some parameter values $g_A > g_B$ and $\varepsilon \in (0,1)$. For this Bernoulli distribution, the mean number of spores and newly germinated bacteria inside the infected cell at time $t \geq 0$ are, respectively,

$$S(t) = \varepsilon e^{-g_A t} + (1-\varepsilon)e^{-g_B t}, \tag{11}$$

$$B_{NGB}(t) = \frac{\varepsilon g_A}{\tilde{\mu}}\left[e^{-g_A t} - e^{-(g_A+\tilde{\mu})t}\right] + \frac{(1-\varepsilon)g_B}{\tilde{\mu}}\left(e^{-g_B t} - e^{-(g_B+\tilde{\mu})t}\right). \tag{12}$$

Cellular Fate: Rupture or Recovery

The CTMC in **Figure 1** has two absorbing states, R and 0, which denote rupture and recovery of the host cell, respectively. In this section we show how to compute the probability of either rupture or recovery of the cell, and the probability density functions of the recovery and rupture times. We also explain how to compute the conditional mean times taken to reach each of the two cellular fates. Finally, we show that the mean number of vegetative bacteria in the infected cell over time is proportional to the probability density function of the rupture time.

Probabilities and Times to Reach Absorbing States

The transient probabilities for the states $j \in \mathbb{N} \cup \{0,\mathrm{R}\}$, obey the following system of differential equations

$$\frac{dp_1}{dt} = g p_{1_{NGB}} - (\lambda+\mu+\gamma)p_1 + 2\mu p_2, \tag{13}$$

$$\frac{dp_j}{dt} = \lambda(j-1)p_{j-1} + \mu(j+1)p_{j+1} - (\lambda+\mu+\gamma)jp_j,\ j \geq 2, \tag{14}$$

$$\frac{dp_0}{dt} = \tilde{\mu} p_{1_{NGB}} + \mu p_1, \tag{15}$$

$$\frac{dp_R}{dt} = \sum_{j=1}^{+\infty} \gamma j p_j\,. \tag{16}$$

The long-term probabilities of recovery or rupture for an infected cell, starting from any state $i \in \mathcal{S}$ can be denoted, respectively, by

$$r_i^0 = \lim_{t\to+\infty} p_{i,0}(t), \quad r_i^R = \lim_{t\to+\infty} p_{i,R}(t)\,. \tag{17}$$

These probabilities can also be expressed in terms of the times it takes the process χ to reach states 0 and R, respectively. In particular, one can denote the times taken for the process to travel from state i to states 0 or R, respectively, by

$$T_i^0 = \inf\{t \geq 0 : X(t) = 0 | X(0) = i\}, \qquad T_i^R = \inf\{t \geq 0 : X(t) = R | X(0) = i\}\,. \tag{18}$$

Since there is a choice of two possible absorbing fates (recovery or rupture), the random variables T_i^0 and T_i^R may be infinite with non-zero probability. That is, the time to recovery, T_i^0, will be infinite if the process ends in the rupture state, and vice versa. Thus, we can write

$$p_{i,0}(t) = \mathbb{P}(T_i^0 \leq t),\ \ p_{i,R}(t) = \mathbb{P}(T_i^R \leq t), \tag{19}$$

and

$$r_i^0 \ = \ \mathbb{P}(T_i^0 < +\infty) = 1 - \mathbb{P}(T_i^R < +\infty) = 1 - r_i^R\,. \tag{20}$$

Carruthers et al. in (16) studied a very similar process to the one described here, for the non-sporulating bacteria *F. tularensis*. In this section we use some of their results. For instance, survival analysis allowed Carruthers et al. to show that the probability density function of the random variable T_1^R (the rupture time starting from state 1, representing a single fully vegetative bacterium), is given by

$$f_{T_1^R}(t) = \frac{\gamma(b-a)^2 e^{-\lambda(b-a)t}}{\left[b - 1 + (1-a)e^{-\lambda(b-a)t}\right]^2},\ t \geq 0, \tag{21}$$

with

$$a = \frac{(\lambda+\mu+\gamma) - \sqrt{(\lambda+\mu+\gamma)^2 - 4\mu\lambda}}{2\lambda}, \qquad b = \frac{(\lambda+\mu+\gamma) + \sqrt{(\lambda+\mu+\gamma)^2 - 4\mu\lambda}}{2\lambda}. \tag{22}$$

They also showed that the probability that a cell eventually ruptures, starting in state 1 with one vegetative bacterium, is $r_1^R = 1 - a$, and the probability that a cell eventually recovers is $r_1^0 = a$. One can adapt these results to our model with the help of first-step analysis, to find the probabilities of rupture and recovery starting with one initial spore in state 1_S. If a phagocyte is infected with a spore at time $t = 0$, then at some time point, the initial spore will germinate, transitioning to the intermediate state of newly germinated bacterium (NGB). At some later time, the NGB will either die, with probability $\tilde{\mu}/(\tilde{\mu}+g)$, or will mature into a vegetative bacterium with probability $g/(\tilde{\mu}+g)$. Since it is certain that the first step of the process will be the transition from spore to newly germinated bacterium, the probability of eventual recovery or rupture starting from state 1_S is the same as the probability of recovery or rupture starting from state 1_{NGB}. The only way that the process will eventually reach the rupture state is if the newly germinated bacterium matures into a vegetative bacterium, and then the cell eventually ruptures starting from state 1. On the other hand, the cell can

recover if either the newly germinated bacterium dies before it matures into a vegetative bacterium, or the newly germinated bacterium matures and the cell eventually recovers starting from state 1. In particular, given a germination rate g for the initial spore, the probabilities for rupture and recovery, starting from state 1_S, are given by

$$r^R_{1_S}(g) = r^R_{1_{NGB}}(g) = \frac{g}{\tilde{\mu}+g} \;\; r^R_1 = \frac{g(1-a)}{\tilde{\mu}+g}, \qquad (23)$$
$$r^0_{1_S}(g) = r^0_{1_{NGB}}(g) = \frac{\tilde{\mu}}{\tilde{\mu}+g} + \frac{g}{\tilde{\mu}+g} \;\; r^0_1 = \frac{\tilde{\mu}+ga}{\tilde{\mu}+g}.$$

One can also derive the probability density functions of the time to rupture and the time to recovery. These shed light on the distribution of times to rupture and recovery across cells that have been infected with one spore. In what follows, we denote the random variable for the time to transition from state i to state j by $T^j_i = \inf\{t \geq 0 : X(t) = j | X(0) = i\}$, and the probability density function for this random variable is denoted by $f_{T^j_i}(t)$.

First, we consider the total time for the initial intracellular spore to germinate and mature into a vegetative bacterium, given by the random variable $T^1_{1_S}$, with probability density function, $f_{T^1_{1_S}}(t)$. This function will be needed later to calculate the probability density function of the times to rupture and recovery. To find the function $f_{T^1_{1_S}}(t)$, let $F_{T^1_{1_S}}(t) = \mathbb{P}(T^1_{1_S} \leq t)$ be the probability that the cell contains an intracellular vegetative bacterium by time t, given that the germination rate of the initial spore is equal to g. Let us consider a small time step Δt, such that only one transition can occur in the interval $(t, t+\Delta t)$. If the cell has not entered state 1, representing a vegetative bacterium, before time t, then it will only be possible to enter state 1 before time $t+\Delta t$ if the cell is in state 1_{NGB} at time t. Furthermore, if the phagocyte contains a newly germinated bacterium at time t, then the probability to transition to a vegetative bacterium between time t and $t+\Delta t$ is $g\Delta t$. Hence,

$$F_{T^1_{1_S}}(t+\Delta t) = F_{T^1_{1_S}}(t) + p_{1_{NGB}}(t) g \Delta t\,. \qquad (24)$$

The function $F_{T^1_{1_S}}(t)$ is the cumulative distribution function of the random variable, $T^1_{1_S}$, for the time that the process takes to reach a vegetative bacterium, starting from one spore. Therefore, the probability density function for this random variable is

$$f_{T^1_{1_S}}(t;g) = \frac{dF_{T^1_{1_S}}(t)}{dt} = g\, p_{1_{NGB}}(t) = \frac{g^2}{\tilde{\mu}}(e^{-gt} - e^{-(g+\tilde{\mu})t}),\; t \geq 0, \quad (25)$$

where we have written explicitly that $f_{T^1_{1_S}}$ is a function of the germination rate g. It can be verified that

$$\mathbb{P}(T^1_{1_S} < +\infty) = \int_0^{+\infty} f_{T^1_{1_S}}(t;g)dt = \frac{g}{g+\tilde{\mu}}, \qquad (26)$$

which is the probability that the process will eventually reach state 1, or equivalently, the probability that a spore will mature into a vegetative bacterium instead of being cleared by the infected cell.

Time to Recovery

Here we show how to compute the probability density function, $f_{T^0_{1_S}}(t)$, for the time to recovery of an infected cell starting with one spore. Following the same approach as above, the probability density function for the random variable $T^0_{1_S}$, given that the spore has germination rate g, is

$$f_{T^0_{1_S}}(t;g) = \frac{dF_{T^0_{1_S}}(t)}{dt} = \tilde{\mu} p_{1_{NGB}}(t) + \mu p_1(t) \qquad (27)$$
$$= g(e^{-gt} - e^{-(g+\tilde{\mu})t}) + \mu \int_0^t f_{T^1_{1_S}}(s;g) p_{1,1}(t-s) ds,\; t \geq 0,$$

where $p_{1,1}(t)$ is the probability that the process, starting in state 1, is in state 1 at time t, and can be derived from results in (16) as follows,

$$p_{1,1}(t) = \frac{(a-b)^2 e^{-\lambda(b-a)t}}{(ae^{-\lambda(b-a)t} - b)^2}, \qquad (28)$$

with a and b defined in Eq. (22). We note that the probability density function for the time to recovery from state 1 can be written as

$$f_{T^0_1}(t) = \mu p_{1,1}(t)\,. \qquad (29)$$

When the germination rate follows a continuous Gaussian distribution (see Eq. (7)), the probability density of the recovery time starting from state 1_S is given by

$$f_{T^0_{1_S}}(t) = \int_0^{+\infty} f_G(g) f_{T^0_{1_S}}(t;g) dg\,. \qquad (30)$$

Alternatively, if the germination rate follows a discrete Bernoulli distribution (see Eq. (10)), the probability density of recovery times is given by

$$f_{T^0_{1_S}}(t) = \varepsilon f_{T^0_{1_S}}(t;g_A) + (1-\varepsilon) f_{T^0_{1_S}}(t;g_B)\,. \qquad (31)$$

The probability density of recovery times yields the distribution of recovery times across cells, since each phagocyte is assumed to be initially infected by a single spore. To gain insights into the expected time of recovery, one can also compute the conditional mean time to recovery of an infected cell, which is the expected time to recovery, given that the cell eventually recovers. This is denoted by $\mathbb{E}[T^0_{1_S} | T^0_{1_S} < +\infty]$, given that the eventual recovery of a cell is equivalent to its recovery time being finite. For any initial state $i \in \mathcal{S}$, one can define the restricted mean time to recovery as $\tau^0_i = \mathbb{E}[T^0_i \cdot \delta_{T^0_i < +\infty}]$, where δ_A is equal to 1 if A is satisfied and 0 otherwise. Then the conditional mean time to recovery, starting from state $i \in \mathcal{S}$, is defined by

$$\mathbb{E}[T^0_i | T^0_i < +\infty] = \frac{\tau^0_i}{r^0_i}\,. \qquad (32)$$

Hence, in order to calculate the conditional expectation, $\mathbb{E}[T^0_{1_S} | T^0_{1_S} < +\infty]$, one must restrict the sample space of $T^0_{1_S}$ to finite values, and divide by the probability that the recovery time is finite. The set of finite recovery times can be partitioned into the set where $T^1_{1_S} = +\infty$ and the set where $T^1_{1_S} < +\infty$. In other words, the cell can either recover without ever containing vegetative bacteria, or the cell can recover after having contained at least one vegetative bacterium. Therefore, the restricted mean time to recovery can be written as follows

$$\tau^0_{1_S} = \mathbb{E}[T^0_{1_S} \cdot \delta_{T^1_{1_S}=+\infty}] + \mathbb{E}[T^0_{1_S} \cdot \delta_{T^1_{1_S}<+\infty} \cdot \delta_{T^0_1<+\infty}] .$$

Using the fact that $T^0_{1_S} \cdot \delta_{T^1_{1_S}<+\infty} \cdot \delta_{T^0_1<+\infty} = (T^1_{1_S} + T^0_1) \cdot \delta_{T^1_{1_S}<+\infty} \cdot \delta_{T^0_1<+\infty}$, and that $T^1_{1_S}$ and T^0_1 are independent, one finds that the restricted mean time to recovery for a cell infected with one spore with germination rate g, is

$$\begin{aligned}\tau^0_{1_S}(g) &= \left(\frac{1}{g} + \frac{1}{\tilde{\mu}+g}\right)\frac{\tilde{\mu}}{\tilde{\mu}+g} + \mathbb{E}[T^1_{1_S} \cdot \delta_{T^1_{1_S}<+\infty}]\mathbb{P}(T^0_1 < +\infty) \\ &\quad + \mathbb{E}[T^0_1 \cdot \delta_{T^0_1<+\infty}]\mathbb{P}(T^1_{1_S} < +\infty) \\ &= \frac{\tilde{\mu}(\tilde{\mu}+2g)}{g(\tilde{\mu}+g)^2} + a\int_0^{+\infty} t f_{T^1_{1_S}}(t;g)dt \\ &\quad + \frac{g}{g+\tilde{\mu}}\int_0^{+\infty} t f_{T^0_1}(t)\,dt \\ &= \frac{g}{\tilde{\mu}+g}\left[\frac{(\tilde{\mu}+2g)(ga+\tilde{\mu})}{g^2(\tilde{\mu}+g)} + \frac{1}{\lambda}\log\left(\frac{b}{b-a}\right)\right], \end{aligned} \tag{33}$$

where we have made use of Eqs. (25), (26), and (29). The values a and b are defined in Eq. (22).

With this restricted mean time at hand, and the probability of recovery in Eq. (23), one can use Eq. (32) to find the conditional mean time until recovery for the two different distributions of the germination rate. In particular, when the germination rate follows a continuous Gaussian distribution in Eq. (7), the conditional mean time to recovery for an infected cell starting with one spore is given by

$$\mathbb{E}[T^0_{1_S} | T^0_{1_S} < +\infty] = \frac{\int_0^{+\infty} f_G(g)\tau^0_{1_S}(g)\ \ dg}{\int_0^{+\infty} f_G(g) r^0_{1_S}(g)\ \ dg}. \tag{34}$$

Alternatively, if one considers the discrete Bernoulli distribution for the germination rate, the conditional mean time to recovery is given by

$$\mathbb{E}[T^0_{1_S} | T^0_{1_S} < +\infty] = \frac{\varepsilon\tau^0_{1_S}(g_A) + (1-\varepsilon)\tau^0_{1_S}(g_B)}{\varepsilon r^0_{1_S}(g_A) + (1-\varepsilon) r^0_{1_S}(g_B)}. \tag{35}$$

Time to Rupture

The time taken for the initial phagocytosed spore to transition into a vegetative bacterium is given by the random variable $T^1_{1_S}$ and the time from vegetative bacterium to rupture is denoted by T^R_1. Thus, the total time between the cell engulfing a spore, and the time of rupture, is $T^R_{1_S} = T^1_{1_S} + T^R_1$. The corresponding probability density function for $T^1_{1_S}$ was given by $f_{T^1_{1_S}}(t;g)$ in Eq. (25), and the probability density function for the rupture time starting from one vegetative bacterium was given by $f_{T^R_1}(t)$ in Eq. (21). One can convolve these two functions to find the probability density function for the total time to rupture, giving

$$f_{T^R_{1_S}}(t;g) = \int_0^t f_{T^1_{1_S}}(s;g) f_{T^R_1}(t-s)\ \ ds . \tag{36}$$

When the germination rate across spores follows the continuous Gaussian distribution, the density of rupture times is given by

$$f_{T^R_{1_S}}(t) = \int_0^{+\infty} f_G(g) f_{T^R_{1_S}}(t;g)\ \ dg . \tag{37}$$

Alternatively, in the discrete Bernoulli case, the density of rupture times is given by

$$f_{T^R_{1_S}}(t) = \varepsilon f_{T^R_{1_S}}(t;g_A) + (1-\varepsilon) f_{T^R_{1_S}}(t;g_B) . \tag{38}$$

As done previously for recovery, one can also calculate the conditional mean time to rupture, denoted $\mathbb{E}[T^R_{1_S} | T^R_{1_S} < +\infty]$ and defined similarly to Eq. (32) with 0 replaced by R. Since the random variables $T^1_{1_S}$ and T^R_1 are independent, it can be shown that the restricted mean time to rupture, for a cell initially infected with a spore with germination rate g, is

$$\begin{aligned}\tau^R_{1_S}(g) &= \mathbb{E}[T^1_{1_S} \cdot \delta_{T^1_{1_S}<+\infty}]\mathbb{P}(T^R_1 < +\infty) \\ &\quad + \mathbb{E}[T^R_1 \cdot \delta_{T^R_1<+\infty}]\mathbb{P}(T^1_{1_S} < +\infty) \\ &= (1-a)\int_0^{+\infty} t f_{T^1_{1_S}}(t;g)\,dt + \frac{g}{g+\tilde{\mu}}\int_0^{+\infty} t f_{T^R_1}(t)\,dt \\ &= \frac{g}{\tilde{\mu}+g}\left[\frac{(\tilde{\mu}+2g)(1-a)}{g(\tilde{\mu}+g)} + \frac{1}{\lambda}\log\left(\frac{b-a}{b-1}\right)\right], \end{aligned} \tag{39}$$

where we have made use of Eqs. (25), (26) and (21). The values a and b are defined in Eq. (22).

Given this restricted mean time, and the probability of rupture from Eq. (23), one can use Eq. (32), with 0 replaced by R, to find the conditional mean time until rupture, for the two different distributions of the germination rate. In particular, when the germination rate across spores follows a continuous Gaussian distribution, the conditional mean time to rupture is of the same form as Eq. (34), with 0 replaced by R. Similarly, if one considers the discrete Bernoulli distribution for the germination rate, the conditional mean time to rupture is of the same form as Eq. (35), with 0 replaced by R.

Number of Intracellular Vegetative Bacteria

Given a particular germination rate g for the phagocytosed spore, we denote the mean number of intracellular vegetative bacteria at time t by $B_v(t;g)$, where

$$B_v(t;g) = \sum_{j=1}^{+\infty} j\, p_j(t) = \frac{1}{\gamma}\frac{dp_R(t)}{dt}, \tag{40}$$

with the second equality arising from Eq. (16). Since $p_R(t)$ represents the cumulative distribution function of the rupture time starting with one spore, this means that the average number of vegetative bacteria is proportional to the probability density function of the rupture time. That is, the mean number of vegetative bacteria at time t, given germination rate g, is

$$B_v(t;g) = \frac{f_{T^R_{1_S}}(t;g)}{\gamma}. \tag{41}$$

Once this is averaged over the possible values of the germination rate, g, for either germination rate distribution hypothesis, the mean number of vegetative bacteria inside a cell at time t, is given by

$$B_v(t) = \frac{f_{T^R_{1_S}}(t)}{\gamma}, \tag{42}$$

where $f_{T^R_{1_S}}(t)$ is defined in Eqs. (37) and (38) for the two germination rate distributions.

Rupture Size Distribution

For an infected cell described by the CTMC χ it is possible to find the probability distribution of its *rupture size*, which is the number of bacteria released into the extracellular environment from the infected cell. If the time for the process to enter state 0 is finite, then the rupture size is equal to 0, indicating that the host cell recovers and does not release any bacteria. On the other hand, if the time to reach state R, denoted by $T^R_{1_S}$, is finite, and $X(T^R_{1_S} - \Delta t) = n$ for small and positive Δt, this means that the process transitions into the rupture state from state $n \in \mathbb{N}$. This corresponds to the death and rupture of the host cell, and the release of n bacteria into the extracellular environment. Let R^n_i denote the probability that the cell will release n bacteria in total, given that the process starts at state $i \in \mathcal{S}$. This is defined as

$$R^n_i = \begin{cases} \mathbb{P}(T^0_i < +\infty), & \text{for } n = 0, \\ \mathbb{P}((T^R_i < +\infty) \text{ and } (X(T^R_i - \Delta t) = n)), & \text{for } n \in \mathbb{N}. \end{cases} \tag{43}$$

With this definition, R^0_i is the probability that the cell recovers, so $r^0_i = R^0_i$. For $n \in \mathbb{N}$, R^n_i is the probability that the cell ruptures and releases n bacteria, so the overall probability of rupture is $r^R_i = \Sigma^{+\infty}_{n=1} R^n_i$. For states $i \in \mathbb{N}$, the probabilities R^n_i do not depend on the germination rate, g. However for $i \in \{1_S, 1_{NGB}\}$, these probabilities do depend on the germination rate, and so will be denoted by $R^n_{1_S}(g)$ and $R^n_{1_{NGB}}(g)$.

We now follow the method of Karlin and Tavare from (24) to find the probabilities R^n_1. If the cell begins with a vegetative bacterium, so that $X(0) = 1$, then for a small time interval $\Delta t \to 0$, one has,

$$\begin{aligned} &\mathbb{P}((X(t) = n) \text{ and } (t < T^R_1 \le t + \Delta t)) = \mathbb{P}(X(t) = n) \\ &\times \mathbb{P}(t < T^R_1 \le t + \Delta t \mid X(t) = n) = p_{1,n}(t) n\gamma\Delta t, \end{aligned} \tag{44}$$

since if the cell contains n bacteria at time t, the probability of rupture between time t and $t + \Delta t$ is $n\gamma\Delta t$. An expression for $p_{1,n}(t)$, which is the probability that a cell contains n bacteria at time t, given that it contains one bacterium at time 0, was given by Carruthers et al. in (16),

$$p_{1,n}(t) = \frac{(b-a)^2 e^{-\lambda(b-a)t}(e^{-\lambda(b-a)t} - 1)^{n-1}}{(ae^{-\lambda(b-a)t} - b)^{n+1}}, \quad n \ge 1. \tag{45}$$

Hence for an infected phagocyte starting with one vegetative bacterium, the probability that the cell releases $n \ge 1$ bacteria, R^n_1, is then

$$R^n_1 = \int_0^{+\infty} p_{1,n}(t) n\gamma \; dt = \frac{(1-a)(b-1)}{b^n}, \tag{46}$$

with a and b defined in Eq. (22). Moreover, a first-step argument allows one to obtain the probability $R^n_{1_S}(g)$ from R^n_1. For $n \ge 1$ and germination rate g, one has

$$R^n_{1_S}(g) = R^n_{1_{NGB}}(g) = \frac{g}{\tilde{\mu} + g} R^n_1. \tag{47}$$

When the germination rate across spores follows the continuous Gaussian distribution, the probability that the rupture size of a cell initially infected with one spore is equal to $n \in \mathbb{N}$ bacteria, is given by

$$R^n_{1_S} = R^n_1 \int_0^{+\infty} f_G(g) \frac{g}{\tilde{\mu} + g} dg. \tag{48}$$

Alternatively, if one considers the discrete Bernoulli distribution for the germination rate, the probability that the rupture size of a cell initially infected with one spore is equal to $n \in \mathbb{N}$ bacteria, is given by

$$R^n_{1_S} = R^n_1 \left(\frac{\varepsilon g_A}{\tilde{\mu} + g_A} + \frac{(1-\varepsilon) g_B}{\tilde{\mu} + g_B} \right). \tag{49}$$

PARAMETER CALIBRATION

In this section we make use of experimental data from an *in vitro* study by Kang et al. (23), which was discussed by Pantha et al. in (15) and used to calibrate their ODE model. In the experiment 10^6 murine peritoneal macrophages were incubated with different numbers of Sterne 34F2 strain spores for 30 minutes, during which time phagocytosis occurred (23). The ratio of spores to cells in the solution at the start of the incubation period is called the multiplicity of infection (MOI) and in this case the four MOIs considered were spore to macrophage ratios of 1:1, 1:2, 1:10 and 1:20, corresponding to the initial number of spores in the solution of 10^6, 5×10^5, 10^5, and 5×10^4. At the end of 30 minutes, the solutions were washed, so no extracellular spores remained, and no more spores were phagocytosed after this time. Then the solutions were incubated with an antibacterial agent called gentamicin for 30 minutes to remove any extracellular bacteria. At various time points after this, samples of parallel replicates of the experiment were washed and the number of intracellular spores and bacteria determined. The data from this experiment is given in (15, Tables 2 and 3). While in reality cells could phagocytose more than one spore each in this experiment, this is less likely to happen when the average number of spores per cell in the solution is low. Therefore we only use the data for MOI 1:2, 1:10 and 1:20 to perform the parameter calibration, since these low MOIs will be more consistent with our model assumption in the *Materials and Methods* section that each macrophage engulfs at most one spore, leading to the initial condition for our model in **Figure 1**. Still, once we have obtained posterior samples of the parameters, we will compare our model predictions to the MOI 1:1 data, as a qualitative validation.

The experiment described above was also performed using a germination-deficient strain of *B. anthracis* spores, in which spore germination is inhibited. The average spore counts at one hour from

two duplicate samples that used the germination-deficient strain are provided in **Table 1**. The number of spores of the germination-deficient strain would have remained unchanged between 0.5 hours and 1 hour, because they will not have germinated, and all extracellular spores were removed by washing at 0.5 hours, so there would have been no more phagocytosis after this time. Thus, if one assumes that there is no difference in spore phagocytosis rates between the germination-deficient and Sterne strains, one concludes that the spore counts for the germination-deficient strain are a good representation of the total number of Sterne strain spores that would have been phagocytosed during the first 0.5 hours of the experiment for each MOI. In the parameter calibration for their Phase II subsystem model, Pantha et al. used these numbers of intracellular spores from the germination-deficient experiment as the initial condition for the intracellular dynamics. The justification given is that germination does not seem to be a dominating process at 0.5 hours, so the number of spores of the Sterne strain at 0.5 hours will be similar to the number of germination-deficient spores at the same time point. Therefore, we make here the same assumption that no germination of the Sterne strain spores has occurred before 0.5 hours. Because of this, our estimates for g_A and g_B in the discrete Bernoulli model should be interpreted with this 30-min delay in mind, and in particular our estimated germination rates might be slightly overestimated as a result. However, this delay could possibly be interpreted as a time-lag after phagocytosis for the activation of germination to occur. Furthermore, after learning about the parameters with ABC-SMC inference, we will see that even the spores with a quicker germination rate (type A) take on average longer than an hour to germinate, so it seems (*a posteriori*) reasonable to assume that germination does not happen in the first 30 minutes of the experiment.

In the data from the experiment, time $t = 0$ corresponds to the start of the incubation period of cells and spores (23). On the other hand, our model considers a single host cell that begins with one intracellular spore, and $t = 0$ is assumed to be the start of the germination process of this spore. Since phagocytosis only occurs during the first 0.5 hours of the experiment, and we assume that germination does not occur until after the first 0.5 hours of the experiment, we do not explicitly include phagocytosis in our mathematical model. Instead, we modify the time points so that $t = 0.5$ in the experiment corresponds to $t = 0$ in our model. That is, we take the number of intracellular spores from the germination-deficient experiment, given in **Table 1**, to be our initial conditions for $t = 0$, and use the data as it is shown in **Table 2**.

We have performed Approximate Bayesian Computation Sequential Monte Carlo (ABC-SMC) (22) to estimate our intracellular model parameters. This method involves choosing prior distributions for the model parameters and then carrying out multiple iterations of the ABC algorithm. At each iteration, parameter values are sampled from the posterior distribution of the previous iteration and are perturbed with a kernel function. Here we use a component-wise uniform perturbation kernel, so that each component of the parameter set is perturbed independently in a uniform interval. The perturbed parameter set is then used to obtain a model prediction and is accepted if the distance between the model and the data falls below the distance threshold for that iteration. We have used a sequence of decreasing distance thresholds, such that the distance threshold at each iteration is the median of the distances from the accepted parameter sets in the previous iteration. In this manner, one obtains a set of distributions for the parameters that converge to the posterior distribution. In the parameter calibration results that follow, we are considering a posterior sample of size 10^3 from the final iteration of the ABC algorithm.

In this section we present the results obtained from using the MOI 1:2, 1:10 and 1:20 data sets to obtain posterior parameter distributions, first for the continuous model of germination rate, and then for the discrete Bernoulli model. The authors of (15) used these data to estimate different parameter sets for each MOI, mentioning that for lower MOIs the smaller average intracellular burden could give a better environment for spores to germinate and bacteria to replicate, leading to larger values of the parameters. However, since we are assuming that every infected cell begins with only a single intracellular spore, this means that the cellular

TABLE 1 | Data taken from (15, Table 2), giving the average number of intracellular spores of two replicates of the experiment counted at 1 hour when using spores of a germination-deficient strain of anthrax.

Intracellular germination-deficient spore count at one hour	
MOI 1:1	377500
MOI 1:2	139000
MOI 1:10	30500
MOI 1:20	13925

The number of spores of the germination-deficient strain should have remained unchanged between 0.5 hours and 1 hour, because they cannot germinate, and all extracellular spores were removed by washing at 0.5 hours, so there would have been no more phagocytosis after this time. We note that the value for MOI 1:10 reported in (15, Table 2) was inconsistent with that observed in (15, Figure 2), so the second one is used here, since it is more consistent with the trajectory over time for the spore counts in (15, Figure 2) for MOI 1:10, and also so that our predictions are comparable with those in (15).

TABLE 2 | Data for the number of intracellular spores and bacteria present at different time points, which have been used to perform ABC-SMC.

	Time (hours)	0	0.5	2.5	4.5	23.5
MOI 1:2	Number of intracellular spores	139000	105000	11400	10250	29500
	Number of intracellular bacteria	0	23000	67100	52250	20000
MOI 1:10	Number of intracellular spores	30500	27000	12000	9100	2750
	Number of intracellular bacteria	0	9500	14000	7650	2500
MOI 1:20	Number of intracellular spores	13925	7900	6450	3100	300
	Number of intracellular bacteria	0	6000	2250	1750	1000

*These data have been taken from (15, Tables 2 and 3), although time points are shifted by 30 min to account for the first phase of the experiment where phagocytosis occurs. The initial conditions (t = 0) correspond to the values reported in **Table 1** from the germination-deficient experiment. The counts at each time point are averages of two experimental replicates.*

TABLE 3 | Prior distributions used in the ABC-SMC for the model with continuous heterogeneity of germination rate (top) and the model with two types of spores (bottom).

Parameter	Units	Description	Prior distribution
Model with continuous germination rate distribution			
μ_g	h^{-1}	Mean of the normal distribution for G	$\log_{10}\mu_g \sim$ U(-2, 1)
σ_g	h^{-1}	Standard deviation of the normal distribution for G	$\log_{10}\sigma_g \sim$ U(-2, 0.15)
$\tilde{\mu}$	h^{-1}	Death rate of newly germinated bacteria	$\log_{10}\tilde{\mu} \sim$ U(-4, 1)
λ	$(bacteria{\cdot}h)^{-1}$	Replication rate of vegetative bacteria	See **Figure 2** (top)
μ	$(bacteria{\cdot}h)^{-1}$	Death rate of vegetative bacteria	$\log_{10}\mu \sim$ U(-4, 1)
γ	$(bacteria{\cdot}h)^{-1}$	Rupture rate	See **Figure 2** (top)
Model with discrete Bernoulli germination rate distribution			
ε	–	Probability that a given spore is of type A	$\varepsilon \sim$ U(0, 1)
g_A	h^{-1}	Germination and maturation rate of spores of type A	$\log_{10}g_A \sim$ U(-4, 1)
g_B	h^{-1}	Germination and maturation rate of spores of type B	$\log_{10}g_B \sim$ U(-4, 1)
$\tilde{\mu}$	h^{-1}	Death rate of newly germinated bacteria	$\log_{10}\tilde{\mu} \sim$ U(-4, 1)
λ	$(bacteria{\cdot}h)^{-1}$	Replication rate of vegetative bacteria	See **Figure 2** (bottom)
μ	$(bacteria{\cdot}h)^{-1}$	Death rate of vegetative bacteria	$\log_{10}\mu \sim$ U(-4, 1)
γ	$(bacteria{\cdot}h)^{-1}$	Rupture rate	See **Figure 2** (bottom)

MOI is assumed to be identical across all data sets, and in this way the MOI is simply a measure of the system size. Hence we believe that the parameters considered in our stochastic model may not depend on the initial conditions given by the MOI. Thus, we use the three data sets together to obtain a single set of estimates for the parameters, aiming to give a reasonable fit to the data sets with a significantly smaller number of parameters in our model.

In the experiment, the samples were washed each time before counting the number of intracellular spores and bacteria. Therefore, any bacteria released from a ruptured macrophage would be removed during the washing process and would not contribute to the number of bacteria observed in the data. Thus, to compare our model with the data, we use our model to calculate the per cell mean number of intracellular spores and bacteria over time. For the model with continuous distribution for the germination rate, Eqs. (8), (9), and (42) define the expected number of spores and bacteria in a cell at time t, given that it contained a spore at time $t = 0$. Since each infected cell is assumed to be independent, we can multiply these by the number of initial spores, to obtain the mean number of spores and bacteria present inside the population of cells at time t. Let S_0 be the total number of initial intracellular spores for the population of cells at time $t = 0$. Then for the hypothesis of continuous heterogeneity in the germination rate, the total expected number of spores in all cells at time t is given by

$$S^*(t) = \frac{S_0}{Z}\exp\left(\frac{\sigma_g^2 t^2}{2} - \mu_g t\right)\Phi\left(\frac{\mu_g - \sigma_g^2 t}{\sigma_g}\right). \quad (50)$$

The total (in all cells) expected number of intracellular bacteria at time t hours, including newly germinated bacteria and vegetative bacteria, is given by

$$B^*(t) = S_0\left[\frac{1}{\tilde{\mu}Z}(1-e^{-\tilde{\mu}t})\left(\frac{\sigma_g}{\sqrt{2\pi}}e^{-\frac{\mu_g^2}{2\sigma_g^2}} + (\mu_g - \sigma_g^2 t)\exp\left(\frac{\sigma_g^2 t^2}{2} - \mu_g t\right)\Phi\left(\frac{\mu_g - \sigma_g^2 t}{\sigma_g}\right)\right) + B_v(t)\right] \quad (51)$$

where $B_v(t)$ is the expected number of vegetative bacteria in one cell at time t, defined in Eq. (42). Similarly, for the model with two discrete germination rates, the model predictions for number of spores and bacteria are given by

$$S^*(t) = S_0\left[\varepsilon e^{-g_A t} + (1-\varepsilon)e^{-g_B t}\right], \quad (52)$$

and

$$B^*(t) = S_0\left[\frac{\varepsilon g_A}{\tilde{\mu}}\left(e^{-g_A t} - e^{-(g_A+\tilde{\mu})t}\right) + \frac{(1-\varepsilon)g_B}{\tilde{\mu}}\left(e^{-g_B t} - e^{-(g_B+\tilde{\mu})t}\right) + B_v(t)\right], \quad (53)$$

where $B_v(t)$ is the expected number of vegetative bacteria in one cell at time t. In the ABC-SMC, we compare the data to the model outputs given by $S^*(t)$ and $B^*(t)$, with the initial number of spores equal to $S_0 = 139000$ for MOI 1:2, $S_0 = 30500$ for MOI 1:10 and $S_0 = 13925$ for MOI 1:20.

In the model with continuous germination rate distribution, where the germination rate follows a truncated normal distribution, $G \sim N_{(0,+\infty)}(\mu_g, \sigma_g^2)$, the parameters characterising germination that we aim to estimate are the mean germination rate μ_g, and its standard deviation σ_g. In the discrete Bernoulli model, the parameters characterising germination that we aim to estimate are the probability ε that a given spore is of type A, and the two germination rates, g_A and g_B. The rest of the parameters that we wish to estimate are common to both versions of the model: the death rate of newly germinated bacteria, $\tilde{\mu}$, the replication rate of vegetative bacteria, λ, the death rate of vegetative bacteria, μ, and the rupture rate, γ. To calibrate these parameters, we performed ABC-SMC, and compared our model with the numbers of intracellular spores and bacteria over time from the experiment by Kang et al. (23). Similarly to (25), we make use of the Euclidean distance for the logarithm of the predicted values and observed data, given by

$$d(model\ prediction,\ Data) = \sqrt{\sum_{i\in\{2,10,20\}}\sum_{t\in T}\left(\log\left(\frac{S_i^*(t)}{s_i(t)}\right)\right)^2 + \left(\log\left(\frac{B_i^*(t)}{b_i(t)}\right)\right)^2}, \quad (54)$$

where $T = \{0.5,2.5,4.5,23.5\}$, $S_i^*(t)$ and $B_i^*(t)$ are the respective model predictions for the number of spores and bacteria at time t for MOI 1:i, and $s_i(t)$, $b_i(t)$ are the respective observed number of spores and bacteria at time t, given by the data for MOI 1:i.

To perform ABC-SMC, one needs to choose prior distributions from which to sample the parameter values at the first iteration. To inform the selection of some of these prior distributions, we leverage data from Akoachere et al. (26), who observed that after infecting murine macrophages with *B. anthracis* Sterne strain spores at a spore to macrophage ratio (MOI) of 20:1, 20% of cells had ruptured by 3.5 hours, and 90% had ruptured by 7 hours. We preliminarily fit our model to these data and use the results to estimate a potential prior distribution for the replication and rupture rates, λ and γ. We are unable to learn much about the other model parameters in this way, since only two data points based on a single experiment are available. However, this small amount of data regarding the rupture time of cells does allow us to gain preliminary knowledge of λ and γ, since in our model the rupture rate is proportional to the size of the bacterial population, which in turn depends on the replication rate of the bacteria. Therefore, we use the distributions obtained for λ and γ as prior distributions in the subsequent fitting to the Kang et al. data. These distributions seem to agree well for the two different germination hypotheses and are shown in **Figure 2**. We also show that these preliminary estimates lead to a good representation of the rupture dynamics. Furthermore, the median of the distribution for λ is around 0.9 $(bacteria{\cdot}h)^{-1}$, which is consistent with the doubling time of 0.78 hours, measured by Kalns et al. (27) and used as an estimate in the within-host model by Day et al. (17). For more details on how the prior distributions in **Figure 2** were obtained, see the **Supplementary Material**. Uniform prior distributions are considered for the remaining parameters, as reported in **Table 3**. We note that these parameters are log-transformed because the prior range spans multiple orders of magnitude. For the model with two types of spores, we fixed $g_A > g_B$ to represent that, without any loss of generality, type A spores are the ones with a faster germination rate. In order to sample parameter values g_A and g_B with priors reported in **Table 3**, and under the constraint $g_A > g_B$, we follow the ideas from (28).

Figure 3 shows the posterior histograms obtained by performing ABC-SMC for the two hypotheses considered, using the Kang et al. data, while summary statistics for these posteriors are reported in **Table 4**. By comparing the posterior histograms in blue with the prior distributions in red, we can see that it has been possible to learn significantly about most of the parameters for both hypotheses. For the model with continuous germination rate distribution, the value of μ_g, corresponding to the most likely value for the germination rate of a given spore, is likely to be between 10^{-2} and 10^{-1} h^{-1}. For the model with two germination rates, the value of ε is likely to be between 0.5 and 1, so that the majority of the spores will germinate with rate g_A, which is likely to be of the order of 10^{-1} h^{-1}, and the rest will germinate with rate g_B, which is likely to be of the order of 10^{-2} h^{-1}. For both hypotheses we learn that the death rate of newly germinated bacteria, $\tilde{\mu}$, is likely to be very small. For the model with continuous heterogeneity in the germination rate, the posterior histograms for λ and γ show that the value of these parameters that produce a good match between this model and the Kang et al. data, are similar to the values that gave a good fit to the data from Akoachere et al. that we used to inform our priors for these parameters. For the model with two types of spores, the posterior histogram for λ is shifted slightly to the left from the prior distribution. For both hypotheses we have been able to learn significantly about the death rate of bacteria, μ and these accepted values are usually larger than the corresponding values for the replication rate, λ. This is shown in the posterior histograms for the ratio between the birth and death rate of bacteria, λ/μ, which mostly contain values less than 1, indicating that the bacteria are likely to die more quickly than they replicate.

Model predictions were obtained for each accepted parameter set from the ABC-SMC. **Figure 4** shows the pointwise 95% credible intervals of these time-courses, which indicate the uncertainty in the mean number of intracellular spores and bacteria from the model, due to the range of accepted parameter sets. The solid lines show the model output for the parameter sets with the smallest distance to the data, referred to as the best model predictions. For the model with continuous germination rate distribution, the predictions are close to the data at some time points, but overall do not seem to explain the data very well, since the peak of intracellular bacteria in the model predictions seems to be lower than the peak indicated by the data. On the other hand, the predictions of the model with two types of spores show a fairly good agreement with all data sets. It seems that this latter model, with two discrete germination rates, is better able to describe the pattern of biphasic decay in the number of spores seen in the data. This model also explains the bacterial data significantly better than the hypothesis of continuous germination rate distribution.

Only the data for MOI 1:2, 1:10 and 1:20 were used in the ABC-SMC analysis to calibrate model parameters. Still, we compare our calibrated model predictions with the MOI 1:1 data, by making use of the corresponding initial condition for this MOI, $S_0 = 377500$. For the model with two types of spores, the best prediction is very close to the MOI 1:1 data at some time points, which demonstrates a possibility that the model could be used to extrapolate between MOIs. However, the assumption that each cell will only phagocytose a single spore is less viable at MOI 1:1, so the model may need to be adapted slightly in order to explain the dynamics at higher MOIs. We note that the long-term behaviour of spores for MOI 1:1, where a significant unexpected increase in the number of spores over time is observed, was not explained by Kang et al. (23) or Pantha et al. (15), and cannot be mechanistically explained by our intracellular infection model either, regardless of the choice of parameters.

Figure 4 also shows, as dashed lines, the predictions of the model by Pantha et al. using their parameter estimates. When we compare these to our predictions, it seems that with the consideration of heterogeneity in the germination rate, our model provides a better explanation for the number of intracellular spores seen in the data, especially for MOI 1:10

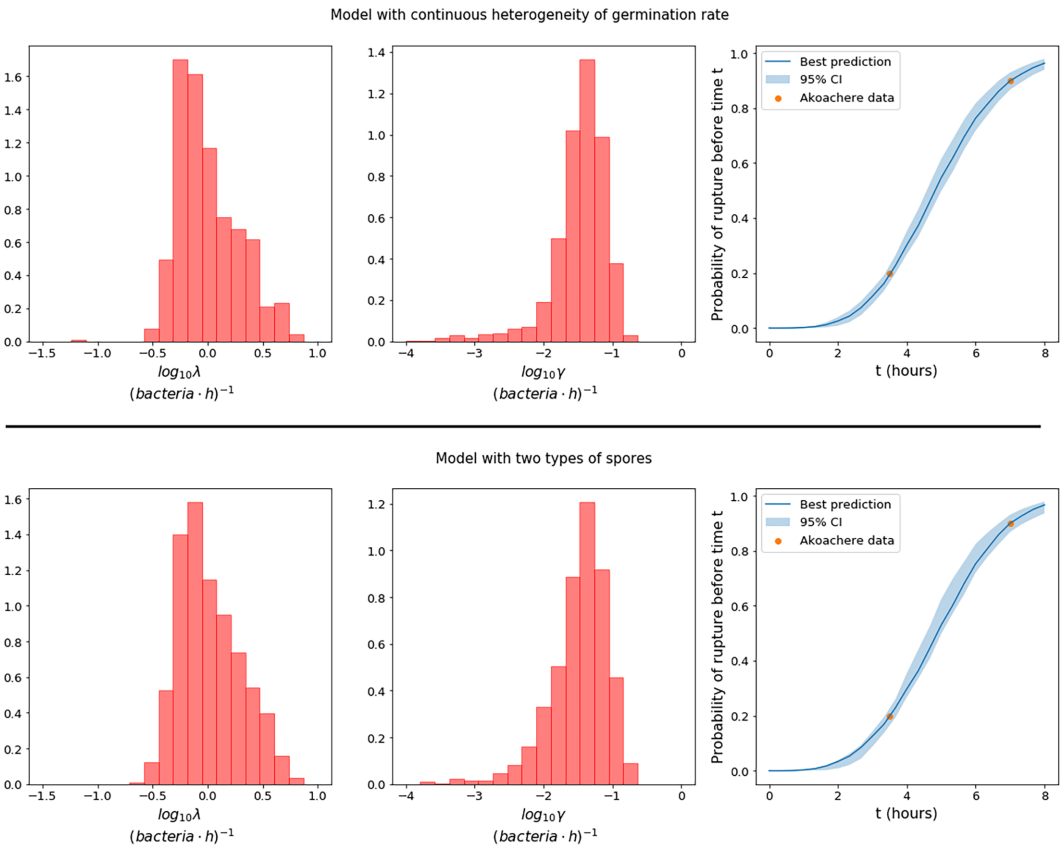


FIGURE 2 | Prior distributions for the replication rate, λ, and the rupture rate γ, in the model with continuous heterogeneity of germination rate (top row), and the model with two types of spores (bottom row), obtained by using observations from Akoachere et al. (26) of the proportion of dead macrophages at two different time points. The plots on the right show the best predictions (solid line) of the fraction of cells that would be expected to rupture before time t in an experiment with MOI 20:1, compared to the data from Akoachere et al. (26). The pointwise 95% credible intervals (shaded region) are shown to represent the uncertainty in the parameter values.

and MOI 1:20. However, the fit of our model to the bacterial data does not look quite as good as the predictions by Pantha et al. This is not surprising, since Pantha et al. allow different parameter values for each data set, meaning that their model has more parameters and higher complexity, giving it more freedom to fit the data. To compare the goodness-of-fit of the models, we use the Corrected Akaike's Information Criterion (AIC_C), which penalises models with a higher number of parameters if there is not enough improvement in the goodness-of-fit to warrant the additional complexity. Roughly speaking, lower values of the AIC_C indicate a better fit to the observed data. We have calculated the value of AIC_C for the model by Pantha et al. and the two versions of our model with different distributions for the germination rate. For details about how the AIC_C was calculated see the **Supplementary Material**. For our models the value of AIC_C was calculated using the parameter set that gave the smallest distance in the ABC-SMC, obtaining a value of AIC_C = 1.42 for the model in which the germination rate is a continuous random variable, and a value of AIC_C = -3.8 for the one with two types of spores. For the model by Pantha et al., the value of AIC_C was calculated using the parameter estimates reported in (15, Table 7), giving a value of AIC_C = 176.73. Our models have a lower AIC_C than the model by Pantha et al., mainly because we used the same parameter estimates for each MOI, whereas Pantha et al. obtained separate estimates for each MOI, meaning that they have many more free parameters. Even with the inclusion of heterogeneity in the germination rate, the version of our model with continuously distributed germination rates is not able to properly capture all the data. Instead, a bi-modal model of heterogeneity in the germination rate is needed to explain both the spore and the bacterial data. This is supported by the fact that the AIC_C value is lower for the discrete Bernoulli hypothesis, even though it has one more parameter. This indicates that this model may explain the data better than the model with continuously distributed germination rates.

RESULTS

The results of the parameter calibration suggest that the hypothesis of germination rate heterogeneity with two discrete types of spores is better supported by the data than the model

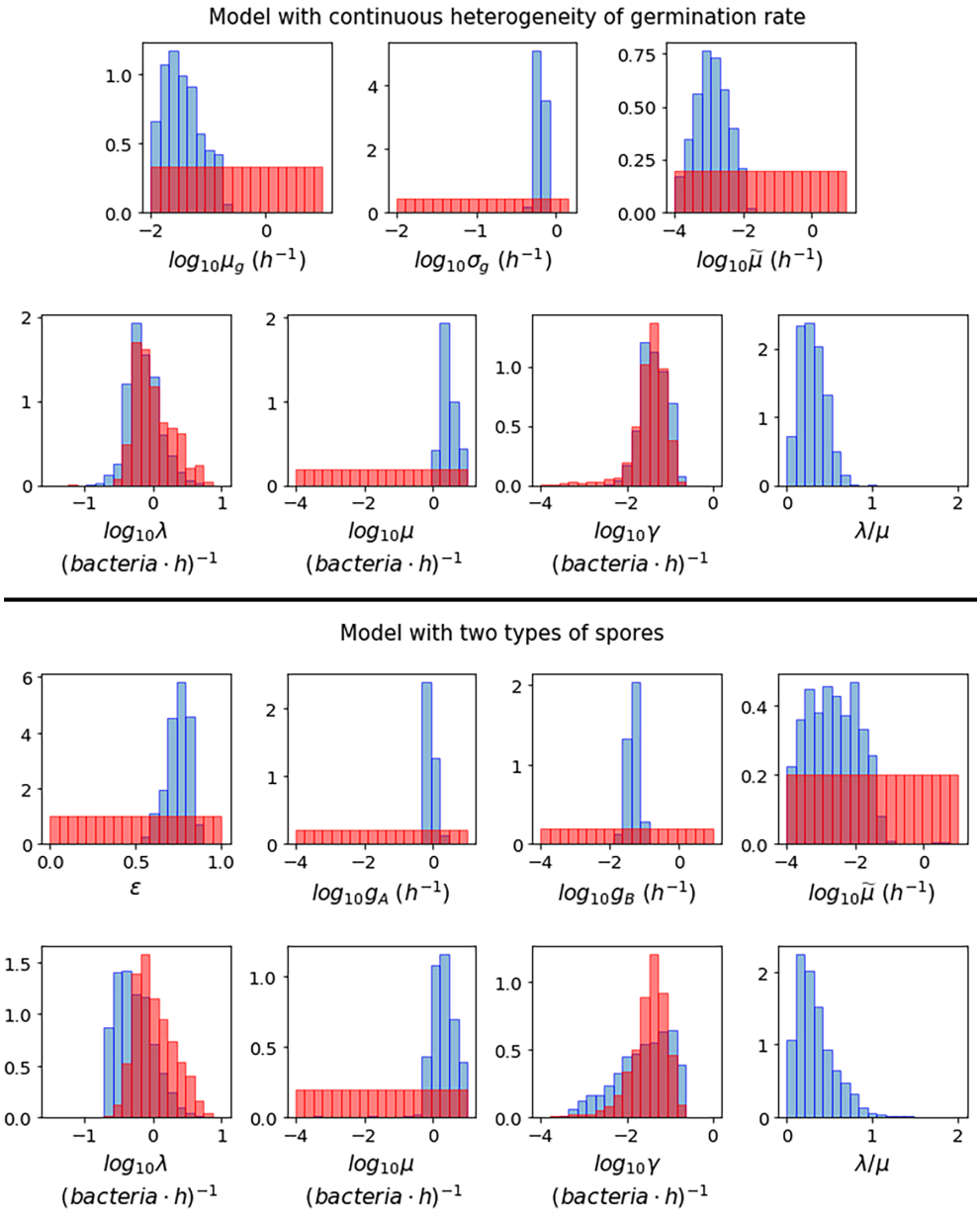


FIGURE 3 | Prior distributions (red) and posterior histograms (blue) when performing ABC-SMC for the model with continuous heterogeneity of germination rate (top), and the model with two types of spores (bottom), using data from (23) of the number of intracellular spores and bacteria at different time points for MOIs 1:2, 1:10 and 1:20.

with continuous heterogeneity. The model with two types of spores is able to describe the biphasic decay seen in the spore data, especially for MOIs 1:10 and 1:20, as well as the behaviour of the bacterial data. Therefore, in this section we will only focus on the model with two types of spores. The set of parameter values that gave the smallest distance to the data in the ABC-SMC is provided in **Table 5**. We have used these parameter values to calculate the various descriptors of the model discussed in the *Materials and Methods* section, and have also investigated the effect of the uncertainty in the parameter values indicated by the posterior distributions obtained from ABC-SMC.

In the *Spore Germination* section, we found expressions for the probabilities, $p_{1_S}(t)$ and $p_{1_{NGB}}(t)$, that an infected macrophage will contain a spore or newly germinated bacterium, respectively, at time $t \geq 0$, given that the macrophage contained a spore with germination rate g at time 0. Note that since in our model the macrophage is assumed to only contain at most one spore or newly germinated bacterium at any one time, these probabilities

TABLE 4 | Summary statistics for the posterior sample of each parameter, shown in blue in **Figure 3**, for the model with continuous heterogeneity of germination rate (top) and two types of spores (bottom).

Parameter	Units	Min	Median	Mean	Max	95% Credible Interval
Model with continuous germination rate distribution						
μ_g	h^{-1}	1.00×10^{-2}	3.30×10^{-2}	3.60×10^{-2}	2.38×10^{-1}	$(1.16 \times 10^{-2}, 1.60 \times 10^{-1})$
σ_g	h^{-1}	4.36×10^{-1}	6.27×10^{-1}	6.23×10^{-1}	8.43×10^{-1}	$(5.01 \times 10^{-1}, 7.65 \times 10^{-1})$
$\bar{\mu}$	h^{-1}	1.01×10^{-4}	1.18×10^{-3}	1.20×10^{-3}	1.73×10^{-2}	$(1.50 \times 10^{-4}, 9.58 \times 10^{-3})$
λ	$(bacteria \cdot h)^{-1}$	1.35×10^{-1}	6.93×10^{-1}	7.28×10^{-1}	4.96×10^{0}	$(2.76 \times 10^{-1}, 2.43 \times 10^{0})$
μ	$(bacteria \cdot h)^{-1}$	9.70×10^{-1}	2.55×10^{0}	2.81×10^{0}	9.66×10^{0}	$(1.31 \times 10^{0}, 7.49 \times 10^{0})$
γ	$(bacteria \cdot h)^{-1}$	3.60×10^{-3}	3.94×10^{-2}	4.15×10^{-2}	1.75×10^{-1}	$(1.07 \times 10^{-2}, 1.33 \times 10^{-1})$
Model with discrete Bernoulli germination rate distribution						
ε	–	5.29×10^{-1}	7.54×10^{-1}	7.45×10^{-1}	9.06×10^{-1}	$(5.93 \times 10^{-1}, 8.48 \times 10^{-1})$
g_A	h^{-1}	4.86×10^{-1}	8.16×10^{-1}	8.50×10^{-1}	2.74×10^{0}	$(5.37 \times 10^{-1}, 1.74 \times 10^{0})$
g_B	h^{-1}	1.18×10^{-3}	4.90×10^{-2}	4.70×10^{-2}	1.04×10^{-1}	$(2.21 \times 10^{-2}, 8.76 \times 10^{-2})$
$\bar{\mu}$	h^{-1}	1.02×10^{-4}	2.16×10^{-3}	2.27×10^{-3}	5.03×10^{0}	$(1.27 \times 10^{-4}, 4.25 \times 10^{-2})$
λ	$(bacteria \cdot h)^{-1}$	2.01×10^{-1}	4.94×10^{-1}	5.34×10^{-1}	4.28×10^{0}	$(2.14 \times 10^{-1}, 1.95 \times 10^{0})$
μ	$(bacteria \cdot h)^{-1}$	4.97×10^{-4}	1.93×10^{0}	2.00×10^{0}	9.98×10^{0}	$(5.82 \times 10^{-1}, 8.16 \times 10^{0})$
γ	$(bacteria \cdot h)^{-1}$	4.13×10^{-4}	2.97×10^{-2}	2.34×10^{-2}	2.24×10^{-1}	$(9.63 \times 10^{-4}, 1.89 \times 10^{-1})$

are equal to the mean number of spores and newly germinated bacteria inside the macrophage at time $t \geq 0$. In the *Cellular Fate: Rupture or Recovery* section we also explained how to calculate the mean number of vegetative bacteria, $B_v(t; g)$, in an infected macrophage at time $t \geq 0$, given that the macrophage contained a spore with germination rate g at time 0. One can then consider a population of S_0 independent infected cells, each containing a single spore at time 0. Assuming that each initial spore can have one of two possible germination rates (either rate g_A with probability ε, or rate g_B with rate 1 - ε), one can calculate a time course for the total mean number of intracellular spores, newly germinated bacteria, and vegetative bacteria, for a population of infected cells, split into the populations arising from each of the two types of spores. This is depicted in **Figure 5**, where the first column corresponds to the populations arising from spores of type A, the second column corresponds to the populations arising from spores of type B, and the third column shows the sum of the populations from both types of spores. The solid lines indicate the means for the estimated parameter values in **Table 5**, while the shaded regions indicate the pointwise 95% credible intervals for these means, when the uncertainty in the parameter values from the posterior is considered. The first two plots show two very different timescales for the dynamics of each kind of spore, and when these populations are added together in the third plot, one can observe the biphasic behaviour in the number of spores that is observed in the data from (23). The blue curve here indicates the prediction from the model for the total mean number of intracellular spores over time. The other two curves indicate the predictions for the mean number of intracellular newly germinated bacteria (orange), and vegetative bacteria (green), so that when these are added together, one obtains the prediction for the total number of intracellular bacteria, as shown in the predictions from the parameter calibration in **Figure 4**. The top row of plots in **Figure 5** corresponds to an initial condition of $S_0 = 30500$ spores, equal to the initial condition from the MOI 1:10 data used in the ABC-SMC. On the other hand, the bottom row corresponds to an initial condition of $S_0 = 100$. In both cases, the coloured dots indicate the sizes of the different populations over time, from a stochastic simulation of the model starting with S_0 spores. The results from the simulations show that when there are many initial spores (top row), the behaviour is very deterministic, but when the number of initial spores is relatively small (bottom row), there is much more stochasticity. This stochasticity could be relevant in *in vivo* settings, where infection might depend on a small group of spores germinating and producing a relatively small number of bacteria, as discussed in the *Discussion* section.

In the *Cellular Fate: Rupture or Recovery* section we explained how to calculate the probability density functions for the times to recovery and rupture of a macrophage initially infected with one spore. The probability density functions for the time to rupture are plotted along the top row of **Figure 6**, for the inferred parameter values in **Table 5**. From left to right, the first two plots show the density functions for the time to rupture of an infected cell containing a spore of type A and type B, respectively. We observe very different rupture timescales for each kind of spore. The third plot shows these densities on the same plot, when they are scaled by the relative frequencies of each germination rate, so that the sum of these two densities gives the overall probability density function for the rupture time of a cell infected with a single spore. Note that this probability density function does not integrate to 1, but instead the probability of rupture. Nevertheless, one can divide the density by the probability of rupture, giving the conditional density function of rupture time, shown as a solid line in the fourth plot. Also shown on the fourth plot is a histogram of the finite rupture times from 10^6 stochastic simulations of the model in **Figure 1**. On the bottom row of **Figure 6** are the analogous functions for the recovery time. Interestingly, the conditional probability density functions for the rupture and recovery times are almost identical. This is likely due to the fact that these timescales are heavily dominated by the germination time of the spore, and once the spore has germinated, rupture or recovery of the phagocyte happens relatively quickly. One can also compute the conditional mean times to rupture or recovery of a macrophage infected with a single spore, which are the means of the rightmost histograms in **Figure 6**. For the parameter

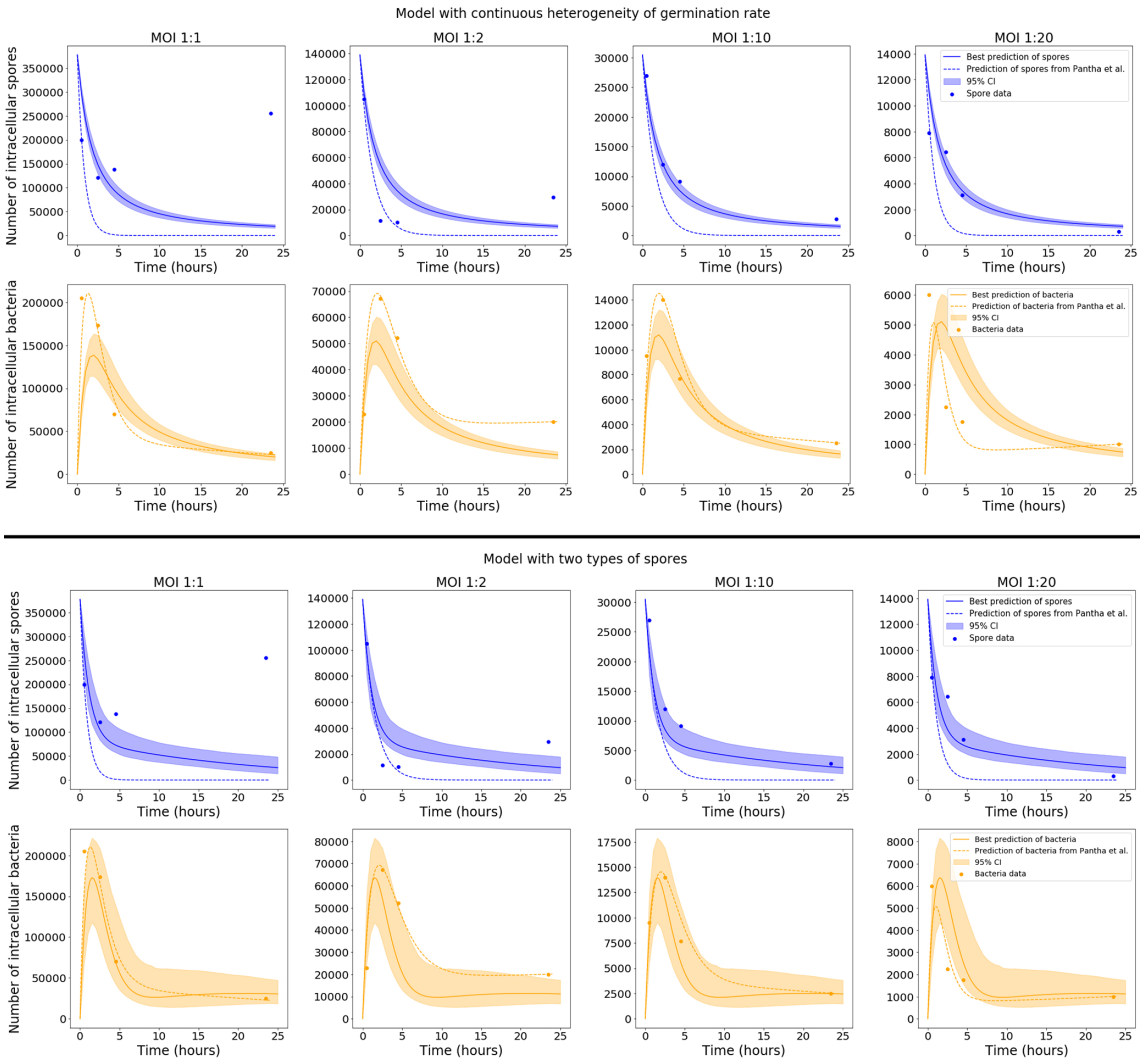


FIGURE 4 | Plots showing the best predictions (solid lines) and pointwise 95% credible intervals (shaded regions) of the time-courses for the mean number of intracellular spores and bacteria, for the model with continuous heterogeneity of germination rate (top), and the model with two types of spores (bottom). The best predictions are the model outputs obtained by using the accepted parameter set with the smallest distance to the data, and the pointwise 95% credible intervals show the uncertainty in those predictions, given the range of parameter values in the 10^3 accepted parameter sets from the final iteration of ABC-SMC. The predictions from the model by Pantha et al., using separate sets of parameter estimates for each MOI, are shown as dashed lines. For our model, only the data for MOI 1:2, 1:10 and 1:20 were used in the ABC-SMC to calibrate model parameters. The comparison of the model predictions with the MOI 1:1 data is shown here as a qualitative model validation.

values in **Table 5**, the conditional mean rupture and recovery times are approximately 11.3 hours and 11.6 hours, respectively. However, the uncertainty in the parameter values, shown in the posterior distributions, leads to uncertainty in these timescales. For instance, this is indicated by the range of conditional mean times until rupture shown in the plot on the right of **Figure 7**.

The probability distribution for the number of bacteria released by an infected macrophage, for the parameter values in **Table 5**, is provided in **Figure 7** (left). The probability that no bacteria are released by the macrophage is predicted to be 0.96, which suggests that most macrophages will be able to recover and eliminate the intracellular infection, and that we would expect only 4% of infected macrophages to eventually rupture and release bacteria. The results also indicate that when macrophages do rupture, they will only release a few bacteria, with an average of 1.6 bacteria released from a macrophage that ruptures. This is consistent with the fact that a high dose of spores is required for infection, reported to be between 8×10^3 and 5×10^4 spores (29). Cote et al. (30) explain that although macrophages kill most of the germinated bacteria that they encounter, a low percentage of bacteria survive the antimicrobial environment in the macrophage and escape to begin the extracellular infection. This is further supported by findings from Jones et al. (31), who observed that after infection of a

TABLE 5 | Parameter values that gave the smallest distance between the two types of spores model and the data from (23) in the ABC-SMC analysis.

Parameter	Units	Description	Value
ε	–	Probability that a given spore is of type A	0.778846
g_A	h^{-1}	Germination and maturation rate of spores of type A	0.894274
g_B	h^{-1}	Germination and maturation rate of spores of type B	0.046794
$\bar{\mu}$	h^{-1}	Death rate of newly germinated bacteria	0.003502
λ	$(bacteria \cdot h)^{-1}$	Replication rate of vegetative bacteria	0.643111
μ	$(bacteria \cdot h)^{-1}$	Death rate of vegetative bacteria	1.637989
γ	$(bacteria \cdot h)^{-1}$	Rupture rate	0.043792

guinea pig with 10^7 spores, 99% of the germinated spores were killed within an hour, but the 1% that survived managed to replicate extracellularly and ultimately reached a concentration of 10^8 bacteria/ml in the blood at the time of death.

Table 5 reports our best prediction for the parameter values according to the distance in the ABC-SMC. Yet, the advantage of a Bayesian approach in the parameter calibration is that it quantifies the uncertainty in the parameter values, which translates into the uncertainty in the descriptors of the model. The scatter plot on the right of **Figure 7** shows the probability of rupture plotted against the average of the conditional rupture size distribution, for each parameter set in the posterior sample shown in **Figure 3** (bottom), with the lines indicating the corresponding values for the parameter set in **Table 5**. There is a positive correlation between these two descriptors, indicating that if macrophages are more likely to rupture, they are also likely to release more bacteria when they do rupture. Parameter sets leading to a very small probability of rupture will likely correspond to death rates much larger than the replication rate. Conversely, parameter sets leading to a larger probability of rupture correspond to bacterial death rates closer to the replication rate. This would allow for greater bacterial replication on average before rupture, and in turn a larger average rupture size. The colour of the points on this scatter plot indicates the conditional mean time until rupture for each parameter set. This illustrates the uncertainty of the timescale for rupture, with the possible mean rupture times from the posterior distributions ranging from around 7.5 to 25 hours. Furthermore, it is possible to find pairs of parameter sets that give differing average rupture times, but with similar probabilities of rupture and conditional average rupture sizes.

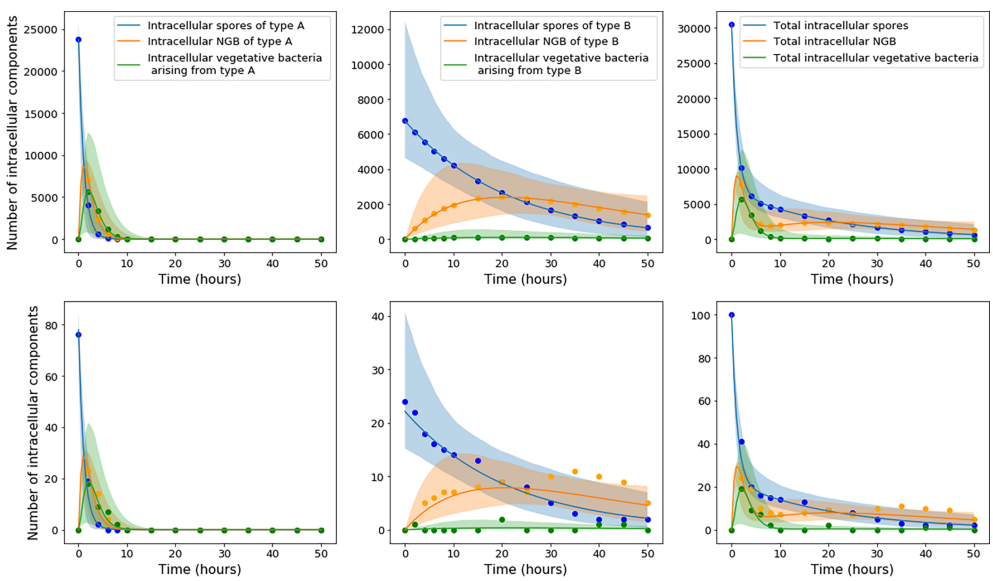


FIGURE 5 | The top row of plots corresponds to a population of $S_0 = 30500$ infected cells, each containing a single spore at time 0, whereas the bottom row corresponds to an initial condition of $S_0 = 100$. Left: The mean number of type A spores, $\varepsilon S_0 p_{1_S}(t; g = g_A)$, type A newly germinated bacteria, $\varepsilon S_0 p_{1_{NGB}}(t; g = g_A)$, and vegetative bacteria, $\varepsilon S_0 B_V(t; g_A) = \varepsilon S_0 \sum_{i=1}^{\infty} i\, p_i(t; g = g_A)$, arising from the type A spores in the infected macrophages. Centre: The analogous functions for the populations arising from the initial spores with germination rate g_B. Right: The overall mean number of spores, $S_0 S(t)$, newly germinated bacteria, $S_0 B_{NGB}(t)$, and vegetative bacteria, $S_0 B_v(t)$, obtained by adding together the populations for each type of spore. The solid lines indicate the means for the estimated parameter values in **Table 5**, while the shaded regions indicate the pointwise 95% credible intervals for these means, when the uncertainty in the parameter values from the posterior distributions is taken into account. The equations used to compute these curves were Eqs. (5), (6), (11), (12), (41), and (42). The dots show values for the size of the different populations over time from a single stochastic simulation beginning with S_0 spores.

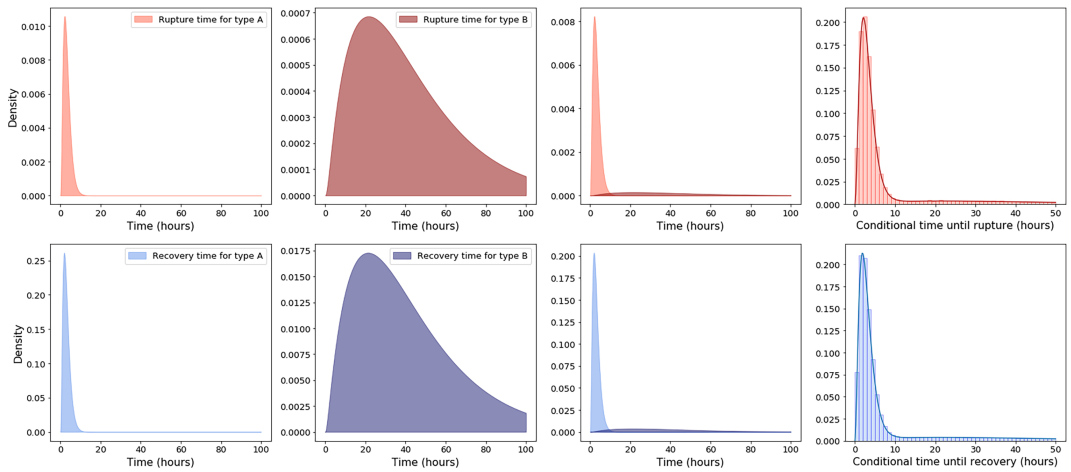


FIGURE 6 | Top row, from left to right, the first two plots show the probability density functions for the rupture time of a macrophage infected with a spore of type A and type B, respectively, given by $f_{T^R_{1_S}}(t; g_A)$ and $f_{T^R_{1_S}}(t; g_B)$. The third plot shows these densities on the same plot, when they are scaled by the relative frequencies of each germination rate: $\varepsilon f_{T^R_{1_S}}(t; g_A)$ and $(1-\varepsilon) f_{T^R_{1_S}}(t; g_B)$. The fourth plot shows as a solid line the probability density function for the rupture time of a macrophage infected with a single spore, conditioned on rupture occurring, which is given by $f_{T^R_{1_S}}(t)/r^R_{1_S}$. Also shown on the fourth plot is a histogram of the finite rupture times from 10^6 stochastic simulations of the model in **Figure 1**. Plots on the bottom row correspond to the analogous densities for the time to recovery of an infected macrophage. The estimated parameter values in **Table 5** were used to compute these functions.

DISCUSSION

We propose a stochastic model for the dynamics of *B. anthracis* spores and bacteria inside an infected phagocyte. One of the main features of our model is the consideration of heterogeneity in the germination rate of spores. Two hypotheses were considered to characterise this heterogeneity. The first hypothesis was that the germination rate is continuously distributed in a population of spores and follows a truncated normal distribution. The second hypothesis was that the spore population can be split into two kinds that germinate at different rates. We carried out parameter calibration, for each hypothesis, by means of Approximate Bayesian Computation Sequential Monte Carlo (ABC-SMC) (22). Our results suggest that the discrete germination hypothesis is better supported by the data, since the model with this distribution of germination rates allows us to account for the biphasic decline seen in the spore counts, as well as the observed behaviour of the bacterial counts. This assumption of two types of spores also agrees with experimental evidence showing that in some *Bacillus* spore populations, a subset of the spores germinate much more slowly than the average spore, and are termed superdormant (20). This leads to qualitatively different predictions for the mean number of spores over time in a population of *in vitro* cells, compared to previous theoretical predictions made in (15), as shown in **Figure 4**. Although our posterior estimated values of the germination rate of spores of type A are similar to the germination rates predicted in (15), our model predicts that a subset of spores will germinate much more slowly than this. We note that the "toy" discrete distribution with two rates, g_A and g_B, considered here, is a first step and more complex descriptions of the germination rate heterogeneity will be considered in future work. Experimental quantification of germination times would allow us to explore and potentially validate these more complex models. A limitation of our model is that the same rate is considered for each step of the germination-maturation process, compared to the model by Pantha et al., where a separate rate, m, was considered for the maturation step. However, Pantha et al. did not calibrate this rate m, and instead performed a sensitivity analysis. Since we are limited in the available experimental data, it would be difficult to calibrate different rates for each germination-maturation step. However, if further data were to become available which allowed one to distinguish between newly germinated and vegetative bacteria, then a separate rate could be incorporated for the maturation step of the germination process. This would also allow for more complicated distributions to account for heterogeneity in the germination and maturation rates.

Another important feature of our model is the consideration of rupture of infected phagocytes. This means that different behaviours can be described by our model compared to the model by Pantha et al., since in our model there is a chance that some intracellular bacteria will survive the microbicidal environment of the phagocyte and cause the cell to rupture. In the experiment by Kang et al., if bacteria were released into the extracellular medium then they would have been washed away before intracellular numbers of spores and bacteria were measured, and hence the decrease of intracellular bacteria seen in the data may not have been purely due to macrophage-induced killing of bacteria but may have been due to the release of intracellular bacteria from dying cells. Further data including information about macrophage rupture versus survival would be needed in order to disentangle these processes. Our stochastic model (see **Figure 1**) has allowed us to compute the probability that an infected cell will eliminate the infection and recover, and the probability that an infected cell will rupture and release its bacterial content. We have also computed the mean time for an

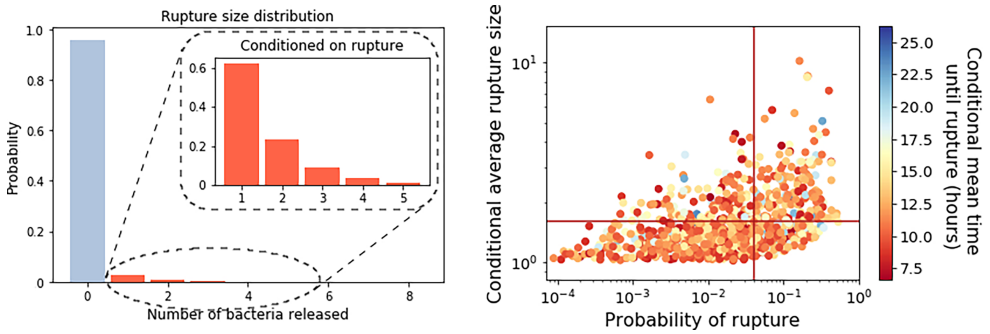


FIGURE 7 | Left: The best predicted rupture size distribution for the model with two types of spores, computed using Eq. (49), with the estimated parameter values from **Table 5**. Inset is the conditional rupture size distribution, for the number of bacteria released by a macrophage infected with a single spore, given that it ruptures rather than recovers. Right: Scatter plot of the probability of rupture against the expected rupture size (conditioned on rupture occurring), for each parameter set in the posterior distribution. Lines indicate the values that correspond to the parameter set from **Table 5**. The colour of the points indicates the conditional mean time to rupture for each parameter set considered.

infected cell to reach one of these two fates, conditioned on the event occurring. The probability distribution for the number of bacteria released by an infected macrophage has also been calculated. By calibrating the parameters using *in vitro* experimental data, the rupture size distribution shown in **Figure 7** is able to capture the fact that the majority of spores taken up by macrophages are likely to be eliminated by the host cell, releasing no bacteria, but a few germinated spores may survive phagocytosis, leading to death of the host cell and release of a small number of bacteria. This is in agreement with recent experimental work (13). Although we have parametrised our model with data from a study that used macrophages, there is also evidence to suggest that dendritic cells play a role in the early infection stages of anthrax (10). Therefore, it will be important to also consider this cell type in future. If *in vitro* infection data for dendritic cells becomes available, it would be relevant to re-parametrise the model with such data to investigate the differences between the roles of the two host cell types in anthrax disease. For example (32) indicates that dendritic cells may not be as capable as macrophages in their abilities to reduce bacterial numbers.

The intracellular model presented here could be used in mechanistic within-host models of anthrax infection to describe the dynamics of *B. anthracis* within the lung and lymph nodes of an individual, following inhalation of some initial dose of spores, such as the one by Day et al. in (17). The stochastic nature of the intracellular model presented here could allow such within-host models to consider inter-phagocyte variability in rupture size by incorporating the rupture size distribution into the within-host infection dynamics. Heterogeneity of the rupture size has been shown to be important in a similar model for the pathogen *F. tularensis* (25). Furthermore, a within-host model could be linked to dose-response data and used to investigate the individual infection risk given an initial inhalational dose. A standard approach in dose-response assessment is the use of single-hit models. These models assume that when an individual is infected with a pathogen, the organisms act independently in the host so that the probability that any one organism in the initial dose produces an eventual infection is independent of the size of the dose, and the probability of infection is equivalent to the probability that at least one of the organisms in the initial dose will lead to an infection. For example, in the simple exponential model, for an average initial dose D, the probability of infection is given by, $I(D; r) = 1 - e^{-rD}$, where r is the probability that a single organism will produce a response. This exponential model can be fitted to data in order to estimate r. **Figure 8** shows the exponential dose-response curve compared to the Altboum et al. guinea-pig Vollum strain dose-response data set (33). For this data set the value of r that gives the best fit is around $r = 3.31 \times 10^{-5}$. However, the specific value of r obtained varies greatly, depending on the data set used to calibrate the dose-response curve. In particular, different anthrax dose-response data sets lead to a range for the ID_{50} (the dose of spores such that the probability of infection is equal to 0.5) of about 10^3 to 10^5. These differences could be due to changes in susceptibility for each animal species used or in virulence for each anthrax strain used (34).

Dose-response models for anthrax already exist, such as the competing risks model developed by Brookmeyer et al. (35). This is an example of a single-hit model in which r is taken to be, $r = \frac{\lambda}{\lambda+\theta}$, where λ is the germination rate of spores and θ the rate at which spores are cleared from the lungs. The hypothesis of the competing risks model for anthrax is that if a single spore survives ingestion by a macrophage and successfully germinates without getting cleared, then the resulting bacterium will be certain to cause an infection. However, there is evidence that neutrophils can kill vegetative anthrax bacteria (36), which means that once a spore in an infected cell has germinated and a bacterium is released, there is no guarantee that the bacterium will survive and cause infection. Another simplification used in the competing risks model is that germination and clearance of spores are both assumed to be exponential processes. However, this might not be the case. For instance, in our intracellular model, as well as the one by Pantha et al., the consideration of the newly germinated bacterium means that the total germination-maturation time is non-exponential.

The competing risks model involves parameters for the two competing processes of spore germination and spore clearance

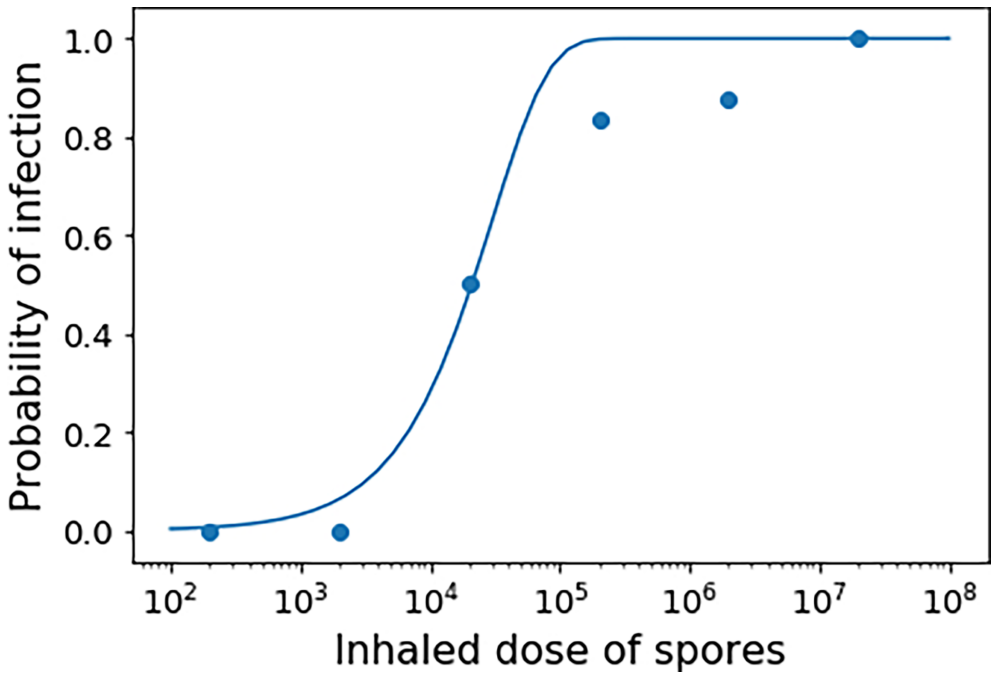


FIGURE 8 | Exponential dose-response curve fit to the Altboum et al. guinea-pig Vollum strain dose-response data set (33). Here the probability that one inhaled spore will cause a response is set to $r = 3.31 \times 10^{-5}$.

but does not explicitly consider macrophage rupture or intracellular bacterial dynamics. Hence, a fully mechanistic model, such as the intracellular model proposed here, could allow one to construct more detailed dose-response approaches which go beyond the simple competing risks assumptions, and explore the stochasticity of the rupture events, and the possibility that even if a few bacteria are generated, infection might not occur if these few bacteria are killed. **Figure 9** shows a "toy" representation of the possible fates of each spore in the early stages of inhalational anthrax infection in the lungs and lymph nodes. After inhalation, these spores must be transported through the respiratory system in order to reach the alveoli of the lungs, where they have the chance to cause an infection. Hence the initial dose of spores that is delivered to this area of the lungs may be much smaller than the original inhaled dose (37). Since the delivered dose may be very small compared to the exposed dose, the stochastic nature of the intracellular dynamics illustrated in **Figure 5** can become important. For the model by Brookmeyer et al. (35), the transport dynamics of spores through the respiratory system would be included in the clearance rate, θ. Alternatively, to explicitly account for the fact that some of the inhaled spores will not be delivered to the alveolated region of the lung, and to differentiate this from the clearance of spores by macrophages, one can assume that each inhaled spore has some probability, ϕ, of being deposited in the alveoli (18).

We assume that a spore deposited in the alveoli will be ingested by a host phagocyte, which will then migrate to the lymph nodes and either recover or rupture, releasing $n \in \mathbb{N}$ bacteria with probability $R^n_{1_S}$ calculated from the intracellular model. When bacteria are released, each extracellular bacterium can be killed by host immune cells with probability $p < 1$, or proliferate extracellularly with probability 1 - p. Each inhaled spore has two possible fates: the response state and the clearance state. We assume that a spore will reach the response state if it manages to lead to a population of $M \in \mathbb{N}$ bacteria in the lymph node. On the other hand, the infection from an inhaled spore can be cleared in one of three ways: if the spore is not deposited in the alveolated region of the lungs, if the spore is phagocytosed but the infected cell recovers rather than ruptures, or if the infected cell ruptures but the population of bacteria released from the cell becomes extinct before reaching the threshold M. From these assumptions, r is the probability that the fate of a single inhaled spore is the response state rather than the clearance state, and we can construct a formula for r that takes into account all of the mechanisms in the intracellular model, and the possibility that the few bacteria released from a rupturing cell may be killed by the host immune defences before they are able to proliferate to a sufficient number to cause a response. In this case, the probability that infection is established by a single inhaled organism can be given by

$$r = \begin{cases} \phi\left(\sum_{n=1}^{M-1} R^n_{1_S} \frac{1-\left(\frac{p}{1-p}\right)^n}{1-\left(\frac{p}{1-p}\right)^M} + \sum_{n=M}^{+\infty} R^n_{1_S}\right), & \text{if } p \neq 0.5, \\ \phi\left(\sum_{n=1}^{M-1} R^n_{1_S} \frac{n}{M} + \sum_{n=M}^{+\infty} R^n_{1_S}\right), & \text{if } p = 0.5. \end{cases}$$

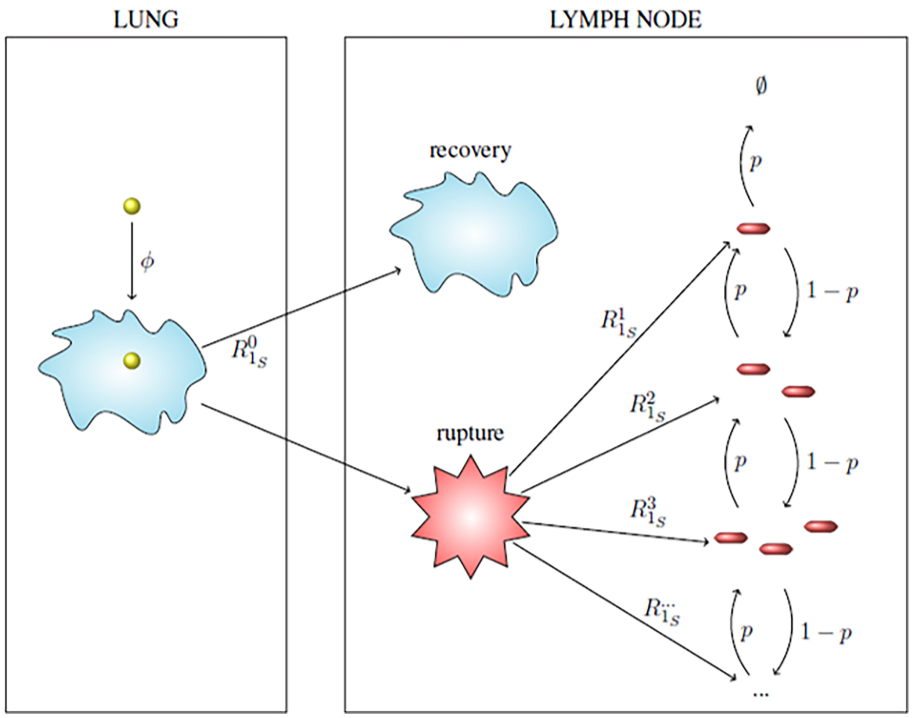


FIGURE 9 | Toy model for the possible fates of a single spore in the early stages of inhalational anthrax infection. Within the lung, an inhaled spore (yellow ball) becomes deposited with probability ϕ and is ingested by a host phagocyte. The intracellular spore germinates and the phagocyte might kill the germinated bacterium or the bacterium may survive the antimicrobial environment, replicate and cause the phagocyte to rupture, as described in the intracellular model. The infected phagocyte migrates to the lymph nodes and either recovers or ruptures and releases some bacteria, according to the probabilities calculated with the intracellular model, $R^n_{1_S}$. The population of bacteria released follows a birth and death process, where an extracellular bacterium may be killed by host immune cells with probability p or a bacterium will replicate extracellularly with probability 1 - p, until either there are no extracellular bacteria remaining, or the population of extracellular bacteria reaches a threshold M.

The competing risks hypothesis is equivalent to taking $M = 1$. A value of M greater than one allows some bacteria to die extracellularly in the early stages of infection. We note that the parameter calibration in our intracellular model incorporates parameter uncertainty, encoded in the posterior distributions (see **Figure 3**). Thus, different parameter sets will lead to different predicted rupture size distributions. Furthermore, the deposition probability, ϕ, depends on a number of different factors, such as breathing rate, which will vary between individuals and depends on the level of physical exertion at the time of exposure to the spores. In a dose response model for Q fever (a bacterial infection caused by *Coxiella burnetii*), Heppell et al. in (38) approximate a distribution for the probability of deposition, ϕ, using the Multiple-Path Particle Dosimetry Model (MPPD) software package. Here, we have considered a wide range of values for the deposition probability, between $\phi = 10^{-4}$ and $\phi = 0.5$. In **Figure 10** we show a wide range of parameter sets involving the value p, the deposition probability ϕ, the threshold value M, and the average rupture size derived from the intracellular model, which all lead to the same value of $r = 3.31 \times 10^{-5}$. This is the value of r obtained by fitting the exponential dose-response curve to the data in **Figure 8**. In this exploration, we are not concerned with exact parameter values, since these will strongly depend on the dose-response data set used to calibrate them. Our aim is to illustrate that, depending on the rupture size distribution used, we can fit the dose-response data with a large number of combinations of the parameters M, ϕ and p. In general, as the threshold value M increases, and the average rupture size and deposition probability ϕ decrease, the required value of p decreases, since more replication will be required to reach the desired threshold M.

Our results show the potential to make use of mechanistic intracellular models and dose-response data sets to nest with within-host infection dynamics. If additional experimental data sets were available, a stochastic within-host model could be developed, similar to recent models for *F. tularensis* infection proposed in Refs (16, 25). This approach would not only allow one to compute the probability of a host response to infection, but also the timescale of symptom onset. In order to characterise the timescale of the response, Brookmeyer et al. consider an exponentially distributed delay between the first spore germination and symptom onset, to account for bacterial growth. However, a limitation of this approach is that in considering this delay to have the same distribution for every dose, their model does not incorporate the fact that a higher number of germinating spores would be expected to replicate and produce toxins faster, causing the time to response to be shorter. Wilkening modified the model

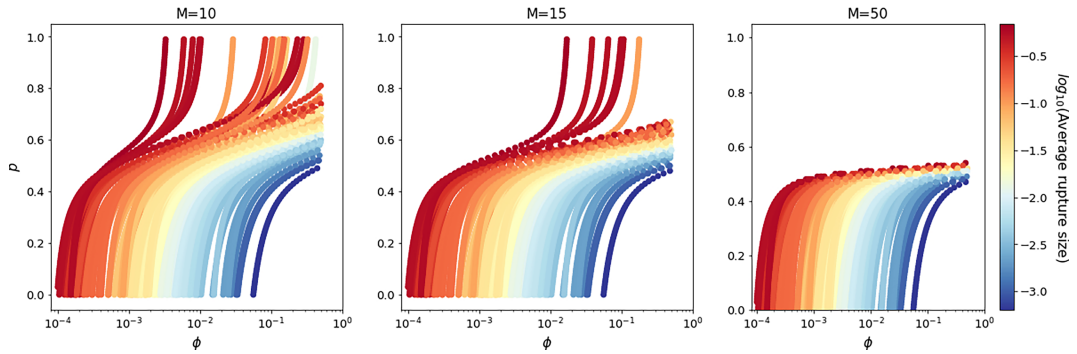


FIGURE 10 | Combination of parameter values for p and ϕ that give $r = 3.31 \times 10^{-5}$, for three different values of the threshold number of bacteria, $M = 10, 15, 50$. The combinations are shown for different rupture size distributions, sampled from the posterior distribution of the intracellular model. The average of each rupture size distribution is indicated by the colour of the points.

developed by Brookmeyer et al. to address this issue by including a dose-dependent log-normal distribution to represent the bacterial growth phase (39). Since the data in **Figure 8** only provides the probability of infection from exposure to different doses of spores, we have considered a discrete-time process for the replication and death of extracellular bacteria, where the value of p represents how likely bacteria are to die in the early stages of infection (36). However, until more data sets are available, it is not possible to estimate specific within-host parameter values, since different combinations of parameter values can provide the same value of r and hence the same dose-response curve. Further experimental work, such as *in vivo* data of the number of spores and bacteria in the lungs and lymph nodes over time, would be needed to calibrate the parameters in a within-host model (or mechanistic dose-response model) of inhalational anthrax infection. Data on anthrax toxin production and stability would also be required in a within-host model of anthrax infection, since the toxins are a key component of pathogenesis. Future work will focus on including the role of toxins in a mathematical model of inhalational anthrax infection.

A restrictive assumption of the intracellular model introduced here is that it only considers a macrophage infection by a single phagocytosed spore. However at higher exposures, it is possible that some macrophages may phagocytose more than one spore. Thus, it would be necessary to include this consideration in the intracellular model. Furthermore, we proposed a linear death rate of intracellular bacteria to keep the model analytically tractable. There is evidence that macrophages with a low intracellular bacterial burden are much more efficient at killing bacteria than those with a higher burden (2, 23, 30). We could generalise our model to include a non-linear, density-dependent death rate of intracellular bacteria, similar to the burden-dependent killing function used in (15). Our model predicts that the intracellular burden of a cell initially infected with a single spore will remain very low, so the inclusion of a density-dependent death rate may be more appropriate when considering higher multiplicities of infection, when each macrophage will phagocytose more than one spore.

To calibrate the parameters of our model, we made use of experimental measurements from a study by Kang et al. (23), of mouse peritoneal macrophage infection with the attenuated non-capsule-producing Sterne strain of *B. anthracis*. This enabled us to mathematically describe a system which can be characterised easily in the laboratory. In fact, one of the more commonly used animal models of anthrax is the AJ mouse model infected with Sterne strain (40). Ideally, for modelling human inhalational anthrax, spores from a fully virulent strain and cells more similar to a human alveolar macrophage would be used. However, this type of data for anthrax is extremely limited. The more clinically relevant alveolar macrophage is complex to isolate and culture, so in the same way that mice are used as a surrogate for primates, peritoneal macrophages are used as a surrogate for the lung's resident phagocytes. Moreover, the Sterne strain is often used in laboratory settings since it poses a reduced infection risk to laboratory workers, and research with virulent strains of *B. anthracis*, such as the Ames strain, requires enhanced biosafety laboratories (40). However, it can be generally agreed that its value has limitations when modelling disease (40). The capsule is known to protect extracellular bacteria from phagocytosis (14, 41), and thus should be considered when modelling the extracellular dynamics of anthrax infection, which has not been explicitly modelled here. It is possible that the capsule also plays a role in protecting emerging intracellular bacteria from the antimicrobial environment of the host cell, since germinating spores are able to quickly produce and coat themselves in the capsule (42, 43). However, it has been shown that macrophages are still able to kill intracellular bacteria even when they are from a strain that is coated in an antiphagocytic capsule, like the Ames strain (44), and the capsule does not seem to be fully protective against the bactericidal activity of macrophages (45). Therefore, re-parametrising the intracellular model using data from a fully virulent strain would be extremely useful in determining whether the capsule has a significant effect on the intracellular dynamics and fate of a phagocytosed spore.

In conclusion, we have developed and analysed a novel stochastic mathematical model of the intracellular bacterial

dynamics of a macrophage infected with a single anthrax spore. By calibrating the model with experimental data, we have found support for a discrete Bernoulli distribution of the spore germination rate, which provides independent evidence for the role of superdormant spores (20, 21). This is both of clinical and biological interest. From a clinical perspective, it indicates the importance to maintain antibiotic dosing for long periods, given the potential for the slow germinating spores to contribute to the characteristic persistence of spores in the lungs after inhalational exposure (46). From a biological perspective, it demonstrates that there might be selective pressure for spores to distribute their germination rates in a heterogeneous manner. This might protect spore populations by ensuring that a reservoir of spores is maintained in case of accidental germination in environments not suitable for growth (47). The results of our calibrated model also predict, in agreement with experimental findings, that many macrophages may be able to recover and resolve the bacterial infection, provided their initial intracellular burden is low. Yet, our results predict a non-zero but low risk of cellular rupture, leading to the release of bacteria from the cell. We believe the intracellular stochastic model proposed here will pave the way to an extension to *in vivo* infection settings and thus, to improve within-host dynamics models.

SUPPLEMENTAL DATA

The **Supplementary Material** contains a detailed explanation of how the data in **Table 1** can be used to find an estimate for the phagocytosis rate of spores. It also contains additional details of the preliminary parameter fitting to the Akoachere et al. data in (26), which allowed us to estimate a prior distribution for the bacteria replication and infected macrophage rupture rates. Finally, the derivation for the likelihood that we have used in the formula for Akaike's Information Criterion (AIC) has been provided, since it was used to compare the goodness-of-fit between the different germination rate hypotheses considered.

DATA AVAILABILITY STATEMENT

Computer codes (in Python) for reproducing the results in **Figures 4–8**, and **10** are available in Williams et al. (48).

AUTHOR CONTRIBUTIONS

All authors conceived the idea and contributed to develop the mathematical models. BW, ML-G, and GL carried out the analysis of the summary statistics in the *Materials and Methods* section. BW and ML-G carried out the Bayesian statistical analysis to carry out parameter calibration with experimental data sets, and designed the results in the *Results* section. BW developed the numerical codes and prepared all the figures. BW and ML-G wrote the first version of the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

BW is supported by an EPSRC CASE studentship, project reference 2345914, in partnership with Dstl under contract number DSTLX-1000142022.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2021.688257/full#supplementary-material

REFERENCES

1. Setlow P. Spore Germination. *Curr Opin Microbiol* (2003) 6:550–6. doi: 10.1016/j.mib.2003.10.001
2. Cote CK, Welkos SL, Bozue J. Key Aspects of the Molecular and Cellular Basis of Inhalational Anthrax. *Microbes Infect* (2011) 13:1146–55. doi: 10.1016/j.micinf.2011.07.005
3. Goel AK. Anthrax: A Disease of Biowarfare and Public Health Importance. *World J Clin Cases: WJCC* (2015) 3:20. doi: 10.12998/wjcc.v3.i1.20
4. Powell JD, Hutchison JR, Hess BM, Straub TM. Bacillus Anthracis Spores Germinate Extracellularly at Air–Liquid Interface in an *In Vitro* Lung Model Under Serum-Free Conditions. *J Appl Microbiol* (2015) 119:711–23. doi: 10.1111/jam.12872
5. Guidi-Rontani C. The Alveolar Macrophage: The Trojan Horse of Bacillus Anthracis. *Trends Microbiol* (2002) 10:405–9. doi: 10.1016/S0966-842X(02)02422-8
6. Guidi-Rontani C, Weber-Levy M, Labruyère E, Mock M. Germination of Bacillus Anthracis Spores Within Alveolar Macrophages. *Mol Microbiol* (1999) 31:9–17. doi: 10.1046/j.1365-2958.1999.01137.x
7. Dixon TC, Fadl AA, Koehler TM, Swanson JA, Hanna PC. Early Bacillus Anthracis–Macrophage Interactions: Intracellular Survival and Escape. *Cell Microbiol* (2000) 2:453–63. doi: 10.1046/j.1462-5822.2000.00067.x
8. Brittingham KC, Ruthel G, Panchal RG, Fuller CL, Ribot WJ, Hoover TA, et al. Dendritic Cells Endocytose Bacillus Anthracis Spores: Implications for Anthrax Pathogenesis. *J Immunol* (2005) 174:5545–52. doi: 10.4049/jimmunol.174.9.5545
9. Cleret A, Quesnel-Hellmann A, Mathieu J, Vidal D, Tournier JN. Resident CD11c+ Lung Cells are Impaired by Anthrax Toxins After Spore Infection. *J Infect Dis* (2006) 194:86–94. doi: 10.1086/504686
10. Cleret A, Quesnel-Hellmann A, Vallon-Eberhard A, Verrier B, Jung S, Vidal D, et al. Lung Dendritic Cells Rapidly Mediate Anthrax Spore Entry Through the Pulmonary Route. *J Immunol* (2007) 178:7994–8001. doi: 10.4049/jimmunol.178.12.7994
11. Guarner J, Jernigan JA, Shieh WJ, Tatti K, Flannagan LM, Stephens DS, et al. Pathology and Pathogenesis of Bioterrorism-Related Inhalational Anthrax. *Am J Pathol* (2003) 163:701–9. doi: 10.1016/S0002-9440(10)63697-8
12. Moayeri M, Leppla SH, Vrentas C, Pomerantsev AP, Liu S. Anthrax Pathogenesis. *Annu Rev Microbiol* (2015) 69:185–208. doi: 10.1146/annurev-micro-091014-104523
13. Banks DJ, Barnajian M, Maldonado-Arocho FJ, Sanchez AM, Bradley KA. Anthrax Toxin Receptor 2 Mediates Bacillus Anthracis Killing of Macrophages Following Spore Challenge. *Cell Microbiol* (2005) 7:1173–85. doi: 10.1111/j.1462-5822.2005.00545.x
14. Sharma S, Bhatnagar R, Gaur D. Bacillus Anthracis Poly-Λ-D-Glutamate Capsule Inhibits Opsonic Phagocytosis by Impeding Complement Activation. *Front Immunol* (2020) 11:462. doi: 10.3389/fimmu.2020.00462
15. Pantha B, Cross A, Lenhart S, Day J. Modeling the Macrophage-Anthrax Spore Interaction: Implications for Early Host-Pathogen Interactions. *Math Biosci* (2018) 305:18–28. doi: 10.1016/j.mbs.2018.08.010
16. Carruthers J, Lythe G, López-García M, Gillard J, Laws TR, Lukaszewski R, et al. Stochastic Dynamics of Francisella Tularensis Infection and Replication. *PloS Comput Biol* (2020) 16:e1007752. doi: 10.1371/journal.pcbi.1007752

17. Day J, Friedman A, Schlesinger LS. Modeling the Host Response to Inhalation Anthrax. *J Theor Biol* (2011) 276:199–208. doi: 10.1016/j.jtbi.2011.01.054
18. Pratt A, Bennett E, Gillard J, Leach S, Hall I. Dose–response Modeling: Extrapolating From Experimental Data to Real-World Populations. *Risk Anal* (2020) 41(1):67–78. doi: 10.1111/risa.13597
19. Ruthel G, Ribot WJ, Bavari S, Hoover TA. Time-Lapse Confocal Imaging of Development of Bacillus Anthracis in Macrophages. *J Infect Dis* (2004) 189:1313–6. doi: 10.1086/382656
20. Setlow P. Summer Meeting 2013–When the Sleepers Wake: The Germination of Spores of Bacillus Species. *J Appl Microbiol* (2013) 115:1251–68. doi: 10.1111/jam.12343
21. Setlow P. Germination of Spores of Bacillus Species: What We Know and do Not Know. *J Bacteriol* (2014) 196:1297–305. doi: 10.1128/JB.01455-13
22. Toni T, Welch D, Strelkowa N, Ipsen A, Stumpf MP. Approximate Bayesian Computation Scheme for Parameter Inference and Model Selection in Dynamical Systems. *J R Soc Interface* (2009) 6:187–202. doi: 10.1098/rsif.2008.0172
23. Kang TJ, Fenton MJ, Weiner MA, Hibbs S, Basu S, Baillie L, et al. Murine Macrophages Kill the Vegetative Form of Bacillus Anthracis. *Infect Immun* (2005) 73:7495–501. doi: 10.1128/IAI.73.11.7495-7501.2005
24. Karlin S, Tavaré S. Linear Birth and Death Processes With Killing. *J Appl Probability* (1982) 19:477–87. doi: 10.1017/S0021900200036986
25. Carruthers J, Lopez-Garcia M J, Gillard J, Laws T R, Lythe G, Molina-Paris C. A Novel Stochastic Multi-Scale Model of Francisella Tularensis Infection to Predict Risk of Infection in a Laboratory. *Front Microbiol* (2018) 9:1165. doi: 10.3389/fmicb.2018.01165
26. Akoachere M, Squires RC, Nour AM, Angelov L, Brojatsch J, Abel-Santos E. Identification of an *In Vivo* Inhibitor of Bacillus Anthracis Spore Germination. *J Biol Chem* (2007) 282:12112–8. doi: 10.1074/jbc.M611432200
27. Kalns J, Morris J, Eggers J, Kiel J. Delayed Treatment With Doxycycline has Limited Effect on Anthrax Infection in BLK57/B6 Mice. *Biochem Biophys Res Commun* (2002) 297:506–9. doi: 10.1016/S0006-291X(02)02226-X
28. Goggans PM, Cao L, Henderson RW. Assigning Priors for Parameters Constrained to a Simplex Region, in: *AIP Conference Proceedings*, Vol. 1636. American Institute of Physics (2014). pp. 94–9.
29. Oyston PC, Sjöstedt A, Titball RW. Tularaemia: Bioterrorism Defence Renews Interest in Francisella Tularensis. *Nat Rev Microbiol* (2004) 2:967–78. doi: 10.1038/nrmicro1045
30. Cote CK, DiMezzo TL, Banks DJ, France B, Bradley KA, Welkos SL. Early Interactions Between Fully Virulent Bacillus Anthracis and Macrophages That Influence the Balance Between Spore Clearance and Development of a Lethal Infection. *Microbes Infect* (2008) 10:613–9. doi: 10.1016/j.micinf.2008.02.006
31. Jones WI, Klein F, Walker JS, Mahlandt BG, Dobbs JP, Lincoln RE. *In Vivo* Growth and Distribution of Anthrax Bacilli in Resistant, Susceptible, and Immunized Hosts. *J Bacteriol* (1967) 94:600–8. doi: 10.1128/jb.94.3.600-608.1967
32. Shetron-Rama LM, Herring-Palmer AC, Huffnagle GB, Hanna P. Transport of Bacillus Anthracis From the Lungs to the Draining Lymph Nodes Is a Rapid Process Facilitated by Cd11c+ Cells. *Microbial Pathogenesis* (2010) 49:38–46. doi: 10.1016/j.micpath.2010.02.004
33. Altboum Z, Gozes Y, Barnea A, Pass A, White M, Kobiler D. Postexposure Prophylaxis Against Anthrax: Evaluation of Various Treatment Regimens in Intranasally Infected Guinea Pigs. *Infect Immun* (2002) 70:6231–41. doi: 10.1128/IAI.70.11.6231-6241.2002
34. Haas CN. On the Risk of Mortality to Primates Exposed to Anthrax Spores. *Risk Anal* (2002) 22:189–93. doi: 10.1111/0272-4332.00028
35. Brookmeyer R, Johnson E, Barry S. Modelling the Incubation Period of Anthrax. *Stat Med* (2005) 24:531–42. doi: 10.1002/sim.2033
36. Liu JZ, Ali SR, Bier E, Nizet V. Innate Immune Interactions Between Bacillus Anthracis and Host Neutrophils. *Front Cell Infect Microbiol* (2018) 8:2. doi: 10.3389/fcimb.2018.00002
37. Weir MH, Haas CN. A Model for *In-Vivo* Delivered Dose Estimation for Inhaled Bacillus Anthracis Spores in Humans With Interspecies Extrapolation. *Environ Sci Technol* (2011) 45:5828–33. doi: 10.1021/es200901e
38. Heppell CW, Egan JR, Hall I. A Human Time Dose Response Model for Q Fever. *Epidemics* (2017) 21:30–8. doi: 10.1016/j.epidem.2017.06.001
39. Wilkening DA. Modeling the Incubation Period of Inhalational Anthrax. *Med Decision Making* (2008) 28:593–605. doi: 10.1177/0272989X08315245
40. Twenhafel N. Pathology of Inhalational Anthrax Animal Models. *Vet Pathol* (2010) 47:819–30. doi: 10.1177/0300985810378112
41. Jang J, Cho M, Chun JH, Cho MH, Park J, Oh HB, et al. The Poly-Λ-D-Glutamic Acid Capsule of Bacillus Anthracis Enhances Lethal Toxin Activity. *Infect Immun* (2011) 79:3846–54. doi: 10.1128/IAI.01145-10
42. Ezzell J Jr, Welkos S. The Capsule of Bacillus Anthracis, a Review. *J Appl Microbiol* (1999) 87:250–0. doi: 10.1046/j.1365-2672.1999.00881.x
43. WHO. *Anthrax in Humans and Animals*. World Health Organization (2008).
44. Welkos S, Friedlander A, Weeks S, Little S, Mendelson I. *In-Vitro* Characterisation of the Phagocytosis and Fate of Anthrax Spores in Macrophages and the Effects of Anti-PA Antibody. *J Med Microbiol* (2002) 51:821–31. doi: 10.1099/0022-1317-51-10-821
45. Gimenez AP, Wu YZ, Paya M, Delclaux C, Touqui L, Goossens PL. High Bactericidal Efficiency of Type IIA Phospholipase A2 Against Bacillus Anthracis and Inhibition of Its Secretion by the Lethal Toxin. *J Immunol* (2004) 173:521–30. doi: 10.4049/jimmunol.173.1.521
46. Jenkins SA, Xu Y. Characterization of Bacillus Anthracis Persistence *In Vivo*. *PloS One* (2013) 8:e66177. doi: 10.1371/journal.pone.0066177
47. Titball R, Manchee R. Factors Affecting the Germination of Spores of Bacillus Anthracis. *J Appl Bacteriol* (1987) 62:269–73. doi: 10.1111/j.1365-2672.1987.tb02408.x
48. Williams B, López-Garcia M, Gillard JJ, Laws TR, Lythe G, Carruthers J, et al. A Stochastic Intracellular Model of Anthrax Infection With Spore Germination Heterogeneity (Computer Codes). [Dataset]. (2021). doi: 10.5518/1026

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.

frontiers in Immunology

PERSPECTIVE
published: 07 September 2021
doi: 10.3389/fimmu.2021.738388

Literature Mining and Mechanistic Graphical Modelling to Improve mRNA Vaccine Platforms

Lorena Leonardelli[1‡], Giuseppe Lofano[2‡], Gianluca Selvaggio[1], Silvia Parolo[1], Stefano Giampiccolo[1], Danilo Tomasoni[1], Enrico Domenici[1,3], Corrado Priami[1,4], Haifeng Song[2], Duccio Medini[5], Luca Marchetti[1,3†§] and Emilio Siena[6†§]*

[1] Fondazione The Microsoft Research – University of Trento Centre for Computational and Systems Biology (COSBI), Rovereto, Italy, [2] Preclinical, GSK, Rockville, MD, United States, [3] Department of Cellular, Computational and Integrative Biology (CIBIO), University of Trento, Povo, Italy, [4] Department of Computer Science, University of Pisa, Pisa, Italy, [5] Toscana Life Sciences Foundation, Siena, Italy, [6] Data Science and Computational Vaccinology, GSK, Siena, Italy

OPEN ACCESS

Edited by:
Jurjen Tel,
Eindhoven University of Technology, Netherlands

Reviewed by:
Kashish Chetal,
Massachusetts General Hospital and Harvard Medical School, United States
Jennifer Oyler-Yaniv,
Harvard Medical School, United States

***Correspondence:**
Duccio Medini
d.medini@toscanalifesciences.org

†ORCID:
Luca Marchetti
orcid.org/0000-0001-9043-7705

‡These authors share first authorship

§These authors share last authorship

Specialty section:
This article was submitted to Systems Immunology, a section of the journal Frontiers in Immunology

Received: 08 July 2021
Accepted: 23 August 2021
Published: 07 September 2021

Citation:
Leonardelli L, Lofano G, Selvaggio G, Parolo S, Giampiccolo S, Tomasoni D, Domenici E, Priami C, Song H, Medini D, Marchetti L and Siena E (2021) Literature Mining and Mechanistic Graphical Modelling to Improve mRNA Vaccine Platforms. Front. Immunol. 12:738388. doi: 10.3389/fimmu.2021.738388

RNA vaccines represent a milestone in the history of vaccinology. They provide several advantages over more traditional approaches to vaccine development, showing strong immunogenicity and an overall favorable safety profile. While preclinical testing has provided some key insights on how RNA vaccines interact with the innate immune system, their mechanism of action appears to be fragmented amid the literature, making it difficult to formulate new hypotheses to be tested in clinical settings and ultimately improve this technology platform. Here, we propose a systems biology approach, based on the combination of literature mining and mechanistic graphical modeling, to consolidate existing knowledge around mRNA vaccines mode of action and enhance the translatability of preclinical hypotheses into clinical evidence. A Natural Language Processing (NLP) pipeline for automated knowledge extraction retrieved key biological evidences that were joined into an interactive mechanistic graphical model representing the chain of immune events induced by mRNA vaccines administration. The achieved mechanistic graphical model will help the design of future experiments, foster the generation of new hypotheses and set the basis for the development of mathematical models capable of simulating and predicting the immune response to mRNA vaccines.

Keywords: mRNA vaccines, natural language processing, graphical modeling, scientific literature mining, mechanisms of action

INTRODUCTION

Since December 2019 SARS-CoV-2 virus has spread across the globe, becoming a pandemic threat and claiming millions of lives. The contagiousness combined with the mortality rate triggered unprecedented efforts to quickly design and develop a vaccine. The first two vaccines that received emergency use authorization by EMA and FDA to prevent COVID-19 disease in humans are based on messenger RNA (mRNA), a relatively new vaccine platform with several advantages over more traditional approaches for vaccine design (1–7). Because of their unique features in terms of manufacturability, mechanism of action and ability to induce potent immune responses, mRNA

vaccines represent an important advancement in the history of vaccinology to defeat infectious diseases. mRNA vaccines are based on the concept that, starting from the amino acid sequence of the antigen of interest, it is possible to design a related mRNA sequence that is employed by the cells of the body as template to express the antigen *in situ*. mRNA vaccines are known to stimulate both arms of the humoral and cellular immunity (7), however their mechanism of action is still partially understood. In this work, we applied a systems biology approach (**Figure 1**) to dissect and elucidate: the delivery of mRNA vaccines, the antigen expression and the resulting vaccine-specific immune responses.

The strategy depicted in **Figure 1** applies to several biological processes and sets the basis for building mathematical models starting from already established knowledge, mainly stored within the scientific literature. The ability for an investigator to process the vast amount of publications through automated literature mining and natural language processing (NLP) algorithms is instrumental to efficiently collect relevant information and define a comprehensive picture. Within each publication, the heterogeneity of data sources (e.g. *in vitro* evidence, animal models, clinical trials, etc.) requires an additional effort to integrate the retrieved information into an interactive platform that can be queried by the users. Databases are usually the preferred solution to store parameters and other numerical information, but as per RNA vaccines some processes may particularly benefit from a mechanistic graphical model that would help identify and fill in the knowledge gaps. Indeed, literature derived data could be complemented with proprietary knowledge and used to draft a mathematical model of the underlying biology. Accordingly, progressive and iterative refinement of the mechanistic process through targeted wet-lab experiments in animal models contributes to the model evaluation. The finalized model is eventually applied to specific scenarios of interest to predict, investigate and support drug development.

Although mRNA vaccines have been the focus of several studies in the last decade, they are usually employed with different components (encoded antigen, delivery system, mRNA architecture) and investigations focused on different arms of the immune response or different biological sites. Consequently, the generated knowledge tends to be rather specific to each individual platform, not always generalizable and sometimes even fragmented. This is what motivated our effort to collect and consolidate all publicly available scientific evidences, related to mRNA vaccine mode of action, into an interactive mechanistic graphical model (**Figure 2**, link to https://www.cosbi.eu/fx/9839203 dynamic figure) tracing the stages of the immune response to mRNA vaccines, from the innate immune activation at the site of injection up to the adaptive response, measurable in the peripheral blood system several days after vaccination. The proposed mechanistic graphical model facilitates the interpretation of what has been discovered so far and fosters interactions among investigators with very different backgrounds, such as immunologists and data scientists.

LITERATURE MINING AND INFORMATION PROCESSING

Automated knowledge extraction from text is the very definition of text-mining, a process becoming more and more a necessity in researching the everyday growing amount of available body of text (8). Text mining does not just find the documents reporting the searched information, but systematically and efficiently

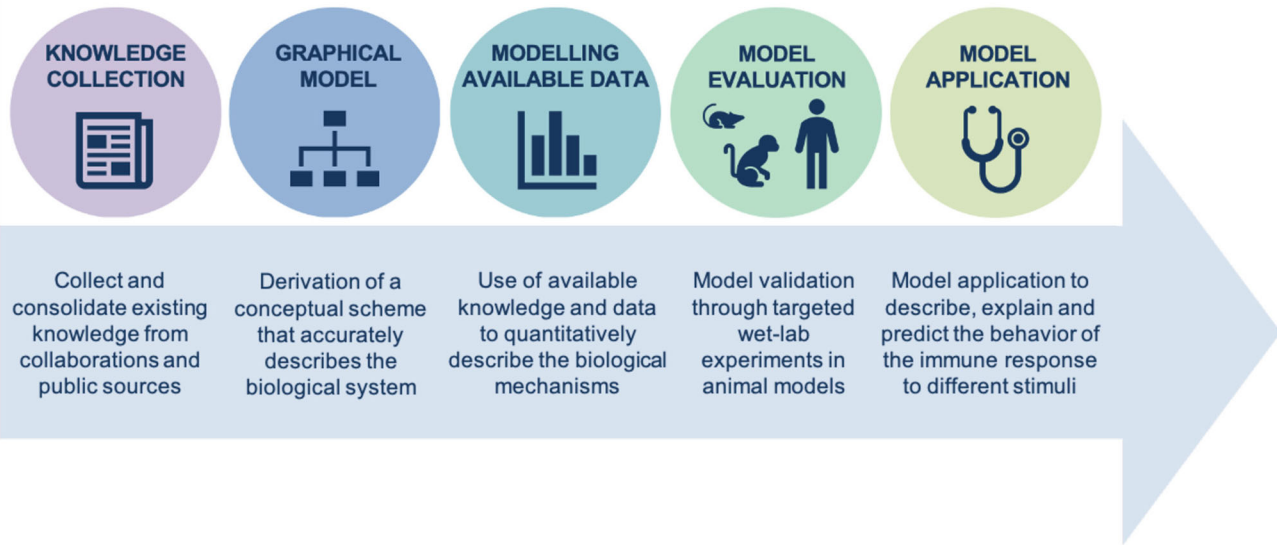


FIGURE 1 | *In silico* modeling of the immune system. Integrative systems biology approach to properly describe, explain, and predict the behavior of any biological mechanism, which in our specific case study is the immune response to mRNA-based vaccines.

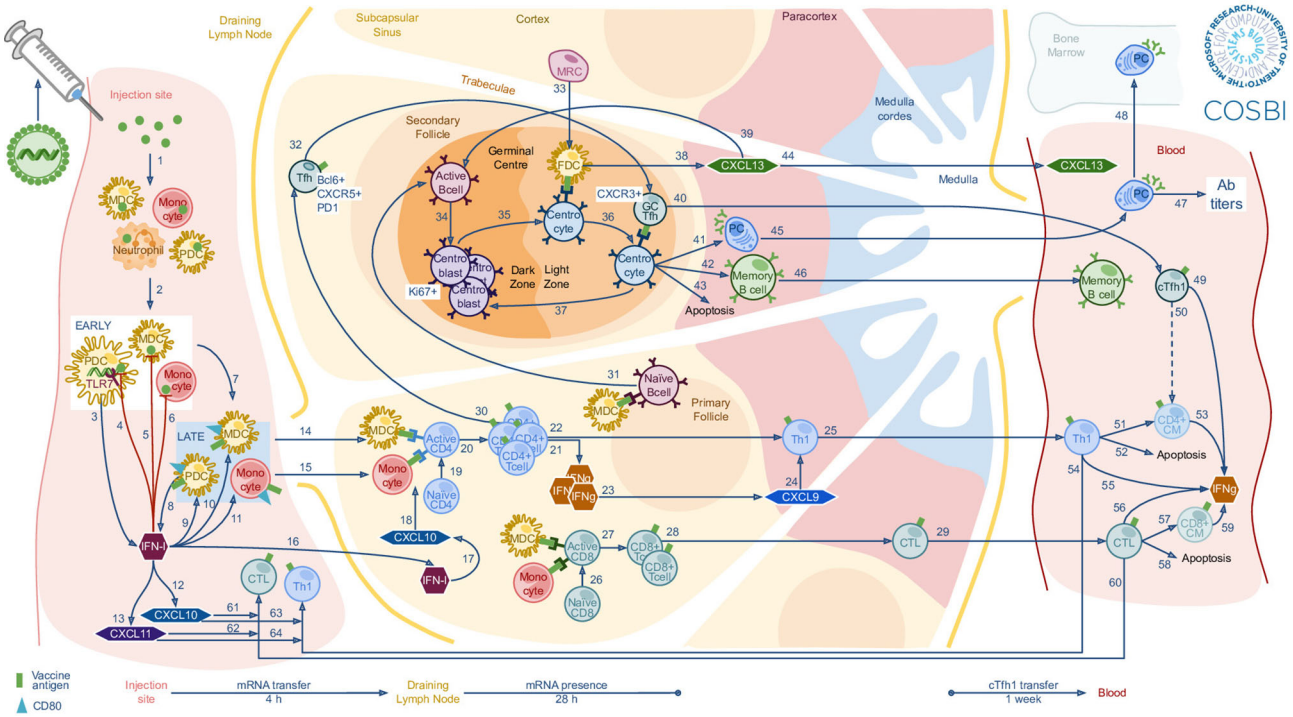


FIGURE 2 | Mechanistic graphical model of the mRNA-based vaccine biological mechanism of action. Lipid nanoparticles encapsulating mRNA molecules are intradermal or intramuscular injected triggering immune signals that reach the draining lymph node, where other immune reactions generate the specific antibodies, eventually transferred into the circulatory system. This figure is proposed in a dynamic view at the link https://www.cosbi.eu/fx/9839203, where each connection (blue arrows) is documented by literature references as well as the involved cells measured in mouse and NHP experiments.

examines those contents, a process that performed manually would be unfeasible, unrealistic and error prone. Several tools have been already developed for extracting biomedical information (9, 10), based on sophisticated NLP algorithms to read and analyze the huge number of scientific publications. Over time those methods have been largely improved, becoming suitable for annotating large experimental datasets and merging data from several studies across the world, which may lead to the discovery of global trends within the existing literature (11, 12).

In this part of the study we investigated the biological perturbations, at the level of immune cells and immunoactive peptides (cytokines/chemokines and cell-surface markers), induced by mRNA vaccines through an NLP-guided literature mining process (10). The initial search for mRNA vaccines-associated literature identified 361 scientific articles in PubMed, 245 of which were automatically annotated as providing experimental evidences generated in animal models. Given the relevance of non-human primates (NHP) as a model for human immunology studies, the pipeline selected 17 NHP-related papers, among which 6 were defined as eligible sources of information and manually confirmed (**Figure 2**). The 6 retrieved scientific publications have been extremely useful to understand the available information about mRNA vaccine mode of action in NHP and one described the SAM platform. To be as inclusive as possible, mice-derived data and information were also taken into consideration. These consisted in a set of 9 self-amplifying mRNA vaccines studies, 5 of which provided eligible quantitative information (**Table 1**). Manually curating the 361 papers of interest since the beginning would have been highly time consuming, while the properly instructed NLP pipeline was able to provide the answers to our questions in a matter of few hours, leaving to the scientists a manageable amount of information to explore.

Scientific literature mining is foundational for the generation of new hypotheses as well as for driving future research and designing new studies. At this point, data and facts of different type and format are gathered and need to be integrated, pressuring data-integration methods to be efficient, in order to explicitly represent the deeply connected big picture the scientist is looking for (24).

MECHANISTIC GRAPHICAL MODELLING

Figure 2 is a static version of the mechanistic review of the mechanisms of action of RNA vaccines, with a dynamic design made available at the link https://www.cosbi.eu/fx/9839203. Given the known similarities and discrepancies between the mouse and NHP models, we represented separately mouse and NHP data, each of them accessible by clicking on the respective black animal

TABLE 1 | List of relevant scientific literature studying mRNA-based vaccines in mouse and NHP animal models.

PMID	Title	Journal	Year	First Author	Animal model	Vaccine features
30936432 (13)	Visualization of early events in mRNA vaccine delivery in non-human primates *via* PET-CT and near-infrared imaging.	Nat Biomed Eng	2019	Lindsay KE	NHP	Yellow fever (YF) prME mRNA vaccine complexed with lipid derivatives
29739835 (4)	Nucleoside-modified mRNA vaccines induce potent T follicular helper and germinal center B cell responses.	J. Exp. Med.	2018	Pardi N	mice, NHP	3 vaccines: mRNA-LNPs encoding HIV-1 envelope (Env), ZIKV prM-E, and influenza virus hemagglutinin (HA)
28958578 (14)	Efficient Targeting and Activation of Antigen-Presenting Cells *In Vivo* after Modified mRNA Vaccine Administration in Rhesus Macaques.	Mol. Ther.	2017	Liang F	NHP	LNP-mRNA encoding hemagglutinin (HA) of H10N8 influenza A virus (H10)
29181005 (15)	Induction of Robust B Cell Responses after Influenza mRNA Vaccination Is Accompanied by Circulating Hemagglutinin-Specific ICOS+ PD-1+ CXCR3+ T Follicular Helper Cells.	Front Immunol	2017	Lindgren G	NHP	LNP-mRNA encoding hemagglutinin (HA) of H10N8 influenza A virus (H10)
25234719 (16)	Potent immune responses in rhesus macaques induced by nonviral delivery of a self-amplifying RNA vaccine expressing HIV type 1 envelope with a cationic nanoemulsion.	J. Infect. Dis.	2015	Bogers WM	NHP	HIV-SAM encoding Env encapsulated in CNE
29263884 (17)	Unmodified mRNA in LNPs constitutes a competitive technology for prophylactic vaccines.	NPJ Vaccines	2017	Lutz J	NHP	LNP-mRNA encoding rabies or influenza antigens
26468547 (18)	Induction of Broad-Based Immunity and Protective Efficacy by Self-amplifying mRNA Vaccines Encoding Influenza Virus Hemagglutinin.	J. Virol.	2016	Brazzoli M	Ferrets, mice	SAM cationic nanoemulsion (CNE) vaccines expressing influenza virus HA
27525409 (19)	Self-Amplifying mRNA Vaccines Expressing Multiple Conserved Influenza Antigens Confer Protection against Homologous and Heterosubtypic Viral Challenge.	PLoS ONE	2016	Magini D	mice	LNP-SAM encoding 2 antigens of influenza virus (NP and M1)
28416600 (20)	Induction of an IFN-Mediated Antiviral Response by a Self-Amplifying RNA Vaccine: Implications for Vaccine Design.	J. Immunol.	2017	Pepini T	mice	LNP-SAM encoding the respiratory syncytial virus (RSV) F protein
31227353 (21)	Co-administration of GM-CSF expressing RNA is a powerful tool to enhance potency of SAM-based vaccines.	Vaccine	2019	Manara C	mice	CNE-SAM encoding the Influenza A virus nucleoprotein (NP)
31290323 (22)	Mannosylation of LNP Results in Improved Potency for Self-Amplifying RNA (SAM) Vaccines.	ACS Infect Dis	2019	Goswami R	mice	LNP-SAM encoding influenza H1N1 antigen HA
26173587 (23)	CD8 T-cell priming upon mRNA vaccination is restricted to bone-marrow-derived antigen-presenting cells and may involve antigen transfer from myocytes.	Immunology	2015	Lazzaro S	mice	LNP-SAM encoding influenza H1N1 antigen HA

The query was performed on December 3rd 2019 searching the following databases: Pubmed, Clinical Trials and USpatent (last update 2019-12-02), EUpatent (last update 2019-11-30).

shape in the left upper corner of the model, overlapping the shared information when possible. The immunization process reported in **Figure 2** covers the basics of biology, documented by the references attached to the arrows connecting each element of the model, which have been manually searched while integrating the novelties apported by the mRNA-based vaccine technology, discovered by literature-mining instead.

The mechanistic graphical model starts with either intramuscular or intradermal injection of mRNA, delivered through lipid nanoparticles (LNPs), which showed a more persistent protein expression than systemic intravenous delivery (4). RNA-LNPs enter the cytosol of local neutrophils, monocytes and dendritic cells (DCs), where mRNA expression begins. Immune cells subsequently migrate to the draining lymph node where they orchestrate the T cell and germinal center responses (14). A critical aspect is represented by the way the mRNA vaccine interacts with the sensors in the host's cells (25). The signaling strength of the exogenous mRNA vector in activating pattern recognition receptors (PPRs), like RIG-I, MDA5 and members of the Toll-like receptors family, is relative to the mRNA species (26–29). This signal is *de facto* a self-adjuvanticity property in the SAM platform that should, at least in principle, be beneficial for the generation of potent immune responses. However, the activation of PPRs is typically associated with the production of type I interferons (IFNs) by plasmacytoid DCs (pDCs) (14) and the induction of the anti-viral state, a condition that has been proposed to severely limit the antibody titers, yet not necessarily impacting the vaccine efficacy (20, 30–32).

The mRNA-based vaccines have shown to most likely leave the injection site through vaccine loaded myeloid DCs (mDCs) making their way to the draining lymph node (dLN) (14), where the concentration of type I IFNs increases, inducing CXCL10, crucial to keep T cells and DCs in proximity, enhancing the chances of T cell activation. After encountering the antigen, both naïve CD4+ and CD8+ become activated and differentiate in mature CD4+ and CD8+ T cells, respectively (14). When CD8+ T cells encounter the antigen and differentiate into short-lived effector cytotoxic T lymphocytes (CTLs), they migrate to the peripheral tissues and to the sites of inflammation (17). In addition, RNA vaccines induce strong Tfh cell responses, which govern the germinal center reactions, including somatic hypermutation, affinity maturation, isotype switching and differentiation of the antigen-specific B cells (15). mRNA vaccines also induce CXCL13, a chemokine responsible for directing B cells efficiently into the follicles (15). From the germinal centers, antigen-specific B cells may differentiate into plasma cells, which home in the bone marrow and continuously

secrete antigen-specific antibodies in the blood, or memory B cells, which recirculate in the blood until further antigen encounter (15).

Subsequently, circulating CD4+ and CD8+ T cells may undergo through two possible fates: apoptosis and survival into memory. Indeed, Th1 cells and CTLs may give rise to central memory (CM) CD4+ and CD8+ T cells, respectively. Moreover, depending on the mRNA vaccine doses, both Th1 cells and CTLs have been observed to produce IFNg (15), which is usually associated with strong anti-viral responses, underlying the efficacy of RNA-based therapeutics (15, 17).

DISCUSSION

mRNA vaccines are emerging as one of the most promising technologies in vaccinology. Several pharmaceutical companies and research institutes are working at the development of new platforms for mRNA-based antigen delivery, trying to identify and characterize those parameters that are required for a safe and protective vaccine response. This, combined with the recent advancements in *omics* technologies, has resulted in the accumulation of a vast amount of data and information. With this work, we built from this unique opportunity and applied text-mining algorithms to screen and analyze scientific literature, with the aim of collecting all available experimentally validated evidence related to the mechanisms of action of mRNA vaccines. We processed and used this information to collect, in a centralized and structured fashion, the various stages of the immune response to mRNA vaccines across different organs and tissues. This knowledge base collection comes in the form of an interactive mechanistic graphical model, which allows to explore the different arms of the immune response to this kind of vaccines.

To maximize the simplicity and interpretability, the model was built using the minimal set of graphical elements, consisting of arrows (representing cells, chemotaxis, transformation or activity) and symbols representing relevant cells and immunoactive peptides. Ideally, a mechanistic graphical model should be readily available to the different stakeholders, should provide some level of interactivity for exploration of the underlying data and be readily upgradable to incorporate new evidences and information. Indeed, an interactive version of the model was made accessible, even remotely, using a web browser, by a proprietary javascript framework. Hosting a copy of the mechanistic graphical model on a remote server ensures future updates to be readily available, avoiding the need of sharing files and risk of misalignments among different versions. Furthermore, all the elements of the mechanistic graphical model can be clicked upon to access the original reference describing that specific evidence. The knowledge gathered during the development of the mechanistic graphical model provides an updated description of the biological phenomena underlying the immune response to mRNA vaccines. The graphical modelling platform can also facilitate the interaction among scientists from different areas, highlight potential gaps in data availability and knowledge and guide the design of new experiments. The natural evolution of this work would be to leverage on the quantitative information acquired during the process (e.g., kinetic parameters, cells concentrations, etc.) to develop a mathematical model describing the immune response, across different biological compartments (e.g. injection site and lymph node) over time (manuscript in preparation). Ideally, this will help highlighting the key elements of an immune response to mRNA vaccines that are responsible for a protective response or, conversely, elements that may lead to suboptimal responses or undesirable effects. Provided an accurate mathematical model is achieved, this could be used to support the experimental design, by allowing to simulate multiple scenarios and predict their outcome. An example could be that of predicting a presumably safe dose range in a dose finding, first time in human clinical study.

As mentioned before, modelling, and prospectively predicting, the behavior of the immune system is a highly challenging task. Consequently, the presented mechanistic graphical model should not be regarded as an endpoint but rather as a milestone within a broader modelling roadmap that we propose as a promising strategy to achieve a better understanding of the human immune system and how it responds to vaccination. The design of future studies for unraveling mRNA mechanism of action will have new pace through the use of literature mining and mathematical modelling, brought together by the power of modern technology.

AUTHOR CONTRIBUTIONS

LL, GL, GS and SP computed the results presented in the paper. DT and SG provided technical support in implementing the interactive version of the graphical model. DM, ES and LM conceived the study. ES, GL, LL and LM wrote the manuscript. LL and GL share first authorship. ES and LM share last authorship. All authors contributed to the article and approved the submitted version.

FUNDING

This research project was funded by GlaxoSmithKline Biologicals SA.

REFERENCES

1. Thess A, Grund S, Mui BL, Hope MJ, Baumhof P, Fotin-Mleczek M, et al. Sequence-Engineered mRNA Without Chemical Nucleoside Modifications Enables an Effective Protein Therapy in Large Animals. *Mol Ther* (2015) 23:1456–64. doi: 10.1038/mt.2015.103
2. Karikó K, Muramatsu H, Welsh FA, Ludwig J, Kato H, Akira S, et al. Incorporation of Pseudouridine Into mRNA Yields Superior Nonimmunogenic Vector With Increased Translational Capacity and Biological Stability. *Mol Ther* (2008) 16:1833–40. doi: 10.1038/mt.2008.200
3. Kauffman KJ, Webber MJ, Anderson DG. Materials for non-Viral Intracellular Delivery of Messenger RNA Therapeutics. *J Control Release* (2016) 240:227–34. doi: 10.1016/j.jconrel.2015.12.032
4. Pardi N, Hogan MJ, Naradikian MS, Parkhouse K, Cain DW, Jones L, et al. Nucleoside-Modified mRNA Vaccines Induce Potent T Follicular Helper and Germinal Center B Cell Responses. *J Exp Med* (2018) 215:1571–88. doi: 10.1084/jem.20171450

5. Guan S, Rosenecker J. Nanotechnologies in Delivery of mRNA Therapeutics Using Nonviral Vector-Based Delivery Systems. *Gene Ther* (2017) 24:133–43. doi: 10.1038/gt.2017.5
6. Ulmer JB, Geall AJ. Recent Innovations in mRNA Vaccines. *Curr Opin Immunol* (2016) 41:18–22. doi: 10.1016/j.coi.2016.05.008
7. Pardi N, Hogan MJ, Porter FW, Weissman D. mRNA Vaccines - a New Era in Vaccinology. *Nat Rev Drug Discov* (2018) 17:261–79. doi: 10.1038/nrd.2017.243
8. Renganathan V. Text Mining in Biomedical Domain With Emphasis on Document Clustering. *Healthc Inform Res* (2017) 23:141–6. doi: 10.4258/hir.2017.23.3.141
9. Ananiadou S MJ. *Text Mining for Biology and Biomedicine*. Artech House Publishers; Illustrated Edition (2006).
10. Ramponi A, Giampiccolo S, Tomasoni D, Priami C, Lombardo R. High-Precision Biomedical Relation Extraction for Reducing Human Curation Efforts in Industrial Applications. *IEEE Access* (2020) 8:150999–1011. doi: 10.1109/access.2020.3014862
11. Schuster J, Superdock M, Agudelo A, Stey P, Padbury J, Sarkar IN, et al. Machine Learning Approach to Literature Mining for the Genetics of Complex Diseases. *Database* (2019) 2019:1–12. doi: 10.1093/database/baz124
12. Simon C, Davidsen K, Hansen C, Seymour E, Barnkob MB, Olsen LR. BioReader: A Text Mining Tool for Performing Classification of Biomedical Literature. *BMC Bioinf* (2019) 19:57. doi: 10.1186/s12859-019-2607-x
13. Lindsay KE, Bhosle SM, Zurla C, Beyersdorf J, Rogers KA, Vanover D, et al. Visualization of Early Events in mRNA Vaccine Delivery in Non-Human Primates *via* PET-CT and Near-Infrared Imaging. *Nat BioMed Eng* (2019) 3:371–80. doi: 10.1038/s41551-019-0378-3
14. Liang F, Lindgren G, Lin A, Thompson EA, Ols S, Röhss J, et al. Efficient Targeting and Activation of Antigen-Presenting Cells In Vivo After Modified mRNA Vaccine Administration in Rhesus Macaques. *Mol Ther* (2017) 25:2635–47. doi: 10.1016/j.ymthe.2017.08.006
15. Lindgren G, Ols S, Liang F, Thompson EA, Lin A, Hellgren F, et al. Induction of Robust B Cell Responses After Influenza mRNA Vaccination Is Accompanied by Circulating Hemagglutinin-Specific ICOS+ PD-1+ CXCR3+ T Follicular Helper Cells. *Front Immunol* (2017) 8:1539. doi: 10.3389/fimmu.2017.01539
16. Bogers WM, Oostermeijer H, Mooij P, Koopman G, Verschoor EJ, Davis D, et al. Potent Immune Responses in Rhesus Macaques Induced by Nonviral Delivery of a Self-Amplifying RNA Vaccine Expressing HIV Type 1 Envelope With a Cationic Nanoemulsion. *J Infect Dis* (2015) 211:947–55. doi: 10.1093/infdis/jiu522
17. Lutz J, Lazzaro S, Habbeddine M, Schmidt KE, Baumhof P, Mui BL, et al. Unmodified mRNA in LNPs Constitutes a Competitive Technology for Prophylactic Vaccines. *NPJ Vaccines* (2017) 2:29. doi: 10.1038/s41541-017-0032-6
18. Brazzoli M, Magini D, Bonci A, Buccato S, Giovani C, Kratzer R, et al. Induction of Broad-Based Immunity and Protective Efficacy by Self-Amplifying mRNA Vaccines Encoding Influenza Virus Hemagglutinin. *J Virol* (2016) 90:332–44. doi: 10.1128/JVI.01786-15
19. Magini D, Giovani C, Mangiavacchi C, Maccari S, Cecchi R, Ulmer JB, et al. Self-Amplifying mRNA Vaccines Expressing Multiple Conserved Influenza Antigens Confer Protection Against Homologous and Heterosubtypic Viral Challenge. *PLoS One* (2016) 11:e0161193. doi: 10.1371/journal.pone.0161193
20. Pepini T, Pulichino A-P, Carsillo T, Carlson AL, Sari-Sarraf F, Ramsauer K, et al. Induction of an IFN-Mediated Antiviral Response by a Self-Amplifying RNA Vaccine: Implications for Vaccine Design. *J Immunol* (2017) 198:4012–24. doi: 10.4049/jimmunol.1601877
21. Manara C, Brazzoli M, Piccioli D, Taccone M, D'Oro U, Maione D, et al. Co-Administration of GM-CSF Expressing RNA Is a Powerful Tool to Enhance Potency of SAM-Based Vaccines. *Vaccine* (2019) 37:4204–13. doi: 10.1016/j.vaccine.2019.04.028
22. Goswami R, Chatzikleanthous D, Lou G, Giusti F, Bonci A, Taccone M, et al. Mannosylation of LNP Results in Improved Potency for Self-Amplifying RNA (SAM) Vaccines. *ACS Infect Dis* (2019) 5:1546–58. doi: 10.1021/acsinfecdis.9b00084
23. Lazzaro S, Giovani C, Mangiavacchi S, Magini D, Maione D, Baudner B, et al. CD8 T-Cell Priming Upon mRNA Vaccination Is Restricted to Bone-Marrow-Derived Antigen-Presenting Cells and may Involve Antigen Transfer From Myocytes. *Immunology* (2015) 146:312–26. doi: 10.1111/imm.12505
24. Zhao S, Su C, Lu Z, Wang F. Recent Advances in Biomedical Literature Mining. *Briefings Bioinf* (2020) 22(3):1–19. doi: 10.1093/bib/bbaa057
25. Chen N, Xia P, Li S, Zhang T, Wang TT, Zhu J. RNA Sensors of the Innate Immune System and Their Detection of Pathogens. *IUBMB Life* (2017) 69:297–304. doi: 10.1002/iub.1625
26. Hornung V, Ellegast J, Kim S, Brzózka K, Jung A, Kato H, et al. 5'-Triphosphate RNA is the Ligand for RIG-I. *Science* (2006) 314:994–7. doi: 10.1126/science.1132505
27. Pichlmair A, Schulz O, Tan C-P, Rehwinkel J, Kato H, Takeuchi O, et al. Activation of MDA5 Requires Higher-Order RNA Structures Generated During Virus Infection. *J Virol* (2009) 83:10761–9. doi: 10.1128/jvi.00770-09
28. Diebold SS, Kaisho T, Hemmi H, Akira S, Reis e Sousa C. Innate Antiviral Responses by Means of TLR7-Mediated Recognition of Single-Stranded RNA. *Science* (2004) 303:1529–31. doi: 10.1126/science.1093616
29. Edwards DK, Jasny E, Yoon H, Horscroft N, Schanen B, Geter T, et al. Adjuvant Effects of a Sequence-Engineered mRNA Vaccine: Translational Profiling Demonstrates Similar Human and Murine Innate Response. *J Trans Med* (2017) 15:1. doi: 10.1186/s12967-016-1111-6
30. Kim YG, Baltabekova AZ, Zhiyenbay EE, Aksambayeva AS, Shagyrova ZS, Khannanov R, et al. Recombinant Vaccinia Virus-Coded Interferon Inhibitor B18R: Expression, Refolding and a Use in a Mammalian Expression System With a RNA-Vector. *PLoS One* (2017) 12:e0189308. doi: 10.1371/journal.pone.0189308
31. Beissert T, Koste L, Perkovic M, Walzer KC, Erbar S, Selmi A, et al. Improvement of In Vivo Expression of Genes Delivered by Self-Amplifying RNA Using Vaccinia Virus Immune Evasion Proteins. *Hum Gene Ther* (2017) 28:1138–46. doi: 10.1089/hum.2017.121
32. Pollard C, Rejman J, De Haes W, Verrier B, Van Gulck E, Naessens T, et al. Type I IFN Counteracts the Induction of Antigen-Specific Immune Responses by Lipid-Based Delivery of mRNA Vaccines. *Mol Ther* (2013) 21:251–9. doi: 10.1038/mt.2012.202

Conflict of Interest: GL, HS, DM, and ES were all employees of the GSK group of companies at the time of the study. The "Fondazione The Microsoft Research – University of Trento Centre for Computational and Systems Biology (COSBI)" institute received financial remuneration for conducting the activates described in this study.

The authors declare that this study received funding from GlaxoSmithKline Biologicals SA. The funder had the following involvement in the study: study design, interpretation of data, the writing of this article and the decision to submit it for publication.

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: 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.

frontiers in Immunology

ORIGINAL RESEARCH
published: 14 October 2021
doi: 10.3389/fimmu.2021.704089

Sustained Drug Treatment Alters the Gut Microbiota in Rheumatoid Arthritis

Liyan Mei[1†], Zhihua Yang[1†], Xiaolin Zhang[1†], Zehao Liu[1], Maojie Wang[1,2], Xiaodong Wu[1], Xiumin Chen[1], Qingchun Huang[1] and Runyue Huang[1,3,4,5*]*

[1] The Second Affiliated Hospital of Guangzhou University of Chinese Medicine (Guangdong Provincial Hospital of Chinese Medicine), Guangzhou, China, [2] Center for Molecular Medicine, University Medical Center Utrecht, Utrecht, Netherlands, [3] State Key Laboratory of Dampness Syndrome of Chinese Medicine (The Second Affiliated Hospital of Guangzhou University of Chinese Medicine), Guangzhou, China, [4] Guangdong Provincial Key Laboratory of Clinical Research on Traditional Chinese Medicine Syndrome, Guangzhou, China, [5] Guangdong-Hong Kong-Macau Joint Lab on Chinese Medicine and Immune Disease Research, Guangzhou University of Chinese Medicine, Guangzhou, China

OPEN ACCESS

Edited by:
Aridaman Pandit, University Medical Center Utrecht, Netherlands

Reviewed by:
Baskar Balakrishnan, Mayo Clinic, United States
Andrew Yung Fong Li Yim, Amsterdam University Medical Center, Netherlands

***Correspondence:**
Runyue Huang
ryhuang@gzucm.edu.cn
Qingchun Huang
qch1963@163.com

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Molecular Innate Immunity, a section of the journal Frontiers in Immunology

Received: *01 May 2021*
Accepted: *23 September 2021*
Published: *14 October 2021*

Citation:
Mei L, Yang Z, Zhang X, Liu Z, Wang M, Wu X, Chen X, Huang Q and Huang R (2021) Sustained Drug Treatment Alters the Gut Microbiota in Rheumatoid Arthritis. Front. Immunol. 12:704089. doi: 10.3389/fimmu.2021.704089

Several studies have investigated the causative role of the microbiome in the development of rheumatoid arthritis (RA), but changes in the gut microbiome in RA patients during drug treatment have been less well studied. Here, we tracked the longitudinal changes in gut bacteria in 22 RA patients who were randomized into two groups and treated with Huayu-Qiangshen-Tongbi formula (HQT) plus methotrexate (MTX) or leflunomide (LEF) plus MTX. There were differences in the gut microbiome between untreated (at baseline) RA patients and healthy controls, with 37 species being more abundant in the RA patients and 21 species (including *Clostridium celatum*) being less abundant. Regarding the functional analysis, vitamin K2 biosynthesis was associated with RA-enriched bacteria. Additionally, in RA patients, alterations in gut microbial species appeared to be associated with RA-related clinical indicators through changing various gut microbiome functional pathways. The clinical efficacy of the two treatments was further observed to be similar, but the response trends of RA-related clinical indices in the two treatment groups differed. For example, HQT treatment affected the erythrocyte sedimentation rate (ESR), while LEF treatment affected the C-reactive protein (CRP) level. Further, 11 species and 9 metabolic pathways significantly changed over time in the HQT group (including *C. celatum*, which increased), while only 4 species and 2 metabolic pathways significantly changed over time in the LEF group. In summary, we studied the alterations in the gut microbiome of RA patients being treated with HQT or LEF. The results provide useful information on the role of the gut microbiota in the pathogenesis of RA, and they also provide potentially effective directions for developing new RA treatments.

Keywords: rheumatoid arthritis, gut microbiota, drug treatment, methotrexate, traditional Chinese medicine, leflunomide

INTRODUCTION

Rheumatoid arthritis (RA) is a common chronic systemic inflammatory autoimmune disease characterized by the production of autoantibodies that target various molecules (1). Painful joint swelling and morning stiffness (2) are typical clinical manifestations of RA, which severely impair physical function and quality of life. The pathogenesis of RA is complex and involves several risk factors, including susceptibility genes, gender, and environmental factors. Increasingly, studies are revealing that microbiota alterations (3–5), particularly changes in the gut and oral microbiomes, are important environmental risk factors in RA development (6).

Driven by advances in shotgun sequencing technology, the burgeoning field of metagenomics has begun to uncover the impact of microbes on RA. *Haemophilus* spp. was decreased and *Lactobacillus salivarius* was increased in RA patients compared to healthy controls (HCs) in the gut, saliva, or dental microbial communities, but microbiome dysbiosis was partially resolved by RA DMARDs treatment (7). A recent study reported a decreased abundance of *Enterobacter*, *Odoribacter*, *Lactobacillus*, and *Alloprevotella* and an increased abundance of the genera *Bacteroides* and *Escherichia-Shigella* in a RA cohort in China (8), while there was an increased abundance of the genus *Prevotella* (e.g., *Prevotella denticola*) in Japanese RA patients. *Blautia*, *Akkermansia*, and *Clostridiales* were increased in anti-citrullinated peptide antibody (ACPA)-positive RA patients compared to ACPA-negative RA patients (9).

Additionally, several studies have indicated that periodontitis is closely associated with RA (10). Specific oral bacteria that are causative agents of periodontal disease have been shown to significantly influence the progression of RA. Members of the oral microbiome, such as *P. gingivalis*, *Prevotella intermedia*, and *Aggregatibacter actinomycetemcomitans*, play roles in the onset of RA *via* mechanisms such as direct or indirect modulation of citrullination and altering T cell-mediated adaptive immunity.

Increasing experimental and clinical evidence also suggests that the gut microbiome composition and function are affected by RA treatment. For example, a study by Picchianti-Diamanti et al. (11) showed that the tumor necrosis factor alpha (TNF-α) inhibitor etanercept can alter microbial communities and at least partially improve the microbiota in RA patients. Similarly, in mice with collagen-induced arthritis, etanercept reduced the abundance of *Escherichia/Shigella* and increased the abundance of *Lactobacillus, Clostridium XIVa*, and *Tannerella* (12). Additionally, antibiotic-induced partial depletion of the gut microbiota aggravated arthritis symptoms in a RA mouse model (13). Regarding natural compounds, *Clematis* triterpenoid saponins can alleviate arthritis-associated gut microbial dysbiosis and thereby improve arthritic disease indices (14). Recently, the emergence of data on microbiomes associated with different treatments has provided new insights regarding the associations among RA treatments, the gut microbiota, and clinical outcomes.

Huayu-Qiangshen-Tongbi formula (HQT) is a traditional Chinese medicine (TCM) formula adapted from classical Chinese medicine that has been widely used in clinical practice for the treatment of RA. In particular, the clinical application of HQT in combination with DMARDs is highly effective in the treatment of RA. Interestingly, we have also confirmed the effectiveness and possible mechanism of HQT in the treatment of RA through clinical trials and experimental studies. In the early stages, we implemented a retrospective clinical study and an investigator-initiated randomized clinical study with methotrexate (MTX) as the baseline drug and leflunomide (LEF) as the control, both of which suggested that "HQT+MTX" may have similar or even better clinical efficacy and tolerability than "LEF+MTX". Importantly, the "HQT+MTX" group had fewer adverse events compared to the "HQT+LEF" group (15, 16). Along with the successful clinical trial of HQT, we have also conducted in-depth research on the mechanism of HQT for the treatment of RA and found that the mechanisms underlying the therapeutic effects of HQT on RA are closely related to its modulation of lncRNA uc.477 and miR-19b (17). In addition, pharmacological studies have also confirmed that HQT can treat RA by anti-inflammatory and regulating the body's bone metabolism (18). In this study, the herbal composition and dosage of HQT were detailed in a previous study (16, 18).

Recent studies have sought to determine the gut microbiome factors that promote RA development. However, few studies have tracked the gut microbiota alterations in RA patients during treatment. In this study, we tracked the gut microbiota alterations in RA patients being treated with either HQT formula plus MTX or LEF plus MTX over 6 months. We aimed to assess the effects of the different treatments on the gut microbiome in RA patients.

MATERIALS AND METHODS

Participant Recruitment and Ethics Statement

RA patients and healthy donors (44 participants) were recruited from the Second Affiliated Hospital of Guangzhou University of Chinese Medicine (Guangdong Provincial Hospital of Chinese Medicine) between August 2016 and September 2018. These included 22 RA patients who fulfilled the 2010 revised criteria of the American College of Rheumatology (ACR) for RA (19) and 22 ethnicity-, sex-, and age-matched individuals with no personal or family history of rheumatic diseases, who served as healthy controls (HCs) (**Table S1**).

After baseline sample collection, the 22 RA patients were randomized into two groups; 13 RA patients received the traditional Chinese medicine Huayu-Qiangshen-Tongbi (HQT) decoction (once every 2 days), while the other 9 RA patients received oral leflunomide (LEF; 20 mg/day). All patients received oral methotrexate (MTX; 10–15 mg/week). Each patient was continuously treated for 6 months. Fecal and blood samples were collected at baseline and in the first (M1), third (M3), and sixth (M6) months after treatment, along with detecting other RA-related clinical indices.

This study was reviewed and approved by the ethics committee of Guangdong Provincial Hospital of Chinese

Medicine (no. B2016-076-01) and registered with the World Health Organization clinical trial registry (no. ChiCTR-INR-16009031). All participants provided their written informed consent to participate in this study before baseline sample collection. The 22 RA patients were part of a clinical study published in September 2020 (16).

Metagenome Sequencing and Analysis

A total of 110 fecal samples from the 44 participants underwent metagenomic sequencing. Total bacterial DNA was extracted from the fecal samples using a NucleoSpin® Soil kit (Macherey-Nagel, Düren, Germany) following the manufacturer's instructions. A HiSeq X Ten sequencer (Illumina, San Diego, CA, USA) yielded 1,036.69 Gb of paired-end reads. After quality control and removal of host (human) reads, a mean ± SD of 8.18 ± 0.82 Gb of reads/sample remained for bioinformatics analysis. Taxonomic and functional profiling of the metagenomes were performed using MetaPhlAn2 (v2.0) and HUMAnN2 (v0.11.2), respectively. Detailed methods and parameters are provided in the **Supplementary Material**.

The origin of each species was determined using Integrated Microbial Genomes (IMG) data (http://img.jgi.doe.gov/cgi-bin/w/main.cgi). First, the bacterial species "Host Name" had to be "Homo sapiens". Second, if the "Isolation" field contained oral components, e.g., "saliva", "dental plaque", or "nasopharynx", the species was considered to be of oral origin. If it contained intestinal components, e.g., "feces" or "gastrointestinal tract", the species was considered to be of fecal origin. In other cases, the species was considered to be of other or unknown origin.

Statistical Analysis

The alpha diversity of the groups was estimated at the gene family and species level. Beta diversity between groups was estimated based on Bray–Curtis distance at the gene family level using the vegdist function in the vegan R package. Permutational multivariate analysis of variance (PERMANOVA) was also performed on the gene family profiles using the adonis function in the vegan R package, and the permuted P value was based on 9,999 permutations. Significantly affected taxa and MetaCyc pathways between the groups were identified by Wilcoxon rank-sum test, based on $P < 0.05$. Significantly affected Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways were identified using the reporter score (20, 21) based on |reporter score| > 1.65. The Jonckheere-Terpstra test was used to investigate the trends in RA-associated bacterial species over time. The correlations among the relative abundances of species, gut microbiome KEGG functional pathways, and RA-related clinical indices were calculated by Spearman's rank correlation analysis and visualized using the ComplexHeatmap R package. Cross-validated random forest modeling (using randomForest 4.6-14 R package) was performed based on the relative abundances of species in the samples (21). In 5 trials of 10 fold cross-validated model, all bacterial species were sorted according to the importance of their variable and added to the model in turn. Then the cross-validation error curves were averaged and the minimum error in the averaged curve plus the s.d. at the point was used as the cut-off for feature selection. All bacterial species set with an error less than the cut-off were listed and the set with the smallest number of bacterial species was selected as the optimal set. Enterotype analysis of samples was performed based on the genus relative abundance profile as described by Arumugam et al. (22). Jensen–Shannon divergence (JSD) distance metric of samples was calculated, and clustered using the PAM clustering algorithm. The optimal number of clusters was determine by Calinski–Harabaze (CH). Principal component analysis (PCA) was preformed to visualize samples distances using 'ade4' package in R. All statistical analyses were carried out in R (v3.5.0).

RESULTS

General Characteristics of Study Cohorts

The study cohorts comprised 22 HCs and 22 RA patients, and their baseline characteristics are shown in **Table 1**. No significant differences in age, gender, or body mass index (BMI) were observed between the HC and RA groups. As expected, the RA diagnostic markers rheumatoid factor (RF) and anti-cyclic citrullinated peptide (anti-CCP) were significantly higher in RA patients than HCs. The inflammatory markers C-reactive protein (CRP) and erythrocyte sedimentation rate (ESR) were also significantly higher in RA patients than HCs. In contrast, no significant differences were observed in the other blood parameters except alanine transaminase (AST), which was lower in RA patients though still in the normal range in both groups.

The baseline characteristics of the 13 MTX+HQT-treated and 9 MTX+LEF-treated RA patients in the study cohort are shown in **Table 2**. There were no significant differences in age, gender, height, and BMI between the HQT and LEF groups of RA patients, except for weight (P=0.044). Undoubtedly, the differences in RA-specific diagnostic markers, inflammatory markers, and other blood markers (RF, Anti-CCP, hs-CRP, ESR, WBC, HB, BUN, Cr, PLT, ALT, AST) between the two treatment groups were also not statistically significant.

Alterations in Gut Microbes of RA Patients

Based on the Shannon index, there was no significant difference in alpha diversity between the RA and HC groups. There was no significant difference in the number of species (**Figure S1A** and **Table S8**). There were also no significant differences in microbial community composition based on principal coordinate analysis (PCoA) and PERMANOVA at the gene family level (**Figure S1B** and **Table S9**). Furthermore, for the separate clusters of PCoA1 in **Figure S1B**, we found by individual gut type analysis that the main reason was the different gut types among individuals.Two clusters, one with *g_Prevotella* as the dominant species (HC=3; RA=6) and the other with *g_Bacteroides* as the dominant species (HC=19; RA=16). The results of Fisher's exact test showed p=0.4566 for different gut types in HC and RA individuals, indicating that there was no significant difference in the frequency of different gut types in the two groups (**Figure S1C**).

TABLE 1 | Baseline characteristics of healthy controls (HC; n = 22) and rheumatoid arthritis (RA) patients (n = 22).

Clinical factor	HCs; n = 22	RA patients; n = 22	Adjusted P value
Age, year*	48.5 ± 13.3	48.2 ± 11.5	1.000
Gender, female§	19 (86.4%)	19 (86.4%)	1.000
Height, cm*	158.7 ± 6.2	160.4 ± 5.4	0.327
Weight, kg*	52.2 ± 6.8	55.8 ± 6.4	0.203
BMI, kg/m^2*	20.7 ± 1.8	21.7 ± 2.3	0.224
WBC count, 10^9/L*	6.2 ± 1.3	7.4 ± 2.6	0.203
HB, g/L*	124.8 ± 17.2	127.6 ± 39.3	0.554
BUN, mmol/L*	4.8 ± 1.4	4.6 ± 1.4	0.687
Cr, μmol/L*	66.1 ± 14.7	62.9 ± 10.7	0.687
PLT count, 10^9/L*	266.9 ± 61.8	303.5 ± 103.6	0.229
ALT, U/L*	18.3 ± 18.3	14.4 ± 11.0	0.283
AST, U/L*	20.7 ± 7.1	15.2 ± 2.9	0.002
hs-CRP, mmol/L*	1.2 ± 1.6	23.8 ± 20.3	<0.001
ESR, mm/h*	22.4 ± 14.8	59.5 ± 27.2	<0.001
RF, IU/mL*	7.5 ± 2.8	201.6 ± 269.1	<0.001
Anti-CCP, U/mL*	0.6 ± 0.2	102.1 ± 80.2	<0.001

**mean ± SD, §n (%).*

BMI, body mass index; WBC, white blood cell; HB, hemoglobin; BUN, blood urea nitrogen; Cr, creatinine; PLT, platelet; ALT, alanine transaminase; AST, aspartate transaminase; hs-CRP, high-sensitivity C-reactive protein; ESR, erythrocyte sedimentation rate; RF, rheumatoid factor; anti-CCP, anti-cyclic citrullinated peptides.

At baseline, 12 phyla, 225 genera, and 656 species were identified. The dominant phyla were Bacteroidetes (69.92% ± 13.95%), Firmicutes (20.88% ± 11.71%), and Proteobacteria (5.48% ± 6.47%) (**Figure S1D** and **Table S10**), and the dominant genera were *Bacteroides* (46.41% ± 22.23%), *Prevotella* (12.18% ± 22.40%), *Alistipes* (6.19% ± 6.70%), *Faecalibacterium* (3.79% ± 3.84%), and *Eubacterium* (3.06% ± 2.82%) (**Figure S1E** and **Table S11**). The number of different taxa at each taxonomic level was 1 phylum, 2 classes, 7 orders, 12 families, 26 genera, and 58 species between the HC and RA groups (based on the Wilcoxon rank-sum test). Among them, for 26 significant differences genera, 16 genera were enriched in RA patients, e.g. *Neisseria, Haemophilus, Veillonella, Campylobacter*, and 10 genera were enriched in HC, e.g. *Xanthomonas, Enterococcus, Megasphaera* (**Figure S2** and **Table S3**). The 58 species with significantly different were all low abundance species (mean relative abundance <2%) (**Figure 1A** and **Table S3**). 37 species were enriched in RA patients, including *Streptococcus* spp., *Haemophilus* spp., and *Neisseria* spp., while the remaining 21 species were depleted, including *Clostridium celatum, Enterococcus faecalis*, and *Fusobacterium varium*. Most of the RA-enriched species belonged to *Proteobacteria* (19, 51.35%), followed by *Firmicutes* (14, 37.84%). Most of the HC-enriched species belonged to *Firmicutes* (12, 57.14%). However, for *Prevotella* spp., no significant differences were observed between the HC and RA (**Figure S3**). Regarding the origin of the species, 63.16% of the RA-enriched species were residents of the human oral cavity such as the genera *Streptococcus*, *Neisseria*, and *Haemophilus*, while 47.62% of the HC-enriched species were residents of the human gut.

Regarding gut microbial function, there were 18 significantly different MetaCyc metabolic pathways in the RA patients

TABLE 2 | Baseline characteristics of MTX+HQT-treated (n = 13) and MTX+LEF-treated (n = 9) RA patients.

Clinical factor	MTX+HQT; n = 13	MTX+LEF; n = 9	Adjusted P value
Age, year*	50.4 ± 10.4	45.1 ± 12.8	0.300
Gender, female§	11 (84.6%)	8 (88.9%)	0.787
Height, cm*	161.2 ± 6.6	159.2 ± 3.1	0.416
Weight, kg*	58.1 ± 4.4	52.6 ± 7.6	0.044
BMI, kg/m^2*	22.4± 2.0	20.7 ± 2.5	0.087
WBC count, 10^9/L*	7.4 ± 2.8	7.4 ± 2.4	0.988
HB, g/L*	133.1 ± 50.0	119.7 ± 13.5	0.444
BUN, mmol/L*	4.8 ± 1.4	4.3 ± 1.5	0.370
Cr, μmol/L*	64.0 ± 11.2	61.3 ± 10.3	0.578
PLT count, 10^9/L*	296.2 ± 122.0	314.2 ± 75.2	0.698
ALT, U/L*	14.9 ± 13.4	13.7± 6.7	0.799
AST, U/L*	15.0 ± 2.6	15.6 ± 3.5	0.674
hs-CRP, mmol/L*	22.5 ± 17.7	25.7 ± 24.7	0.724
ESR, mm/h*	67.9 ± 21.9	47.4 ± 30.7	0.082
RF, IU/mL*	222.4 ± 302.0	171.6 ± 227.1	0.674
Anti-CCP, U/mL*	97.0± 87.8	108.9 ± 73.4	0.746

**mean ± SD, §n (%).*

BMI, body mass index; WBC, white blood cell; HB, hemoglobin; BUN, blood urea nitrogen; Cr, creatinine; PLT, platelet; ALT, alanine transaminase; AST, aspartate transaminase; hs-CRP, high-sensitivity C-reactive protein; ESR, erythrocyte sedimentation rate; RF, rheumatoid factor; anti-CCP, anti-cyclic citrullinated peptides.

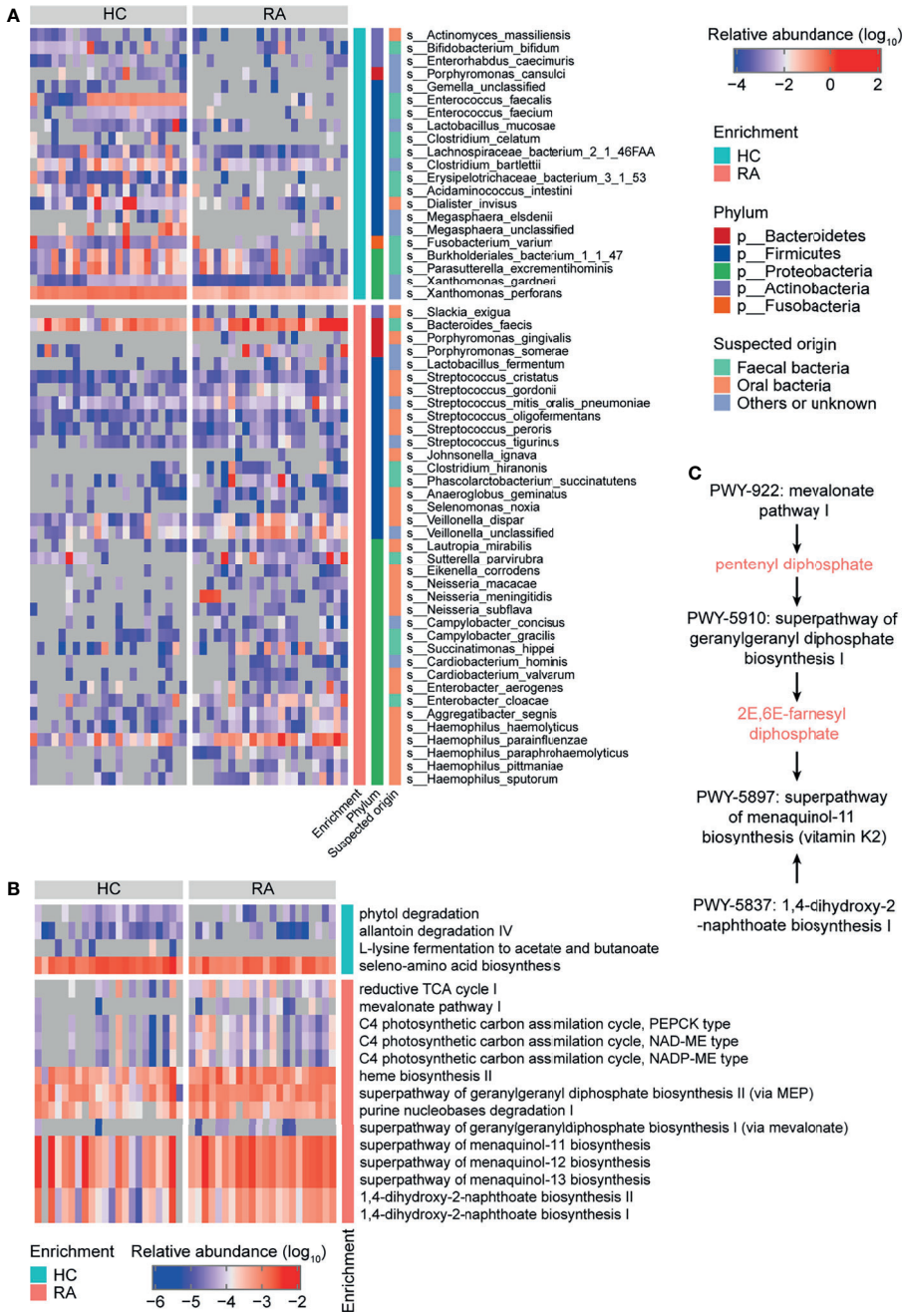


FIGURE 1 | Taxonomy and function of gut microbes that varied between rheumatoid arthritis (RA) patients and healthy controls (HCs) at baseline. Heatmaps of differentially abundant **(A)** bacterial species and **(B)** MetaCyc metabolic pathways in the RA group compared to the HC group. **(C)** MetaCyc vitamin K biosynthesis-related pathways.

compared to HCs (**Figure 1B** and **Table S4**). Downregulated pathways included allantoin degradation IV (anaerobic) (PWY0-41), L-lysine fermentation to acetate and butanoate (P163-PWY), and seleno-amino acid biosynthesis (PWY-6936). Upregulated pathways included purine nucleobases degradation I (anaerobic) (P164-PWY), mevalonate pathway I (PWY-922), geranylgeranyl diphosphate biosynthesis (PWY-5121 and PWY-5910), menaquinol biosynthesis (PWY-5897, PWY-5898, and PWY-5899), and 1,4-dihydroxy-2-naphthoate biosynthesis (PWY-5791 and PWY-5837) (**Figure 1B** and **Table S4**). The PWY-922 pathway is responsible for the synthesis of pentenyl diphosphate, which is then metabolized to 2E,6E-farnesyl diphosphate *via* the PWY-5910 pathway. 2E,6E-farnesyl diphosphate, along with the products of the PWY-5837 pathway,

are raw materials for vitamin K2 biosynthesis. Thus, the PWY-922, PWY-5910, and PWY-5837 pathways, which are involved in vitamin K2 biosynthesis in gut microbes, were upregulated in RA patients (**Figure 1C**). These results further indicate that the gut microbial composition and function differed between RA patients and HCs.

Associations of Gut Microbiota With Clinical Indices

Based on the above analysis, changes in gut microbes and differences in gut microbiome functional pathways between RA patients and the HCs were observed. To investigate the relationships among the clinical indices, the gut microbiome species and the gut microbiome KEGG functional pathways at baseline in the RA and HC groups, correlation coefficients were calculated using the method reported by Pedersenet et al. (23).

There were 11 bacterial species and 24 KEGG pathways associated with the clinical indices (**Figure 2** and **Table S7**).

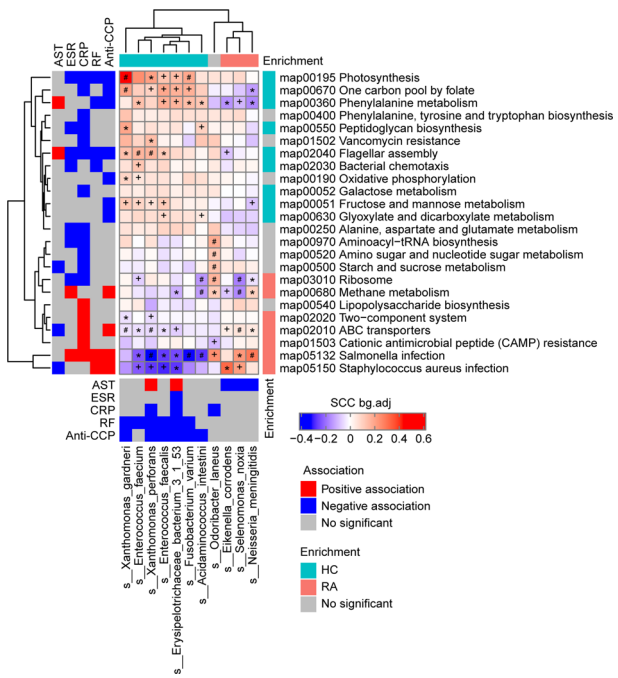


FIGURE 2 | Associations among the phenotypes (rheumatoid arthritis RA-related clinical indices), gut microbiome species, and gut microbiome functions (KEGG pathways) in the RA patients and healthy controls (HCs) at baseline. Left and bottom panels show significant correlations between the RA-related clinical indices and KEGG pathways or bacterial species, respectively (blue, negative association; red, positive association; grey, no significant association). Top and right panels show the treatment group-specific enrichment of bacterial species and KEGG pathways, respectively (blue, HC-enriched; red, RA-enriched; grey, no significant difference). Middle panel shows the significant associations between the clinical index-related bacterial species and clinical index-related KEGG pathways (blue, negative; red, positive). +, FDR < 0.05; *, FDR < 0.01; #, FDR < 0.001, according to the Wilcoxon rank-sum test. AST, aspartate transaminase; ESR, erythrocyte sedimentation rate; CRP, high-sensitivity C-reactive protein; RF, rheumatoid factor;Anti-CCP, anti-cyclic citrullinated peptides.

9 HC-enriched pathways and 7 HC-enriched species were negatively correlated with ESR, CRP, RF, or anti-CCP, while 7 RA-enriched pathways were positively correlated with at least one of these clinical indices. In contrast, these species and pathways exhibited the opposite correlations or no correlation with AST. In detail, the phenylalanine metabolism pathway (map00360) was identified as being positively correlated with AST and negatively correlated with anti-CCP and RF, and it was positively correlated with five HC-enriched species (*Enterococcus faecium*, *E. faecalis, Acidaminococcus intestini*, *Erysipelotrichaceae bacterium* 3-1-53, and *F. varium*) and negatively correlated with three RA-enriched species (*Eikenella corrodens*, *S. noxia*, and *Neisseria meningitidis*). Of note, *E. bacterium* 3-1-53 was positively correlated with AST, while *E. corrodens*, *S. noxia*, and *N. meningitidis* were negatively correlated with AST, and *E. bacterium* 3-1-53, *E. faecalis*, and *F. varium* were negatively correlated with RF and anti-CCP. Thus, gut microbes may affect phenylalanine metabolism and thereby influence AST and RF/anti-CCP (via mechanisms that act in opposite directions for the former *vs* the latter two). In addition, the *Salmonella* infection pathway (map05132) was identified as being positively correlated with ESR, CRP, RF, and anti-CCP, and it was negatively correlated with six HC-enriched species (*E. faecium*, *Xanthomonas perforans*, *E. faecalis, E. bacterium* 3-1-53, *Acidaminococcus intestinii*, and *F. varium*) and positively correlated with two RA-enriched species (*S. noxia* and *N. meningitidis*). Of note, *X. perforans, E. faecalis, E. bacterium 3-1-53*, and *F. varium* were negatively correlated with RF and anti-CCP, and the mechanism may involve the *Salmonella* infection pathway. These results suggest that gut microbes may influence clinical indices in the host.

Different Effects of the Two Treatments on Clinical Indicators

Furtherly, to assess the effects of different therapies on the gut microbiome in RA patients, the 13 RA patients in the HQT group received traditional Chinese medicine plus MTX and the 9 patients in the LEF group received LEF plus MTX. The RA patients underwent clinical assessment (including arthritis severity scores and inflammation-related indices) and blood and fecal sample collection at baseline and three times during the 6-month treatment period (**Figure 3A**).

Among the 22 RA patients who were treated for 6 months, most of the clinical indices were continuously and effectively improved over time compared to baseline in both the HQT and LEF groups (**Figures 3B**, **S4** and **Table S1**), including disease activity score for 28 joints based on the C-reactive protein level (DAS28-CRP), morning stiffness duration (MS), joint tenderness score (JTS), joint swelling score (JSS), RPJ, visual analogue scale for disability (VAS-D), and visual analogue scale for pain (VAS-P).

However, the two treatments influenced the clinical indices in different ways. In the HQT group, JSS, JTS, and MS were significantly reduced at month 3 compared to baseline, while these clinical indices were only significantly reduced in the LEF group at month 6. Thus, these clinical indices were improved

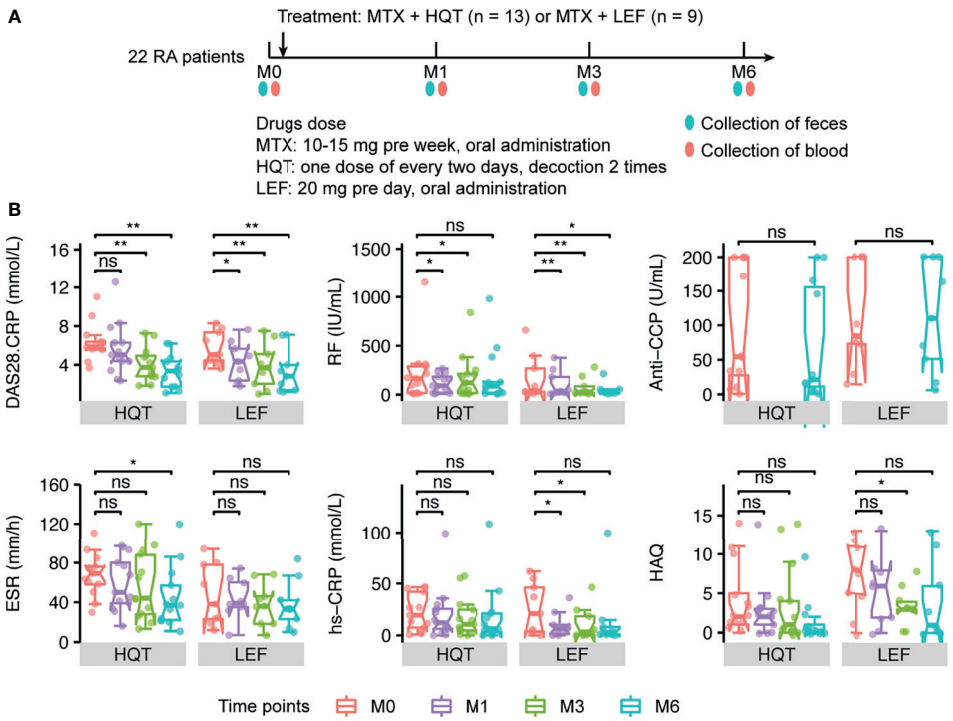


FIGURE 3 | Different effects of the two treatments on primary clinical indicators **(A)** Sample collection time points in the two treatment groups. **(B)** Changes between time points in RA-related primary clinical indices in HQT and LEF groups. *P < 0.05; **P < 0.01, ns, no significance, between different time points according to Wilcoxon rank-sum test. DAS28-CRP, disease activity score for 28 joints based on the C-reactive protein level; RF, rheumatoid factor; Anti-CCP, anti-cyclic citrullinated peptides; CRP, high-sensitivity C-reactive protein; ESR, erythrocyte sedimentation rate; HAQ, health assessment questionnaires.

earlier in the HQT group than in the LEF group. Additionally, in the LEF group, RF, DAS28-CRP, and RPJ were decreased at month 1 until the end of the study while, in the HQT group, RF was significantly decreased at month 1, though not at month 6, and DAS28-CRP and RPJ were significantly decreased at months 3 and 6. Moreover, ESR did not significantly change with LEF treatment, but significantly decreased at month 6 in the HQT group. In contrast, CRP was altered by LEF, but HQT had no effect. Furthermore, in the HQT group, VAS-P was decreased at month 1 until the end of the study while, in the LEF group, it only significantly decreased at month 3 until the end of the study. Health Assessment Questionnaires (HAQ) scores were not significantly changed in the HQT group, but significantly decreased at month 3 in the LEF group. Neither treatment affected anti-CCP levels (**Figures 3B**, **S4** and **Table S1**). Although many clinical indices significantly differed at various time points compared to baseline in the same treatment group, there were no significant differences between the two treatment groups. In summary, the response trends of RA-related clinical indices in the two treatment groups differed, but the clinical efficacy of the two treatments was similar.

Effects of Treatment on the Microbiome

Based on longitudinally tracking the gut microbiota during treatment, the gut microbial diversity (alpha or beta) did not significantly differ over time, except for beta diversity being greater between baseline and month 6 than between baseline and month 1 in the LEF group (**Figures S5A, B** and **Table S12**). In the HQT group, 11 species and 9 MetaCyc metabolic pathways significantly changed over time (based on the Jonckheere–Terpstra test), with 4 species (*C. somerae*, *Haemophilus aegyptius*, *Dialister succinatiphilus*, and *C. celatum*) being restored by the late stage of HQT treatment (**Figure 4A** and **Table S5**). Additionally, as HQT treatment progressed, the abundances of *Roseburiaw inulinivorans*, *Turicibacter sanguinis*, and *Pasteurella bettyae* significantly increased while *Clostridium symbiosum* and *Clostridiales bacterium* 1-7-47FAA significantly decreased (**Figure 4A** and **Table S5**). In the HQT group, bacterial purine degradation (PWY0-1297 and PWY-6353) was decreased and amino acid biosynthesis (VALSYN-PWY and ILEUSYN-PWY) was increased over time (**Figure 4B** and **Table S6**). In the LEF group, only 4 species and 2 MetaCyc metabolic pathways significantly changed over time (based on the Jonckheere–Terpstra test), but their abundances were very low in both groups (**Figure S6**, **Tables S13** and **S14**). However, the *Prevotella* spp. known as the dominant gut microbiota in RA patients was not significantly different in both the HQT-treated and LEF-treated groups. Thus, there were few changes over time in the relative abundance of gut microbiota species in the LEF group, but many changes in the HQT group.

To determine how distinct the gut microbiota in the LEF and HQT groups were, the relative abundance of each species

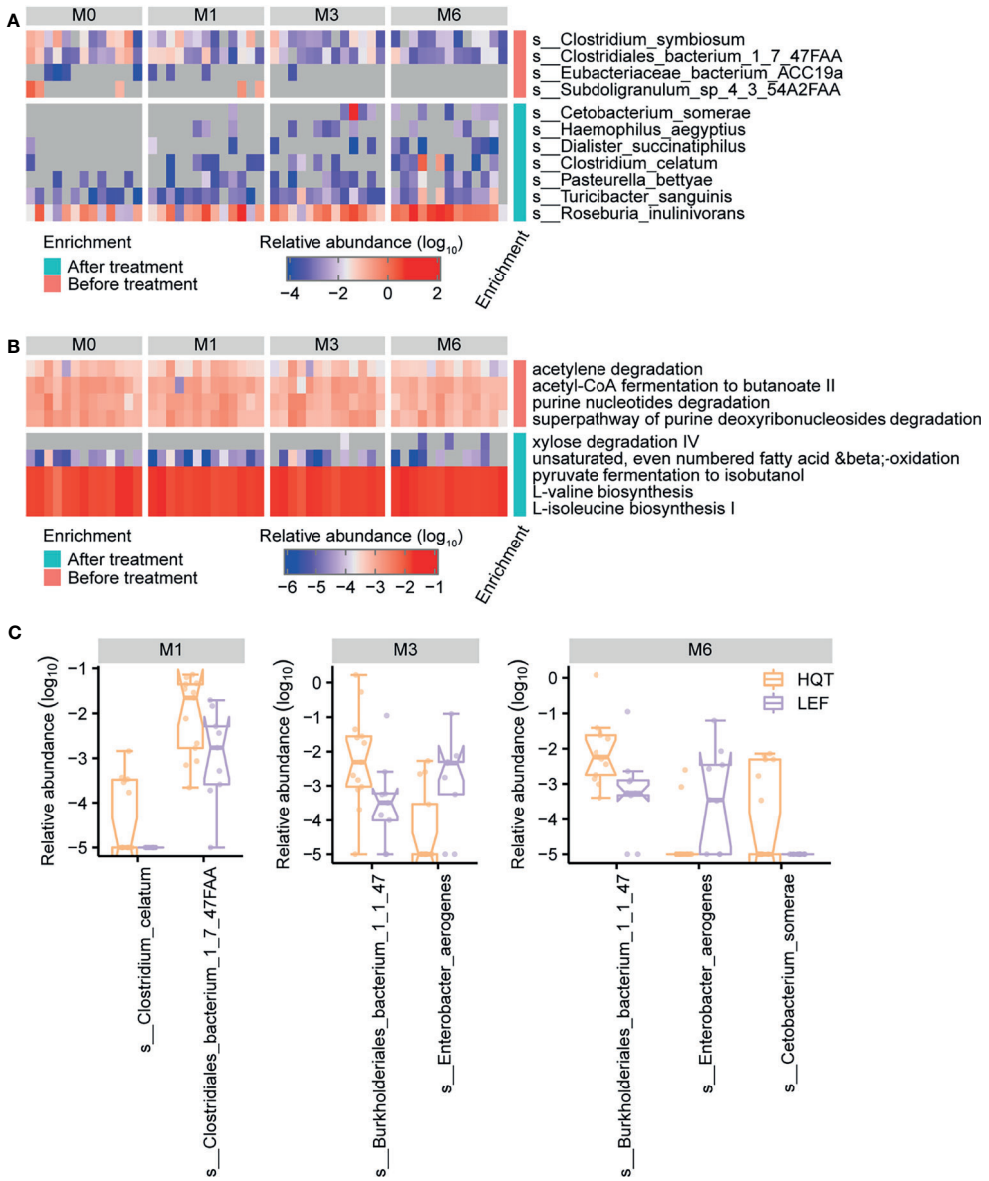


FIGURE 4 | Trends over time in gut microbial taxonomy and function after treatment with Huayu-Qiangshen-Tongbi (HQT) or leflunomide (LEF). Significant changes in **(A)** species and **(B)** MetaCyc metabolic pathways in HQT group over time. $P < 0.05$ according to Jonckheere–Terpstra test. **(C)** Comparisons of relative abundance of various species at various time points between the HQT and LEF groups. $P < 0.05$ according to Wilcoxon rank-sum test.

between the LEF and HQT groups at each time point was assessed using the Wilcoxon rank-sum test. In the HQT group compared to the LEF group, *C. bacterium* 1-7-47FAA was increased at month 1, *Burkholderiales bacterium* 1-1-47 was increased at months 3 and 6, and *Enterobacter aerogenes* was decreased at months 3 and 6 (**Figure 4C**).

DISCUSSION

In this study, gut microbiome metagenome sequencing provided rich microbial data, including data on the microbial compositions and functions, in participants with diverse clinical phenotypes (in terms of RA-related clinical indices). We reconfirmed that certain bacteria (including oral microbiome bacteria) in the gut were associated with RA. More changes over time in the gut microbiome were observed during traditional Chinese medicine (HQT) treatment than LEF treatment. This study provides a deep understanding of the gut microbiome of RA patients during LEF or HQT treatment, and it may help to develop more efficient and safe treatments for RA.

The diversity of the gut microbiome in RA patients is a matter of ongoing debate. Some studies have reported significant differences in microbiome diversity between RA patients and

healthy individuals (8, 9), but others have indicated no differences (7, 24). Our results indicated that the microbial diversity did not significantly differ between RA patients and healthy individuals, which may be caused by the small sample size and/or the involvement of RA patients with different phenotypes.

The species identified as being significantly different between RA and HCs vary among studies. Kishikawa et al. and Scher et al. reported that multiple species in the genus *Prevotella* were increased in RA patients in Japanese (24) and US (25) cohorts, respectively. In a Chinese cohort, Sun et al. found that the genera *Bacteroides* and *Escherichia-Shigella* were more abundant in RA patients (8), while Chiang et al. reported a higher abundance of the genus *Akkermansia* (9). In the present study, species such as *Streptococcus* spp., *Haemophilus* spp., and *Neisseria* spp. were found to be enriched in RA patients, while *Clostridium celatum*, *Enterococcus faecalis*, and *Fusobacterium varium* were depleted. Most of the RA-enriched species belonged to *Proteobacteria* and *Firmicutes*, and the HC-enriched species belonged to *Firmicutes*. Among them, *Enterococcus faecalis*, Yamamoto et al. found that hyperimmunization with attenuated *E. faecalis* as normal gut microbiota could provide an animal model of chronic polyarthritis (26). Chandradevan et al. and Luo et al. observed *E. faecalis* in both blood cultures and synovial tissue culture from RA patients (27, 28). However, other species and phyla have not been reported in other studies. The fact that the RA-associated species differed among these studies might reflect dietary (29) and geographical (30) variation impacting the gut microbiome composition in the various study samples.

Meanwhile, in this study, we found that most of the RA-enriched bacteria are known to colonize the oral cavity, such as *P. gingivalis*, *Aggregatibacter segnis*, *Streptococcus* spp., *Haemophilus* spp., and *Neisseria* spp. Additionally, oral microbial dysbiosis in RA patients has been reported to promote increased joint inflammation, and the oral bacteria *P. gingivalis*, *A. actinomycetemcomitans*, and *Prevotella nigrescens* are associated with RA pathogenesis (31). Moreover, oral cavity and gut bacteria in RA patients exhibit covariation. Therefore, the inflammatory status in the joints in RA patients may be due to the translocation of oral bacteria to the gut, which is consistent with our findings.

Whole-genome shotgun sequencing of the metagenome provides data on microbial function, allowing identification of altered microbial functions in RA patients relative to healthy individuals. Menaquinone, a type of vitamin K, can be produced by bacteria. Several MetaCyc menaquinone biosynthesis-related pathways were enriched in RA patients, which is consistent with previous research (32). Vitamin K homologs have been shown to affect serum CRP, matrix metalloproteinase (MMP)-3, and DAS28-CRP in RA patients (33). This indicates that the gut microbiota may contribute to the initiation or development of RA by affecting vitamin K biosynthesis.

To identify the associations of the gut microbiome species and gut microbiome functional pathways with RA-related clinical indices, correlation coefficients were calculated using the baseline data from the RA and HC groups. Most of the HC-enriched pathways and species were negatively correlated with ESR, CRP, RF, and anti-CCP, while the RA-enriched pathways and species were positively correlated with these clinical indices. Further, we concluded that gut microbes may affect phenylalanine metabolism and thereby influence the liver function index AST and RA specific index RF/anti-CCP (via mechanisms that act in opposite directions for the former *vs* the latter two). This pathway was positively correlated with HC-enriched *E. bacterium* 3-1-53, and this species was positively correlated with AST and negatively correlated with ESR, CRP, RF, and anti-CCP, which suggested that this species directly influences phenylalanine metabolism and thereby affects ESR, CRP, RF, anti-CCP, and AST. Furthermore, several gut microbial species may affect the clinical indices by influencing the *Salmonella* infection pathway. Among these species, HC-enriched species were negatively correlated with ESR, CRP, RF, and/or anti-CCP and positively correlated with AST, while RA-enriched species were negatively correlated with AST. Associations between the gut microbiota in RA patients and clinical indices were observed in another study (11), but the related gut microbiome functional pathways have rarely been reported.

By assessing the effects of different treatments on RA patients, we found that both treatments, to a certain extent, alleviated RA progression, based on different clinical indices. LEF improved the CRP level, while the traditional Chinese medicine (HQT) improved the ESR level. This result indicates that their mechanisms of action differ. Previous research reported that the gut microbiota was moderately restored in RA patients after DMARD treatment (7), and similar results were found in spondyloarthritis (34) and ankylosing spondylitis (35). We observed that several bacteria were restored after both treatments but the gut microbiota exhibited greater restoration (regarding gut microbial species and functional pathways) in the HQT group than the LEF group. A possible reason for this is that the composition of the traditional Chinese medicine (HQT) is more complicated than LEF and so has broader effects.

In a previous study, the quantification of 10 compounds in HQT extracts was carried out by UPLC-PDA method (18). The 10 compounds are danshensu, 3-caffeoylquinic acid, paeoniflorin, rutin, quercetin, salvigenin, caffeic acid, rosmarinic acid, calycosin and glycyrrhizic acid, respectively. Interestingly, most of these 10 compounds in the HQT extract were shown to have significant antibacterial effects and even to directly modulate the homeostasis of the gut microbiota. For example, Liu et al. suggested that the paeoniflorin derivative can be used as an antibacterial agent to abolish the hemolytic activity of *Staphylococcus aureus* α-toxin (36). Motallebi et al. highlighted that a combination of rutin and florfenicol could act as an alternative strategy to treat bacterial infections (37). The antibacterial effects of quercetin have been demonstrated in several studies (38, 39), and quercetin has even been found to be effective in restoring the intestinal microbiota of mice after antibiotic treatment (40). Meanwhile, chlorogenic acid was found to have antibacterial activity against Foodborne Pathogen *Pseudomonas aeruginosa* and *Salmonella enteritidis* (41), and caffeic acid was found to have antibacterial activity against *Staphylococcus aureus* clinical strains (42). In addition,

the antibacterial effect of *Salvia miltiorrhiza* Bunge (Danshen), a major component herb of HQT, has been widely reported and studied (43–45). Therefore, we speculate that HQT can significantly restore and improve gut microbial species and function in RA patients, possibly due to the fact that HQT contains a variety of compounds with antimicrobial effects.

During drug treatment, there were 11 species and 9 MetaCyc metabolic pathways in the HQT group while 4 species and 2 MetaCyc metabolic pathways in the LEF group changed significantly over time. These species and pathways include *C. somerae, Clostridium symbiosum, Turicibacter sanguinis, C. celatum* and bacterial purine degradation pathways (PWY0-1297 and PWY-6353) et al. In particular, the presence and abundance of an important microbe, *C. celatum*, was restored in the HQT group. Importantly, it was shown to be downregulated in RA patients relative to HC. Currently, *C. celatum* has tended not to be noted in microbiota studies of RA patients. However, for the genus *Clostridium*, Schmidt et al. found that infection with *Clostridioides (Clostridium) difficile* VPI 10463 induced intestinal inflammation and thus reduced the incidence of collagen induced arthritis (CIA) in mice (46). Moreover, we identified several other restored species and functions that are not consistent with previous studies, possibly due to treatment differences. Recent studies have also reported that MTX treatment can alter the gut microbiota of RA patients. Nayak et al. evaluated the differences in gut microbiome between responders (MTX-R) and non-responders (MTX-NR) after MTX treatment for RA, but found that MTX-R exhibited a significant decrease in *Bacteroidetes* relative to MTX-NR, without any significant difference in the other phyla (47). Similarly, the analysis by Artacho et al. found that, at the phylum level, MTX-R was significantly more abundant in OTUs from *Bacteroides* and *Prevotella* genus (*Bacteroidetes* phylum) and less abundant in OTUs from the order *Clostridiales* and the genus *Ruminococcus* (phylum *Firmicutes*) (48). However, These microbiota were not found to change in our study of drug therapy, which may be influenced by various factors such as diet and environment. Therefore, in this study, we speculate that some of the alterations in the gut microbiota of the HQT and LEF groups were also caused by MTX.

Despite our promising findings, several limitations should be considered when interpreting the results. Accumulating studies have demonstrated that gut microbial dysbiosis can occur in RA patients compared to healthy individuals, but differences in gut microbial compositions were also observed among RA patients with different clinical phenotypes, such as ACPA seropositivity and cytokine levels (TNF-α, IL-6, and IL-17A) (8, 9). Sun et al. also reported that some basic characteristics of RA patients (including age, gender and enterotypes) alter the gut microbiota (8). The effects of geographical location on the gut microbiome community structure (30) also need to be considered. Therefore, future metagenome research on RA populations should be designed to minimize the impact of these known factors that introduce bias. Given the known heterogeneity of the gut microbiome among individuals, exploring the role of the gut microbiome in the etiology and pathogenesis of RA requires larger sample sizes to ensure sufficient statistical power. Unfortunately, current RA microbiome studies (49, 50), including the present study, are limited by their small sample sizes. Therefore, larger longitudinal cohort studies, which can provide more precise results, are required to confirm the current research results to fully understand the etiology of RA.

Our results further confirm that the initiation and/or development of RA is accompanied by alterations in part of the gut microbiota and also demonstrate that the microbiome composition and functions change during treatment, including treatment with traditional Chinese medicine. *C. celatum* was depleted in RA patients relative to HCs and it was restored in RA patients by HQT treatment. Additionally, vitamin K biosynthesis may act as a newly identified bridge between the gut microbiome and RA. Further, we observed that the two treatments had similar clinical efficacy, but the response trends of RA-related clinical indices differed between treatments. Moreover, the abundances of specific gut microbiome species in RA patients were associated with various serological and clinical indices. However, validation of these potential RA-related microbial markers using large independent cohorts is required. This is also a deficiency of the current research, and we need to consider increasing the number of samples and discussing more optimal experimental design options in depth in future studies. However, the data from this exploratory phase of the study may provide a reference for further large cohort studies. Further studies on the role of the gut microbiome in RA should incorporate other omics technologies including metatranscriptomics and metabolomics and other microbiomes such as mycobiomes and viromes. This research provides useful resources for the future development of new therapeutic strategies for RA.

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://db.cngb.org/cnsa/, CNP0001832.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the ethics committee of Guangdong Provincial Hospital of Chinese Medicine. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

RH and QH designed the study and supervised all parts of the study. MW and ZY contributed to the clinical trial and the

collection of samples and clinical data. LM conducted analyses and wrote the manuscript. All authors approved the final version.

FUNDING

This study was supported by National Natural Science Foundation of China (No.81774218, No.81804041), Natural Science Foundation of Guangdong Province (No. 2021A1515011593, No. 2021A1515011477), the grant from the Clinical Research Project of Guangdong Provincial Hospital of Chinese Medicine (No. YN10101906, YN2018ML08), Guangdong-Hong Kong-Macau Joint Lab on Chinese Medicine and Immune Disease Research (2020B1212030006), Guangdong Provincial Key laboratory of Chinese Medicine for Prevention and Treatment of Refractory Chronic Diseases(2018) (No. 2018B030322012), the grant from Guangzhou Basic Research Program (No.202102010256), as well as grants from Guangdong Provincial Hospital of Chinese Medicine (No. MB2019ZZ07). The study was also funded by State Key Laboratory Project of Dampness Syndrome of Chinese Medicine (No. SZ2020ZZ17), the Key Research Project of Guangzhou University of Chinese Medicine (No. XK2019021), and the Key-Area Research and Development Program of Guangdong Province (No. 2020B1111100010).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2021.704089/full#supplementary-material

REFERENCES

1. Weyand CM, Goronzy JJ. Immunometabolism in the Development of Rheumatoid Arthritis. *Immunol Rev* (2020) 294:177–87. doi: 10.1111/imr.12838
2. Smolen JS, Aletaha D, Barton A, Burmester GR, Emery P, Firestein GS, et al. Rheumatoid Arthritis. *Nat Rev Dis Primers* (2018) 4:18001. doi: 10.1038/nrdp.2018.1
3. Shamriz O, Mizrahi H, Werbner M, Shoenfeld Y, Avni O, Koren O. Microbiota at the Crossroads of Autoimmunity. *Autoimmun Rev* (2016) 15:859–69. doi: 10.1016/j.autrev.2016.07.012
4. Taneja V. Arthritis Susceptibility and the Gut Microbiome. *FEBS Lett* (2014) 588:4244–9. doi: 10.1016/j.febslet.2014.05.034
5. Correa JD, Fernandes GR, Calderaro DC, Mendonca S, Silva JM, Albiero ML, et al. Oral Microbial Dysbiosis Linked to Worsened Periodontal Condition in Rheumatoid Arthritis Patients. *Sci Rep* (2019) 9:8379. doi: 10.1038/s41598-019-44674-6
6. Scher JU, Littman DR, Abramson SB. Microbiome in Inflammatory Arthritis and Human Rheumatic Diseases. *Arthritis Rheumatol* (2016) 68:35–45. doi: 10.1002/art.39259
7. Zhang X, Zhang D, Jia H, Feng Q, Wang D, Liang D, et al. The Oral and Gut Microbiomes Are Perturbed in Rheumatoid Arthritis and Partly Normalized After Treatment. *Nat Med* (2015) 21:895–905. doi: 10.1038/nm.3914
8. Sun Y, Chen Q, Lin P, Xu R, He D, Ji W, et al. Characteristics of Gut Microbiota in Patients With Rheumatoid Arthritis in Shanghai, China. *Front Cell Infect Microbiol* (2019) 9:369. doi: 10.3389/fcimb.2019.00369
9. Chiang HI, Li JR, Liu CC, Liu PY, Chen HH, Chen YM, et al. An Association of Gut Microbiota With Different Phenotypes in Chinese Patients With Rheumatoid Arthritis. *J Clin Med* (2019) 8:1770. doi: 10.3390/jcm8111770
10. Mercado F, Marshall RI, Klestov AC, Bartold PM. Is There a Relationship Between Rheumatoid Arthritis and Periodontal Disease? *J Clin Periodontol* (2000) 27:267–72. doi: 10.1034/j.1600-051x.2000.027004267.x
11. Picchianti-Diamanti A, Panebianco C, Salemi S, Sorgi ML, Di Rosa R, Tropea A, et al. Analysis of Gut Microbiota in Rheumatoid Arthritis Patients: Disease-Related Dysbiosis and Modifications Induced by Etanercept. *Int J Mol Sci* (2018) 19:2938. doi: 10.3390/ijms19102938
12. Wang B, He Y, Tang J, Ou Q, Lin J. Alteration of the Gut Microbiota in Tumor Necrosis Factor-Alpha Antagonist-Treated Collagen-Induced Arthritis Mice. *Int J Rheum Dis* (2020) 23:472–9. doi: 10.1111/1756-185X.13802
13. Dorozynska I, Majewska-Szczepanik M, Marcinska K, Szczepanik M. Partial Depletion of Natural Gut Flora by Antibiotic Aggravates Collagen Induced Arthritis (CIA) in Mice. *Pharmacol Rep* (2014) 66:250–5. doi: 10.1016/j.pharep.2013.09.007
14. Guo LX, Wang HY, Liu XD, Zheng JY, Tang Q, Wang XN, et al. Saponins From Clematis Mandshurica Rupr. Regulates Gut Microbiota and Its Metabolites During Alleviation of Collagen-Induced Arthritis in Rats. *Pharmacol Res* (2019) 149:104459. doi: 10.1016/j.phrs.2019.104459
15. Lv Y, Chen X, Huang R, Zhao Y, Wu J, Chen XM, et al. Clinical Analysis of Chinese Medicine Compound Huayu-Qiangshen-Tongbi Decoction Combined With Methotrexate for the Treatment of Chinese Patients With Rheumatoid Arthritis: A Retrospective Study. *Zhongguo Zhong Xi Yi Jie He Za Zhi* (2019) 39:547–52. doi: 10.7661/j.cjim.20190329.110
16. Wu J, Chen X, Lv Y, Gao K, Liu Z, Zhao Y, et al. Chinese Herbal Formula Huayu-Qiangshen-Tongbi Decoction Compared With Leflunomide in Combination With Methotrexate in Patients With Active Rheumatoid Arthritis: An Open-Label, Randomized, Controlled, Pilot Study. *Front Med (Lausanne)* (2020) 7:484. doi: 10.3389/fmed.2020.00484
17. Wang M, Mei L, Liu Z, Tang X, Wu X, Chen X, et al. The Mechanism of Chinese Herbal Formula HQT in the Treatment of Rheumatoid Arthritis Is Related to Its Regulation of lncRNA Uc.477 and miR-19b. *J Leukoc Biol* (2020) 108:519–29. doi: 10.1002/JLB.3MA0620-441RRRR
18. Wang Z, Linghu KG, Hu Y, Zuo H, Yi H, Xiong SH, et al. Deciphering the Pharmacological Mechanisms of the Huayu-Qiangshen-Tongbi Formula Through Integrating Network Pharmacology and *In Vitro* Pharmacological Investigation. *Front Pharmacol* (2019) 10:1065. doi: 10.3389/fphar.2019.01065
19. Kay J, Upchurch KS. ACR/EULAR 2010 Rheumatoid Arthritis Classification Criteria. *Rheumatol (Oxford)* (2012) 51 Suppl 6:i5–9. doi: 10.1093/rheumatology/kes279
20. Patil KR, Nielsen J. Uncovering Transcriptional Regulation of Metabolism by Using Metabolic Network Topology. *Proc Natl Acad Sci USA* (2005) 102:2685–9. doi: 10.1073/pnas.0406811102
21. Feng Q, Liang S, Jia H, Stadlmayr A, Tang L, Lan Z, et al. Gut Microbiome Development Along the Colorectal Adenoma-Carcinoma Sequence. *Nat Commun* (2015) 6:6528. doi: 10.1038/ncomms7528
22. Arumugam M, Raes J, Pelletier E, Le Paslier D, Yamada T, Mende DR, et al. Enterotypes of the Human Gut Microbiome. *Nature* (2011) 473:174–80. doi: 10.1038/nature09944
23. Pedersen HK, Forslund SK, Gudmundsdottir V, Petersen AO, Hildebrand F, Hyotylainen T, et al. A Computational Framework to Integrate High-Throughput '-Omics' Datasets for the Identification of Potential Mechanistic Links. *Nat Protoc* (2018) 13:2781–800. doi: 10.1038/s41596-018-0064-z
24. Kishikawa T, Maeda Y, Nii T, Motooka D, Matsumoto Y, Matsushita M, et al. Metagenome-Wide Association Study of Gut Microbiome Revealed Novel Aetiology of Rheumatoid Arthritis in the Japanese Population. *Ann Rheum Dis* (2020) 79:103–11. doi: 10.1136/annrheumdis-2019-215743
25. Scher JU, Sczesnak A, Longman RS, Segata N, Ubeda C, Bielski C, et al. Expansion of Intestinal Prevotella Copri Correlates With Enhanced Susceptibility to Arthritis. *ELIFE* (2013) 2:e1202. doi: 10.7554/eLife.01202
26. Yamamoto H, Konishi Y, Mineo S, Sekiya M, Kohno T, Kohno M. Characteristics of Polyarthritis in Rabbits by Hyperimmunization With Attenuated Enterococcus Faecalis. *Exp Toxicol Pathol* (2000) 52:247–55. doi: 10.1016/S0940-2993(00)80041-X

27. Chandradevan R, Takeda H, Lim T, Patel N. Mycobacterium Tuberculosis Concealed by Enterococcal Sacroiliitis. *IDCases* (2020) 21:e858. doi: 10.1016/j.idcr.2020.e00858
28. Luo JM, Guo L, Chen H, Yang PF, Xiong R, Peng Y, et al. A Study of Pre-Operative Presence of Micro-Organisms in Affected Knee Joints of Rheumatoid Arthritis Patients Who Need Total Knee Arthroplasty. *KNEE* (2017) 24:409–18. doi: 10.1016/j.knee.2016.11.001
29. Philippou E, Petersson SD, Rodomar C, Nikiphorou E. Rheumatoid Arthritis and Dietary Interventions: Systematic Review of Clinical Trials. *Nutr Rev* (2021) 79:410–28. doi: 10.1093/nutrit/nuaa033
30. He Y, Wu W, Zheng HM, Li P, McDonald D, Sheng HF, et al. Regional Variation Limits Applications of Healthy Gut Microbiome Reference Ranges and Disease Models. *Nat Med* (2018) 24:1532–5. doi: 10.1038/s41591-018-0164-x
31. du Teil EM, Gabarrini G, Harmsen H, Westra J, van Winkelhoff AJ, van Dijl JM. Talk to Your Gut: The Oral-Gut Microbiome Axis and Its Immunomodulatory Role in the Etiology of Rheumatoid Arthritis. *FEMS Microbiol Rev* (2019) 43:1–18. doi: 10.1093/femsre/fuy035
32. Jeong Y, Kim JW, You HJ, Park SJ, Lee J, Ju JH, et al. Gut Microbial Composition and Function Are Altered in Patients With Early Rheumatoid Arthritis. *J Clin Med* (2019) 8:693. doi: 10.3390/jcm8050693
33. Ebina K, Shi K, Hirao M, Kaneshiro S, Morimoto T, Koizumi K, et al. Vitamin K2 Administration Is Associated With Decreased Disease Activity in Patients With Rheumatoid Arthritis. *MOD Rheumatol* (2013) 23:1001–7. doi: 10.3109/s10165-012-0789-4
34. Bazin T, Hooks KB, Barnetche T, Truchetet ME, Enaud R, Richez C, et al. Microbiota Composition May Predict Anti-Tnf Alpha Response in Spondyloarthritis Patients: An Exploratory Study. *Sci Rep* (2018) 8:5446. doi: 10.1038/s41598-018-23571-4
35. Yin J, Sternes PR, Wang M, Song J, Morrison M, Li T, et al. Shotgun Metagenomics Reveals an Enrichment of Potentially Cross-Reactive Bacterial Epitopes in Ankylosing Spondylitis Patients, as Well as the Effects of TNFi Therapy Upon Microbiome Composition. *Ann Rheum Dis* (2020) 79:132–40. doi: 10.1136/annrheumdis-2019-215763
36. Liu X, Zhang Y, Li Z, Zhang P, Sun YJ, Wu YJ. Paeoniflorin Derivative in Paeoniae Radix Aqueous Extract Suppresses Alpha-Toxin of Staphylococcus Aureus. *Front Microbiol* (2021) 12:649390. doi: 10.3389/fmicb.2021.649390
37. Motallebi M, Khorsandi K, Sepahy AA, Chamani E, Hosseinzadeh R. Effect of Rutin as Flavonoid Compound on Photodynamic Inactivation Against P. Aeruginosa and S. Aureus. *Photodiagnosis Photodyn Ther* (2020) 32:102074. doi: 10.1016/j.pdpdt.2020.102074
38. Yang X, Zhang W, Zhao Z, Li N, Mou Z, Sun D, et al. Quercetin Loading CdSe/ZnS Nanoparticles as Efficient Antibacterial and Anticancer Materials. *J Inorg Biochem* (2017) 167:36–48. doi: 10.1016/j.jinorgbio.2016.11.023
39. Wang S, Yao J, Zhou B, Yang J, Chaudry MT, Wang M, et al. Bacteriostatic Effect of Quercetin as an Antibiotic Alternative *In Vivo* and Its Antibacterial Mechanism *In Vitro*. *J Food Prot* (2018) 81:68–78. doi: 10.4315/0362-028X.JFP-17-214
40. Shi T, Bian X, Yao Z, Wang Y, Gao W, Guo C. Quercetin Improves Gut Dysbiosis in Antibiotic-Treated Mice. *Food Funct* (2020) 11:8003–13. doi: 10.1039/D0FO01439G
41. Su M, Liu F, Luo Z, Wu H, Zhang X, Wang D, et al. The Antibacterial Activity and Mechanism of Chlorogenic Acid Against Foodborne Pathogen Pseudomonas Aeruginosa. *Foodborne Pathog Dis* (2019) 16:823–30. doi: 10.1089/fpd.2019.2678
42. Kepa M, Miklasinska-Majdanik M, Wojtyczka RD, Idzik D, Korzeniowski K, Smolen-Dzirba J, et al. Antimicrobial Potential of Caffeic Acid Against Staphylococcus Aureus Clinical Strains. *BioMed Res Int* (2018) 2018:7413504. doi: 10.1155/2018/7413504
43. Kong WJ, Zhang SS, Zhao YL, Wu MQ, Chen P, Wu XR, et al. Combination of Chemical Fingerprint and Bioactivity Evaluation to Explore the Antibacterial Components of Salvia Miltiorrhizae. *Sci Rep* (2017) 7:8112. doi: 10.1038/s41598-017-08377-0
44. Lai D, Li J, Zhao S, Gu G, Gong X, Proksch P, et al. Chromone and Isocoumarin Derivatives From the Endophytic Fungus Xylomelasma Sp. Samif07, and Their Antibacterial and Antioxidant Activities. *Nat Prod Res* (2019) 33:1–5. doi: 10.1080/14786419.2019.1696333
45. Liu QQ, Han J, Zuo GY, Wang GC, Tang HS. Potentiation Activity of Multiple Antibacterial Agents by Salvianolate From the Chinese Medicine Danshen Against Methicillin-Resistant Staphylococcus Aureus (MRSA). *J Pharmacol Sci* (2016) 131:13–7. doi: 10.1016/j.jphs.2015.10.009
46. Schmidt CJ, Wenndorf K, Ebbers M, Volzke J, Muller M, Strubing J, et al. Infection With Clostridioides Difficile Attenuated Collagen-Induced Arthritis in Mice and Involved Mesenteric Treg and Th2 Polarization. *Front Immunol* (2020) 11:571049. doi: 10.3389/fimmu.2020.571049
47. Nayak RR, Alexander M, Deshpande I, Stapleton-Gray K, Rimal B, Patterson AD, et al. Methotrexate Impacts Conserved Pathways in Diverse Human Gut Bacteria Leading to Decreased Host Immune Activation. *Cell Host Microbe* (2021) 29:362–77. doi: 10.1016/j.chom.2020.12.008
48. Artacho A, Isaac S, Nayak R, Flor-Duro A, Alexander M, Koo I, et al. The Pretreatment Gut Microbiome Is Associated With Lack of Response to Methotrexate in New-Onset Rheumatoid Arthritis. *Arthritis Rheumatol* (2021) 73:931–42. doi: 10.1002/art.41622
49. Eriksson K, Fei G, Lundmark A, Benchimol D, Lee L, Hu Y, et al. Periodontal Health and Oral Microbiota in Patients With Rheumatoid Arthritis. *J Clin Med* (2019) 8:630. doi: 10.3390/jcm8050630
50. Scher JU, Joshua V, Artacho A, Abdollahi-Roodsaz S, Ockinger J, Kullberg S, et al. The Lung Microbiota in Early Rheumatoid Arthritis and Autoimmunity. *Microbiome* (2016) 4:60. doi: 10.1186/s40168-016-0206-x

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.

frontiers in Immunology

ORIGINAL RESEARCH
published: 05 November 2021
doi: 10.3389/fimmu.2021.743559

Continuous Modeling of T CD4 Lymphocyte Activation and Function

David Martínez-Méndez[1], *Luis Mendoza*[2,3], *Carlos Villarreal*[1,3*] *and Leonor Huerta*[2*]

[1] Instituto de Física, Universidad Nacional Autónoma de México, Mexico City, Mexico, [2] Instituto de Investigaciones Biomédicas, Universidad Nacional Autónoma de México, Mexico City, Mexico, [3] Centro de Ciencias de la Complejidad, Universidad Nacional Autónoma de México, Mexico City, Mexico

T CD4+ cells are central to the adaptive immune response against pathogens. Their activation is induced by the engagement of the T-cell receptor by antigens, and of co-stimulatory receptors by molecules also expressed on antigen presenting cells. Then, a complex network of intracellular events reinforce, diversify and regulate the initial signals, including dynamic metabolic processes that strongly influence both the activation state and the differentiation to effector cell phenotypes. The regulation of cell metabolism is controlled by the nutrient sensor adenosine monophosphate-activated protein kinase (AMPK), which drives the balance between oxidative phosphorylation (OXPHOS) and glycolysis. Herein, we put forward a 51-node continuous mathematical model that describes the temporal evolution of the early events of activation, integrating a circuit of metabolic regulation into the main routes of signaling. The model simulates the induction of anergy due to defective co-stimulation, the CTLA-4 checkpoint blockade, and the differentiation to effector phenotypes induced by external cytokines. It also describes the adjustment of the OXPHOS-glycolysis equilibrium by the action of AMPK as the effector function of the T cell develops. The development of a transient phase of increased OXPHOS before induction of a sustained glycolytic phase during differentiation to the Th1, Th2 and Th17 phenotypes is shown. In contrast, during Treg differentiation, glycolysis is subsequently reduced as cell metabolism is predominantly polarized towards OXPHOS. These observations are in agreement with experimental data suggesting that OXPHOS produces an ATP reservoir before glycolysis boosts the production of metabolites needed for protein synthesis, cell function, and growth.

Keywords: T CD4 cells, metabolism, T cell receptor, lymphocyte activation, CTLA-4, mTOR, regulatory network, mathematical model

OPEN ACCESS

Edited by:
Federica Eduati,
Eindhoven University of Technology, Netherlands

Reviewed by:
Richard H. Scheuermann,
J. Craig Venter Institute (La Jolla), United States
Kashish Chetal,
Massachusetts General Hospital and Harvard Medical School, United States
James Faeder,
University of Pittsburgh, United States

***Correspondence:**
Leonor Huerta
leonorhh@biomedicas.unam.mx
Carlos Villarreal
carlos@fisica.unam.mx

Specialty section:
This article was submitted to Systems Immunology, a section of the journal Frontiers in Immunology

Received: *18 July 2021*
Accepted: *05 October 2021*
Published: *05 November 2021*

Citation:
Martínez-Méndez D, Mendoza L, Villarreal C and Huerta L (2021) Continuous Modeling of T CD4 Lymphocyte Activation and Function. Front. Immunol. 12:743559. doi: 10.3389/fimmu.2021.743559

1 INTRODUCTION

The activation of T CD4 lymphocytes is triggered through the proper binding of the T-cell receptor (TCR) to specific antigens presented in the context of the major histocompatibility complex (MHC) on antigen presenting cells (APC), and of co-stimulatory molecules like CD28 with ligands such as CD80 and CD86 (jointly denoted as CD80/86), also displayed on the APC membrane. Activation involves the coordinated activity of a plethora of intra- and extra-cellular biochemical mediators

forming a network that reinforces, amplifies, diversifies, and regulates the initial antigenic and co-stimulatory signals (1–9). It is known that TCR activation is a progressive process, since MHC–peptide molecules serially engage several TCRs, amplifying the magnitude of intracellular signals that eventually cross a certain activation threshold (10). Furthermore, a minimal interaction half-time between the TCR and MHC is required for productive TCR signaling (10–13). Under optimal stimulation, activation ultimately leads to cell proliferation and differentiation into particular effector cell phenotypes, which are active against diverse antigens. In contrast, binding of antigen to the TCR in the absence of CD28 ligation conduces to a state of anergy. In such a state, T cells are unable to produce interleukin 2 (IL-2) or proliferate on subsequent stimulations (14–16).

Activation of T CD4 cells by antigen and co-stimulatory molecules triggers inhibitory pathways that regulate the whole process, among which the cytotoxic T-lymphocyte antigen 4 (CTLA-4), expressed on the surface of activated T cells, has been the most extensively investigated (17, 18). CTLA-4 is partially homologous to CD28 and binds to the same ligands (CD80/CD86) on the APC, although with a much higher affinity than the latter. Thus, upregulation of this molecule on activated cells results in the competition between CD28 and CTLA-4 for binding to CD80/CD86 (19, 20). The displacement of CD28 by CTLA-4 induces of a state of cellular arrest known as checkpoint blockade (21–24).

The activation process intrinsically involves the function of metabolic mediators whose activity is necessary to fulfill the bioenergetic and biosynthetic demands of increased cell proliferation and function (9, 25, 26). Metabolism of resting naïve cells depends on the tricarboxyic acid (TCA) cycle linked to oxidative phosphorylation (OXPHOS), a highly efficient but slow route for ATP generation. Upon activation, T cells rapidly shift to a predominant glycolytic metabolism, a less efficient process of ATP generation which, however, produces essential molecular intermediates for the generation of metabolites required for growth and proliferation [reviewed in (27)]. In spite of abundant information on the metabolic shift toward glycolysis, evidence suggests that OXPHOS is also induced in early states of activation, since AMPK is activated by signals from the TCR, CD28 and Ca^{2+} (2, 25, 28–30) (**Figure 1**). A general theoretical model is necessary to integrate the pathways of signaling from the TCR and costimulatory molecules with those from metabolic controllers, in order to understand how the balance between glycolysis and OXPHOS is established and how it is adjusted as effector functions arise.

A number of mathematical models have been proposed to analyze the intricate organization underlying the general mechanisms involved in T cell response (31–40), as well as the role of central actors participating in the regulation of the immune response (41–43). In this context, we previously addressed the early intracellular events in T cell activation and subsequent cell differentiation by analyzing the dynamics of a 46-node hybrid Boolean model. The model was based on a network composed by a central module simulating the activation induced by priming of TCR and CD28, another module describing the CTLA-4-mediated regulation of activation, and four modules corresponding to the events inducing differentiation to the Th1, Th2, Th17, and Treg phenotypes (40).

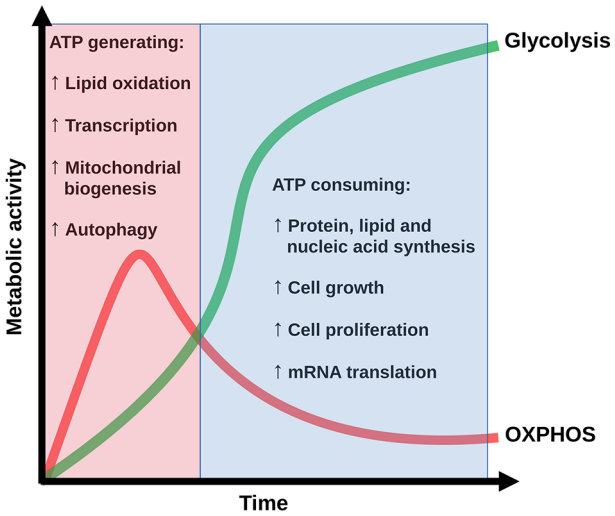


FIGURE 1 | Conceptualization of the time evolution of the OXPHOS-glycolysis shift along the bioenergetic and biosynthetic profile of T CD4 cells under antigenic stimulation. Adapted from (25).

With the purpose of modeling the mutual regulatory mechanisms of T CD4 lymphocyte activation and metabolism, in the present work we put forth a sub-network simulating the main processes of cellular metabolic control by AMPK; this module has been subsequently incorporated into the formerly described network to achieve an integrated scheme of the immune and metabolic processes driving the early events of T cell activation. In addition, two nodes have been introduced to represent the time-dependent priming of the TCR by MHC-antigen, as well as the competitive engagement of CD28 or CTLA-4 to the CD80/CD86 complex. As a preliminary step, the resulting 51-node network was characterized in terms of a set of discrete Boolean rules determining the fundamental interactive topology of the system. However, a more realistic description should take into account that the expression levels, concentrations, and parameters of the system may display any value within a continuous range limited only by functionality constraints. Thus, we performed the translation of the discrete interactive Boolean rules to the continuous domain through an algorithmic approach based on fuzzy logic. Fuzzy logic is a theory aimed to provide formal foundation to approximate reasoning (44, 45). Applied to biological systems, fuzzy propositions describe cases in which a cell displays intermediate levels of expression/activity of elements, so that they do not necessarily belong to a specific phenotype (46). The fuzzy logic rules were introduced as inputs into a system of ordinary differential equations to describe the overall network dynamics. The continuous analysis allows the introduction of variable degrees of activating stimulus and the description of gradual changes of the output elements reflecting activation. It

also allows to assess the influence of different time-scales of activity of key components of the signaling network.

2 METHODS

2.1 Inference and Integration of a Network Module of Metabolism Control

A concise network of the main components controlling T cell metabolism was constructed based on experimental information. A central actor in the lymphocyte metabolic activity is the AMPK complex, which is capable of sensing the intracellular AMP/ATP ratio, which represents the T cell energy pool availability, and of regulating the main metabolic pathways leading to the production of energy reserves (OXPHOS) or to the rapid generation of metabolites and structural proteins (glycolysis) (47–52). It has been demonstrated that two main effects related to metabolism take place upon TCR stimulation and CD28 co-stimulation (53). First, an increase in the basal activity of oxidative phosphorylation (OXPHOS) arises promoted by the action of the nutrient sensor AMPK, which is activated directly by the PI3k-AKT axis and calcium release (54). Afterwards, mTORC1 is activated with the consequent inhibition of AMPK and the activation of glycolysis (55).

In the network, AMPK is activated in several ways: signaling from the TCR and CD28 *via* the PI3k-AKT axis, calcium release, a high AMP/ATP ratio, the activation of the serine–threonine liver kinase B1 (LKB1) and the Foxp3 transcription factor (7, 47, 48, 56). Although there are several signaling intermediates proposed in these pathways, they were mathematically implied due to the lineal nature of the signaling trajectory in order to obtain a set of simplified logical propositions.

The balance between OXPHOS and glycolysis depends primordially on a negative feedback loop between AMPK and mTORC1, which plays the role of a metabolic polarization switch (50, 52, 57, 58). In the network presented here, TCR and CD28 activation, an elevated AMP/ATP ratio, LKB1 and Foxp3 activate AMPK, thus inhibiting the activity of mTORC1, finally leading to OXPHOS and inhibition of glycolysis.

Conversely, under a low AMP/ATP ratio, AMPK is inhibited, which in turns activates mTORC1 function and additional inhibition of AMPK, leading to glycolysis (**Figure 2**). The metabolic module was then incorporated into the previously reported Boolean network of T CD4 cell activation through links associated to AMPK and mTOR (**Figure 3**). The network

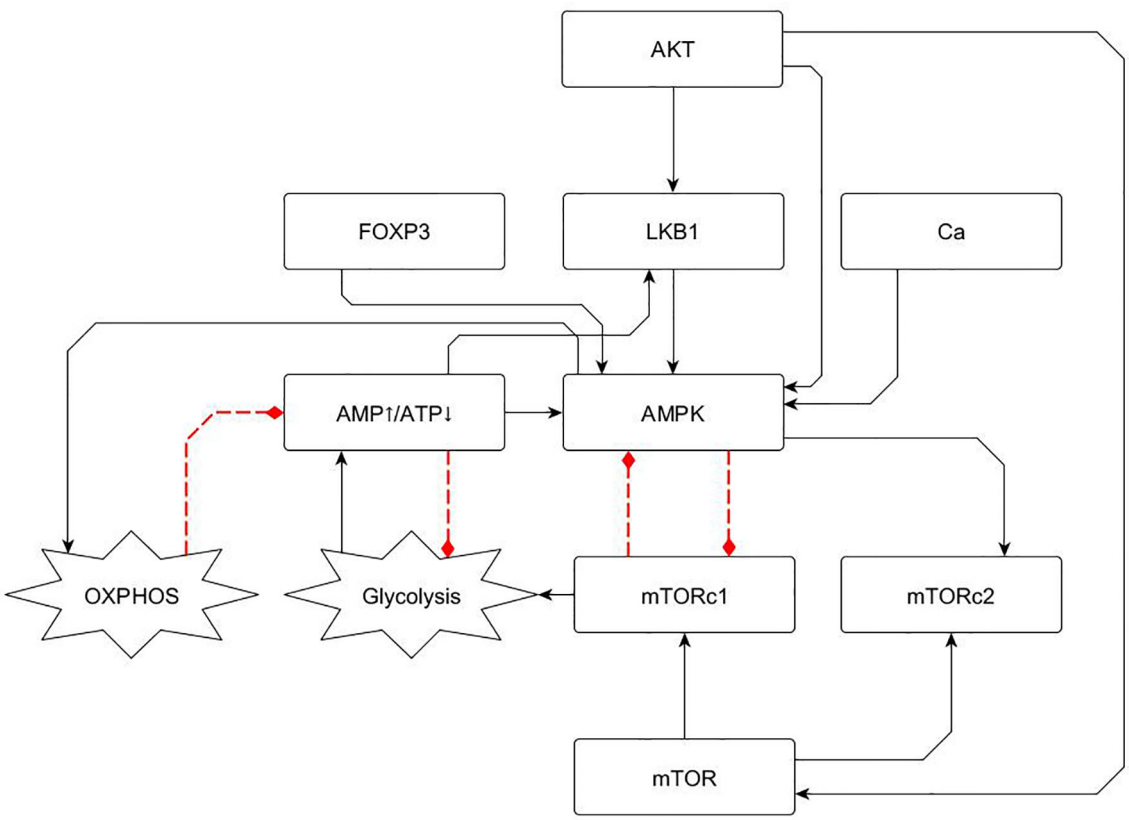


FIGURE 2 | Metabolism module including basic elements involved in glycolysis and OXPHOS regulation, as described in section 2.1. AMPK is a central energy sensor of the AMP/ATP ratio and displays a negative feedback loop with MTORC1. This loop defines a switch driving either OXPHOS or glycolytic activity. Continuous and dotted lines represent activator and inhibitory pathways, respectively.

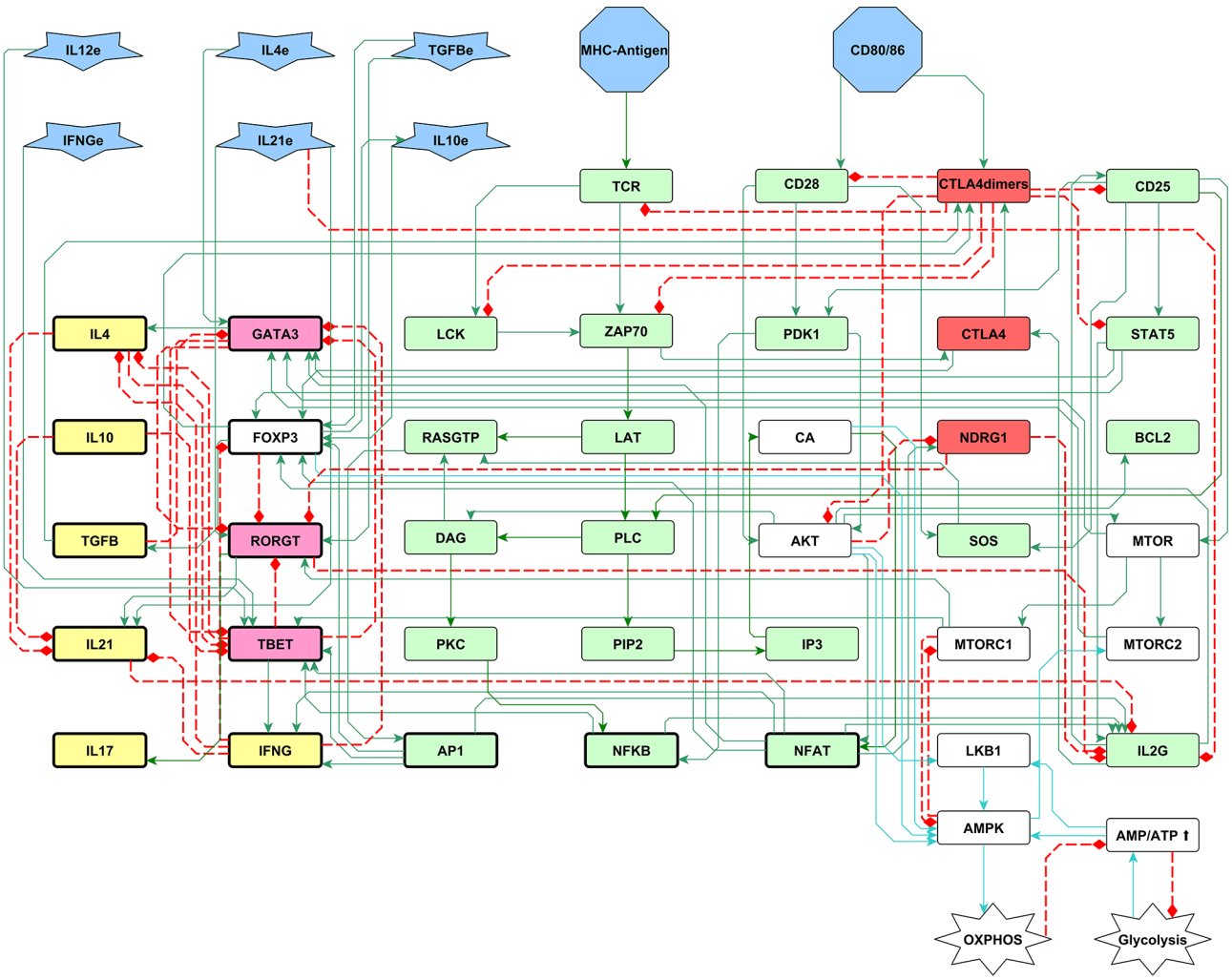


FIGURE 3 | A 51-node network of the early biochemical interactions induced by TCR activation, co-stimulation, and phenotype-inducing cytokines. Inputs of the network are the MHC-antigen, the co-stimulatory molecules CD80/86, and external cytokines (blue polygons). The set of interacting rules defining the dynamical system for Boolean modeling is shown in **Supplementary Material 1**. Boolean rules were translated to continuous functions as shown in Material and Methods, and the complete set of ordinary differential equations is provided in **Supplementary Material 2**. Green and red rectangles represent stimulatory and inhibitory nodes of the activation core, respectively. Pink and yellow rectangles represent phenotype-inducing transcription factors and cytokines produced endogenously, respectively. White rectangles represent the metabolism module including AMPK as a central regulatory element, as shown in figure 2. Construction of the network can be consulted in (40). LAT: LAT-Gads-SLP-76 complex; IL2G: IL-2 gene; L4e, IFNγe, IL10e, TGFβe and IL21e: exogenous cytokines.

previously constructed encompasses modules corresponding to an activation core derived from TCR and CD28 signaling, a feedback loop for IL-2 production though its high-affinity receptor (CD25), the role of the anergy factor NDRG1 (controlled by AKT), activation of the checkpoint by CTLA-4, and the induction of the effector phenotypes by external cytokines (59, 60). The construction of the network can be consulted in (40).

2.2 Boolean Approach

The interactions involved in the metabolism module were formalized as Boolean propositions as shown in **Table 1**. After incorporation into the general T-cell activation network proposed in (40), an exhaustive analysis was performed to analyze its general behavior and congruence with our previously published results (**Supplementary Material 3**). The robustness of the integrated model was also verified by introducing random noise in the initial states and measuring the distance between transition states and attractors. Robustness was also tested by either inducing perturbations in the network structure by random bit flipping of the Boolean functions or by random permutation of the output values (data not shown). The set of functions that defines the whole activation network are shown in **Supplementary Material 1** and the model structure can be consulted in Wolfram Notebook format in https://github.com/DrDavidMM/TCD4cell-activation-model-supplementary-material-.git.

TABLE 1 | Boolean rules for the metabolism module. Here, ∨⟶ or, ∧⟶ and, while ¬⟶ not.

VARIABLE	TRANSITION RULE
MTOR (*t* + 1) =	*CD25* (*t*) ∨ AKT (*t*)
MTORC1 (*t* + 1) =	*MTOR* (*t*) ∧ ¬*AMPK* (*t*)
MTORC2 (*t* + 1) =	[*MTOR* (*t*) ∧ *AMPK* (*t*)] ∨[*MTOR* (*t*) ∧ *IL4e* (*t*)]
LKB1 (*t* +1) =	*AKT* (*t*) ∧ *AMP/ATP* (*t*)
AMPK (*t* + 1) =	[*LKB1*(*t*) ∧ ¬*MTORC1* (*t*)] ∨[*CA* (*t*) ∧ *AMP/ATP* (*t*) ∧ ¬*MTORC1* (*t*)] ∨[*AKT* (*t*) ∧ *AMP/ATP*(*t*) ∧ ¬*MTORC1*(*t*)] ∨*FOXP3* (*t*)
Gycolysis(*t* + 1) =	*MTORC1* (*t*) ∧ ¬ *AMP/ATP* (*t*)
OXPHOS(*t* + 1) =	*AMPK* (*t*)
AMP/ATP(*t* + 1) =	*Glycolysis* (*t*) ∧ ¬*OXPHOS* (*t*)

2.3 Continuous Fuzzy Logic Approach

Fuzzy logic is characterized by a graded approach, so that the degree to which an object exhibits a given property is specified by a membership (or characteristic) function with truth values ranging from total falsity, $\mu[w_k] = 0$, to totally true, $\mu[w_k] = 1$. Here, w_k denotes a fuzzy logic proposition describing the interactions of node k with the rest of network nodes. By assuming that the state of the regulatory network at time t is described by the set $\{q1(t),\ldots q_n(t)\}$, $\mu[w_k]$, may be represented by a sigmoid function with continuous variation in the interval [0,1]:

$$\mu[w_k] = \frac{1}{1 + e^{-\beta(w_k(q_1,\ldots,q_n) - w_{thr})}}, \tag{1}$$

where w_{thr} is an activation threshold, hereby considered as $w_{thr} = 1/2$, and β is a saturation rate (46, 61).

The Boolean interaction rules $W_k[q_1(t), q_2(t), \ldots q_n(t)]$ were translated into fuzzy logic expressions

$$W_k[q_1(t), q_2(t), \ldots\ q_n(t)] \rightarrow w_k[q_1(t), q_2(t), \ldots\ q_n(t)]\,.$$

This procedure can be straightforwardly implemented by replacing the Boolean logic connectors 'and', 'or' and 'not' by their fuzzy counterparts according to the following scheme:

Boolean	Fuzzy Logic
q and *p*	*q*·*p*
q or *p*	*q* + *p* − *q*·*p*
not *p*	1 − p

An example of the translation from Boolean to a fuzzy framework is:

$$W[p, q, r] = (q\ or\ p)\ and\ (not\ r) \rightarrow$$
$$w[p, q, r] = (q + p - q \cdot p) \cdot (1 - r),$$

Within the continuous scheme, the dynamical behavior is determined by a set of ordinary differential equations describing the temporal change of the activity level of the network components. For the *k - th* node, this is written as

$$\frac{dq_k}{dt} = \mu[w_k(q_1, \ldots, q_n)] - d_k q_k, \tag{2}$$

Where d_k is the decay rate of node k. In this work, we assume that and that the default value of $d_k = 1$, unless otherwise stated. It is important to notice that in absence of an input ($w_k = 0$), the activity of node k decays exponentially, that is, $q_k \sim e^{-d_k t}$; therefore, the parameter $\tau_k = 1/d_k$ represents a characteristic expression time, so that $d_k > 1$ ($d_k < 1$) gives rise to a relatively rapid (slow) decay of the activity of the element k of the network. This kind of analysis allows to assess the influence of different time-scales of activity of the elements comprising the regulatory network.

In the continuous scheme, the equilibrium states (attractors) of the system are defined by the condition $dq_k/dt = 0$. This condition implies that, for a specific set of initial values $\{q_1(0),\ldots, q_n(0)\}$, the system dynamically evolves until reaching steady-state values given by:

$$q_k^{st} = \frac{1}{d_k}\mu[w_k(q_1^{st}, \ldots, q_n^{st})]\,. \tag{3}$$

This latter expression shows that the resulting set of asymptotic states, $\{q_1^{st}, \ldots q_n^{st}\}$, is determined, besides the initial values $q_k(0)$, by the actual values of the decay rates d_k. A consequence of the former results is that the emergent behavior of the system may conduce to alternative dynamic patterns depending on the specific values of *i*) the set of initial concentrations, dosages, or expression levels, $\{q_k(0)\}$, and *ii*) differences in either stimulation times, τ_{stim}, or characteristic expression times of the network components given by their decay rates, $\tau_k \sim 1/d_k$ (46, 61, 62).

2.4 TCR-Antigen, CD28-CD80/86, and CD28-CTLA-4 Interactions

The extent and time span of stimulation of T CD4 cells due to MHC-antigen presentation to TCR and of CD80/86 binding to CD28, was broadly modeled by introducing two input nodes whose priming activity only lasts for a limited lapse of time τ_{stim}. This was described by means of functions with a step-like behavior associated to an initial and constant avidity strength, $A_{MHC/A}$ and A_{CD8086}, suffering an abrupt decay at time $t = \tau_{stim}$ by a factor $D_{MHC/A}$ and D_{CD8086}, respectively. The time variations of avidity can be related to changes in the number of TCR-MHC-peptide complexes, the presence of adhesion molecules, TCR internalization or degradation after initial engagement, etc. (10, 62–67). Now, by introducing the step-function $H(t - \tau_{stim})$ defined by

$$H(t - \tau_{stim}) = \begin{cases} 0, & if\ t - \tau_{stim} < 0 \\ 1, & if\ t - \tau_{stim} > 0 \end{cases}$$

then the avidity variations are written as follows:

$$A_{MHC/A}(t) = A_{MHC/A} - D_{MHC/A}H(t - \tau_{MHC/A}) \tag{4}$$

$$A_{CD8086}(t) = A_{CD8086} - D_{CD8086}H(t - \tau_{CD8086}), \tag{5}$$

where $D_{MHC/A}$ and D_{CD8086} represent the magnitude of detachment reduction of TCR and CD28 from their ligands for times longer than $\tau_{MHC/A}$ and τ_{CD8086}, respectively.

This approach does not only allow to analyze the behavior of the cell as a function of the extent of stimulation, but also the

description of the competitive action between CD28 and CTLA-4 for binding to CD80/86. This is performed through the downstream interactions of CD28 with CTLA-4 (see **Figure 3**). Upon activation, CTLA-4 may down-regulate the engagement of CD28 with CD80/86. However, its inhibitory capacity depends on its decay rate, d_{CTLA4}, which should be relatively small (<1), in order to have an expression time long enough to overwhelm the influence of factors that promote the transcription factors activity. In the model, longer interaction times of the antigen with the TCR and CD28 allows a sustained activation state before being arrested due to the activity of CTLA-4. This process is simulated by introducing diverse values for the decay rate of CTLA-4, d_{CTLA4}, above and below the default value d_{CTLA4}=1, combined with different temporal duration of antigen attachment, τ_{stim} (**Figure 5**).

2.5 Numerical Methods

The set of differential equations that defines the dynamical system is shown in the **Supplementary Material 2**. The model equations are presented in Wolfram Notebook format in https://github.com/DrDavidMM/TCD4cell-activation-model-supplementary-material-.git.

For the computation of the differential equations system, Wolfram Mathematica 12.2.0.0 and open-source R studio have been used with the packages BoolNet, deSolve and ggplot2. For visual display of the interaction network, we use yEd graph editor 3.20.1 from yWorks. A link to the wolfram cloud public code has been added in the readme file in the GitHub repository.

3 RESULTS

A module describing the main processes of cellular metabolic control by AMPK was constructed as described in **Section 2.1** and **Figure 2**, and incorporated into a general T-cell activation network previously described in (40). In addition, two input nodes were introduced to represent the MHC-antigen and CD80/CD86 complexes, the presence of which activates the TCR and either CD28 or CTLA-4, respectively. The whole 51-node network was then reformulated as a set of logical rules and introduced in a system of ordinary differential equations to describe the activation dynamics (**Figure 3**).

An exhaustive analysis of the continuous model was performed to determine the congruence with our previously published results using the hybrid Boolean model **(see section 2.2).** The analysis showed that the attractors obtained in the hybrid Boolean approach can be recovered by the continuous model under certain restrictive conditions, equivalent to assume that the network variables may acquire only values corresponding to null or full expression (0 or 1).

The continuous model allows the introduction of variable levels of stimulating conditions (for example, duration of antigenic priming) and characteristic expression times (inverse decay rates) of the elements constituting the regulatory network. Likewise, it outlines gradual changes in the output elements (like AP-1, NFAT and NFκB transcription factors, and type of metabolism).

3.1 Initial Stimulation Conditions: From Naive to Antigen-Primed T Cells

We have assumed that T-cells initially are in a naive state, that is, TCR and CD28 are expressed and not activated. These conditions were implemented by considering that at time $t = 0$, TCR = 0, and CD28 = 0. MHC-antigen and CD80/86 act as inputs activating the TCR and CD28, respectively (although with the course of time CD28 may be displaced by CTLA-4). Optimal stimulation by MHC-antigen and CD80/86 was represented as $A_{MHC/A} = 1$, and $A_{CD8086} = 1$. Similarly, sub-optimal stimulation was represented by introducing values smaller than unity for either $A_{MHC/A}$ or A_{CD8086} (or both). As shown below, the level of stimulation defined by these parameters can produce activation or anergy. Since the metabolic profile of naive T-cells is characterized by a basal level of OXPHOS, modeling was performed assuming an initial value of OXPHOS = 0.2 and a high AMP/ATP ratio = 1, consistent with a significant activity of the nutrient sensor AMPK = 1.

3.2 The AMPK-mTOR Axis Promotes Metabolic Polarization During T CD4 Activation

Under optimal stimulating conditions, the network dynamics conduces to alternative states of sustained activation or CTLA-4-mediated arrest, each differing in their metabolic profile (**Figure 4**). The presence of MHC-antigen and CD80/86 (at time $t = 0$) induces the activity of TCR and CD28 (**Figure 4A**). These interactions were maintained at a maximal level at stimulation times $\tau_{stim} = \tau_{MHC/A} = \tau_{CD8086} = 15$ units, until disengagement was induced at longer times $t > \tau_{stim}$. During the stimulation time, dimerized CTLA-4 is transiently expressed at a low level, which however, is not sufficient to compete with CD28 for binding to CD80/86. Under these conditions, activation leads to the expression of the AP1, NFAT, and NFκB transcription factors at later times (**Figure 4B**) and, accordingly, transcription of the IL-2 gene (IL2G) and MTORC1 are also fully expressed (**Figure 4C**).

The predicted dynamic behavior of the activation process is coherent with trends inferred from the network interactive relationships depicted in **Figures 2** and **3**. As indicated above, we assumed an initial naive state characterized by a low basal level of OXPHOS, a large AMP/ATP ratio, and activity of AMPK (**Figure 4D**). Upon T cell stimulation by TCR and co-stimulatory molecules, AMPK is further activated by Ca^{2+}, LKB1 and AKT, boosting an early increase of OXPHOS. After reaching a peak of activity, OXPHOS is undermined due to the decrease of the AMP concentration and the afterward contribution of mTORC1, which impairs the repressive activity of AMPK and leads metabolism polarization towards glycolysis. Glycolysis contributes to the synthesis of cell-growth metabolites and maintenance of the ATP pool. This predicted behavior is congruent with experimental data showing that AMPK is activated early after T cell stimulation, which indicates that the

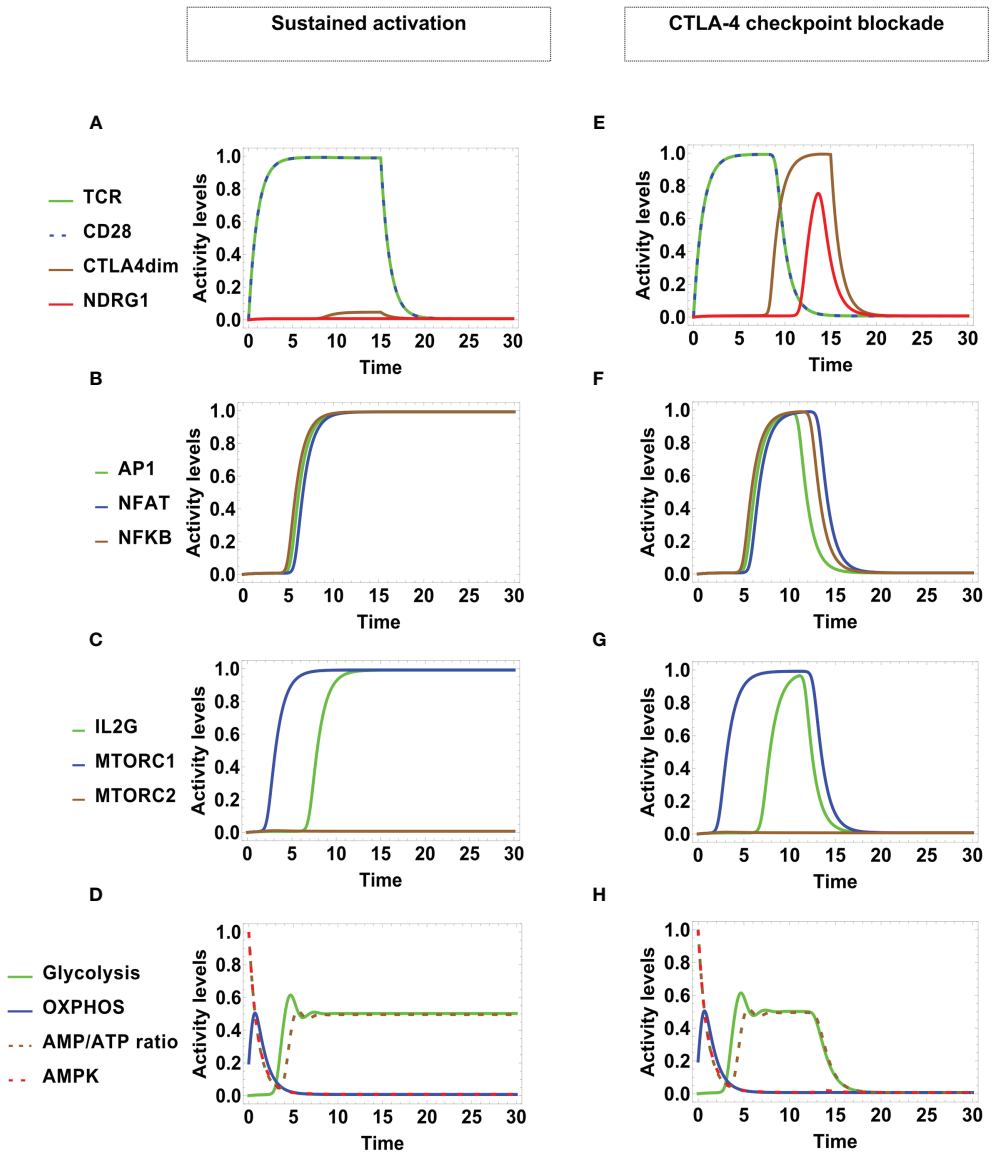


FIGURE 4 | Dynamics of transcription factors and metabolism under optimal engagement conditions of TCR with the MHC-antigen complex, and CD28 with CD80/86, during a stimulation time $\tau_{MHC/A} = \tau_{CD8086} = 15$ units. Left panel: Sustained T-cell activation associates to low-level activity of CTLA-4 (with high decay rated $d_{CTLA4} = 5$). **(A)** TCR and CD28 are fully activated as far as antigenic stimulation persists, leading to **(B)** sustained expression of the transcription factors AP-1, NFAT, and NFκB, and **(C)** sustained expression of mTORC1 and IL-2. **(D)** Metabolic profile: After initial activation, the AMP/ATP ratio decreases, leading to a reduction of the activity of the nutrient sensor AMPK and a temporary increment of OXPHOS. With the course of time, OXPHOS decays and glycolytic activity increases up to a steady level, in parallel with the AMP/ATP ratio. Right panel: Checkpoint blockade associated to high-level activity of CTLA-4 (with low decay rate $d_{CTLA4} = 0.5$). **(E)** Initially, TCR and CD28 are fully activated. With the course of time, the expression level of dimerized CTLA-4 increases, displacing CD28 from co-stimulatory molecules, with the concomitant induction of anergy. This is manifested as **(F)** transitory expression and down-regulation of the activation transcription factors AP-1, NFAT, NFκB, and **(G)** transitory expression and down-regulation of mTORC1 and IL-2. **(H)** Metabolic profile: At the beginning, the metabolic activity shows an identical pattern as in the case of sustained activation; eventually, the inhibitory action of CTLA-4 induces the decay of glycolysis.

engagement of mitochondrial metabolism is important for exiting quiescence (68), whereas the expression of enzymes pertaining to the glycolytic pathway is dispensable at earlier times (69). Thus, the model is in agreement with the proposal that AMPK activation ensures sufficient ATP availability to progress through full activation (53, 70). It can be observed that, even if TCR and CD28 were stimulated during a limited time span, the production of the AP-1, NFAT and NFκB transcription factors is held longer.

3.3 CTLA-4 Checkpoint Blockade

After TCR and CD28 stimulation, in due course the expression level of dimerized CTLA-4 increases, displacing CD28 from co-stimulatory molecules and leading to a state of activation arrest,

or checkpoint. Since arrested cells have decreased levels of protein synthesis and expansion, the action of CTLA-4 may also have implications on the regulation of metabolism (71). To model this process, we assumed initial optimal engagement of TCR with the MHC-antigen complex, and CD28 with CD80/86, during an antigen presentation time $\tau_{MHC/A} = \tau_{CD8086} = 15$ units, combined with a high-level activity of CTLA-4, induced by a low decay rate $d_{CTLA4} = 0.5$.

Similar to the previous case, TCR and CD28 are fully activated due antigenic stimulation (**Figure 4E**). However, with the course of time the inhibitory action of CTLA-4 and the anergy factor NDRG1 are expresses, coinciding with the diminution of the levels of activity of TCR and CD28. This leads to an only transitory expression of the AP-1, NFAT, NFκB transcription factors (**Figure 4F**), the IL-2 gene, and mTORC1 activity (**Figure 4G**). The initial metabolic activity shows a pattern similar to that displayed in sustained activation. Nevertheless, the regulatory action of CTLA-4 eventually induces the decay of glycolysis (**Figure 4H**).

3.4 Influence of Stimulation Time and CTLA-4 Activity on Sustained or Regulated Activation

The model predicts that cell activation depends on whether the extent of the stimulation time of TCR and CD28, τ_{stim}, is long enough to overcome the regulatory activity of CTLA-4, which persists during a characteristic time, defined by $\tau_{CTLA4} = 1/d_{CTLA4}$. The functionality of CTLA-4 resides in its continuous turnover, cellular location, and membrane delivery (18). In combination with activation-promoting factors, in the present model the decay rate of functional CTLA-4 is able to encompass these processes. In the simulation shown in **Figure 4**, sustained activation ensued by assuming that $\tau_{MHC/A} = \tau_{CD8086} = 15$, and $d_{CTLA4} = 5$, whereas regulated activation was associated to $\tau_{MHC/A} = \tau_{CD8086} = 15$, and $d_{CTLA4} = 0.5$.

An analysis of the values of d_{CTLA4} leading to states of sustained or regulated activation as a function of the stimulation time, τ_{stim}, is presented in **Figure 5**. We observe that no activation arises for $\tau_{stim} < 7$ units, while sustained activation ensues for $7 \leq \tau_{stim} \leq 10$ units, and regulated activation occurs for $\tau_{stim} > 10$ units. The stage of no activation is associated with insufficient priming time by TCR and CD28. Sustained activation, independent of d_{CTLA4}, is due to the fact that signaling from TCR and CD28 propagates downstream during a certain time without activating CTLA-4. On the other hand, regulated activation is driven by two conditions: first, that τ_{stim} is long enough, and second, that CTLA-4 activity persists longer than those of activation-inducing elements ($\tau_{CTLA4} = 1/d_{CTLA4} > 1$). In other words, CTLA-4 should be functional long enough to overcome a threshold level and its activity should be maintained to perform inhibition. Therefore, regulated activation persists for even longer values of τ_{stim} during which d_{CTLA4} displays a quasi-periodic behavior with slight variations centered at $d_{CTLA4} \sim 0.5$. The oscillation of the regulation threshold can be explained by considering that the expression of CTLA-4 depends on activation of the TCR and CD28. Once expressed, CTLA-4 initiates inhibition of signaling, which reduces its own expression and therefore allows activation again; this induces a stimulation-inhibition cycle which is downstream-propagated throughout the network. As a consequence, the threshold value of d_{CTLA4} leading to regulation will depend on the expression level attained by activation inducers at a given phase of the cycle.

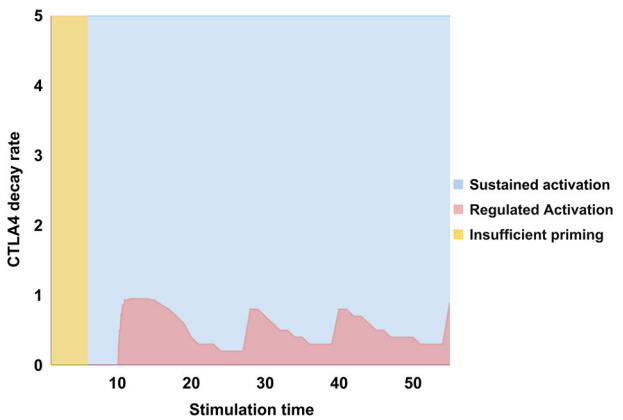


FIGURE 5 | Values of the decay rate of CTLA-4, d_{CTLA4}, as a function of the stimulation time, $\tau_{stim} = t_{CD8086} = t_{MHC/A}$, associated to states of no activation (yellow), sustained activation (blue), or regulated activation (pink). For $\tau_{stim} < 7$ units no activation arises; sustained activation ensues for $7 \leq \tau_{stim} \leq 10$, and regulated activation for $\tau_{stim} > 10$. The first stage implies that a minimum stimulation time is necessary to boost activation; the second one, that CTLA-4 requires a minimal time of activation to be induced and perform inhibition; the third stage reveals that regulated activation only arises at later stimulation times and when the CTLA-4 decay rate is low, that is, $d_{CTLA4} < 1$. Therefore, CTLA-4 should be induced for a time long enough to overcome a threshold level and its activity should be maintained to perform inhibition. The oscillatory behavior of the regulation threshold is associated to the inhibitory action of CTLA-4 which depends in turn on TCR and CD28 activation. This induces a stimulation-inhibition cycle which is downstream-propagated throughout the network. Consequently, the threshold d_{CTLA4} leading to regulation is determined by the expression level attained by activation inducers at a given phase of this cycle.

3.5 Incomplete Activation Promotes T Cell Anergy With Abnormal Metabolic Profiles

Variable levels of stimulation may originate from the progressive engagement of TCR to MHC-antigen complexes, the lack of a minimal half-life of interaction, diverse levels of antigen concentration, mutations in both cell receptors, etc. (63, 72, 73). Incomplete stimulation leads T cells to anergic states (14, 15, 74) and has been associated with mechanisms of tolerance "(adaptive tolerance") or antigen presentation failure. On the other hand, defective CD28 co-stimulation allows the expression of the N-Myc Downstream Regulated 1 (NDRG1) protein, leading to anergy (75–77). The effect of different levels of stimulation on the downstream expression of the network components was determined. To simulate these phenomena, we considered initial conditions in which either TCR or CD28 were subjected to a "strong" or a "weak" stimulation. A first case

considers $A_{MHC/A}$ = 1 and A_{CD8086} =0.5, and a second one, $A_{MHC/A}$ = 0.5 and A_{CD8086} = 1.

The left-hand side of **Figure 6** shows the activation dynamics arising from a strong signaling from TCR and a weak co-stimulation through CD28. TCR and CD28 show full and partial activation, respectively. Low levels of AP-1 and interleukin 2 are transiently expressed. NDRG1 is expressed at a later time, coinciding with the activity decay of TCR and CD28 **(Figure 6A)**. As before, this conduces to a full but transient expression of NFAT and NFκB, and a very small level of AP-1 (**Figure 6B**). A similar behavior if shown by mTORC1, and IL-2 is strongly suppressed (**Figure 6C**). Notably, the metabolic pattern is identical to that obtained in the case of regulated activation although in this case activity of CTLA-4 is not induced (**Figure 4E** and **Figure 6D**). The former results are consistent with reports indicating that TCR binding in the absence of CD28 ligation results in either apoptosis or a state of anergy that does not involve CTLA-4. Such anergic T cells are unable to produce IL-2 or proliferate on subsequent stimulation, even in the presence of co-stimulation (74, 78).

In the right-hand side of **Figure 6** we show the alternative dynamics arising from a weak signaling of TCR, but a strong co-stimulation by CD28. In this case, TCR and CD28 are transiently expressed at respective low and high levels, but they subsequently decay independently of the action of CTLA-4 or the anergy factor NDRG1 (**Figure 6E**). On the other hand, no transcription factors are induced (**Figure 6F**) neither IL-2, although mTORC1 is transiently expressed (**Figure 6G**). Here, the metabolism profile (**Figure 6H**) also coincides with that obtained for regulated activation **(Figure 4H)**.

3.6 Effector T CD4 Cell Phenotypes Display Specific Metabolism Requirements

It is widely documented that the different effector phenotypes such as Th1, Th2, Th17, and Treg present different energy requirements (56, 79, 80), and studies aimed to elucidate the intricate links between lymphocyte activation and metabolic reprogramming are needed (27). We simulated the individual conditions required for the differentiation of naive T cells into Th1, Th2, Th17 and Treg phenotypes, determined by the presence of the appropriate external cytokines.

Differentiation toward the effector Th cell lineages Th1, Th2, and Th17 is known to be reliant on mTOR activity, while inhibition of mTOR with rapamycin has been shown to favor Treg cell differentiation (56, 80, 81). mTOR complex1 (mTORC1) is formed with the scaffolding protein regulatory associated protein of mTOR (RAPTOR), while mTOR complex 2 (mTORC2) uses Rapamycin-insensitive companion of mammalian target of rapamycin (RICTOR) as a scaffold. All effector lineages, including Th2 cells, require mTORC1 activation (82, 83). Treg cells are a particular case in which the cell uses glycolysis to grow size and replicate; however, at stable stages they predominantly express Foxp3, which is a direct activator of AMPK and OXPHOS activity. This establishes metabolism as a key factor for the correct function of Treg cells (56).

Modeling of differentiation shows that, after activation, production of the characteristic cytokines starts first for Th1 and Th17 (**Figures 7A, C**), while the Th2 cytokines are induced at later times (**Figure 7B**). As shown before, OXPHOS is transiently upregulated upon activation (**Figures 4D** and **7D**). Next, the three cell phenotypes develops a predominant and stable glycolytic profile. Instead, Treg differentiation shows the production of IL-10 and TGFβ at later times compared to Th1 and Th17 cytokines (**Figure 7E**). Interestingly, the simulation shows that glycolytic activity is followed by a strong and sustained polarization to OXPHOS, corresponding with AMPK activity (**Figure 7F**). This metabolic behavior could allow Treg cells to increase the pool of amino acids required for the synthesis of structural proteins, cell growth for cell clonal expansion, and the production of diverse metabolites necessary to carry out function. However, as the expression of Foxp3 stimulates AMPK, it induces again a polarization to OXPHOS leading cells to a stable regulatory stage. On the other hand, the model is in agreement with the observation that the functional form of CTLA-4 (CTLA-4dim) is upregulated and is constitutively expressed on Treg cells (84) (**Figure 7E**). Thus, the model effectively integrates the pathways that lead to the adjustment of the metabolic profile of different effector phenotypes.

4 DISCUSSION

Understanding the function of T CD4 lymphocytes requires dynamic models able to integrate the diversity and connectivity of external and internal signals, the role of their variable levels of expression/function, and the system regulatory mechanisms. Currently, abundant experimental information allows the construction of models taking into account a basic, hierarchic organization of components by identification of those governing the main functional outcomes. Here, we put forward a regulatory network with continuous interactive rules, including a core module encompassing the downstream signals after TCR and CD28 activation along with the antagonist action of CTLA-4, a module including the main components participating in the control of T-cell metabolism, and four differentiation modules associated with the effect of exogenous cytokines leading to effector phenotypes (**Figure 3**). The model reproduces the time course of the main early events of T cell activation, anergy due to the lack of co-stimulation, CTLA-4 checkpoint blockade, cell differentiation, and intrinsic metabolic changes.

The continuous model describes the OXPHOS-glycolysis dynamic adjustment as T cell effector functions develop. Remarkably, the model predicts a transient phase of increased OXPHOS at the onset of activation, followed by its subsequent decrease along with the induction of a glycolytic phase (**Figure 4**). This profile was obtained during differentiation to the Th1, Th2, and Th17 phenotypes (**Figure 7**). These observations are in agreement with experimental data

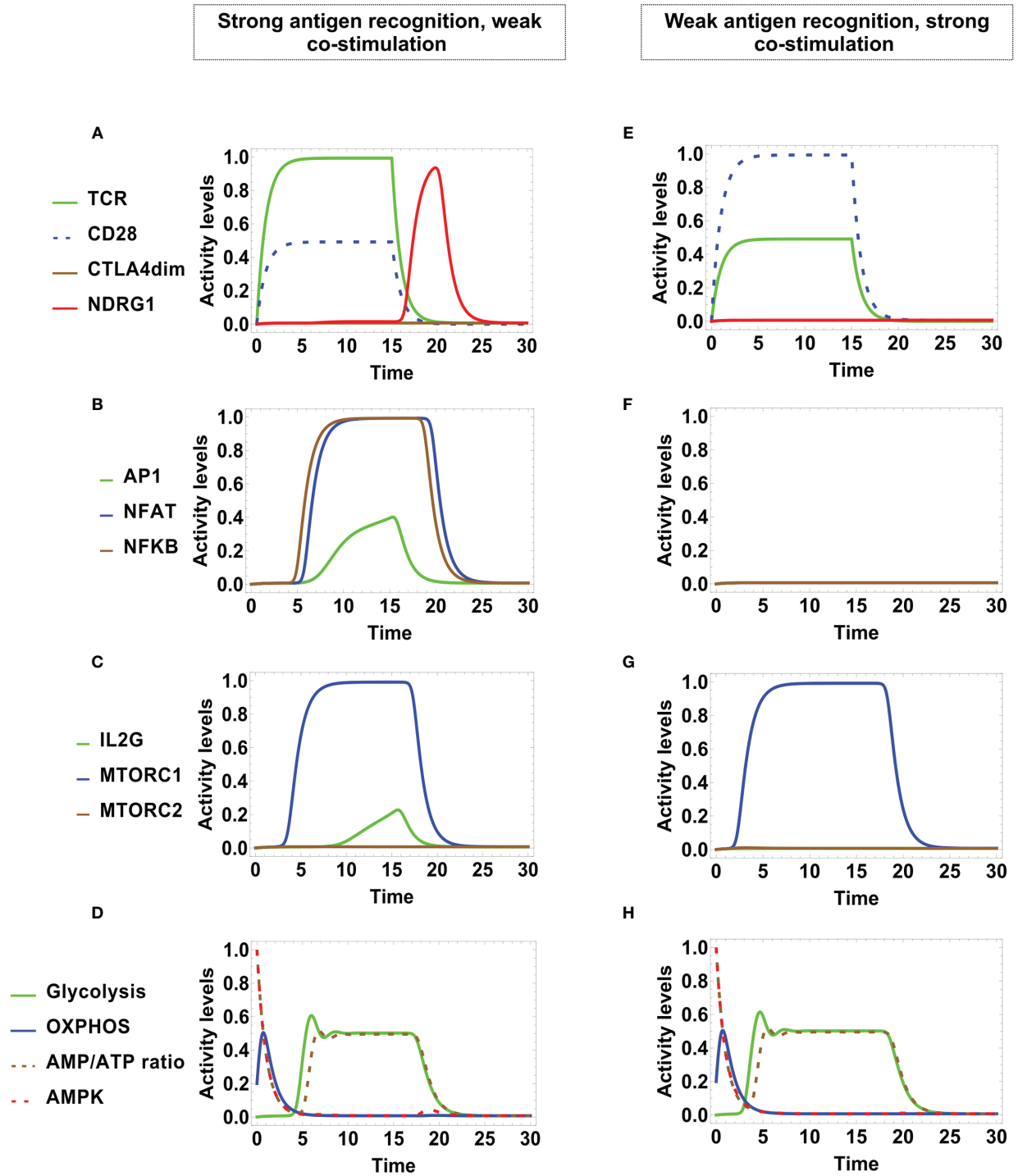


FIGURE 6 | Dynamics of transcription factors and metabolism after incomplete stimulation during an activation time $\tau_{MHC/A} = \tau_{CD8086} = 15$ units. Left panel: Anergy is induced by strong TCR stimulation and weak CD28 co-stimulation ($A_{MHC/A} = 1$ and $A_{CD8086} = 0.5$). **(A)** TCR and CD28 are transiently activated, decaying both at time $\tau_{MHC/A} = \tau_{CD8086}$. At this time the anergy factor NDRG1 is temporarily expressed. **(B)** NFAT, NFκB, and AP-1 are down-regulated by the inhibitory action of NDRG1. **(C)** Similarly, mTORC1 and IL2 are only transiently expressed. **(D)** The metabolic profile is similar to that associated to regulated activation induced by CTLA-4 (**Figure 4H**). Right panel: Anergy is induced by weak TCR stimulation and strong CD28 co-stimulation ($A_{MHC/A} = 0.5$ and $A_{CD8086} = 1$. **(E)** TCR and CD28 show full and low -level activation, respectively, both decaying at time $\tau_{MHC/A} = \tau_{CD8086}$; however, the anergy factor NDRG1 remains unexpressed. **(F)**. NFAT, NFκB, and AP-1 are unexpressed. **(G)** mTORC1 is fully activated, but decays along the antigenic stimulation. **(H)** The metabolic profile is similar to that associated to regulated activation induced by CTLA-4 (**Figure 4H**).

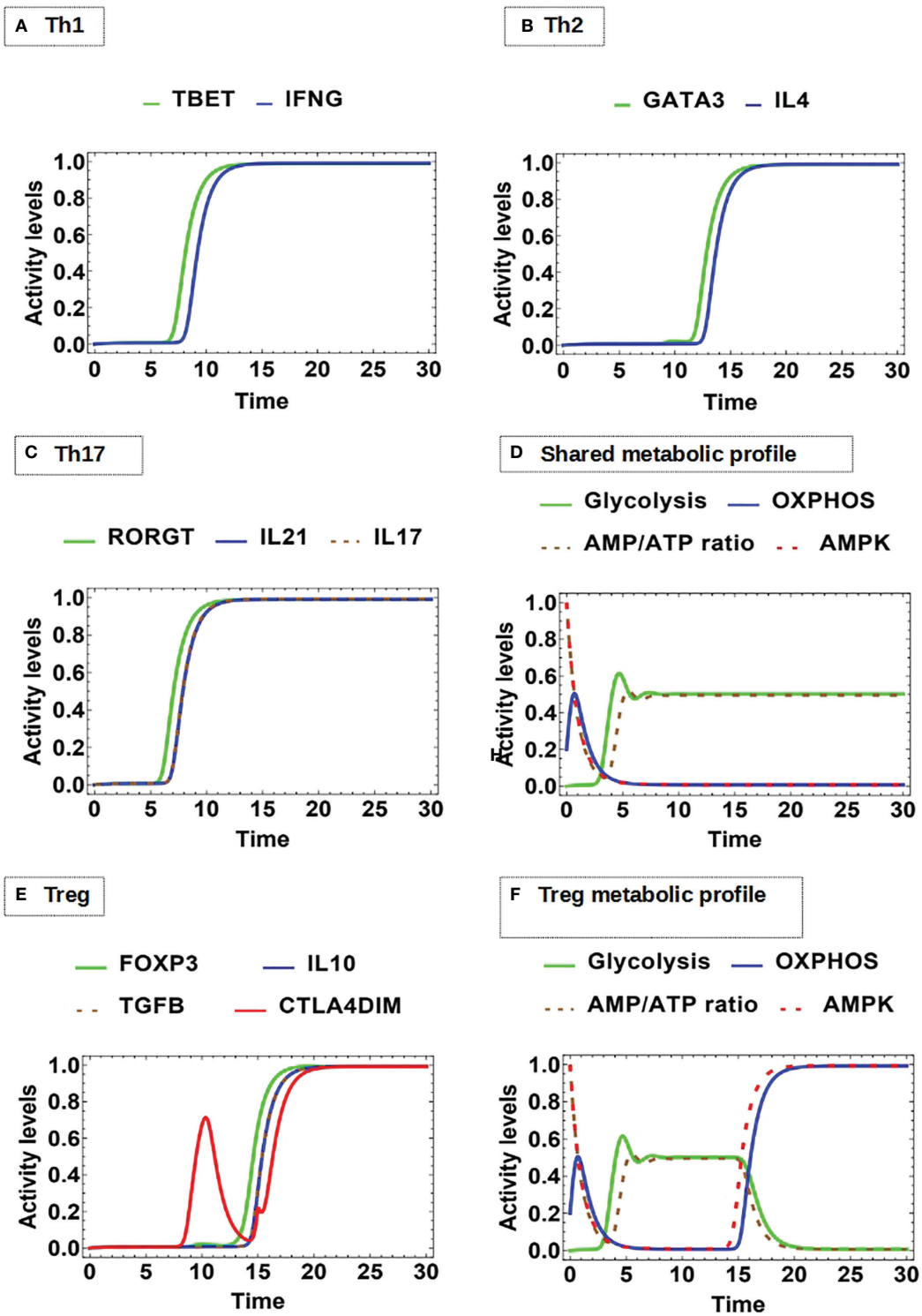


FIGURE 7 | Dynamics of master transcription factors and interleukin production by effector cells under optimal antigenic recognition and low-level activity of CTLA-4: **(A)** Th1 profile: Sustained expression of T-bet and INF-gamma production, **(B)** Th2 profile: Sustained expression of GATA3 and IL-4 production, **(C)** Th17 profile: Sustained expression of RORγT, as well as Il-17 and IL-21 production. **(D)** The Th1, Th2, and Th17 phenotypes display an initial transient OXPHOS phase and then a stable glycolytic metabolism. **(E)** Treg profile: Sustained expression of Foxp3 and dimeric CTLA-4 with joint production of IL-10 and TGF-β. **(F)** Treg displays an initial transient OXPHOS phase followed by glycolytic metabolism; however, with the course of time this is replaced by a metabolism based on OXPHOS.

suggesting that OXPHOS produces an ATP reservoir before glycolysis (a much less efficient, but quicker process of ATP production) rises up the levels of metabolites needed for protein synthesis, cell function and growth (53, 70). A possible explanation for this phenomenon is the need for cells to generate a reserve of energy before committing to glycolysis, which implies anabolism in order to produce metabolites for the support of proliferation and function (53). Our model also proposes that in order for the cells to stabilize glycolytic activity, it is necessary to reach an energy balance between the feeding of the metabolic pathways positively regulated by glycolysis, and the AMP/ATP ratio available in the intracellular medium. On the other hand, the model does not predict differences in the metabolism kinetics of Th1, Th2 and Th17 phenotypes. This effect may depend on additional microenvironmental conditions that may be included in the network. In contrast, the glycolysis phase is subsequently reduced during differentiation of Treg cells, as their metabolism is eventually polarized towards OXPHOS (**Figure 7**). The master transcription factor of the Treg lineage (Foxp3) represses glycolytic gene transcription through the upregulation of CTLA-4 and AMPK (56, 85–87). It is suggested that OXPHOS favors the suppressor function of Treg cells, although they may reenter glycolysis when they initiate clonal expansion or migratory activities (51, 56). The continuous model might simulate oscillations of metabolism in changing microenvironmental conditions (work in progress).

The incorporation of two nodes representing the MHC-antigen complex and CD80/86 as inputs of the system allows the assignment of variable levels of functional stimulation and co-stimulation (**Figure 3**). Inhibition of the anergy factor NDRG1 by CD28 as well as IL-2 downstream signaling is required for lymphocyte activation (76) and thus, weak co-stimulation drives the cell into anergy (14, 15, 74, 88). Modeling shows that, when TCR is strongly stimulated and co-stimulation through CD28 is weak, the anergy factor NDRG1 is expressed and anergy emerges. In the anergic state glycolysis is inhibited. In the opposite case, when the TCR signal is weak and CD28 is strongly activated, cells reach a final state in which AP-1, NFAT and NFκB are not produced and the IL-2 gene is not activated. However, activity of mTORC1 is induced along with a transient glycolysis state (**Figure 6**). A study by Zheng et al. has shown that cells induced to anergy display transient increases of expression of the amino acid transporter CD98 and the transferrin receptor CD71, both membrane molecules associated to glycolysis (89). Experimentally, it would be expected to observe a transient activation of the AKT-mTORC1 circuit, along with the activity of enzymes pertaining to the glycolytic pathway in anergic cells. Thus, the model is congruent with the induction of anergy and abnormal metabolic profiles as a result of incomplete stimulation.

The inhibitory action of CTLA-4 is induced after lymphocyte activation and implicitly leads to anergy by competition with CD28 for binding to CD80/86. CTLA-4 has much higher affinity for CD80/86 than CD28. In this panorama, CTLA-4 increases its expression in the cell membrane when the immunological synapse has been maintained during sufficient time, raising the threshold for T-cell responses (10, 90–93). In our model, CTLA-4 is inherently induced by activation signaling; however, the model predicts an initial window over which only sustained activation is observed, because short activation times may be sufficient to activate the T cell but not CTLA-4. The window would allow the initial full activation of the network before turning-on regulatory mechanisms. Thereafter, the inhibitory effect of CTLA-4 depends on the combination of its characteristic time of activity (defined by $\tau_{CTLA4} = 1/d_{CTLA4}$) and the extent of TCR/CD28 stimulation (**Figure 5**). Thus, a high decay rate allows a sustained T-cell activation. Instead, when the CTLA-4 decay rate is lower (and thus its function is allowed) checkpoint blockade ensues (**Figure 4**). Interestingly, the model predicts that the value of the threshold for the inhibitory action of CTLA-4 oscillates as a function of the TCR and CD28 activation time (**Figure 5**). A relationship between the intensity of antigen stimulation and the activity of CTLA-4 has been suggested by studies showing that the level of translocation of CTLA-4 to the immunological synapse was sensitive to variations in the strength of the TCR signal, and suggesting that CTLA-4 preferentially inhibits the T cell response under conditions of potent TCR-peptide/MHC interactions. Results of modeling agrees with this hypothesis. Diminishing the advantage of highly responding cells, CTLA-4 would avoid that these cells might quickly out-compete clones with weaker responses, allowing for greater representation of cells bearing medium/low affinity TCR's. In this view, CTLA-4 would favor a greater diversity of the T cell response to a complex set of antigens, by acting at early stages of the cell activation. This effect could be important in the elaboration of a protective T cell response (94).

As expected, activation of CTLA-4 contributes to negatively regulate glycolysis, bringing general metabolism to baseline levels. This is consistent with experimental studies reporting that CTLA-4 blocks the positive regulation of glycolysis through the inhibition of CD28 and AKT pathways (95–97). The glycolysis decline of anergic cells is similar, but not identical, to that displayed during the CTLA-4 checkpoint blockade. The checkpoint blockade is induced by the inhibition of signaling at several levels in the network and is the quickest pathway of glycolysis inhibition (**Figure 4**). In the case of anergy induced by weak antigen stimulation or co-stimulation, glycolysis decay is due to the lack of activity of the IL-2/CD25 axis (**Figures 6C, G**) induced by incomplete stimulation and the activity of NDRG1, respectively (**Figure 3**). In agreement with Frauwirth and cols., CTLA-4 is not activated in these conditions (78) (**Figures 6A, E**).

The continuous model proposed here may constitute a relevant step for the comprehensive integration of experimental information on the mechanism behind activation and function of T CD4 cells. Additional factors with a role in the response to diverse stimuli and microenvironmental conditions may be incorporated, including different levels of ligand interactions and times of activity of key players. It is also useful for the understanding of the emerging metabolic requirements of different functional stages of T CD4 cells.

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

DM-M, CV, LM, and LH contributed to the conception of the model. DM-M and LH designed the regulatory network. DM-M and CV constructed the logical propositions, conducted numerical experiments, and performed the analysis of the system dynamics. All authors contributed to the interpretation of results. All authors participated in manuscript writing.

FUNDING

This work was supported by Programa de Apoyo a Proyectos de Investigación e Innovación Tecnológica of the Universidad Nacional Autónoma de México (grant number IN215820 to LH, and IN202721 to LM) and a postdoctoral fellowship from CONACYT (CVU number 555239 to DM-M).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2021.743559/full#supplementary-material

REFERENCES

1. Zheng Y, Manzotti CN, Liu M, Burke F, Mead KI, Sansom DM. CD86 and CD80 Differentially Modulate the Suppressive Function of Human Regulatory T Cells. *J Immunol* (2004) 172:2778–84. doi: 10.4049/jimmunol.172.5.2778
2. Tamás P, Hawley SA, Clarke RG, Mustard KJ, Green K, Hardie DG, et al. Regulation of the Energy Sensor AMP-Activated Protein Kinase by Antigen Receptor and Ca2+ in T Lymphocytes. *J Exp Med* (2006) 203:1665–70. doi: 10.1084/jem.20052469
3. Fooksman. Functional Anatomy of T Cell Activation and Synapse Formation. *Cisco Networking Acad locator* (2016) 1:79–105. doi: 10.1146/annurev-immunol-030409-101308.Functional
4. Chen L, Flies DB. Molecular Mechanisms of T Cell Co-Stimulation and Co-Inhibition. *Nat Rev Immunol* (2013) 13:227–42. doi: 10.1038/nri3405
5. Man K, Kallies A. Synchronizing Transcriptional Control of T Cell Metabolism and Function. *Nat Rev Immunol* (2015) 15:574–84. doi: 10.1038/nri3874
6. Brzostek J, Gascoigne NR, Rybakin V. Cell Type-Specific Regulation of Immunological Synapse Dynamics by B7 Ligand Recognition. *Front Immunol* (2016) 7:24. doi: 10.3389/fimmu.2016.00024
7. Palmer CS, Hussain T, Duette G, Weller TJ, Ostrowski M, Sada-Ovalle I, et al. Regulators of Glucose Metabolism in CD4+ and CD8+ T Cells. *Int Rev Immunol* (2016) 35:477–88. doi: 10.3109/08830185.2015.1082178
8. Gaud G, Lesourne R, Love PE. Regulatory Mechanisms in T Cell Receptor Signalling. *Nat Rev Immunol* (2018) 18:485–97. doi: 10.1038/s41577-018-0020-8
9. Menk AV, Scharping NE, Moreci RS, Zeng X, Guy C, Salvatore S, et al. Early TCR Signaling Induces Rapid Aerobic Glycolysis Enabling Distinct Acute T Cell Effector Functions. *Cell Rep* (2018) 22:1509–21. doi: 10.1016/j.celrep.2018.01.040
10. Alegre ML, Frauwirth KA. Thompson CB. T-Cell Regulation by CD28 and CTLA-4. *Nat Rev Immunol* (2001) 1:220–8. doi: 10.1038/35105024
11. Huang Y, Wange RL. T Cell Receptor Signaling: Beyond Complex Complexes. *J Biol Chem* (2004) 279:28827–30. doi: 10.1074/jbc.R400012200
12. Sansom DM. Moving CTLA-4 From the Trash to Recycling. *Science* (2015) 349:377–8. doi: 10.1126/science.aac7888
13. Cui J, Yu J, Xu H, Zou Y, Zhang H, Chen S, et al. Autophagy-Lysosome Inhibitor Chloroquine Prevents CTLA-4 Degradation of T Cells and Attenuates Acute Rejection in Murine Skin and Heart Transplantation. *Theranostics* (2020) 10:8051–60. doi: 10.7150/thno.43507
14. Appleman LJ, Boussiotis VA. T Cell Anergy and Costimulation. *Immunol Rev* (2003) 192:161–80. doi: 10.1034/j.1600-065X.2003.00009.x
15. Schwartz RH. T Cell Anergy. *Annu Rev Immunol* (2003) 21:305–34. doi: 10.1146/annurev.immunol.21.120601.141110
16. Duré M, Macian F. IL-2 Signaling Prevents T Cell Anergy by Inhibiting the Expression of Anergy-Inducing Genes. *Mol Immunol* (2009) 46:999–1006. doi: 10.1016/j.molimm.2008.09.029
17. Choi JM, Kim SH, Shin JH, Gibson T, Yoon BS, Lee DH, et al. Transduction of the Cytoplasmic Domain of CTLA-4 Inhibits TcR-Specific Activation Signals and Prevents Collagen-Induced Arthritis. *Proc Natl Acad Sci USA* (2008) 105:19875–80. doi: 10.1073/pnas.0805198105
18. Walker LS, Sansom DM. Confusing Signals: Recent Progress in CTLA-4 Biology. *Trends Immunol* (2015) 36:63–70. doi: 10.1016/j.it.2014.12.001
19. Vandenborre K, Van Gool SW, Kasran A, Ceuppens JL, Boogaerts MA, Vandenberghe P. Interaction of CTLA-4 (CD152) With CD80 or CD86 Inhibits Human T-Cell Activation. *Immunology* (1999) 98:413–21. doi: 10.1046/j.1365-2567.1999.00888.x
20. Kovacs B, Parry RV, Ma Z, Fan E, Shivers DK, Freiberg BA, et al. Ligation of CD28 by Its Natural Ligand CD86 in the Absence of TCR Stimulation Induces Lipid Raft Polarization in Human CD4 T Cells. *J Immunol (Baltimore Md. 1950)* (2005) 175:7848–54. doi: 10.4049/jimmunol.175.12.7848
21. Schneider H, Valk E, Leung R, Rudd CE. CTLA-4 Activation of Phosphatidylinositol 3-Kinase (PI 3-K) and Protein Kinase B (PKB/AKT) Sustains T-Cell Anergy Without Cell Death. *PloS One* (2008) 3:e3842. doi: 10.1371/journal.pone.0003842
22. Buchbinder EI, Desai A. CTLA-4 and PD-1 Pathways: Similarities, Differences, and Implications of Their Inhibition. *Am J Clin Oncol* (2016) 39:98–106. doi: 10.1097/COC.0000000000000239
23. Chikuma S. CTLA-4, an Essential Immune-Checkpoint for T-Cell Activation. *Curr Top Microbiol Immunol* (2017) 410:99–126. doi: 10.1007/82_2017_61
24. Brunner-Weinzierl MC, Rudd CE. CTLA-4 and PD-1 Control of T-Cell Motility and Migration: Implications for Tumor Immunotherapy. *Front Immunol* (2018) 9:2737. doi: 10.3389/fimmu.2018.02737
25. MacIver NJ, Jacobs SR, Wieman HL, Wofford JA, Coloff JL, Rathmell JC. Glucose Metabolism in Lymphocytes Is a Regulated Process With Significant Effects on Immune Cell Function and Survival. *J Leukocyte Biol* (2008) 84:949–57. doi: 10.1189/jlb.0108024
26. Michalek RD, Gerriets VA, Jacobs SR, Macintyre AN, MacIver NJ, Mason EF, et al. Cutting Edge: Distinct Glycolytic and Lipid Oxidative Metabolic Programs Are Essential for Effector and Regulatory CD4+ T Cell Subsets. *J Immunol (Baltimore Md. 1950)* (2011) 186:3299–303. doi: 10.4049/jimmunol.1003613
27. Pearce EL, Poffenberger MC, Chang CH, Jones RG. Fueling Immunity: Insights Into Metabolism and Lymphocyte Function. *Science* (2013) 342:210–21. doi: 10.1126/science.1242454
28. Zheng Y, Collins SL, Lutz MA, Allen AN, Kole TP, Zarek PE, et al. A Role for Mammalian Target of Rapamycin in Regulating T Cell Activation Versus Anergy. *J Immunol (Baltimore Md. 1950)* (2007) 178:2163–70. doi: 10.4049/jimmunol.178.4.2163
29. Blagih J, Coulombe F, Vincent EE, Dupuy F, Galicia-Vázquez G, Yurchenko E, et al. The Energy Sensor AMPK Regulates T Cell Metabolic Adaptation and Effector Responses InVivo. *Immunity* (2015) 42:41–54. doi: 10.1016/j.immuni.2014.12.030
30. Steinberg GR, Carling D. AMP-Activated Protein Kinase: The Current Landscape for Drug Development. *Nat Rev Drug Discovery* (2019) 18:527–51. doi: 10.1038/s41573-019-0019-2
31. Mendoza L. A Network Model for the Control of the Differentiation Process in Th Cells. *Biosystems* (2006) 84:101–14. doi: 10.1016/j.biosystems.2005.10.004

32. Saez-Rodriguez J, Simeoni L, Lindquist JA, Hemenway R, Bommhardt U, Arndt B, et al. A Logical Model Provides Insights Into T Cell Receptor Signaling. *PloS Comput Biol* (2007) 3:e163. doi: 10.1371/journal.pcbi.0030163
33. Naldi A, Carneiro J, Chaouiya C, Thieffry D. Diversity and Plasticity of Th Cell Types Predicted From Regulatory Network Modelling. *PloS Comput Biol* (2010) 6:e1000912. doi: 10.1371/journal.pcbi.1000912
34. Carbo A, Hontecillas R, Kronsteiner B, Viladomiu M, Pedragosa M, Lu P, et al. Systems Modeling of Molecular Mechanisms Controlling Cytokine-Driven CD4+ T Cell Differentiation and Phenotype Plasticity. *PloS Comput Biol* (2013) 9:13–5. doi: 10.1371/journal.pcbi.1003027
35. Martínez-Sosa P, Mendoza L. The Regulatory Network That Controls the Differentiation of T Lymphocytes. *BioSystems* (2013) 113:96–103. doi: 10.1016/j.biosystems.2013.05.007
36. Martinez-Sanchez ME, Mendoza L, Villarreal C, Alvarez-Buylla ER. A Minimal Regulatory Network of Extrinsic and Intrinsic Factors Recovers Observed Patterns of CD4+ T Cell Differentiation and Plasticity. *J Comput Biol* (2015) 11:e1004324. doi: 10.1371/journal.pcbi.1004324
37. Mendoza L, Méndez A. A Dynamical Model of the Regulatory Network Controlling Lymphopoiesis. *Biosystems* (2015) 137:26–33. doi: 10.1016/j.biosystems.2015.09.004
38. Martinez-Sanchez ME, Huerta L, Alvarez-Buylla ER, Luján CV. Role of Cytokine Combinations on CD4+ T Cell Differentiation, Partial Polarization, and Plasticity: Continuous Network Modeling Approach. *Front Physiol* (2018) 9:877. doi: 10.3389/fphys.2018.00877
39. Puniya BL, Todd RG, Mohammed A, Brown DM, Barberis M, Helikar T. A Mechanistic Computational Model Reveals That Plasticity of CD4+ T Cell Differentiation Is a Function of Cytokine Composition and Dosage. *Front Physiol* (2018) 9:878. doi: 10.3389/fphys.2018.00878
40. Martínez-Méndez D, Villarreal C, Mendoza L, Huerta L. An Integrative Network Modeling Approach to T CD4 Cell Activation. *Front Physiol* (2020) 11:380. doi: 10.3389/fphys.2020.00380
41. Sulaimanov N, Klose M, Busch H, Boerries M. Understanding the mTOR Signaling Pathway *via* Mathematical Modeling. *Wiley Interdiscip Reviews: Syst Biol Med* (2017) 9:1–18. doi: 10.1002/wsbm.1379
42. Ganesan A, Arulraj T, Choulli T, Barakat KH. A Mathematical Modelling Tool for Unravelling the Antibody-Mediated Effects on CTLA-4 Interactions. *BMC Med Inf Decision Making* (2018) 18:1–17. doi: 10.1186/s12911-018-0606-x
43. Khailaie S, Rowshanravan B, Robert PA, Waters E, Halliday N, Badillo Herrera JD, et al. Characterization of CTLA4 Trafficking and Implications for Its Function. *Biophys J* (2018) 115:1330–43. doi: 10.1016/j.bpj.2018.08.020
44. Zadeh LA. Fuzzy Logic and Approximate Reasoning. *Synthese* (1975) 30:407–28. doi: 10.1007/BF00485052
45. Novak V, Perfilieva I, Mockor J. *Mathematical Principles of Fuzzy Logic*. New York, NY: Kluwer Academic (1999).
46. Enciso J, Pelayo R, Villarreal C. From Discrete to Continuous Modeling of Lymphocyte Development and Plasticity in Chronic Diseases. *Front Immunol* (2019) 10:1927. doi: 10.3389/fimmu.2019.01927
47. Blagih J, Krawczyk CM, Jones RG. LKB1 and AMPK: Central Regulators of Lymphocyte Metabolism and Function. *Immunol Rev* (2012) 249:59–71. doi: 10.1111/j.1600-065X.2012.01157.x
48. Andris F, Leo O. AMPK in Lymphocyte Metabolism and Function. *Int Rev Immunol* (2015) 34:67–81. doi: 10.3109/08830185.2014.969422
49. Kishton RJ, Barnes CE, Nichols AG, Cohen S, Gerriets VA, Siska PJ, et al. AMPK Is Essential to Balance Glycolysis and Mitochondrial Metabolism to Control T-ALL Cell Stress and Survival. *Cell Metab* (2016) 23:649–62. doi: 10.1016/j.cmet.2016.03.008
50. Pezze PD, Ruf S, Sonntag AG, Langelaar-Makkinje M, Hall P, Heberle AM, et al. A Systems Study Reveals Concurrent Activation of AMPK and mTOR by Amino Acids. *Nat Commun* (2016) 7:1–19. doi: 10.1038/ncomms13254
51. Ma EH, Poffenberger MC, Wong AH, Jones RG. The Role of AMPK in T Cell Metabolism and Function. *Curr Opin Immunol* (2017) 46:45–52. doi: 10.1016/j.coi.2017.04.004
52. Cork GK, Thompson J, Slawson C. Real Talk: The Inter-Play Between the mTOR, AMPK, and Hexosamine Biosynthetic Pathways in Cell Signaling. *Front Endocrinol* (2018) 9:522. doi: 10.3389/fendo.2018.00522
53. MacIver NJ, Michalek RD, Rathmell JC. Metabolic Regulation of T Lymphocytes. *Annu Rev Immunol* (2013) 31:259–83. doi: 10.1146/annurev-immunol-032712-095956
54. Fracchia KM, Pai CY, Walsh CM. Modulation of T Cell Metabolism and Function Through Calcium Signaling. *Front Immunol* (2013) 4:324. doi: 10.3389/fimmu.2013.00324
55. Marzec M, Liu X, Kasprzycka M, Witkiewicz A, Raghunath PN, El-Salem M, et al. IL-2- and IL-15-Induced Activation of the Rapamycin-Sensitive Mtorc1 Pathway in Malignant CD4+ T Lymphocytes. *Blood* (2008) 111:2181–9. doi: 10.1182/blood-2007-06-095182
56. Kempkes RW, Joosten I, Koenen HJ, He X. Metabolic Pathways Involved in Regulatory T Cell Functionality. *Front Immunol* (2019) 10:2839. doi: 10.3389/fimmu.2019.02839
57. Perez VL, Parijs LV, Biuckians A, Zheng XX, Strom TB, Abbas AK. Induction of Peripheral T Cell Tolerance *In Vivo* Requires CTLA-4 Engagement. *Immunity* (1997) 6:411–7. doi: 10.1016/S1074-7613(00)80284-8
58. Holczer M, Hajdú B, Lőrincz T, Szarka A, Bánhegyi G, Kapuy O. A Double Negative Feedback Loop Between MTORC1 and AMPK Kinases Guarantees Precise Autophagy Induction Upon Cellular Stress. *Int J Mol Sci* (2019) 20:22–38. doi: 10.3390/ijms20225543
59. Sereti I, Gea-Banacloche J, Kan MY, Hallahan CW, Lane HC. Interleukin 2 Leads to Dose-Dependent Expression of the Alpha Chain of the IL-2 Receptor on CD25-Negative T Lymphocytes in the Absence of Exogenous Antigenic Stimulation. *J Clin Immunol (Orlando Fla.)* (2000) 97:266–76. doi: 10.1006/clim.2000.4929
60. Murphy KM, Stockinger B. Effector T Cell Plasticity: Flexibility in the Face of Changing Circumstances. *Nat Immunol* (2010) 11:674–80. doi: 10.1038/ni.1899
61. Villarreal C, Padilla-Longoria P, Alvarez-Buylla ER. General Theory of Genotype to Phenotype Mapping: Derivation of Epigenetic Landscapes From N-Node Complex Gene Regulatory Networks. *Phys Rev Lett* (2012) 109:1–5. doi: 10.1103/PhysRevLett.109.118102
62. Barberis M, Helikar T, Verbruggen P. Simulation of Stimulation: Cytokine Dosage and Cell Cycle Crosstalk Driving Timing-Dependent T Cell Differentiation. *Front Physiol* (2018) 9:879. doi: 10.3389/fphys.2018.00879
63. Purtic B, Pitcher LA, Van Oers NS, Wülfing C. T Cell Receptor (TCR) Clustering in the Immunological Synapse Integrates TCR and Costimulatory Signaling in Selected T Cells. *Proc Natl Acad Sci United States America* (2005) 102:2904–9. doi: 10.1073/pnas.0406867102
64. Hartman NC, Nye JA, Groves JT. Cluster Size Regulates Protein Sorting in the Immunological Synapse. *Proc Natl Acad Sci USA* (2009) 106:12729–34. doi: 10.1073/pnas.0902621106
65. Dustin ML. The Immunological Synapse. *Cancer Immunol Res* (2014) 2:1023–33. doi: 10.1158/2326-6066.CIR-14-0161
66. Martín-Cófreces NB, Vicente-Manzanares M, Sánchez-Madrid F. Adhesive Interactions Delineate the Topography of the Immune Synapse. *Front Cell Dev Biol* (2018) 6:149. doi: 10.3389/fcell.2018.00149
67. Mayer A, Zhang Y, Perelson AS, Wingreen NS. Regulation of T Cell Expansion by Antigen Presentation Dynamics. *Proc Natl Acad Sci United States America* (2019) 116:5914–9. doi: 10.1073/pnas.1812800116
68. Desdín-Micó G, Soto-Heredero G, Mittelbrunn M. Mitochondrial Activity in T Cells. *Mitochondrion* (2018) 41:51–7. doi: 10.1016/j.mito.2017.10.006
69. Tan H, Yang K, Li Y, Shaw TI, Wang Y, Blanco DB, et al. Integrative Proteomics and Phosphoproteomics Profiling Reveals Dynamic Signaling Networks and Bioenergetics Pathways Underlying T Cell Activation. *Immunity* (2017) 46:488–503. doi: 10.1016/j.immuni.2017.02.010
70. Geltink RI, Kyle RL, Pearce EL. Unraveling the Complex Interplay Between T Cell Metabolism and Function. *Annu Rev Immunol* (2018) 36:461–88. doi: 10.1146/annurev-immunol-042617-053019
71. Qureshi OS, Kaur S, Hou TZ, Jeffery LE, Poulter NS, Briggs Z, et al. Constitutive Clathrin-Mediated Endocytosis of CTLA-4 Persists During T Cell Activation. *J Biol Chem* (2012) 287:9429–40. doi: 10.1074/jbc.M111.304329
72. Valitutti S, Coombs D, Dupré L. The Space and Time Frames of T Cell Activation at the Immunological Synapse. *FEBS Lett* (2010) 584:4851–7. doi: 10.1016/j.febslet.2010.10.010
73. Moran AE. Hogquist KA. T-Cell Receptor Affinity in Thymic Development. *Immunology* (2012) 135:261–7. doi: 10.1111/j.1365-2567.2011.03547.x
74. Wells AD, Walsh MC, Bluestone JA, Turka LA. Signaling Through CD28 and CTLA-4 Controls Two Distinct Forms of T Cell Anergy. *J Clin Invest* (2001) 108:895–904. doi: 10.1172/JCI13220

75. Macián F, García-Cózar F, Im SH, Horton HF, Byrne MC, Rao A. Transcriptional Mechanisms Underlying Lymphocyte Tolerance. *Cell* (2002) 109:719–31. doi: 10.1016/s0092-8674(02)00767-5
76. Oh YM, Park HB, Shin JH, Lee JE, Park HY, Kho DH, et al. Ndrg1 Is a T-Cell Clonal Anergy Factor Negatively Regulated by CD28 Costimulation and Interleukin-2. *Nat Commun* (2015) 6:86–98. doi: 10.1038/ncomms9698
77. Merlot AM, Porter GM, Sahni S, Lim EG, Peres P, Richardson DR. The Metastasis Suppressor, NDRG1, Differentially Modulates the Endoplasmic Reticulum Stress Response. *Biochim Biophys Acta Mol basis Dis* (2019) 1865:2094–110. doi: 10.1016/j.bbadis.2019.04.007
78. Frauwirth KA, Thompson CB. Activation and Inhibition of Lymphocytes by Costimulation. *J Clin Invest* (2002) 109:295–9. doi: 10.1172/JCI0214941
79. Dupage M, Bluestone JA. Harnessing the Plasticity of CD4+ T Cells to Treat Immune-Mediated Disease. *Nat Rev Immunol* (2016) 16:149–63. doi: 10.1038/nri.2015.18
80. Stark JM, Tibbitt CA, Coquet JM. The Metabolic Requirements of Th2 Cell Differentiation. *Front Immunol* (2019) 10:2318. doi: 10.3389/fimmu.2019.02318
81. Darlington PJ, Kirchhof MG, Criado G, Sondhi J, Madrenas J. Hierarchical Regulation of CTLA-4 Dimer-Based Lattice Formation and Its Biological Relevance for T Cell Inactivation. *J Immunol* (2005) 175:996–1004. doi: 10.4049/jimmunol.175.2.996
82. Lee K, Gudapati P, Dragovic S, Spencer C, Joyce S, Killeen N, et al. Mammalian Target of Rapamycin Protein Complex 2 Regulates Differentiation of Th1 and Th2 Cell Subsets *via* Distinct Signaling Pathways. *Immunity* (2010) 32:743–53. doi: 10.1016/j.immuni.2010.06.002
83. Delgoffe GM, Pollizzi KN, Waickman AT, Heikamp E, Meyers DJ, Horton MR, et al. The Kinase mTOR Regulates the Differentiation of Helper T Cells Through the Selective Activation of Signaling by Mtorc1 and Mtorc2. *Nat Immunol* (2011) 12:295–303. doi: 10.1038/ni.2005
84. Mitsuiki N, Schwab C, Grimbacher B. What did We Learn From CTLA-4 Insufficiency on the Human Immune System? *Immunol Rev* (2019) 287:33–49. doi: 10.1111/imr.12721
85. Hori S, Nomura T, Sakaguchi S. Control of Regulatory T Cell Development by the Transcription Factor Foxp3. *Sci (New York NY)* (2003) 299:1057–61. doi: 10.1126/science.1079490
86. Wing K, Onishi Y, Prieto-Martin P, Yamaguchi T, Miyara M, Fehervari Z, et al. Ligation of CD28 by Its Natural Ligand CD86 in the Absence of TCR Stimulation Induces Lipid Raft Polarization in Human CD4 T Cells. *Science* (2008) 322:271–5. doi: 10.1126/science.1160062
87. Walker LSK. Treg and CTLA-4: Two Intertwining Pathways to Immune Tolerance. *J Autoimmun* (2013) 45:49–57. doi: 10.1016/j.jaut.2013.06.006
88. Knoechel B, Lohr J, Zhu S, Wong L, Hu D, Ausubel L, et al. Functional and Molecular Comparison of Anergic and Regulatory T Lymphocytes. *J Immunol (Baltimore Md. 1950)* (2006) 176:6473–83. doi: 10.4049/jimmunol.176.11.6473
89. Zheng Y, Delgoffe GM, Meyer CF, Chan W, Powell JD. Anergic T Cells Are Metabolically Anergic. *J Immunol (Baltimore Md. 1950)* (2009) 183:6095–101. doi: 10.4049/jimmunol.0803510
90. Qi SY, Groves JT, Chakraborty AK. Synaptic Pattern Formation During Immune Recognition. *Proc Natl Acad Sci* (2001) 98:6548–53. doi: 10.1073/pnas.111536798
91. Vicente-Manzanares M, Ortega-Carrion A. Concerning Immune Synapses: A Spatiotemporal Timeline. *F1000Research* (2016) 5:1–11. doi: 10.12688/f1000research.7796.1
92. Calvo V, Izquierdo M. Imaging Polarized Secretory Traffic at the Immune Synapse in Living T Lymphocytes. *Front Immunol* (2018) 9:684. doi: 10.3389/fimmu.2018.00684
93. Martín-Cófreces NB, Sánchez-Madrid F. Sailing to and Docking at the Immune Synapse: Role of Tubulin Dynamics and Molecular Motors. *Front Immunol* (2018) 9:1174. doi: 10.3389/fimmu.2018.01174
94. Egen JG, Allison JP. Cytotoxic T Lymphocyte Antigen-4 Accumulation in the Immunological Synapse Is Regulated by TCR Signal Strength. *Immunity* (2002) 16:23–35. doi: 10.1016/s1074-7613(01)00259-x
95. Wang XB, Zheng CY, Giscombe R, Lefvert AK. Regulation of Surface and Intracellular Expression of CTLA-4 on Human Peripheral T Cells. *Sca J Immunol* (2001) 54:453–8. doi: 10.1046/j.1365-3083.2001.00985.x
96. Hanada M, Feng J, Hemmings BA. Structure, Regulation and Function of PKB/AKT–a Major Therapeutic Target. *Biochim Biophys Acta* (2004) 1697:3–16. doi: 10.1016/j.bbapap.2003.11.009
97. Parry RV, Chemnitz JM, Frauwirth KA, Lanfranco AR, Braunstein I, Kobayashi SV, et al. CTLA-4 and PD-1 Receptors Inhibit T-Cell Activation by Distinct Mechanisms. *Mol Cell Biol* (2005) 25:9543–53. doi: 10.1128/MCB.25.21.9543-9553.2005

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.

frontiers in Immunology

REVIEW
published: 29 November 2021
doi: 10.3389/fimmu.2021.778913

Corona Virus Disease 2019 (COVID-19) as a System-Level Infectious Disease With Distinct Sex Disparities

Modjtaba Emadi-Baygi[1*], *Mahsa Ehsanifard*[1†], *Najmeh Afrashtehpour*[1†], *Mahnaz Norouzi*[2†] *and Zahra Joz-Abbasalian*[3]

[1] Department of Genetics, Faculty of Basic Sciences, Shahrekord University, Shahrekord, Iran, [2] Department of Research and Development, Erythrogen Medical Genetics Lab, Isfahan, Iran, [3] Clinical Laboratory, Sina Hospital, Tehran University of Medical Sciences, Tehran, Iran

The current global pandemic of the Severe Acute Respiratory Syndrome CoronaVirus 2 (SARS-CoV-2) causing COVID-19, has infected millions of people and continues to pose a threat to many more. Angiotensin-Converting Enzyme 2 (ACE2) is an important player of the Renin-Angiotensin System (RAS) expressed on the surface of the lung, heart, kidney, neurons, and endothelial cells, which mediates SARS-CoV-2 entry into the host cells. The cytokine storms of COVID-19 arise from the large recruitment of immune cells because of the dis-synchronized hyperactive immune system, lead to many abnormalities including hyper-inflammation, endotheliopathy, and hypercoagulability that produce multi-organ dysfunction and increased the risk of arterial and venous thrombosis resulting in more severe illness and mortality. We discuss the aberrated interconnectedness and forthcoming crosstalks between immunity, the endothelium, and coagulation, as well as how sex disparities affect the severity and outcome of COVID-19 and harm men especially. Further, our conceptual framework may help to explain why persistent symptoms, such as reduced physical fitness and fatigue during long COVID, may be rooted in the clotting system.

Keywords: COVID-19, interconnectedness, immunity, coagulation, thromboembolism

OPEN ACCESS

Edited by:
Federica Eduati, Eindhoven University of Technology, Netherlands

Reviewed by:
Darragh Duffy, Institut Pasteur, France
Mark E Snyder, University of Pittsburgh, United States

****Correspondence:***
Modjtaba Emadi-Baygi
emadi-m@sku.ac.ir

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Systems Immunology, a section of the journal Frontiers in Immunology

Received: *17 September 2021*
Accepted: *11 November 2021*
Published: *29 November 2021*

Citation:
Emadi-Baygi M, Ehsanifard M, Afrashtehpour N, Norouzi M and Joz-Abbasalian Z (2021) Corona Virus Disease 2019 (COVID-19) as a System-Level Infectious Disease With Distinct Sex Disparities. Front. Immunol. 12:778913. doi: 10.3389/fimmu.2021.778913

INTRODUCTION

Corona Virus Disease 2019 (COVID-19) is an infectious disease caused by SARS-CoV-2, an RNA virus with a crown-like appearance, and spreads rapidly all over the world. It transmits from human-to-human mainly *via* respiratory system (1). COVID-19, known to be a heterogeneous disease that manifests a varying range of symptoms from asymptomatic to severe disease. As an RNA virus, SARS-CoV-2 is a highly mutable virus with a rapidly evolving rate that leads to a various subtypes of the virus. Some characteristics of the SARS-CoV-2 have changed significantly in some evolved subtypes of the virus. Rapid evolution of the virus creates critical changes in SARS-CoV-2 behaviors, including an altered transmission or more severe disease (2). Considering the high rate of alterations in SARS-CoV-2 genome that cause various types of clinical and laboratory manifestations that in turn leads to the progression of COVID-19 towards severe and fatal forms of the disease in some cases, it calls for an urgent need for the identification of aberrated

interconnectedness of biological levels of organization of humans and crosstalks between pathways that establish a malicious circuitry involving in COVID-19 pathogenesis. Interconnected cell signaling pathways (interconnectedness) may effectively and precisely transmit innumerable diverse signals, despite an intrinsic potential for improper levels of cross-talk. Network-level mechanisms insulate pathways from crosstalk and allow cells to process information from their environment and respond in ways to input signals. Evolutionarily, metazoan signaling networks are intricate, with incredible levels of crosstalk between pathways where proteins shared between two pathways provoke one pathway's activity to be modulated by the activity of another. Indeed, crosstalk allows different cell types, each expressing a specific subset of signaling proteins, to trigger distinct outputs when dealt with the same inputs, reacting distinctly to the same environment. To point out, the tissue-specific networks mainly respond distinctly to the inhibition of individual proteins. These findings imply that the intricate interaction between network topology and gene expression that allows various cell types to respond distinctly to the same signals has significant implications for the development of drugs that target signaling processes (3, 4). As COVID-19 clinical manifestations indicate the symptoms, severity of the disease, and corresponding care settings, vary amongst the infected patients (5). One crucial determining factor in COVID-19 severity is the interaction between the virus and the host cells. It is known that the Spike (S) protein of CoronaViruses (CoVs) mediates the binding of the virus to the cell receptors and enables the fusion of the virus with the host cell membrane (6). Of particular interest, SARS-CoV-2 interacts with the RAS *via* ACE2, which was also identified as a functional receptor for Severe Acute Respiratory Syndrome CoronaVirus (SARS-CoV) (7–9). It has been indicated that RAS activation during SARS-CoV-2 infection leads to a number of unfavorable effects, which include vasoconstriction and hypertension, cellular differentiation and growth, endothelial dysfunction, and the formation of Reactive Oxidative Species (ROS) that may ultimately lead to organ damage (10, 11).

In human, ACE2 is expressed in nearly all human organs, such as the upper respiratory tract, alveolar epithelial cells, vascular endothelial cells and macrophages. In addition to acting as the receptor for SARS-CoV-2, ACE2 is a component of RAS which regulates several pathological processes like fibrosis, inflammation, oxidative stress and vasoconstriction (12, 13). Currently, there is a lack of evidence on how aberrated interconnectedness of RAS with the other systems contributed in COVID-19 pathophysiology. It has been hypothesized that the RAS may be involved in the COVID-19 pathogenesis *via* activation of the classic pathway. The broad distribution of ACE2 receptors in the endothelium could potentially allow for widespread effects outside the lung, following SARS-CoV-2 infection (14).

COVID-19 appears as asymptomatic disease or shows mild symptoms in the majority of patients (about 80%) (15, 16). In clinical evaluation, fever, cough, dyspnea, myalgia, and fatigue are the most common symptoms among mildly symptomatic patients. Moreover, uncommon symptoms, including headache, sputum production, hemoptysis, and diarrhea, have been reported in SARS-CoV-2 infection (17, 18). However, the remained proportion of the patients experience severe complications within a short time after infection, such as Acute Respiratory Distress Syndrome (ARDS), Disseminated Intravascular Coagulation (DIC), sepsis followed by organ failure, and death (19, 20). Notably, studies reported elder people are at a greater risk for developing severe forms of COVID-19, and a higher mortality rate was reported among older adults (21). COVID-19 is a complicated multi-system disease that greatly affects the vascular system and hemostasis maintenance. The correlation between the immune system dysfunction and impairment of general hemostasis is well-known and has been addressed in different contexts (1). The innate immune response is the first-line defense of the human body and has a determinant role in protective and/or destructive responses to any infections (22). An effective immune response upon viral infection includes type I InterFeroN (IFN-I) responses and its downstream cascades (23). In SARS-CoV-2, at the first contact with the respiratory mucosa and following SARS-CoV-2 entrance, the production of structural and non-structural proteins of the virus is triggered, subsequently blocking of the interferon response is promoted *via* SARS-CoV-2 N-protein (24). Failed IFN-I response curtails the early viral control and induced the penetration of hyper-inflammatory neutrophils, monocytes and macrophages, which lead to extensive production of pro-inflammatory cytokines and may evoke a cytokine storm that is correlated to the severe manifestations of COVID-19 (25–27).

Clinical evidence of severe patients of COVID-19 indicate that hyper-inflammation, and particular forms of vasculopathy, including Thrombotic MicroAngiopathy (TMA) and intravascular coagulopathy, are frequent features among them. In these cases, an uncontrolled increase of inflammatory cytokines induces vascular hyperpermeability and Multi-Organ Dysfunction Syndrome (MODS) leading to cardiac, hepatic, renal systems' failure, and eventually death (1). Consistent with a state of hypercoagulability linked with a severe inflammatory response, coagulopathy is an important pathophysiological feature of COVID-19, characterized by the elevated fibrinogen levels, Von Willebrand Factor (VWF), and the fibrin degradation product (D-dimer, a fibrin degradation product that its increases are frequently reported in COVID-19 patients) (28). In fact, a failure to retain hemostasis due to pulmonary injury and MODS creates a critical condition in COVID-19 severe patients (29). A more complicated situation has been reported in the presence of coagulopathy, which is rather a prothrombotic character with a high chance of Venous ThromboEmbolism (VTE) among COVID-19 patients in Intensive Care Units (ICUs) (30). It seems that endothelial dysfunction, microvascular thrombosis, and occlusion, or autoimmune mechanisms may contribute to developing coagulopathy in severe SARS-CoV-2 pneumonia (31). In general, COVID-19 severity is not limited to the respiratory tract and shows age and sex tendencies. Besides the age bias, sex bias with higher numbers of cases, greater disease

severity, and higher death rates among men compared to women is one of the interesting features of this disease that might occur due to male-specific factors that increase men's susceptibility to the SARS-CoV-2 infection (32, 33). A growing body of evidence implies that we need to consider COVID-19 as a system-level infectious disease in which aberrated interconnectedness of immune system dysfunction, aberrant inflammatory responses, and prothrombotic conditions and crosstalks between some pathways occurred during COVID-19 pathogenesis that lead to the various manifestations and severity of the disease. Interconnected biological levels of organization of humans are capable of transmiting a multitude of different signals efficiently and accurately, regardless of a built-in potential for unenviable levels of crosstalks (28, 34, 35). With this in mind, we aimed to throw more light on pivotal aberrated interconnectedness of biological levels of organizations and pathways crosstalks in progression of COVID-19, to provide a well-defined insight to recognize COVID-19-associated pathophysiology as a system-level infectious disease.

THE ROLE OF RENIN-ANGIOTENSIN SYSTEM

When SARS-CoV-2 infects cells expressing the surface receptors ACE2 and transmembrane serine protease 2 (TMPRSS2), the active replication and release of the virus causes the host cell to undergo pyroptosis, an inflammatory cell death, and release Damage-Associated Molecular Patterns (DAMPs). In contrast to Angiotensin-Converting Enzyme (ACE), a zinc metalloproteinase and a key regulator of the RAS, ACE2 is an enzyme identified in rodents and humans with a more restricted distribution than ACE (36). In humans, ACE2 is mainly expressed in lung epithelial cells, enterocytes, arterial endothelial cells, and smooth muscle cells in cardiovascular tissues (37). In ACE2/TMPRSS2 dependent cell entry, S protein of SARS-CoV-2 binds to ACE2 to initiate virus entry (38, 39). Once bound, TMPRSS2 cleaves the S protein to allow for membrane fusion. Upon endocytosis, the viral RNA genome is released into the cytoplasm of the host by fusion of the virions with the endosomal membrane (6). This process is the most decisive phase in defining host compatibility and transmissibility of SARS-CoV-2.

The RAS has two pathways, the ACE-dependent pathway (vasoconstrictive side) and the ACE-independent pathway (vasodilative side). Both pathways begin with renin, which is produced by the kidney, converting angiotensinogen from the liver into ANGiotensin I (ANG I). In the ACE-dependent pathway, ANG I is converted to Angiotensin II (ANG II) by ACE. ANG II attaches to its receptor, ANG II Type 1 Receptor (AT1R). This increases Blood Pressure (BP) by causing vasoconstriction and sodium retention. Notwithstanding, in ACE-independent pathway a different enzyme, ACE2, converts ANG I to Angiotensin-1-9 (ANG-1-9) and ANG II to Angiotensin-1-7 (ANG-1-7). ANG-1-7 interacts with two different receptors, Mas and ANG II Type 2 Receptor (AT2R) receptor (40, 41). This pathway works to oppose the actions of the ACE-dependent pathway by causing vasodilation, thereupon lowering BP and providing other cardioprotective effects (40).

As SARS-CoV infection on lung cells leads to the decreased levels of ACE2 (42), it is postulated that SARS-CoV-2 works in the same vein in individuals without pre-existing conditions to reduce ACE2 expression in lung tissue. Markedly, ACE2 is highly expressed in the lung parenchyma, especially in type II pneumocytes (Alveolar Type II (AT2) cells) (43). Type II cells synthesize and release pulmonary surfactant, enriched with a rather unique phospholipid and four surfactant-associated proteins, which is necessary to maintain alveolar structure (44). Furthermore, Type II cells can differentiate to become Alveolar Type I (AT1) cells, a mechanism for replacement of type I cells that are damaged. The SARS-CoV-2 and SARS-CoV-1 viruses perturb alveoli to cause the main pathology in the lung, with increased fluid entry, cell death and inflammation along with reduction in gas exchange and levels of surfactant (45, 46). Moreover, viral infection triggers ACE2 endocytosis, leading to reduced cell surface expression of ACE2 (**Figure 1A**). Conversely, in individuals with comorbidities including diabetes, cardiovascular disease and hypertension, the ACE independent pathway is activated after SARS-CoV-2 infection due to the administrated drugs in these patients that block RAS. Indeed, pharmacological research concerning the SARS-Cov-2 patients and animal studies revealed frequent use of ACE inhibitor and/or Angiotensin II type I receptor blockers lead to a significant increase in ACE2 expression due to rise of angiotensin- (1–7) levels and decline in Ang-II levels in the plasma (47, 48) (**Figure 1B**). In addition, Gottschalk G et al. reported that ACE2 receptor was strongly upregulated in lungs during SARS-CoV-2 infection. As a result, ACE2 is seen as a double-edged sword in which in normal people the activation of ACE-dependent pathway upon SARS-CoV-2 infection leads to reduced expression of ACE2 that in turn results in ARDS and endothelial dysfunction, while in individuals with comorbidities leads to coagulation, inflammation, and damage to lungs and brain (49–51).

Using RNA sequencing of BronchoAlveolar Lavage (BAL) samples from patients with severe COVID-19 and comparing the results with control samples, Garvin et al. provided further detail into dysregulation of the RAS (52). Of note, the BAL fluid showed profound dysregulation of the RAS. Angiotensinogen and renin increased significantly. The degradation of Inhibitor of nuclear factor Kappa-B Kinase subunit gamma (IKK- γ) by the virus-encoded protein results in blocking production of interferon and ACE transcription. Without ACE, and thereby ANG II, ACE2 is upregulated in the lavage samples that in turn provide exceptionally more entry points for the virus. Furthermore, ACE2 converts ANG I to the fragment ANG-1-9. This fragment activates bradykinin receptor signaling that leads to bradykinin storm (52). In the lavage samples, bradykinin receptors 1 and bradykinin receptors 2 were substantially upregulated (53). Meanwhile, because of the reduction of ACE2 expression in response to SARS-CoV-2 infection ostensively in normal individuals, Ang II/AT1 activation *via* ACE-dependent pathway results in inflammation through an

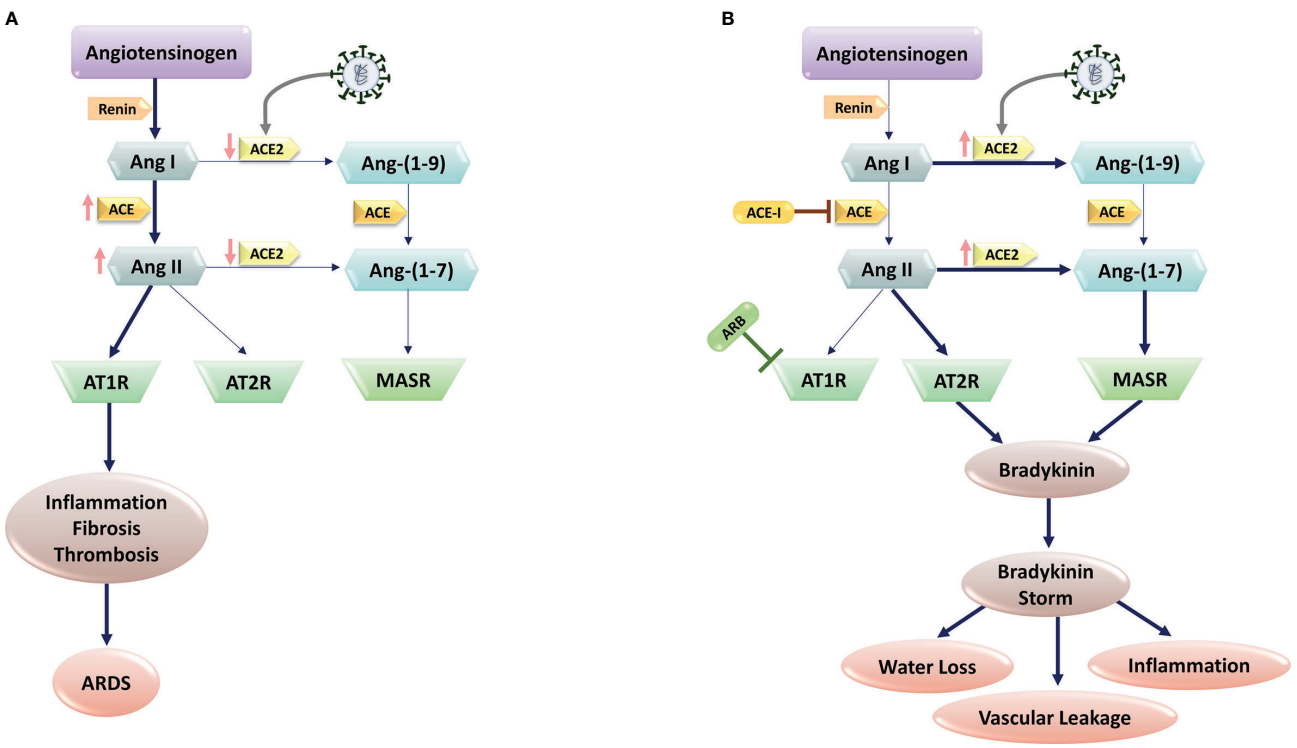


FIGURE 1 | SARS-CoV-2 and Renin-Angiotensin System (RAS) interactions. **(A)** SARS-CoV-2 infection is mediated *via* the human Angiotensin-Converting Enzyme 2 (ACE2), which would cause ACE2 downregulation. Subsequent block of Ang II/Ang–(1-7) metabolism lead to the high levels of Ang II production through ACE upregulation. The imbalanced Ang II/Ang 1–7 causes the suppression of AT2 and Mas receptors with anti-inflammatory properties. In contrast, Ang II and its receptor AT1R activations induce pro-inflammatory effects, fibrosis, and thrombosis, leading to Acute Respiratory Distress Syndrome (ARDS) in lungs. **(B)** In patients with comorbidities, including hypertension, administration of RAS blockers (e.g., ACE-I, and ARBs) suppresses the ACE pathway. In these patients, SARS-CoV-2 infection increase ACE2 expression. Upregulated ACE-2 converts Ang I to Ang–(1-9) and Ang–(1-7), activating the bradykinin pathway *via* MAS and AT2R receptors. Upregulated bradykinin (bradykinin storm) in the lungs of COVID-19 patients would increase SARS-CoV-2 virulence by inducing hypotension, vascular permeability, water loss, and inflammation.

increase in ROS level that in turn leads to a rise in InterLeukin 6 (IL-6) and C-reactive protein levels. Furthermore, Ang II/AT1 activation results in endothelial dysfunction, and in the context of a viral infection, increased endothelial signaling may be the catalyst for initiation of the coagulation cascade in certain individuals. In this pathway, Ang2/AT1 activation results in multiplicative effects on vasoconstriction, BP, endothelial dysfunction, ROS formation, and finally damage to organs (**Figure 1**) (54).

Considering the role of RAS in COVID-19 pathogenesis, the key mechanisms associated with lower COVID-19 severity and mortality in women are: I) Decreased ACE2 methylation in women, resulting in increased ACE2 expression; II) ACE2 is on the short arm of the X chromosome, where up to 30% of genes undergo X inactivation Escape; III) estrogen promotes ACE2 expression. Higher levels of ACE2 could supply a better source to protect tissue after viral entry. Evidence shows ACE2 plays a protective role in chronic pathologies including hypertension, cardiovascular diseases, and acute respiratory distress syndrome, which are the comorbidities representing the risk of worse prognosis in COVID-19. Studies in mouse models support the protective role of ACE2 by showing more severe lung failure upon ACE2 down-regulation that results in overactivation of the Angiotensin (Ang) II/AT1R axis that may explain the multi-organ dysfunction seen in patients (**Figure 1**) (55, 56). However, low estrogens in men lead to the absence of higher levels of ACE2, supporting the ACE pathway in the RAS axis that facilitates disease severity in men with the same viral load as women. In men, androgens increased the expression of TMPRSS2 that supports viral entry, resulting in male increased susceptibility. In contrast, decreased levels of androgens in women may keep TMPRSS2 expression at lower levels, providing an additional defensive determinant against the development of COVID-19 infection (57).

IMMUNE SYSTEM AND INFLAMMATORY RESPONSES

The human immune system protects us against life-threatening and pathogenic agents. As the first line of immune defense, the innate immune system plays an indispensable role against

viruses (58, 59). Germline-encoded Pattern Recognition Receptors (PRRs) are essential immune receptors that trigger antiviral innate immune responses through sensing the conservative structures of viruses. PRRs have a crucial role in immune responses by recognizing Pathogen-Associated Molecular Patterns (PAMPs) that are unique microbial molecules and DAMPs that are self-derived molecules elicited from damaged cells (60, 61). Two PRRs, Toll-Like Receptors (TLRs) and Retinoic acid-Inducible Gene-I [RIG-I Like Receptors (RLRs)], are identified to have a crucial function in sensing viral ssRNA genome and dsRNA replication intermediates. TLRs and RLRs recognize these nucleic acid species and bind certain intracellular adaptor proteins, which activate NF-κB, mitogen-activated protein kinases, and interferon regulatory factors, which control the transcription of genes encoding IFN-I and other inflammatory cytokines, which are critical for virus elimination (62).

PRRs can identify SARS-CoV-2, like other RNA viruses, leading to the activation of IFN-I response (29, 63, 64). To point out, multiple viral structural and non-structural proteins of SARS-CoV impair IFN-I response, resulting in a hyper-inflammatory response. To enumerate, SARS-CoV is able to antagonize the TLR signaling pathway *via* its papain-like protease (65). Indeed, antagonized IFN responses in SARS-CoV-2 infection is likely to occur at various pathways involved in immune responses (**Figure 2**) (1). Like SARS-CoV and Middle East Respiratory Syndrome CoronaVirus (MERS-CoV), vesicle-dependent replication of SARS-CoV-2 genome may cause protecting RNA genome from host detection by cytosolic (e.g., RIG-I) and endosomal (e.g., TLR3/7) PRRs (66, 67). Furthermore, it is possible that INF-I inhibitions cause the antiviral response delay (e.g., by inhibiting TLR3 and TLR7 signaling pathways, virus-encoded antagonists to the IFN responses and auto-antibodies) that in turn it facilitates replication of virus particles and extensive virus cytopathic effects at the early stages of the COVID-19 disease (29, 63, 65, 68–70). Moreover, the infection of ACE2-positive ATII pneumocytes would lead to a significant decrease in the production of pulmonary surfactant and exposing the TLR4. In addition, direct or indirect SARS-CoV-2 binding to TLR4 causes an increase in the expression of ACE2 *via* IFNs and interferon-stimulated genes. Therefore, the virus may directly enter the cell using TLR4 and cause aberrant TLR4 signaling (64). However,

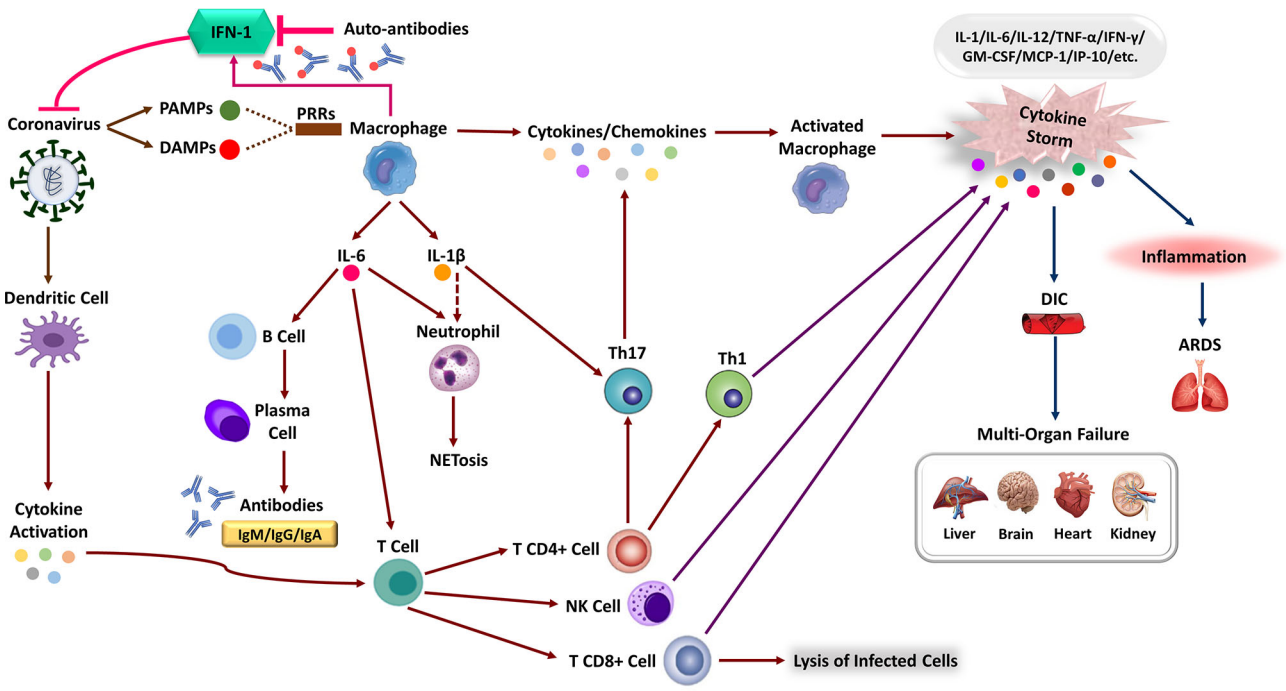


FIGURE 2 | The immune response corporation in covid-19 disease. Type 1 interferon (IFN-I) immune response plays a pivotal role in effective immunity against SARS-CoV-2 infection and rapid viral clearance. Following SARS-CoV-2 infection, the IFN-1 initiates *via* recognizing PAMPs/DAMPs by PRRs of the human immune cells and releasing inflammatory cytokines (e.g., IL-1 beta, IL-6). The expression of numerous inflammatory cytokines leads to activation of neutrophils (NET formation) and attraction of different immune cells toward the site of infection. Macrophages activation, Dendritic Cell (DC) maturation and inflammatory cytokines stimulate the adaptive immune response to join the fight against SARS-CoV-2. Activation of T-cells *via* cytokines promotes multiple T-cell differentiation (i.e., CD4 +, CD8 +, NK cell, Th1) that directly kills virus-infected cells. Activated B-cells produce virus-specific antibodies that would participate in a successive immune response. In patients with severe forms of COVID-19, the delayed IFN-I pathways or neutralizing auto-Abs will lead to immune system overreaction and the generation of high inflammatory cytokine levels. The aberrant induction of the immune system and the production of various pro-inflammatory cytokines (e.g., IL-1, IL-6, MCP-1, TNF-α, and etc.), the so-called "cytokine storm," leads to severe COVID-19 immunopathology. These can cause severe local damage to the lungs (e.g., ARDS) and other organs (e.g., DIC), and in the worst case, can lead to Multi-Organ Failures (MOF) and even death.

Onabajo et al. show that the ACE2 induced by IFN is a truncated isoform of ACE2, deltaACE2, which is nonfunctional in binding the SARS-CoV-2 spike protein (71). Correspondingly, the activity of SARS-CoV-2 non-structural protein 14, which has a (guanine-N7)-methyltransferase activity, results in the efficient escape of viral RNA from detection by the RIG-I receptor. RIG-I is a critical cytosolic RNA sensor that interacts with the mitochondrial antiviral signaling proteins to activate the downstream programs such as type I/III IFN responses (72). This suboptimal innate immune operation by host PRRs during SARS-CoV-2 infection induces non-productive inflammatory responses, resulting in a cytokine storm and disseminated damage to the host. Generally, cytokine storm is a hyperactive immune response characterized by the release of different cytokines, chemokines, and other immune mediators that may hurt host cells. Notably, the majority of cytokine storm mediators demonstrate pleiotropic downstream effects mostly interdependent in their biological activity. Therefore discovering the precise dysregulated inflammatory response implies pathogenesis of the disease has been a major challenge in critical illness like COVID-19 (73). Although reports suggest that like SARS-CoV, SARS-CoV-2 have subversion strategies against innate immune signaling and antiviral interferon response, the exact aberrations of interconnectedness and crosstalks remain poorly understood (66, 72, 74). Commonly, cytokine storm is considered a critical determining factor associated with adverse outcomes of the COVID-19 disease; however, there is variation and sometimes-even discrepancies between studies about the exact detail of this phenomenon. Despite the poorly defined pathophysiology of cytokine storm, widespread acceptance of the term in COVID-19 has motivated to apply potent immunomodulatory therapies such as IL-6 inhibitors and high-dose corticosteroids. Nevertheless,in comparison with the other causes of ARDS (median IL-6 level is 10- to 200-fold higher than COVID-19) the lower levels of circulating cytokines in COVID-19 may not be representative of lung inflammation and need to be determined whether COVID-19 related ARDS phenotype is associated with the cytokine storm (73, 75).

SARS-CoV-2 infection in pulmonary airway epithelial cells triggers local and systemic pathological responses through recruiting macrophages and monocytes, expressing inflammatory cytokines, and inducing adaptive immune responses. Although immune response resolves the infection in most COVID-19 patients, immune system dysfunction leads to hyper-inflammation resulting in a cytokine storm that causes severe lung injury and multi-system damages (1, 76, 77). These pro-inflammatory cytokines stimulate an influx of neutrophils and other myeloid cells into the lung, creating a local hyper-inflammatory response and significant immunopathology (78). The association of hyper-inflammatory response with lung injury, high rate of progression to ARDS, Multi-Organ Failure (MOF), and unfavorable prognosis of severe COVID-19 has been described in several studies (15, 79, 80). In the same way, overproduction of pro-inflammatory cytokines (Tumor Necrosis Factor alpha (TNF-α), IL-1, IL-6, and IL-1β), monocytes, and neutrophils, followed by a sharp decrease in lymphocytes, have been shown in various studies (15, 45, 81).

Clinical studies show major differences in the expression of inflammatory markers and immune phenotypes between moderate and severe cases, along with disease progression (82). An elevated level of immunomodulatory cytokines, including IL-1α, IL-1β, IFN-α, IL-17A, and IL-12p70, constitute COVID-19 signature that exhibits dynamic features associated with clinical manifestations (**Figure 2**) (15, 45). However, a considerable extent of other cytokines and chemokines, including IFN-γ, thrombopoietin (associated with blood clotting aberrations), IL-1, IL-6, IL-8, IL-2, IL-7, IL-10, Granulocyte Colony-Stimulating Factor (G-CSF), Interferon-inducible Protein 10 (IP-10), Monocyte Chemoattractant Protein 1 (MCP-1), Macrophage Inflammatory Protein 1α (MIP-1α) and TNF, C-X-C Ligand 10 (CXCL10), C-C motif Ligand 2 (CCL2) and CCL3, have been reported in severe patients (15). IL-6 is thought to play a pivotal role in pathology of COVID-19, considering its highest levels in non-survivors and critically ill patients. Notwithstanding that the IL-6 levels in these patients continue to increase over time, the mechanism leading to its elevation in severe COVID-19 is not currently clear. It is possible that aberrant activation of virus-specific PRRs like TLR4 triggers some sorts of crosstalks, which then drives IL-6 production (1, 83, 84).

The high levels of IL-6 observed in COVID-19 patients are apt to elicit Neutrophil Extracellular Traps (NETs) which include extracellular DNA fibers, histones, microbicidal proteins, proteases and oxidant enzymes to be released by neutrophils (**Figure 2**) (85, 86). Pathological effects of NETs have been indicated in propagated inflammation and respiratory failure. Although neutrophils are early indicators of SARS-CoV-2 infection, their excessive recruitment causes an unregulated NETs release that contributed to organ damage and mortality in COVID-19 patients (87). A significant increase in the blood neutrophil counts has been indicated in the COVID-19 patients admitted to the ICU. Consistent with SARS-CoV and MERS-CoV, increased levels of pro-inflammatory cytokines may drive an influx of neutrophils and other myeloid cells into the lung. The elevated neutrophil/lymphocyte ratio is considered an early prognostic marker in SARS-CoV-2 infection, which increases in parallel with the severity of the disease. Transcriptional analysis of BAL fluid and peripheral blood mononuclear cells showed that elevated CXCL2 and CXCL8 chemokines contribute to neutrophils recruitment and aggravate the inflammatory response in COVID-19 patients (66, 76). Of note, besides neutrophils, inflammatory chemokines cause the recruitment of other innate immune cells like monocytes, Dendritic Cells (DCs), and Natural Killer (NK) cells. These innate immune cells respond to tissue damage *via* producing several cytokines, including IL-1, IL-6, and TNF. Consequently, various immune cells such as neutrophils, macrophages, and T cells mobilize from the blood circulation into the infected tissue that leads to diffuse alveolar damage, capillary damage, vascular barrier damage, MOF and ultimately death (88–90).

Considering that the association of COVID-19 severity with aberrant inflammatory response has been demonstrated in various

studies, the exact molecular mechanisms causing innate immune response dysregulation are yet to be elucidated (1). In fact, activation of innate immunity against pathogens would benefit the host in two ways. First, effective initiation of innate immune responses would directly kill pathogens. If those responses failed to eliminate the infection, the adaptive immune responses would initiate to provide a second layer of immune protection (61). In SARS-CoV-2 infection, however, inadequate and delayed intracellular innate immune responses failed to prime adaptive immune response for a long time, leading to severe lung disease and MODS. This particular condition in SARS-CoV-2 infection leads to a dis-synchronized innate and adaptive immune response (74), a kind of aberrated interconnedness.

In general, the high specificity of the adaptive immune system enables a highly regulated and targeted immune response to eliminate the infected cells and neutralizing free virions (91, 92). The effective adaptive immune response contains both humoral (B-cells) and cellular (T-cells) responses. The activity of CD8+ T cells and B cells is predominantly regulated by CD4+ T cells. In contrast, the targeted killing of virus-infected cells is accomplished by CD8+ T cells. B cells Antibodies (Abs) block surface proteins and agglutinate virions and thereby prevent infection (**Figure 2**) (93). Humoral immune system produces specific Abs that are expected to neutralize different antigens, limit virus viability and also elicit T-cells to the location of infection by delivering them. But some produced Abs against CoVs result in more complicated situations. Neutralizing antigen-specific Abs against CoVs is mostly targeting spike domain (94). A group of anti-spike Immunoglobulin G (IgG) in acute phase of infection before viral clearance can end in some inflammatory hyper-reactions (95). This phenomenon is due to a shift in polarization of alveolar macrophages and their related cytokines to a pro-inflammatory state. This alteration is mostly because of the binding affinity of anti-spike IgG and receptors presented on macrophages (91). Equally important, some infected individuals with SARS-CoV-2 exhibit a range of B cell population producing auto-antibodies similar to those observed in Systemic Lupus Erythromatose (SLE) (96). These Abs are mostly antinuclear Abs, rheumatoid factor (anti-IgG-Fc Abs), AntiPhosphoLipid Antibodies (APLAbs), and Abs against IFN-I. These Abs suppress the innate immune response *via* IFN-I antagonizing and directly contributes to pathophysiology of COVID-19 (97). Furthermore, auto-antibodies target immune-related proteins including those involved in lymphocyte function and activation, leukocyte trafficking, the type I and type III IFN responses, type II immunity and the acute phase response increase in patients with severe COVID-19. With this in mind, patients with auto-antibodies against IFN-1 experience extended durations of hospital admission due to impairment of virological clearance (98).

Equally important, like SARS-CoV and MERS-CoV infection, a drastic diminish in the number of T cells would ultimately cause an inefficient T cell response in severe COVID-19 patients (99). The critical role of CD4+ and CD8+ T cells in virus clearance has been demonstrated in the immunodeficient mouse model of SARS-CoV (100). Considering the pivotal function of T cell immunity in SARS-CoV infection, T cells appear to have a protective role against SARS-CoV-2 infection. However, the function of T cells in the resolution, and long-term protection against SARS-CoV-2 remains debated. The significant dysregulation of T-cell response has been indicated in severe COVID-19 patients; However, it has not been clear which T-cell activities contribute to the development of the disease severity (101, 102). Lymphopenia is one of the prominent features of severe COVID-19 that have been reported to affect CD4+, CD8+, B, and NK cells in some patients. Coupled with the high IL-6, IL-10, and TNF levels, lymphopenia was observed in severe COVID-19 manifestations (103–105). In particular, different studies have demonstrated that lymphopenia harbors T cells biased in COVID-19, which would lead to hyperactivation of these cells (101). While activation of T cells seems to have positive effects, SARS-CoV-2 may have mechanisms to restrict T cell activation. Regarding the peripheral lymphopenia observed in patients with COVID-19, the autopsy studies showed that the lymphocytic infiltration to the respiratory tract or adhesion to inflamed respiratory vascular endothelium is not excessive (106, 107). Observed lymphopenia in adult COVID-19 patients is possibly multifactorial (108). Lymphopenia in patients with severe disease may be related to cytokine over-production possibly by a direct effect of the cytokines on T cell populations (109, 110) and/or indirect effects through other cell types including dendritic (111) cells and neutrophils (112, 113). Augmented expression of pro-apoptotic molecules (apoptotic loss) (114), probably decreased mobilization of lymphocytes from bone marrow, and immunosenescence may also partake in T cell depletion (108, 115). Moreover, with unknown mechanisms, limited MHC I and MHC II antigen presentation was demonstrated in association with T cell dysregulation in these patients (116). In contrast to the mild SARS-CoV-2 infection that harbors a successful lymphocyte-mediated virus clearance, T cells are functionally exhausted, a state that arises during many chronic infections (117), and express a high level of Programmed cell Death protein 1 (PD-1) and T-cell immunoglobulin mucin-3 in severe COVID-19 patients (117, 118). Besides, the implication of the expression of these markers in acute viral infection relates rather with the activation state than with functional exhaustion (119). Moreover, the observed decrease in IFN-γ and IL-21 production would support CD8+ and CD4+ T exhaustion in SARS-CoV-2 patients. In addition, the functional exhaustion of NK cells has been demonstrated in the persistence COVID-19 condition as seen in cancer and chronic viral infection (99, 120). Nevertheless, this exhaustion is likely to be transient, as the decreased expression of PD-1 has been observed in recovered patients from ICU compared to the severe ICU ones (121).

Eventually, a sex-specific immune transcriptome is documented in humans, resulting in females exhibiting an enhanced adaptive immune response, while males show enhanced aspects of innate immunity (122). Equally important, in the context of the COVID-19, there is a difference in immune responses and in turn in severity between males and females. Male patients had substantial induction of non-classical

monocytes and higher plasma levels of innate immune cytokines such as IL-8 and IL-18. By contrast, female patients during SARS-CoV-2 infection had substantial T cell activation compared to male patients. In particular, poor T cell response was exclusively associated with worse prognosis in male patients, which negatively correlated with patients' age. By contrast, higher levels of innate immune cytokines were associated with worse disease progression in female patients, but not in male patients (123). In general, since an extra X chromosome exists in women, cellular mosaicism created by X inactivation in females may contribute partly to the more efficient immune response against SARS-CoV-2. Besides, more ICU admissions and mortality rates are observed in men with COVID-19. To point out, there are many X-linked genes that are related to the immune system, including PRR (for instance TLR7 and TLR8) and the major TLR signaling regulators (like IL-1 receptor-associated kinase 1), that may have a higher copy number in females (124). To emphasize, the biallelic expression of TLR7, an endosomal innate immune receptor, through X chromosome, may potentially cause a more substantial IFN-I response in the early stages of COVID-19 in female patients. Furthermore, increased IFN-α production triggered by TLR7 ligands as shown by *in vitro* experimental observations are detected more in females and subsequently lead to a more rapid antiviral response (29, 124). Moreover, Female reproductive steroids are anti-inflammatory, reshape competence of immune cells, stimulate Ab production, and promote proliferation and repair of respiratory epithelial cells, suggesting they may protect against COVID-19 symptoms (125). Equally important, steroid hormone receptors play a pivotal role in sex-dependent immunity, as defined by lower T lymphocytes percentage in men compared to women. Meanwhile, multiple pieces of evidence imply that there are associations between testosterone levels (normal male level vs age-related hypogonadism) in COVID-19 pathogenesis, indicating its function as a double-edged sword. Testosterone and dihydrotestosterone mediate their actions *via* the Androgen Receptor (AR), a ligand-dependent nuclear transcription factor (126). High testosterone impact on COVID-19 severity through the TMPRSS2 Connection (**Figure 3**) (127). Taken together, epidemiological data emerging from the COVID-19 pandemic, backed by animal studies and further by preliminary clinical studies in diverse clinical settings, support the notion that high testosterone levels acting *via* the AR attune TMPRSS2 function positively to enhance SARS-CoV-2 S proteins and eventually increase COVID-19 infectivity and severity. Additionally, AR mutations

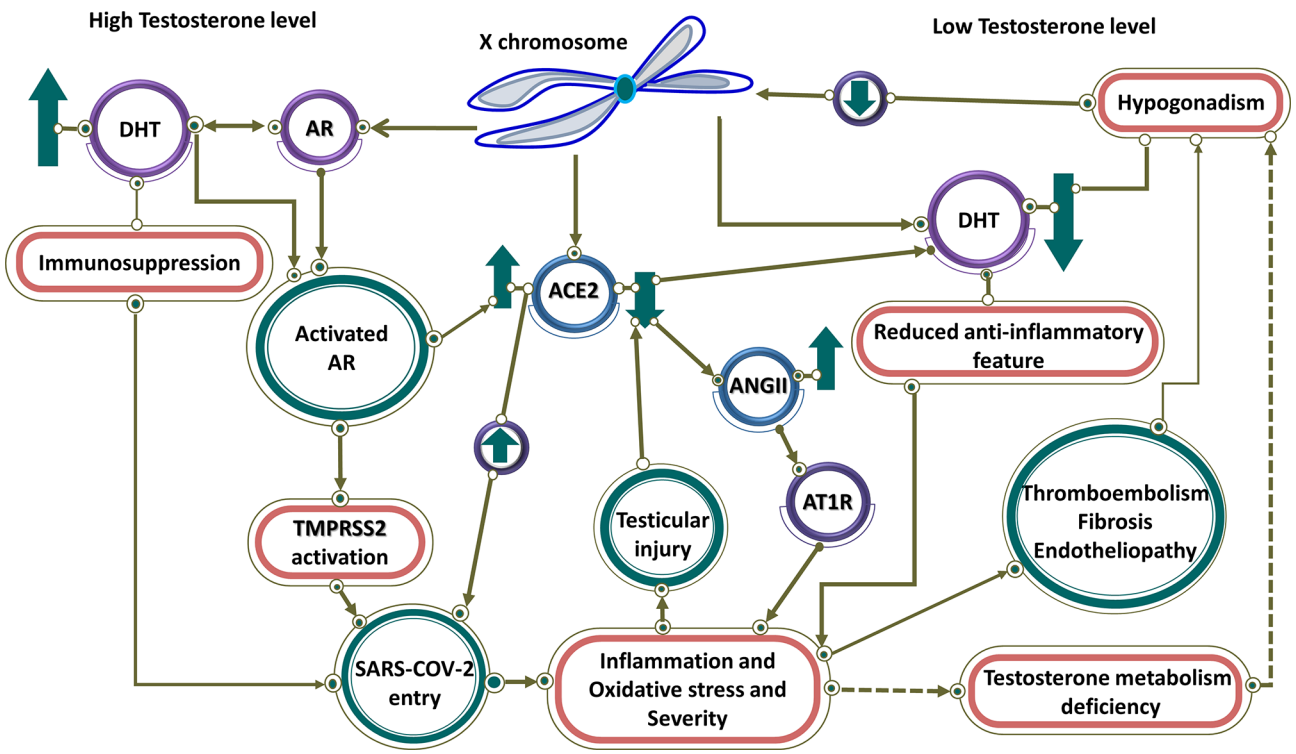


FIGURE 3 | High and low testosterone impact on COVID-19 severity. Controlled by androgen, both TMPRSS2 and ACE2 are regulated by genes on the X chromosome. At high testosterone levels, both receptors are up-regulated, resulting in SARS-CoV-2 entry. Testosterone has an immunosuppressive feature, which leads to more virus entry and disease severity. In contrast, it has linked low testosterone level caused by hypogonadism in men to increased infection of COVID-19. Therefore, in men with hypogonadism, caused by various factors such as reduction of ACE2, testicular damage, and thromboembolism through hyperinflammation caused by the viral infection, the level of testosterone is reduced and the anti-inflammatory effect of testosterone is blocked, resulting in disease severity. Low testosterone results in downregulation of ACE2 that through activation of AR1T by ANG II leads to inflammation and disease severity.

or other gene polymorphisms along the pathway of SARS-CoV-2 pathogenesis may further lead to COVID-19 progression and deterioration. This concept ought to be further explored in future studies (127). However, low testosterone, a characteristic biomarker of aging males with functional hypogonadism, impact on COVID-19 severity through the ACE2 Connection (**Figure 3**) (127). To clarify, in COVID-19, SARS-CoV-2 infection may impact the testes *via* binding to ACE2 expressed in the Sertoli and Leydig cells, provoking infertility and inhibiting testosterone production (128). It emphasized that low testosterone serum level is associated with SARS-CoV-2 infections and COVID-19 severity in critically ill patients through lower immunomodulatory properties of androgen antiviral effects (129). Altogether, low testosterone levels appear to be a major factor for poor prognosis and mortality in COVID-19 male patients. This is likely to be significantly exacerbated in men with co-morbid conditions admitted to the ICU. Further research is needed to prove this concept (127).

By all means, COVID-19 is a complicated multi-system disease that greatly affects the immune system and homeostasis maintenance of the body (**Figure 2**) (130, 131). A serious clinical implication and involvement of aberrated interconnectedness and possibly triggered crosstalks have been observed in COVID-19 pathogenesis. The correlation between the immune system dysfunction and impairment of homeostasis is well-known and has been addressed in different contexts (132). Clinical evidence indicates that hyper-inflammation, and particular forms of vasculopathy, including TMA and intravascular coagulopathy, are frequent features of COVID-19 among severe patients (29). In these cases, an uncontrolled increase of inflammatory cytokines induces vascular hyperpermeability and MODS, leading to failure of cardiac, hepatic, renal systems, and eventually death (1). In fact, a failure to retain hemostasis due to pulmonary injury and MODS creates a critical condition in COVID-19 patients (29, 130). A more complicated situation has been reported in the presence of coagulopathy, which is rather a prothrombotic character with a high chance of VTE among COVID-19 patients in ICUs (30). As a result, the overproduction of inflammatory cytokines and the over-activation of immune cells during SARS-CoV-2 infection promote endothelial dysfunction and vascular permeability. Further investigations are mandatory to explore possible mechanisms behind these aberrated interconnectedness and possibly triggered crosstalks (29, 133).

COAGULOPATHY AND ENDOTHELIAL DYSFUNCTION

Coagulation is a tightly regulated process that involves interactions between numerous blood components called coagulation factors and mechanisms that prevent thrombosis (134, 135). Currently, the so-called 'COVID-19 associated coagulopathy', is considered to be a critical player in the pathophysiology of SARS-CoV-2 infection, especially in its severe form (136–138). Coagulation and inflammation are highly integrated and delicately balanced biological systems with extensive crosstalk to optimize the body's response to any damage and invasion. The impaired interplay of these systems affects both homeostasis and hemostasis, leading to various conditions with varying degrees of excessive inflammation, thrombosis, or bleeding that may lead to tissue damage and MOF (**Figure 4**) (34, 139–141). The hyper-coagulation state in COVID-19 is triggered by the deep and complex inflammatory response to the virus *via* the activation of tissue factors (TFs) on the surface of activated endothelial cells. Bidirectionally, coagulation and inflammation drive an intensifying circle of events by inducing Protease-Activated Receptors (PARs) mediated inflammatory signaling, participation of innate immune pathways, the engaging of the PC–thrombomodulin mechanism as negative regulatory systems, and the playing roles of NETs, and the fibrinolytic system (34, 142). PARs express in many cell types, including immune cells, platelets, endothelial cells, and smooth muscle cells. After proteolytic cleavage by serine proteases such as factor X and thrombi, they activate and induce the production of cytokines and chemokines. PAR-mediated activation of adhesion molecules in endothelial cells results in the production of IL-6 and IL-8 by fibroblasts and monocytes and increased platelet effects *via* a positive feedback loop, amplifying inflammation and procoagulant processes (34, 138, 142).

Mounting evidence shows that coagulopathy events, including immunothrombosis and thromboinflammation, have been observed in COVID-19 (28, 143). Although pieces of literature are controversial and show conflicting findings (144–146), severe COVID-19 patients specially critically ill ones not responding to shock management may show features of systemic hyper-inflammation called Macrophage Activation Syndrome (MAS) or cytokine storm, also known as secondary Hemophagocytic LymphoHistiocytosis (sHLH). It should be noted that the cytokine profile similar to MAS/sHLH has also been observed in COVID-19 patients, especially the increase of IL1β, IL2, IL6, IL17, IL8, TNF and CCL2. Furthermore, like DIC associated with MAS/sHLH, there is evidence that D-dimer levels are elevated in COVID-19 pneumonia which may suggest that the virus-induced hyper-inflammatory pulmonary immunopathology is spreading to the adjacent microcirculation with a broad secondary fibrinolytic activation (**Figure 4**) (144, 147). An increase in D-dimer indicates that COVID-19 patients are in a hypercoagulable state, which can be attributed to the following reasons. First, viral infections are often accompanied by a threatening pro-inflammatory response and inadequate control of the anti-inflammatory response. It can cause endothelial cell dysfunction, leading to excessive thrombin production. Second, the hypoxia found in severe COVID-19 can stimulate thrombosis through both increasing blood viscosity and a hypoxia-inducible transcription factor-dependent signaling pathway. Third, hospitalized severe COVID-19 patients are more likely to have risk factors such as advanced age, underlying diseases, long-term bed rest, and receiving invasive treatment, which are risk factors for hypercoagulability or thrombosis. To point out, the dissection

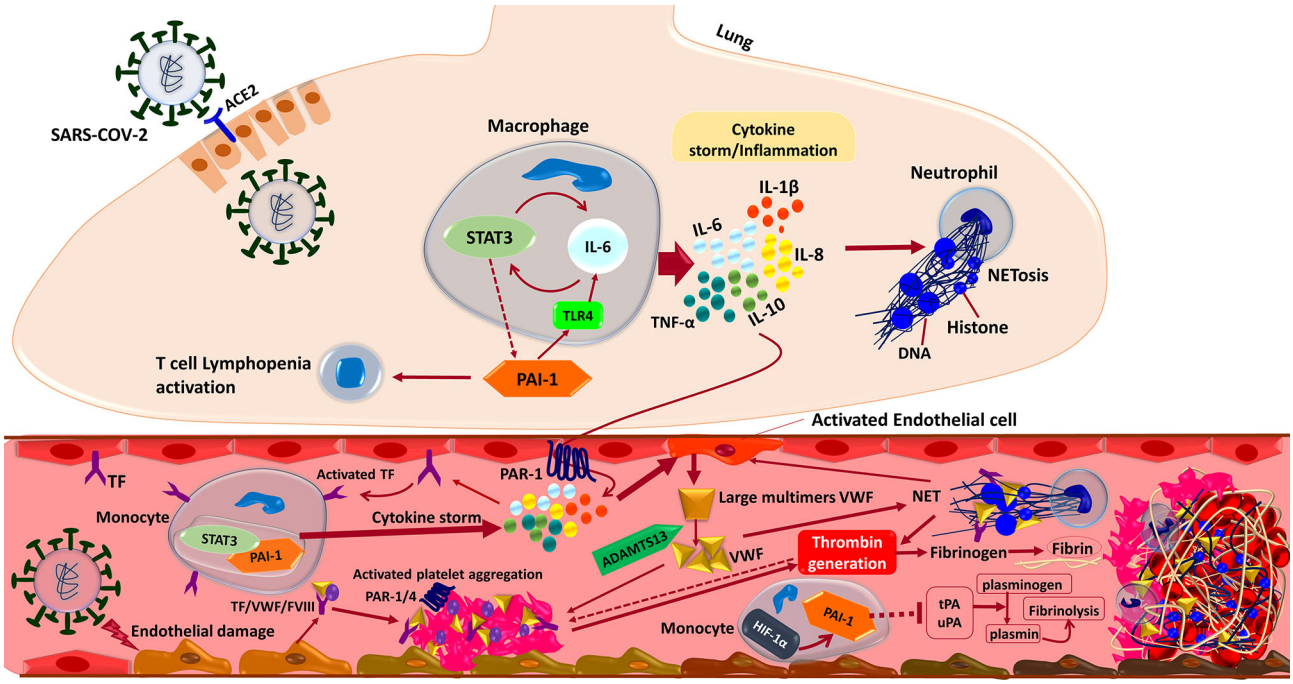


FIGURE 4 | Interactions between the components of the immune system and coagulation cascades that induce thrombosis in COVID-19. In brief, the SARS-CoV-2 infection triggers an inflammatory response in the alveolar lumen by macrophages and neutrophils. Overexpression of PAI-1 and its interaction with TLR4 induce the IL-6 expression and STAT3 activation in the lung of COVID-19 patients. Subsequent production of inflammatory cytokines such as IL-1β, IL-8, IL-6, and TNF-α causes neutrophils to be recruited and release Neutrophil Extracellular Traps (NETs). NETs directly activate factor XII, thereby activating the contact-dependent clotting pathway. The PAR-1 receptor mediates cytokine-dependent Tissue Factor (TF) activation, which stimulates the STAT3/PAI-1 complex in blood's monocytes and enhances inflammatory cytokine production leading to extrinsic coagulation cascade activation and thrombosis formation. Moreover, inflammatory cytokine in combination with the direct binding of SARS-CoV-2 to the endothelial cells triggers Von Willebrand Factor (VWF) secretion cleaved by ADAMTS13 to the regular size. Activated TF/VWF/FVIII complex in cooperation with NETs recruits platelets to adhere to endothelial surfaces and provoke intrinsic coagulation cascades. Altogether these events ultimately driving thrombin formation from circulating prothrombin, which cleaves fibrinogen to fibrin and stimulates thrombosis formation. Besides, suppression of endothelial enzymes and Plasminogen Activators (tPA/uPA) would aggravate the coagulopathy state by preventing effective fibrinolysis. Both tPA and uPA participate in the normal coagulation-plasmin-fibrin pathway, can be inhibited by PAI-1/HIF-1α in severe COVID-19 patients with ARDS.

of the lungs of critically ill patients with COVID-19 shows occlusion of small pulmonary vessels and microthrombosis (148). Fourth, some patients may develop septic coagulopathy or even DIC. Elevated D-dimer is always associated with adverse events. In some clinical patients with SARS-CoV-2 infection, acral gangrene in hypercoagulation state has been described pathophysiologically, indicating that coagulation is the most significant cause of vascular thrombosis and venous gangrene in COVID-19 cases (149). Of note, the serological observations of these COVID-19 patients revealed increased D-dimers associated with antithrombin III deficiency, illustrating wide vascular microthrombi and necrosis in skin biopsy (148, 149). Finally, Similar to COVID-19 non-survivors, DIC is one of the major complications reported in fatal MERS-CoV cases. Markedly, a case of MERS-CoV induced DIC, intracerebral hemorrhage, and MOF appeared two weeks' post-admission in an otherwise stable patient. Furthermore, a fatal case of MERS-CoV was associated with DIC, hyperkalemia, ventricular tachycardia, and cardiac arrest (150–152).

Most compelling evidence shows that endothelial cells can be infected by SARS-CoV-2, leading to endothelialitis. The initiation of inflammation-induced coagulation is mostly mediated by the expression of the TF pathway (CD142). TF is expressed in both mononuclear cells in response to pro-inflammatory cytokines (mainly IL-6) and vascular endothelial cells to promote the conversion of prothrombin into thrombin resulting in fibrin-based blood clots due to the conversion of circulating fibrinogen into fibrin (**Figure 4**) (153–156). To point out, evidence reveals that severe cases of COVID-19 are commonly dependent on a positive feedback loop established between Signal Transducer and Activator of Transcription-3 (STAT3) and Plasminogen Activator Inhibitor-1 (PAI-1) that results in the over-stimulation of the STAT3/PAI-1 signaling network that is shared among diverse disease manifestations and leads to catastrophic consequences (157). To enumerate, PAI-1 is highly expressed in lungs and plasma of COVID-19 patients (158, 159). Furthermore, PAI-1 interacts with TLR4 and triggers the expression of IL-6 that activates STAT3 (157, 158). In turn, STAT3 and PAI-1 augmented thrombosis and coagulopathy in COVID-19 possibly by effectively suppressing urokinase-type Plasminogen Activator (uPA) and tissue-type Plasminogen Activator (tPA) (**Figure 4**) (158–160).

Neutrophils kill and clear invading microorganisms through phagocytosis and degranulation. NETs are a novel antimicrobial strategy of neutrophils that are released from neutrophils into extracellular space to catch and kill invading bacteria to protect the host from infection (161). Regarding the pathophysiology of COVID-19, immunothrombosis and release of NET lead to inflammatory lung/organ damage and thrombosis *via* enhanced NETosis in which neutrophils are more prone to spontaneously form NETs by adopting the low-density phenotype (28, 162–165). Notably, isolated neutrophils from patients with COVID-19 displayed increased NET release at baseline, similar to neutrophils stimulated by phorbol myristate acetate from healthy donors, implying that the environment of the COVID-19 plasma provokes NET formation (**Figure 4**) (28, 165). In the same vein, high levels of NETs detected in the sera of COVID-19 patients can trigger NETs formation *in vitro* in neutrophils collected from healthy volunteers (166, 167). Further, direct stimulation of NETosis by APLAbs of COVID-19 patients (28, 168), as well as the involvement of infectious diseases in the AntiPhospholipid Syndrome (APS) (28, 169), indicates that SARS-CoV-2 infection possibly synergize with APLAbs to provoke the immunothrombotic process. To clarify, Higher APLAbs titers are associated with increased neutrophil and platelet activity and more serious respiratory disease (28, 170). NETs also contribute to acute lung injury by inducing macrophage release of IL-1β, which in turn can reinforce the formation of NET (25, 28, 171). Furthermore, it is known that most of the inflammatory mediators increased in COVID-19 patients regulate the activity of neutrophils through the expression of chemotactic factors (15, 28, 172). Altogether, the evidence indicate that dysregulation of cytokine release might be maintained by NET-mediated crosstalk between neutrophils and macrophages, leading to disordered or disproportionate immunothrombotic status (28).

As a key mediator of hemostasis and plasma glycoprotein, the VWF takes circulating platelets at the vascular injury sites and mediates activation and aggregation of platelets (166, 173). VWF is a multimer that exists in the endothelial Weibel-Palade bodies and releases under the stimulation of cytokines. Proteolytically cleaved by A Disintegrin And Metalloproteinase with ThromboSpondin type 1 motifs, member 13 (ADAMTS13) to regulate the size and activity of VWF multimers, ultra-large VWF multimers are processed into smaller VWF forms to prevent the formation of thrombus (166, 174). To clarify, high molecular weight forms of excessively released VWF may provoke microcirculatory dysfunction through spontaneous binding to platelets resulting in TMA (175, 176). Considering COVID-19, VWF antigen levels (VWF : Ag) and VWF : Ristocetin cofactor activity (VWF : RCo) have significantly increased in moderate and severe cases than in normal controls possibly without change in ADAMTS13 activity (**Figure 4**) (175).

Overall, NETs and VWF/ADAMTS13 axis are essential for thrombosis and inflammation. Given these points, recent evidence shows that the axis is associated with and contributes to the poor prognosis of COVID-19 patients manifested initially by ARDS and then developed to more complex clinical phenotypes, including thrombotic thrombocytopenic purpura-like syndrome, hepatic coagulopathy, MODS, and extensive micro- and macrovascular thrombosis (163, 166, 177). There are increasing global reports considering APLAbs in COVID-19–associated coagulopathy (178). APS is an acquired thrombophilia in which patients develop pathogenic auto-antibodies against phospholipids and phospholipid-binding proteins. Of note, patients in a subtype of APS develop multi-organ thromboses over the shortest periods (Catastrophic APS (CAPS)) like patients with severe COVID-19–associated coagulopathy. Both conditions show acute inflammatory response, cytokine storm, and highly elevated ferritin levels (179, 180). Furthermore, both conditions develop retiform purpura and livedoid rashes representing thrombotic microvascular injury with endothelial damage and cytokine reaction, spread by activation and deposition of complement that predisposes to thrombosis (181). However, thrombocytopenia is not as common as CAPS in COVID-19 (182). Therefore, there are clinical and laboratory similarities between severe COVID-19 and CAPS. However, we need more studies to delineate APL-related mechanisms of thrombosis in COVID-19 (178).

As aforementioned, males and females respond differently to MERS- and SARS-CoV infections in which the conditions are more threatening for males (183–185). Comparatively, there are important differences between men and women for cardiovascular diseases. To point out, men have a higher risk of first and recurrent venous thrombosis than do women (186). Considering COVID-19, different genetic and endocrine mechanisms might influence the mechanisms of coagulopathy and thrombosis in COVID-19 (187). Most compelling evidence shows genetic mechanisms of hemostasis do not differ in men and women. Henceforth, the greater risk of thromboembolism in men, even in the context of COVID-19, does not attribute to any sex-linked difference in genetic predisposition (186, 187). However, there are some pieces of evidence indicating the different role of endocrine mechanisms in regulating hemostasis in males and females. In humans, the plausible general rule is that under normal conditions of exposure to at least normal sex hormone levels, estrogens play a possible positive role and androgens play a negative role in hemostasis (187). Of note, testosterone deficiency in men correlated well with an increase in procoagulant factors and a decrease in anticoagulant factors contributes to the higher risk of thromboembolism in men than in women at any age including a greater frequency in men during elderly and a decreased frequency in women during fertile periods (188–192). To point out, the estrogen deficiency possibly explains why the incidence of thrombosis observed in the elderly is higher than in young women (187, 193). Given these points, a growing body of evidence seems to imply that deficiencies of sex hormones are a detrimental factor in the immune and inflammatory response or predisposition to thrombosis, upon which they present an unfavorable prognostic factor in the severity and outcome of COVID-19 (187, 194). To clarify, evidence imply that primary

hypogonadism in males and secondary hypogonadism in both sexes may have a direct or indirect role in the development of systemic inflammation, endotheliopathy, and thromboembolism in patients with COVID-19 (129, 187, 194, 195). Altogether, pieces of evidence imply that the severity of COVID-19 deteriorates in older ages for both sexes, in men more than in women, possibly because of inflammaging and endothelial dysfunction with aging. However, it is not quite clear if sex hormones have different effects on immunity, inflammation, and thrombotic status, and if different reductions in sex hormones in both sexes affect the severity and outcome of COVID-19 (183–185, 187, 196, 197).

Besides susceptibility to age-dependent diseases, epidemiological studies indicated sexdifferences in incidence and Case Fatality Rates (CFRs) after SARS-CoV infection in humans in which males experience higher CFRs than females (183, 184). Similarly, data from recent MERS-CoV outbreaks showed high incidence and CFRs among men (185).

CONCLUSION

COVID-19, especially in its severe form, is a multisystem syndrome with sex disparities in severity and outcome characterized by a state of immunothrombosis in which a complex interaction of SARS-CoV-2 virus invasion with platelets, leukocytes, endothelial cells, immune response, and the possible involvement of megakaryocytes resulting in endotheliopathy, coagulopathy and thromboinflammation that culminate in more severe complications and mortality. With this in mind, we may claim that the capability of interconnected biological levels of the organization is aberrated significantly by SARS-CoV-2 infection during COVID-19 progression, resulting in pathologically and fatally enhanced levels of crosstalks. A recent systematic review and meta-analysis of over 50 long-term effects of COVID-19 show that fatigue, headache, attention disorder, hair loss, and dyspnea are the five most common symptoms (198). Correspondingly, a more recent study (199) revealed that clotting markers elevated in patients with long COVID syndrome while inflammation markers had returned to normal, showing that the clotting system may be involved in the root cause of long COVID syndrome that helps to explain persistent symptoms such as reduced physical fitness and fatigue. To clarify, we propose that a deeper understanding of the coagulation and inflammation in acute COVID-19 may also be relevant for a better understanding of certain cases of long COVID.

AUTHOR CONTRIBUTIONS

All authors wrote the first draft of the manuscript. MEB and MN critically reviewed and edited the manuscript. ME and NA created figures. MEB supervised all aspects of the work. All authors contributed to the article and approved the submitted version.

FUNDING

This project is supported in part by the Shahrekord university grant to MEB (grant number 1399).

ACKNOWLEDGMENTS

We would like to thank all individuals who cooperated in this study.

REFERENCES

1. Tay MZ, Poh CM, Rénia L, MacAry PA, Ng LFP. The Trinity of COVID-19: Immunity, Inflammation and Intervention. *Nat Rev Immunol* (2020) 20 (6):363–74. doi: 10.1038/s41577-020-0311-8
2. Giovanetti M, Benedetti F, Campisi G, Ciccozzi A, Fabris S, Ceccarelli G, et al. Evolution Patterns of SARS-CoV-2: Snapshot on Its Genome Variants. *Biochem Biophys Res Commun* (2021) 538:88–91. doi: 10.1016/j.bbrc.2020.10.102
3. Haney S, Bardwell L, Nie Q. Ultrasensitive Responses and Specificity in Cell Signaling. *BMC Syst Biol* (2010) 4:119. doi: 10.1186/1752-0509-4-119
4. Rowland MA, Greenbaum JM, Deeds EJ. Crosstalk and the Evolvability of Intracellular Communication. *Nat Commun* (2017) 8(1):1–8. doi: 10.1038/ncomms16009
5. Lin L, Lu L, Cao W, Li T. Hypothesis for Potential Pathogenesis of SARS-CoV-2 Infection–A Review of Immune Changes in Patients With Viral Pneumonia. *Emerg Microbes Infect* (2020) 9(1):727–32. doi: 10.1080/22221751.2020.1746199
6. Hoffmann M, Kleine-Weber H, Schroeder S, Krüger N, Herrler T, Erichsen S, et al. SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell* (2020) 181(2):271–80.e8. doi: 10.1016/j.cell.2020.02.052
7. Zhou P, Yang XL, Wang XG, Hu B, Zhang L, Zhang W, et al. A Pneumonia Outbreak Associated With a New Coronavirus of Probable Bat Origin. *Nature* (2020) 579(7798):270–3. doi: 10.1038/s41586-020-2012-7
8. Amraei R, Rahimi N. COVID-19, Renin-Angiotensin System and Endothelial Dysfunction. *Cells* (2020) 9(7):1652. doi: 10.3390/cells9071652
9. Wulandari L, Hamidah B, Pakpahan C, Damayanti NS, Kurniati ND, Adiatmaja CO, et al. Initial Study on TMPRSS2 P.Val160Met Genetic Variant in COVID-19 Patients. *Hum Genomics* (2021) 15(1):1–9. doi: 10.1186/s40246-021-00330-7
10. Dhanachandra Singh K, Karnik SS. Angiotensin Receptors: Structure, Function, Signaling and Clinical Applications. *J Cell Signal* (2017) 01 (02):1–2. doi: 10.4172/2576-1471.1000111
11. Ni W, Yang X, Yang D, Bao J, Li R, Xiao Y, et al. Role of Angiotensin-Converting Enzyme 2 (ACE2) in COVID-19. *Crit Care* (2020) 24(1):1–10. doi: 10.1186/s13054-020-03120-0
12. Li W, Moore MJ, Vasllieva N, Sui J, Wong SK, Berne MA, et al. Angiotensin-Converting Enzyme 2 Is a Functional Receptor for the SARS Coronavirus. *Nature* (2003) 426(6965):450–4. doi: 10.1038/nature02145
13. Shukla AK, Banerjee M. Angiotensin-Converting-Enzyme 2 and Renin-Angiotensin System Inhibitors in COVID-19: An Update. *High Blood Press Cardiovasc Prev* (2021) 28(2):129–39. doi: 10.1007/s40292-021-00439-9
14. Wiese OJ, Allwood BW, Zemlin AE. COVID-19 and the Renin-Angiotensin System (RAS): A Spark That Sets the Forest Alight? *Med Hypotheses* (2020) 144:110231. doi: 10.1016/j.mehy.2020.110231
15. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical Features of Patients Infected With 2019 Novel Coronavirus in Wuhan, China. *Lancet* (2020) 395(10223):497–506. doi: 10.1016/S0140-6736(20)30183-5

16. Wu Z, McGoogan JM. Characteristics of and Important Lessons From the Coronavirus Disease 2019 (COVID-19) Outbreak in China: Summary of a Report of 72314 Cases From the Chinese Center for Disease Control and Prevention. *JAMA - J Am Med Assoc* (2020) 323(13):1239–42. doi: 10.1001/jama.2020.2648
17. Rodriguez-Morales AJ, Cardona-Ospina JA, Gutiérrez-Ocampo E, Villamizar-Peña R, Holguin-Rivera Y, Escalera-Antezana JP, et al. Clinical, Laboratory and Imaging Features of COVID-19: A Systematic Review and Meta-Analysis. *Travel Med Infect Dis* (2020) 34:101623. doi: 10.1016/j.tmaid.2020.101623
18. Yang W, Cao Q, Qin L, Wang X, Cheng Z, Pan A, et al. Clinical Characteristics and Imaging Manifestations of the 2019 Novel Coronavirus Disease (COVID-19):A Multi-Center Study in Wenzhou City, Zhejiang, China. *J Infect* (2020) 80(4):388–93. doi: 10.1016/j.jinf.2020.02.016
19. Im Kampe EO, Lehfeld AS, Buda S, Buchholz U, Haas W. Surveillance of COVID-19 School Outbreaks, Germany, March to August 2020. *Eurosurveillance* (2020) 25(38):2001645. doi: 10.2807/1560-7917.ES.2020.25.38.2001645
20. Miesbach W, Makris M. COVID-19: Coagulopathy, Risk of Thrombosis, and the Rationale for Anticoagulation. *Clin Appl Thromb* (2020) 26:1–7. doi: 10.1177/1076029620938149
21. Farshbafnadi M, Kamali Zonouzi S, Sabahi M, Dolatshahi M, Aarabi MH. Aging & COVID-19 Susceptibility, Disease Severity, and Clinical Outcomes: The Role of Entangled Risk Factors. *Exp Gerontol* (2021) 154:111507. doi: 10.1016/j.exger.2021.111507
22. Kindler E, Thiel V, Weber F. Interaction of SARS and MERS Coronaviruses With the Antiviral Interferon Response. *Adv Virus Res* (2016) 96:219–43. doi: 10.1016/bs.aivir.2016.08.006
23. Prompetchara E, Ketloy C, Palaga T. Immune Responses in COVID-19 and Potential Vaccines: Lessons Learned From SARS and MERS Epidemic. *Asian Pac J Allergy Immunol* (2020) 38(1):1–9. doi: 10.12932/AP-200220-0772
24. Larenas-Linnemann D, Rodríguez-Pérez N, Arias-Cruz A, Blandón-Vijil MV, Del Río-Navarro BE, Estrada-Cardona A, et al. Enhancing Innate Immunity Against Virus in Times of COVID-19: Trying to Untangle Facts From Fictions. *World Allergy Organ J* (2020) 13(11):100476. doi: 10.1016/j.waojou.2020.100476
25. Channappanavar R, Perlman S. Pathogenic Human Coronavirus Infections: Causes and Consequences of Cytokine Storm and Immunopathology. *Semin Immunopathol* (2017) 39(5):529–39. doi: 10.1007/s00281-017-0629-x
26. Alipoor SD, Mortaz E, Jamaati H, Tabarsi P, Bayram H, Varahram M, et al. COVID-19: Molecular and Cellular Response. *Front Cell Infect Microbiol* (2021) 11:563085. doi: 10.3389/fcimb.2021.563085
27. Lazzaroni MG, Piantoni S, Masneri S, Garrafa E, Martini G, Tincani A, et al. Coagulation Dysfunction in COVID-19: The Interplay Between Inflammation, Viral Infection and the Coagulation System. *Blood Rev* (2021) 46:100745. doi: 10.1016/j.blre.2020.100745
28. Bonaventura A, Vecchié A, Dagna L, Martinod K, Dixon DL, Van Tassell BW, et al. Endothelial Dysfunction and Immunothrombosis as Key Pathogenic Mechanisms in COVID-19. *Nat Rev Immunol* (2021) 21(5):319–29. doi: 10.1038/s41577-021-00536-9
29. Henry BM, Vikse J, Benoit S, Favaloro EJ, Lippi G. Hyperinflammation and Derangement of Renin-Angiotensin-Aldosterone System in COVID-19: A Novel Hypothesis for Clinically Suspected Hypercoagulopathy and Microvascular Immunothrombosis. *Clin Chim Acta* (2020) 507:167–73. doi: 10.1016/j.cca.2020.04.027
30. Kollias A, Kyriakoulis KG, Dimakakos E, Poulakou G, Stergiou GS, Syrigos K. Thromboembolic Risk and Anticoagulant Therapy in COVID-19 Patients: Emerging Evidence and Call for Action. *Br J Haematol* (2020) 189(5):846–7. doi: 10.1111/bjh.16727
31. Zhang Y, Xiao M, Zhang S, Xia P, Cao W, Jiang W, et al. Coagulopathy and Antiphospholipid Antibodies in Patients With Covid-19. *N Engl J Med* (2020) 382(17):e38. doi: 10.1056/NEJMc2007575
32. Farghaly S, Makboul M. Correlation Between Age, Sex, and Severity of Coronavirus Disease-19 Based on Chest Computed Tomography Severity Scoring System. *Egypt J Radiol Nucl Med* (2021) 52(1):23. doi: 10.1186/s43055-021-00408-1
33. Gadi N, Wu SC, Spihlman AP, Moulton VR. What's Sex Got to Do With COVID-19? Gender-Based Differences in the Host Immune Response to Coronaviruses. *Front Immunol* (2020) 11:2147. doi: 10.3389/fimmu.2020.02147
34. Foley JH, Conway EM. Cross Talk Pathways Between Coagulation and Inflammation. *Circ Res* (2016) 118(9):1392–408. doi: 10.1161/CIRCRESAHA.116.306853
35. Kohansal Vajari M, Shirin M, Pourbagheri-Sigaroodi A, Akbari ME, Abolghasemi H, Bashash D. COVID-19-Related Coagulopathy: A Review of Pathophysiology and Pharmaceutical Management. *Cell Biol Int* (2021) 45(9):1832–50. doi: 10.1002/cbin.11623
36. Burrell LM, Johnston CI, Tikellis C, Cooper ME. ACE2, A New Regulator of the Renin-Angiotensin System. *Trends Endocrinol Metab* (2004) 15(4):166–9. doi: 10.1016/j.tem.2004.03.001
37. Hamming I, Timens W, Bulthuis MLC, Lely AT, Navis GJ, van Goor H. Tissue Distribution of ACE2 Protein, the Functional Receptor for SARS Coronavirus. A First Step in Understanding SARS Pathogenesis. *J Pathol* (2004) 203(2):631–7. doi: 10.1002/path.1570
38. Letko M, Marzi A, Munster V. Functional Assessment of Cell Entry and Receptor Usage for SARS-CoV-2 and Other Lineage B Betacoronaviruses. *Nat Microbiol* (2020) 5(4):562–9. doi: 10.1038/s41564-020-0688-y
39. Yan R, Zhang Y, Li Y, Xia L, Guo Y, Zhou Q. Structural Basis for the Recognition of SARS-CoV-2 by Full-Length Human ACE2. *Science (80-)* (2020) 367(6485):1444–8. doi: 10.1126/science.abb2762
40. Santos RAS, Oudit GY, Verano-Braga T, Canta G, Steckelings UM, Bader M. The Renin-Angiotensin System: Going Beyond the Classical Paradigms. *Am J Physiol - Hear Circ Physiol* (2019) 316(5):H958–70. doi: 10.1152/ajpheart.00723.2018
41. Underwood PC, Adler GK. The Renin Angiotensin Aldosterone System and Insulin Resistance in Humans. *Curr Hypertens Rep* (2013) 15(1):59–70. doi: 10.1007/s11906-012-0323-2
42. Kuba K, Imai Y, Rao S, Jiang C, Penninger JM. Lessons From SARS: Control of Acute Lung Failure by the SARS Receptor ACE2. *J Mol Med* (2006) 84(10):814–20. doi: 10.1007/s00109-006-0094-9
43. Zou L, Ruan F, Huang M, Liang L, Huang H, Hong Z, et al. SARS-CoV-2 Viral Load in Upper Respiratory Specimens of Infected Patients. *N Engl J Med* (2020) 382(12):1177–9. doi: 10.1056/NEJMc2001737
44. Han SH, Mallampalli RK. The Role of Surfactant in Lung Disease and Host Defense Against Pulmonary Infections. *Ann Am Thorac Soc* (2015) 12(5):765–74. doi: 10.1513/AnnalsATS.201411-507FR
45. Xu Z, Shi L, Wang Y, Zhang J, Huang L, Zhang C, et al. Pathological Findings of COVID-19 Associated With Acute Respiratory Distress Syndrome. *Lancet Respir Med* (2020) 8(4):420–2. doi: 10.1016/S2213-2600(20)30076-X
46. Gralinski LE, Baric RS. Molecular Pathology of Emerging Coronavirus Infections. *J Pathol* (2015) 235(2):185–95. doi: 10.1002/path.4454
47. Parit R, Jayavel S. Association of ACE Inhibitors and Angiotensin Type II Blockers With ACE2 Overexpression in COVID-19 Comorbidities: A Pathway-Based Analytical Study. *Eur J Pharmacol* (2021) 896:173899. doi: 10.1016/j.ejphar.2021.173899
48. Ferrario CM, Jessup J, Chappell MC, Averill DB, Brosnihan KB, Tallant EA, et al. Effect of Angiotensin-Converting Enzyme Inhibition and Angiotensin II Receptor Blockers on Cardiac Angiotensin-Converting Enzyme 2. *Circulation* (2005) 111(20):2605–10. doi: 10.1161/CIRCULATIONAHA.104.510461
49. Pedrosa MA, Valenzuela R, Garrido-Gil P, Labandeira CM, Navarro G, Franco R, et al. Experimental Data Using Candesartan and Captopril Indicate No Double-Edged Sword Effect in COVID-19. *Clin Sci* (2021) 135(3):465–81. doi: 10.1042/CS20201511
50. Wang K, Gheblawi M, Oudit GY. Angiotensin Converting Enzyme 2: A Double-Edged Sword. *Circulation* (2020) 142(5):426–8. doi: 10.1161/CIRCULATIONAHA.120.047049
51. Onweni CL, Zhang YS, Caulfield T, Hopkins CE, Fairweather DL, Freeman WD. ACEI/ARB Therapy in COVID-19: The Double-Edged Sword of ACE2 and SARS-CoV-2 Viral Docking. *Crit Care* (2020) 24(1):475. doi: 10.1186/s13054-020-03195-9
52. Garvin MR, Alvarez C, Miller JI, Prates ET, Walker AM, Amos BK, et al. A Mechanistic Model and Therapeutic Interventions for Covid-19 Involving a Ras-Mediated Bradykinin Storm. *Elife* (2020) 9:1–16. doi: 10.7554/eLife.59177

53. Simoneaux R, Shafer SL. A RAS and Bradykinin-Mediated Mechanism for COVID-19. *ASA Monit* (2020) 84(11):1–11. doi: 10.1097/01.ASM.0000722064.35978.f6
54. Liu F, Li L, Xu M, Wu J, Luo D, Zhu YS, et al. Prognostic Value of Interleukin-6, C-Reactive Protein, and Procalcitonin in Patients With COVID-19. *J Clin Virol* (2020) 127:104370. doi: 10.1016/j.jcv.2020.104370
55. Hanff TC, Harhay MO, Brown TS, Cohen JB, Mohareb AM. Is There an Association Between COVID-19 Mortality and the Renin-Angiotensin System? A Call for Epidemiologic Investigations. *Clin Infect Dis* (2020) 71(15):870–4. doi: 10.1093/cid/ciaa329
56. Banu N, Panikar SS, Leal LR, Leal AR. Protective Role of ACE2 and Its Downregulation in SARS-CoV-2 Infection Leading to Macrophage Activation Syndrome: Therapeutic Implications. *Life Sci* (2020) 256:117905. doi: 10.1016/j.lfs.2020.117905
57. Foresta C, Rocca MS, Di Nisio A. Gender Susceptibility to COVID-19: A Review of the Putative Role of Sex Hormones and X Chromosome. *J Endocrinol Invest* (2021) 44(5):951–6. doi: 10.1007/s40618-020-01383-6
58. Chumakov K, Avidan MS, Benn CS, Bertozzi SM, Blatt L, Chang AY, et al. Old Vaccines for New Infections: Exploiting Innate Immunity to Control COVID-19 and Prevent Future Pandemics. *Proc Natl Acad Sci USA* (2021) 118(21):e2101718118. doi: 10.1073/pnas.2101718118
59. Marshall JS, Warrington R, Watson W, Kim HL. An Introduction to Immunology and Immunopathology. *Allergy Asthma Clin Immunol* (2018) 14(2):1–10. doi: 10.1186/s13223-018-0278-1
60. Amarante-Mendes GP, Adjemian S, Branco LM, Zanetti LC, Weinlich R, Bortoluci KR. Pattern Recognition Receptors and the Host Cell Death Molecular Machinery. *Front Immunol* (2018) 9:2379. doi: 10.3389/fimmu.2018.02379
61. Suresh R, Mosser DM. Pattern Recognition Receptors in Innate Immunity, Host Defense, and Immunopathology. *Am J Physiol - Adv Physiol Educ* (2013) 37(4):284–91. doi: 10.1152/advan.00058.2013
62. Jensen S, Thomsen AR. Sensing of RNA Viruses: A Review of Innate Immune Receptors Involved in Recognizing RNA Virus Invasion. *J Virol* (2012) 86(6):2900–10. doi: 10.1128/JVI.05738-11
63. Siu KL, Kok KH, Ng MHJ, Poon VKM, Yuen KY, Zheng BJ, et al. Severe Acute Respiratory Syndrome Coronavirus M Protein Inhibits Type I Interferon Production by Impeding Theformation of TRAF3·TANK·Tbk1/Ikkε Complex. *J Biol Chem* (2009) 284(24):16202–9. doi: 10.1074/jbc.M109.008227
64. Aboudounya MM, Heads RJ. COVID-19 and Toll-Like Receptor 4 (TLR4): SARS-CoV-2 May Bind and Activate TLR4 to Increase ACE2 Expression, Facilitating Entry and Causing Hyperinflammation. *Mediators Inflamm* (2021) 2021:8874339. doi: 10.1155/2021/8874339
65. Li SW, Wang CY, Jou YJ, Huang SH, Hsiao LH, Wan L, et al. SARS Coronavirus Papain-Like Protease Inhibits the TLR7 Signaling Pathway Through Removing Lys63-Linked Polyubiquitination of TRAF3 and TRAF6. *Int J Mol Sci* (2016) 17(5):678. doi: 10.3390/ijms17050678
66. Taefehshokr N, Taefehshokr S, Hemmat N, Heit B. Covid-19: Perspectives on Innate Immune Evasion. *Front Immunol* (2020) 11:580641. doi: 10.3389/fimmu.2020.580641
67. Knoops K, Kikkert M, Van Den Worm SHE, Zevenhoven-Dobbe JC, van der Meer Y, Koster AJ, et al. SARS-Coronavirus Replication Is Supported by a Reticulovesicular Network of Modified Endoplasmic Reticulum. *PloS Biol* (2008) 6(9):1957–74. doi: 10.1371/journal.pbio.0060226
68. Blanco-Melo D, Nilsson-Payant BE, Liu WC, Uhl S, Hoagland D, Møller R, et al. Imbalanced Host Response to SARS-CoV-2 Drives Development of COVID-19. *Cell* (2020) 181(5):1036–45.e9. doi: 10.1016/j.cell.2020.04.026
69. Hadjadj J, Yatim N, Barnabei L, Corneau A, Boussier J, Smith N, et al. Impaired Type I Interferon Activity and Inflammatory Responses in Severe COVID-19 Patients. *Science (80-)* (2020) 369(6504):718–24. doi: 10.1126/science.abc6027
70. Zhou W, Wang W. Auto-Antibodies Against Type I IFNs Are Associated With Severe COVID-19 Pneumonia. *Signal Transduct Target Ther* (2021) 6(1):1–2. doi: 10.1038/s41392-021-00514-6
71. Onabajo OO, Banday AR, Stanifer ML, Yan W, Obajemu A, Santer DM, et al. Interferons and Viruses Induce a Novel Truncated ACE2 Isoform and Not the Full-Length SARS-CoV-2 Receptor. *Nat Genet* (2020) 52(12):1283–93. doi: 10.1038/s41588-020-00731-9
72. Yamada T, Sato S, Sotoyama Y, Orba Y, Sawa H, Yamauchi H, et al. RIG-I Triggers a Signaling-Abortive Anti-SARS-CoV-2 Defense in Human Lung Cells. *Nat Immunol* (2021) 22(7):820–8. doi: 10.1038/s41590-021-00942-0
73. Sinha P, Matthay MA, Calfee CS. Is a "Cytokine Storm" Relevant to COVID-19? *JAMA Intern Med* (2020) 180(9):1152–4. doi: 10.1001/jamainternmed.2020.3313
74. Siu KL, Chan CP, Kok KH, Chiu-Yat Woo P, Jin DY. Suppression of Innate Antiviral Response by Severe Acute Respiratory Syndrome Coronavirus M Protein Is Mediated Through the First Transmembrane Domain. *Cell Mol Immunol* (2014) 11(2):141–9. doi: 10.1038/cmi.2013.61
75. The REMAP-CAP Investigators, Derde LPG. Effectiveness of Tocilizumab, Sarilumab, and Anakinra for Critically Ill Patients With COVID-19 The REMAP-CAP COVID-19 Immune Modulation Therapy Domain Randomized Clinical Trial. *medRxiv* (2021) 2021.06.18.21259133. doi: 10.1101/2021.06.18.21259133
76. Narasaraju T, Tang BM, Herrmann M, Muller S, Chow VTK, Radic M. Neutrophilia and NETopathy as Key Pathologic Drivers of Progressive Lung Impairment in Patients With COVID-19. *Front Pharmacol* (2020) 11:870. doi: 10.3389/fphar.2020.00870
77. Chauhan AJ, Wiffen LJ, Brown TP. COVID-19: A Collision of Complement, Coagulation and Inflammatory Pathways. *J Thromb Haemost* (2020) 18(9):2110–7. doi: 10.1111/jth.14981
78. Vabret N, Britton GJ, Gruber C, Hegde S, Kim J, Kuksin M, et al. Immunology of COVID-19: Current State of the Science. *Immunity* (2020) 52(6):910–41. doi: 10.1016/j.immuni.2020.05.002
79. Ruan Q, Yang K, Wang W, Jiang L, Song J. Clinical Predictors of Mortality Due to COVID-19 Based on an Analysis of Data of 150 Patients From Wuhan, China. *Intensive Care Med* (2020) 46(5):846–8. doi: 10.1007/s00134-020-05991-x
80. Gao Y, Li T, Han M, Li X, Wu D, Xu Y, et al. Diagnostic Utility of Clinical Laboratory Data Determinations for Patients With the Severe COVID-19. *J Med Virol* (2020) 92(7):791–6. doi: 10.1002/jmv.25770
81. Qin C, Zhou L, Hu Z, Zhang S, Yang S, Tao Y, et al. Dysregulation of Immune Response in Patients With Coronavirus 2019 (COVID-19) in Wuhan, China. *Clin Infect Dis* (2020) 71(15):762–8. doi: 10.1093/cid/ciaa248
82. Lucas C, Wong P, Klein J, Castro TBR, Silva J, Sundaram M, et al. Longitudinal Analyses Reveal Immunological Misfiring in Severe COVID-19. *Nature* (2020) 584(7821):463–9. doi: 10.1038/s41586-020-2588-y
83. Chen X, Zhao B, Qu Y, Chen Y, Xiong J, Feng Y, et al. Detectable Serum Severe Acute Respiratory Syndrome Coronavirus 2 Viral Load (RNAemia) Is Closely Correlated With Drastically Elevated Interleukin 6 Level in Critically Ill Patients With Coronavirus Disease 2019. *Clin Infect Dis* (2020) 71(8):1937–42. doi: 10.1093/cid/ciaa449
84. Zhang X, Wu K, Wang D, Yue X, Song D, Zhu Y, et al. Nucleocapsid Protein of SARS-CoV Activates Interleukin-6 Expression Through Cellular Transcription Factor NF-κb. *Virology* (2007) 365(2):324–35. doi: 10.1016/j.virol.2007.04.009
85. Merza M, Hartman H, Rahman M, Hwaiz R, Zhang E, Renström E, et al. Neutrophil Extracellular Traps Induce Trypsin Activation, Inflammation, and Tissue Damage in Mice With Severe Acute Pancreatitis. *Gastroenterology* (2015) 149(7):1920–31.e8. doi: 10.1053/j.gastro.2015.08.026
86. Mehta P, McAuley DF, Brown M, Sanchez E, Tattersall RS, Manson JJ. COVID-19: Consider Cytokine Storm Syndromes and Immunosuppression. *Lancet* (2020) 395(10229):1033–4. doi: 10.1016/S0140-6736(20)30628-0
87. Barnes BJ, Adrover JM, Baxter-Stoltzfus A, Borczuk A, Cools-Lartigue J, Crawford JM, et al. Targeting Potential Drivers of COVID-19: Neutrophil Extracellular Traps. *J Exp Med* (2020) 217(6):e20200652. doi: 10.1084/jem.20200652
88. Mangalmurti N, Hunter CA. Cytokine Storms: Understanding COVID-19. *Immunity* (2020) 53(1):19–25. doi: 10.1016/j.immuni.2020.06.017
89. Hosseini A, Hashemi V, Shomali N, Asghari F, Gharibi T, Akbari M, et al. Innate and Adaptive Immune Responses Against Coronavirus. *BioMed Pharmacother* (2020) 132:110859. doi: 10.1016/j.biopha.2020.110859
90. Nile SH, Nile A, Qiu J, Li L, Jia X, Kai G. COVID-19: Pathogenesis, Cytokine Storm and Therapeutic Potential of Interferons. *Cytokine Growth Factor Rev* (2020) 53:66–70. doi: 10.1016/j.cytogfr.2020.05.002

91. Hasan A, Al-Ozairi E, Al-Baqsumi Z, Ahmad R, Al-Mulla F. Cellular and Humoral Immune Responses in Covid-19 and Immunotherapeutic Approaches. *ImmunoTargets Ther* (2021) 10:63–85. doi: 10.2147/ITT.S280706
92. Christiaansen A, Varga SM, Spencer JV. Viral Manipulation of the Host Immune Response. *Curr Opin Immunol* (2015) 36:54–60. doi: 10.1016/j.coi.2015.06.012
93. Channappanavar R, Zhao J, Perlman S. T Cell-Mediated Immune Response to Respiratory Coronaviruses. *Immunol Res* (2014) 59(1–3):118–28. doi: 10.1007/s12026-014-8534-z
94. Cerutti G, Guo Y, Zhou T, Gorman J, Lee M, Rapp M, et al. Potent SARS-CoV-2 Neutralizing Antibodies Directed Against Spike N-Terminal Domain Target a Single Supersite. *Cell Host Microbe* (2021) 29(5):819–33.e7. doi: 10.1016/j.chom.2021.03.005
95. Fu Y, Cheng Y, Wu Y. Understanding SARS-CoV-2-Mediated Inflammatory Responses: From Mechanisms to Potential Therapeutic Tools. *Virol Sin* (2020) 35(3):266–71. doi: 10.1007/s12250-020-00207-4
96. Chang SE, Feng A, Meng W, Apostolidis SA, Mack E, Artandi M, et al. New-Onset IgG Autoantibodies in Hospitalized Patients With COVID-19. *Nat Commun* (2021) 12(1):1–15. doi: 10.1038/s41467-021-25509-3
97. Bastard P, Gervais A, Le Voyer T, Rosain J, Philippot Q, Manry J, et al. Autoantibodies Neutralizing Type I IFNs Are Present in ~4% of Uninfected Individuals Over 70 Years Old and Account for ~20% of COVID-19 Deaths. *Sci Immunol* (2021) 6(62):eabl4340. doi: 10.1126/sciimmunol.abl4340
98. Wang EY, Mao T, Klein J, Dai Y, Huck JD, Jaycox JR, et al. Diverse Functional Autoantibodies in Patients With COVID-19. *Nature* (2021) 595 (7866):283–8. doi: 10.1038/s41586-021-03631-y
99. Taefehshokr N, Taefehshokr S, Heit B. Mechanisms of Dysregulated Humoral and Cellular Immunity by SARS-CoV-2. *Pathogens* (2020) 9 (12):1–21. doi: 10.3390/pathogens9121027
100. Zhao J, Zhao J, Perlman S. T Cell Responses Are Required for Protection From Clinical Disease and for Virus Clearance in Severe Acute Respiratory Syndrome Coronavirus-Infected Mice. *J Virol* (2010) 84(18):9318–25. doi: 10.1128/JVI.01049-10
101. Chen Z, John Wherry E. T Cell Responses in Patients With COVID-19. *Nat Rev Immunol* (2020) 20(9):529–36. doi: 10.1038/s41577-020-0402-6
102. Kalfaoglu B, Almeida-Santos J, Tye CA, Satou Y, Ono M. T-Cell Dysregulation in COVID-19. *Biochem Biophys Res Commun* (2021) 538:204–10. doi: 10.1016/j.bbrc.2020.10.079
103. Kuri-Cervantes L, Pampena MB, Meng W, Rosenfeld AM, Ittner CAG, Weisman AR, et al. Comprehensive Mapping of Immune Perturbations Associated With Severe COVID-19. *Sci Immunol* (2020) 5(49):eabd7114. doi: 10.1126/sciimmunol.abd7114
104. Giamarellos-Bourboulis EJ, Netea MG, Rovina N, Akinosoglou K, Antoniadou A, Antonakos N, et al. Complex Immune Dysregulation in COVID-19 Patients With Severe Respiratory Failure. *Cell Host Microbe* (2020) 27(6):992–1000.e3. doi: 10.1016/j.chom.2020.04.009
105. Mazzoni A, Salvati L, Maggi L, Capone M, Vanni A, Spinicci M, et al. Impaired Immune Cell Cytotoxicity in Severe COVID-19 Is IL-6 Dependent. *J Clin Invest* (2020) 130(9):4694–703. doi: 10.1172/JCI138554
106. Liao M, Liu Y, Yuan J, Wen Y, Xu G, Zhao J, et al. Single-Cell Landscape of Bronchoalveolar Immune Cells in Patients With COVID-19. *Nat Med* (2020) 26(6):842–4. doi: 10.1038/s41591-020-0901-9
107. Wichmann D, Sperhake JP, Lütgehetmann M, Steurer S, Edler C, Heinemann A, et al. Autopsy Findings and Venous Thromboembolism in Patients With COVID-19: A Prospective Cohort Study. *Ann Intern Med* (2020) 173(4):268–77. doi: 10.7326/M20-2003
108. Shrotri M, van Schalkwyk MCI, Post N, Eddy D, Huntley C, Leeman D, et al. T Cell Response to SARS-Cov-2 Infection in Humans: A Systematic Review. *PloS One* (2021) 16(1 January):e0245532. doi: 10.1371/journal.pone.0245532
109. Böttcher JP, Schanz O, Garbers C, Zaremba A, Hegenbarth S, Kurts C, et al. IL-6 Trans-Signaling-Dependent Rapid Development of Cytotoxic CD8+ T Cell Function. *Cell Rep* (2014) 8(5):1318–27. doi: 10.1016/j.celrep.2014.07.008
110. Taga K, Tosato G. IL-10 Inhibits Human T Cell Proliferation and IL-2 Production. *J Immunol (Baltimore Md : 1950)* (1992) 148:1143–8.
111. Waal Malefyt R, Haanen J, Spits H, Koncarolo MG, Te Velde A, Figdor C, et al. Interleukin 10 (Il-10) and Viral Il-10 Strongly Reduce Antigen-Specific Human T Cell Proliferation by Diminishing the Antigen-Presenting Capacity of Monocytes *via* Dowm'egulation of Class H Major Histocompatibility Complex Expression. *J Exp Med* (1991) 174(4):915–24. doi: 10.1084/jem.174.4.915
112. Liu Y, Du X, Chen J, Jin Y, Peng L, Wang HHX, et al. Neutrophil-To-Lymphocyte Ratio as an Independent Risk Factor for Mortality in Hospitalized Patients With COVID-19. *J Infect* (2020) 81(1):e6–12. doi: 10.1016/j.jinf.2020.04.002
113. Ma Y, Shi N, Fan Y, Wang J, Zhao C, Li G, et al. Predictive Value of the Neutrophil-To-Lymphocyte Ratio(NLR) for Diagnosis and Worse Clinical Course of the COVID-19: Findings From Ten Provinces in China. *SSRN Electron J* (2020). doi: 10.2139/ssrn.3569838
114. Adamo S, Chevrier S, Cervia C, Zurbuchen Y, Raeber ME, Yang L, et al. Lymphopenia-Induced T Cell Proliferation Is a Hallmark of Severe COVID-19. *bioRxiv* (2020) 2020.08.04.236521. doi: 10.1101/2020.08.04.236521
115. Vardhana SA, Wolchok JD. The Many Faces of the Anti-COVID Immune Response. *J Exp Med* (2020) 217(6):e20200678. doi: 10.1084/jem.20200678
116. Zhang Y, Zhang J, Chen Y, Luo B, Yuan Y, Huang F, et al. The ORF8 Protein of SARS-CoV-2 Mediates Immune Evasion Through Potently Downregulating MHC-I. *bioRxiv* (2020) 2020.05.24.111823. doi: 10.1101/2020.05.24.111823v1
117. Diao B, Wang C, Tan Y, Chen X, Liu Y, Ning L, et al. Reduction and Functional Exhaustion of T Cells in Patients With Coronavirus Disease 2019 (COVID-19). *Front Immunol* (2020) 11:827. doi: 10.3389/fimmu.2020.00827
118. Yang X, Dai T, Zhou X, Qian H, Guo R, Lei L, et al. Analysis of Adaptive Immune Cell Populations and Phenotypes in the Patients Infected by SARS-CoV-2. *medRxiv* (2020) 2020.03.23.20040675. doi: 10.1101/2020.03.23.20040675v2
119. Rha MS, Jeong HW, Ko JH, Choi SJ, Seo IH, Lee JS, et al. PD-1-Expressing SARS-CoV-2-Specific CD8+ T Cells Are Not Exhausted, But Functional in Patients With COVID-19. *Immunity* (2021) 54(1):44–52.e3. doi: 10.1016/j.immuni.2020.12.002
120. Zheng M, Gao Y, Wang G, Song G, Liu S, Sun D, et al. Functional Exhaustion of Antiviral Lymphocytes in COVID-19 Patients. *Cell Mol Immunol* (2020) 17(5):533–5. doi: 10.1038/s41423-020-0402-2
121. Oja AE, Saris A, Ghandour CA, Kragten NAM, Hogema BM, Nossent EJ, et al. Divergent SARS-CoV-2-Specific T and B Cell Responses in Severe But Not Mild COVID-19. *bioRxiv* (2020) 2020.06.18.159202. doi: 10.1101/2020.06.18.159202v1
122. Bongen E, Lucian H, Khatri A, Fragiadakis GK, Bjornson ZB, Nolan GP, et al. Sex Differences in the Blood Transcriptome Identify Robust Changes in Immune Cell Proportions With Aging and Influenza Infection. *Cell Rep* (2019) 29(7):1961–73.e4. doi: 10.1016/j.celrep.2019.10.019
123. Takahashi T, Ellingson MK, Wong P, Israelow B, Lucas C, Klein J, et al. Sex Differences in Immune Responses That Underlie COVID-19 Disease Outcomes. *Nature* (2020) 588(7837):315–20. doi: 10.1038/s41586-020-2700-3
124. Viveiros A, Rasmuson J, Vu J, Mulvagh SL, Yip CYY, Norris CM, et al. Sex Differences in COVID-19: Candidate Pathways, Genetics of ACE2, and Sex Hormones. *Am J Physiol - Hear Circ Physiol* (2021) 320(1):H296–304. doi: 10.1152/ajpheart.00755.2020
125. Pinna G. Sex and COVID-19: A Protective Role for Reproductive Steroids. *Trends Endocrinol Metab* (2021) 32(1):3–6. doi: 10.1016/j.tem.2020.11.004
126. Chang C, Saltzman A, Yeh S, Young W, Keller E, Lee HJ, et al. Androgen Receptor: An Overview. *Crit Rev Eukaryot Gene Expr* (1995) 5(2):97–125. doi: 10.1615/CritRevEukarGeneExpr.v5.i2.10
127. Younis JS, Skorecki K, Abassi Z. The Double Edge Sword of Testosterone's Role in the COVID-19 Pandemic. *Front Endocrinol (Lausanne)* (2021) 12:607179. doi: 10.3389/fendo.2021.607179
128. Abobaker A, Raba AA. Does COVID-19 Affect Male Fertility? *World J Urol* (2021) 39(3):975–6. doi: 10.1007/s00345-020-03208-w
129. Schroeder M, Schaumburg B, Mueller Z, Parplys A, Jarczak D, Nierhaus A, et al. The Majority of Male Patients With COVID-19 Present Low Testosterone Levels on Admission to Intensive Care in Hamburg, Germany: A Retrospective Cohort Study. *medRxiv* (2020) 2020.05.07.20073817.
130. Wang J, Saguner AM, An J, Ning Y, Yan Y, Li G. Dysfunctional Coagulation in COVID-19: From Cell to Bedside. *Adv Ther* (2020) 37(7):3033–9. doi: 10.1007/s12325-020-01399-7

131. Terpos E, Ntanasis-Stathopoulos I, Elalamy I, Kastritis E, Sergentanis TN, Politou M, et al. Hematological Findings and Complications of COVID-19. *Am J Hematol* (2020) 95(7):834–47. doi: 10.1002/ajh.25829
132. Levi M, van der Poll T. Inflammation and Coagulation. *Crit Care Med* (2010) 38(SUPPL 2):S26–34. doi: 10.1097/CCM.0b013e3181c98d21
133. Smargiassi A, Soldati G, Borghetti A, Scoppettuolo G, Tamburrini E, Testa AC, et al. Lung Ultrasonography for Early Management of Patients With Respiratory Symptoms During COVID-19 Pandemic. *J Ultrasound* (2020) 23 (4):449–56. doi: 10.1007/s40477-020-00501-7
134. Palta S, Saroa R, Palta A. Overview of the Coagulation System. *Indian J Anaesth* (2014) 58(5):515–23. doi: 10.4103/0019-5049.144643
135. Ezihe-Ejiofor JA, Hutchinson N. Anticlotting Mechanisms 1: Physiology and Pathology. *Contin Educ Anaesth Crit Care Pain* (2013) 13(3):87–92. doi: 10.1093/bjaceaccp/mks061
136. Loo J, Spittle DA, Newnham M. COVID-19, Immunothrombosis and Venous Thromboembolism: Biological Mechanisms. *Thorax* (2021) 76 (4):412–20. doi: 10.1136/thoraxjnl-2020-216243
137. Lorini FL, Di Matteo M, Gritti P, Grazioli L, Benigni A, Zacchetti L, et al. Coagulopathy and COVID-19. *Eur Hear J Suppl* (2021) 23(Supplement_E): E95–8. doi: 10.1093/eurheartj/suab100
138. Frantzeskaki F, Armaganidis A, Orfanos SE. Immunothrombosis in Acute Respiratory Distress Syndrome: Cross Talks Between Inflammation and Coagulation. *Respiration* (2017) 93(3):212–25. doi: 10.1159/000453002
139. Engelmann B, Massberg S. Thrombosis as an Intravascular Effector of Innate Immunity. *Nat Rev Immunol* (2013) 13(1):34–45. doi: 10.1038/nri3345
140. Schulz C, Engelmann B, Massberg S. Crossroads of Coagulation and Innate Immunity: The Case of Deep Vein Thrombosis. *J Thromb Haemost* (2013) 11 (SUPPL.1):233–41. doi: 10.1111/jth.12261
141. Newton K, Dixit VM. Signaling in Innate Immunity and Inflammation. *Cold Spring Harb Perspect Biol* (2012) 4(3):a006049. doi: 10.1101/cshperspect.a006049
142. Stark K, Massberg S. Interplay Between Inflammation and Thrombosis in Cardiovascular Pathology. *Nat Rev Cardiol* (2021) 18(9):666–82. doi: 10.1038/s41569-021-00552-1
143. Jayarangaiah A, Kariyanna PT, Chen X, Jayarangaiah A, Kumar A. COVID-19-Associated Coagulopathy: An Exacerbated Immunothrombosis Response. *Clin Appl Thromb* (2020) 26:1076029620943293. doi: 10.1177/1076029620943293
144. McGonagle D, Sharif K, O'Regan A, Bridgewood C. The Role of Cytokines Including Interleukin-6 in COVID-19 Induced Pneumonia and Macrophage Activation Syndrome-Like Disease. *Autoimmun Rev* (2020) 19(6):102537. doi: 10.1016/j.autrev.2020.102537
145. Schnaubelt S, Tihanyi D, Strassl R, Schmidt R, Anders S, Laggner AN, et al. Hemophagocytic Lymphohistiocytosis in COVID-19: Case Reports of a Stepwise Approach. *Med (Baltimore)* (2021) 100(12):e25170. doi: 10.1097/MD.0000000000025170
146. Lorenz G, Moog P, Bachmann Q, La Rosée P, Schneider H, Schlegl M, et al. Title: Cytokine Release Syndrome Is Not Usually Caused by Secondary Hemophagocytic Lymphohistiocytosis in a Cohort of 19 Critically Ill COVID-19 Patients. *Sci Rep* (2020) 10(1):1–11. doi: 10.1038/s41598-020-75260-w
147. Wan S, Yi Q, Fan S, Lv J, Zhang X, Guo L, et al. Characteristics of Lymphocyte Subsets and Cytokines in Peripheral Blood of 123 Hospitalized Patients With 2019 Novel Coronavirus Pneumonia (NCP). *medRxiv* (2020) 2020.02.10.20021832. doi: 10.1101/2020.02.10.20021832v1
148. Bamgboje A, Hong J, Mushiyev S, Pekler G. A 61-Year-Old Man With Sars-Cov-2 Infection and Venous Thrombosis Presenting With Painful Swelling and Gangrene of the Lower Limb Consistent With Phlegmasia Cerulea Dolens. *Am J Case Rep* (2020) 21:1–6. doi: 10.12659/AJCR.928342
149. Wang JS, Pasieka HB, Petronic-Rosic V, Sharif-Askary B, Evans KK. Digital Gangrene as a Sign of Catastrophic Coronavirus Disease 2019-Related Microangiopathy. *Plast Reconstr Surg - Glob Open* (2020) 8(7):e3025. doi: 10.1097/GOX.0000000000003025
150. Zhang L, Yan X, Fan Q, Liu H, Liu X, Liu Z, et al. D-Dimer Levels on Admission to Predict in-Hospital Mortality in Patients With Covid-19. *J Thromb Haemost* (2020) 18(6):1324–9. doi: 10.1111/jth.14859
151. Giannis D, Ziogas IA, Gianni P. Coagulation Disorders in Coronavirus Infected Patients: COVID-19, SARS-CoV-1, MERS-CoV and Lessons From the Past. *J Clin Virol* (2020) 127:104362. doi: 10.1016/j.jcv.2020.104362
152. Iba T, Levi M, Levy JH. Sepsis-Induced Coagulopathy and Disseminated Intravascular Coagulation. *Semin Thromb Hemost* (2020) 46(1):89–95. doi: 10.1055/s-0039-1694995
153. Simmons J, Pittet JF. The Coagulopathy of Acute Sepsis. *Curr Opin Anaesthesiol* (2015) 28(2):227–36. doi: 10.1097/ACO.0000000000000163
154. Iba T, Levy J, Raj A, Warkentin T. Advance in the Management of Sepsis-Induced Coagulopathy and Disseminated Intravascular Coagulation. *J Clin Med* (2019) 8(5):728. doi: 10.3390/jcm8050728
155. Van Der Poll T, Van De Veerdonk FL, Scicluna BP, Netea MG. The Immunopathology of Sepsis and Potential Therapeutic Targets. *Nat Rev Immunol* (2017) 17(7):407–20. doi: 10.1038/nri.2017.36
156. Merad M, Martin JC. Pathological Inflammation in Patients With COVID-19: A Key Role for Monocytes and Macrophages. *Nat Rev Immunol* (2020) 20 (6):355–62. doi: 10.1038/s41577-020-0331-4
157. Matsuyama T, Kubli SP, Yoshinaga SK, Pfeffer K, Mak TW. An Aberrant STAT Pathway Is Central to COVID-19. *Cell Death Differ* (2020) 27 (12):3209–25. doi: 10.1038/s41418-020-00633-7
158. Jafarzadeh A, Nemati M, Jafarzadeh S. Contribution of STAT3 to the Pathogenesis of COVID-19. *Microb Pathog* (2021) 154:104836. doi: 10.1016/j.micpath.2021.104836
159. Zuo Y, Warnock M, Harbaugh A, Yalavarthi S, Gockman K, Zuo M, et al. Plasma Tissue Plasminogen Activator and Plasminogen Activator Inhibitor-1 in Hospitalized COVID-19 Patients. *Sci Rep* (2021) 11 (1):2020.08.29.20184358. doi: 10.1101/2020.08.29.20184358v4
160. Cimmino G, Cirillo P. Tissue Factor: Newer Concepts in Thrombosis and Its Role Beyond Thrombosis and Hemostasis. *Cardiovasc Diagn Ther* (2018) 8 (5):581–93. doi: 10.21037/cdt.2018.10.14
161. Brinkmann V, Reichard U, Goosmann C, Fauler B, Uhlemann Y, Weiss DS, et al. Neutrophil Extracellular Traps Kill Bacteria. *Science (80-)* (2004) 303 (5663):1532–5. doi: 10.1126/science.1092385
162. Leppkes M, Knopf J, Naschberger E, Lindemann A, Singh J, Herrmann I, et al. Vascular Occlusion by Neutrophil Extracellular Traps in COVID-19. *EBioMedicine* (2020) 58:102925. doi: 10.1016/j.ebiom.2020.102925
163. Fernández-Pérez MP, Águila S, Reguilón-Gallego L, los Reyes-García AM, Miñano A, Bravo-Pérez C, et al. Neutrophil Extracellular Traps and Von Willebrand Factor Are Allies That Negatively Influence COVID-19 Outcomes. *Clin Transl Med* (2021) 11(1):e268. doi: 10.1002/ctm2.268
164. Goshua G, Pine AB, Meizlish ML, Chang CH, Zhang H, Bahel P, et al. Endotheliopathy in COVID-19-Associated Coagulopathy: Evidence From a Single-Centre, Cross-Sectional Study. *Lancet Haematol* (2020) 7(8):e575–82. doi: 10.1016/S2352-3026(20)30216-7
165. Middleton EA, He XY, Denorme F, Campbell RA, Ng D, Salvatore SP, et al. Neutrophil Extracellular Traps Contribute to Immunothrombosis in COVID-19 Acute Respiratory Distress Syndrome. *Blood* (2020) 136 (10):1169–79. doi: 10.1182/blood.2020007008
166. Yang J, Wu Z, Long Q, Huang J, Hong T, Liu W, et al. Insights Into Immunothrombosis: The Interplay Among Neutrophil Extracellular Trap, Von Willebrand Factor, and ADAMTS13. *Front Immunol* (2020) 11:610696. doi: 10.3389/fimmu.2020.610696
167. Zuo Y, Yalavarthi S, Shi H, Gockman K, Zuo M, Madison JA, et al. Neutrophil Extracellular Traps in COVID-19. *JCI Insight* (2020) 5(11): e138999. doi: 10.1172/jci.insight.138999
168. Yalavarthi S, Gould TJ, Rao AN, Mazza LF, Morris AE, Núñez-Álvarez C, et al. Release of Neutrophil Extracellular Traps by Neutrophils Stimulated With Antiphospholipid Antibodies: A Newly Identified Mechanism of Thrombosis in the Antiphospholipid Syndrome. *Arthritis Rheumatol* (2015) 67(11):2990–3003. doi: 10.1002/art.39247
169. Mendoza-Pinto C, García-Carrasco M, Cervera R. Role of Infectious Diseases in the Antiphospholipid Syndrome (Including Its Catastrophic Variant). *Curr Rheumatol Rep* (2018) 20(10):62. doi: 10.1007/s11926-018-0773-x
170. Zuo Y, Estes SK, Ali RA, Gandhi AA, Yalavarthi S, Shi H, et al. Prothrombotic Autoantibodies in Serum From Patients Hospitalized With COVID-19. *Sci Transl Med* (2020) 12(570):3876. doi: 10.1126/scitranslmed.abd3876
171. Chousterman BG, Swirski FK, Weber GF. Cytokine Storm and Sepsis Disease Pathogenesis. *Semin Immunopathol* (2017) 39(5):517–28. doi: 10.1007/s00281-017-0639-8

172. Bösmüller H, Traxler S, Bitzer M, Häberle H, Raiser W, Nann D, et al. The Evolution of Pulmonary Pathology in Fatal COVID-19 Disease: An Autopsy Study With Clinical Correlation. *Virchows Arch* (2020) 477(3):349–57. doi: 10.1007/s00428-020-02881-x
173. Löf A, Müller JP, Brehm MA. A Biophysical View on Von Willebrand Factor Activation. *J Cell Physiol* (2018) 233(2):799–810. doi: 10.1002/jcp.25887
174. South K, Lane DA. ADAMTS-13 and Von Willebrand Factor: A Dynamic Duo. *J Thromb Haemost* (2018) 16(1):6–18. doi: 10.1111/jth.13898
175. Francischetti IMB, Toomer K, Zhang Y, Jani J, Siddiqui Z, Brotman DJ, et al. Upregulation of Pulmonary Tissue Factor, Loss of Thrombomodulin and Immunothrombosis in SARS-CoV-2 Infection. *EClinicalMedicine* (2021) 39:101069. doi: 10.1016/j.eclinm.2021.101069
176. Masias C, Cataland SR. The Role of ADAMTS13 Testing in the Diagnosis and Management of Thrombotic Microangiopathies and Thrombosis. *Blood* (2018) 132(9):903–10. doi: 10.1182/blood-2018-02-791533
177. Chang JC. Disseminated Intravascular Coagulation: New Identity as Endotheliopathy-Associated Vascular Microthrombotic Disease Based on In Vivo Hemostasis and Endothelial Molecular Pathogenesis. *Thromb J* (2020) 18(1):1–21. doi: 10.1186/s12959-020-00231-0
178. Gkrouzman E, Barbhaiya M, Erkan D, Lockshin MD. Reality Check on Antiphospholipid Antibodies in COVID-19–Associated Coagulopathy. *Arthritis Rheumatol* (2021) 73(1):173–4. doi: 10.1002/art.41472
179. Rosário C, Zandman-Goddard G, Meyron-Holtz EG, D'Cruz DP, Shoenfeld Y. The Hyperferritinemic Syndrome: Macrophage Activation Syndrome, Still's Disease, Septic Shock and Catastrophic Antiphospholipid Syndrome. *BMC Med* (2013) 11(1):185. doi: 10.1186/1741-7015-11-185
180. Henderson LA, Canna SW, Schulert GS, Volpi S, Lee PY, Kernan KF, et al. On the Alert for Cytokine Storm: Immunopathology in COVID-19. *Arthritis Rheumatol* (2020) 72(7):1059–63. doi: 10.1002/art.41285
181. Magro C, Mulvey JJ, Berlin D, Nuovo G, Salvatore S, Harp J, et al. Complement Associated Microvascular Injury and Thrombosis in the Pathogenesis of Severe COVID-19 Infection: A Report of Five Cases. *Transl Res* (2020) 220:1–13. doi: 10.1016/j.trsl.2020.04.007
182. Connors JM, Levy JH. COVID-19 and Its Implications for Thrombosis and Anticoagulation. *Blood* (2020) 135(23):2033–40. doi: 10.1182/blood.2020006000
183. Karlberg J, Chong DSY, Lai WYY. Do Men Have a Higher Case Fatality Rate of Severe Acute Respiratory Syndrome Than Women Do? *Am J Epidemiol* (2004) 159(3):229–31. doi: 10.1093/aje/kwh056
184. *Leong: SARS in Singapore-Predictors of Disease Severity - Google Scholar*. Available at: https://scholar.google.com/scholar_lookup?title=SARSinSingapore—predictorsofdiseaseseverity&author=HNLeong&author=AEarnest&author=HHLim&publication_year=2006&journal=AnnAcadMedSingapore&volume=35&pages=326-31.
185. Alghamdi IG, Hussain II, Almalki SS, Alghamdi MS, Alghamdi MM, El-Sheemy MA. The Pattern of Middle East Respiratory Syndrome Coronavirus in Saudi Arabia: A Descriptive Epidemiological Analysis of Data From the Saudi Ministry of Health. *Int J Gen Med* (2014) 7:417–23. doi: 10.2147/IJGM.S67061
186. Roach REJ, Cannegieter SC, Lijfering WM. Differential Risks in Men and Women for First and Recurrent Venous Thrombosis: The Role of Genes and Environment. *J Thromb Haemost* (2014) 12(10):1593–600. doi: 10.1111/jth.12678
187. Pivonello R, Auriemma RS, Pivonello C, Isidori AM, Corona G, Colao A, et al. Sex Disparities in COVID-19 Severity and Outcome: Are Men Weaker or Women Stronger? *Neuroendocrinology* (2021) 111(11):1066–85. doi: 10.1159/000513346
188. Erem C, Kocak M, Hacihasanoglu A, Yilmaz M. Blood Coagulation and Fibrinolysis in Male Patient With Hypogonadotropic Hypogonadism: Plasma Factor V and Factor X Activities Increase in Hypogonadotropic Hypogonadism. *J Endocrinol Invest* (2008) 31(6):537–41. doi: 10.1007/BF03346404
189. Roy-O'Reilly M, McCullough LD. Sex Differences in Stroke: The Contribution of Coagulation. *Exp Neurol* (2014) 259:16–27. doi: 10.1016/j.expneurol.2014.02.011
190. Næss IA, Christiansen SC, Romundstad P, Cannegieter SC, Rosendaal FR, Hammerstrøm J. Incidence and Mortality of Venous Thrombosis: A Population-Based Study. *J Thromb Haemost* (2007) 5(4):692–9. doi: 10.1111/j.1538-7836.2007.02450.x
191. Kyrle PA, Minar E, Bialonczyk C, Hirschl M, Weltermann A, Eichinger S. The Risk of Recurrent Venous Thromboembolism in Men and Women. *N Engl J Med* (2004) 350(25):2558–63. doi: 10.1056/NEJMoa032959
192. Salzano A, Demelo-Rodriguez P, Marra AM, Proietti M. A Focused Review of Gender Differences in Antithrombotic Therapy. *Curr Med Chem* (2017) 24(24):2576–88. doi: 10.2174/0929867323666161029223512
193. Previtali E, Bucciarelli P, Passamonti SM, Martinelli I. Risk Factors for Venous and Arterial Thrombosis. *Blood Transfus* (2011) 9(2):120–38. doi: 10.2450/2010.0066-10
194. Rastrelli G, Di Stasi V, Inglese F, Beccaria M, Garuti M, Di Costanzo D, et al. Low Testosterone Levels Predict Clinical Adverse Outcomes in SARS-CoV-2 Pneumonia Patients. *Andrology* (2021) 9(1):88–98. doi: 10.1111/andr.12821
195. Wang Z, Xu X. scRNA-Seq Profiling of Human Testes Reveals the Presence of the ACE2 Receptor, A Target for SARS-CoV-2 Infection in Spermatogonia, Leydig and Sertoli Cells. *Cells* (2020) 9(4):3474. doi: 10.3390/cells9040920
196. Gemmati D, Bramanti B, Serino ML, Secchiero P, Zauli G, Tisato V. COVID-19 and Individual Genetic Susceptibility/Receptivity: Role of ACE1/ACE2 Genes, Immunity, Inflammation and Coagulation. Might the Double X-Chromosome in Females be Protective Against SARS-COV-2 Compared to the Single X-Chromosome in Males? *Int J Mol Sci* (2020) 21(10):3474. doi: 10.3390/ijms21103474
197. Tzoran I, Hoffman R, Monreal M. Hemostasis and Thrombosis in the Oldest Old. *Semin Thromb Hemost* (2018) 44(7):624–31. doi: 10.1055/s-0038-1657779
198. Lopez-Leon S, Wegman-Ostrosky T, Perelman C, Sepulveda R, Rebolledo PA, Cuapio A, et al. More Than 50 Long-Term Effects of COVID-19: A Systematic Review and Meta-Analysis. *Sci Rep* (2021) 11(1):1–12. doi: 10.1038/s41598-021-95565-8
199. Fogarty H, Townsend L, Morrin H, Ahmad A, Comerford C, Karampini E, et al. Persistent Endotheliopathy in the Pathogenesis of Long COVID Syndrome. *J Thromb Haemost* (2021) 19(10):2546–53. doi: 10.1111/jth.15490

Conflict of Interest: MN was employed by Erythrogen Medical Genetics Lab.

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: 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.

frontiers in Immunology

ORIGINAL RESEARCH
published: 12 January 2022
doi: 10.3389/fimmu.2021.742736

Integrative Lipidomics and Metabolomics for System-Level Understanding of the Metabolic Syndrome in Long-Term Treated HIV-Infected Individuals

Sofie Olund Villumsen[1], Rui Benfeitas[2], Andreas Dehlbæk Knudsen[1], Marco Gelpi[1], Julie Høgh[1], Magda Teresa Thomsen[1], Daniel Murray[3], Henrik Ullum[4], Ujjwal Neogi[5,6] and Susanne Dam Nielsen[1]*

[1] Department of Infectious Diseases, Copenhagen University Hospital Rigshospitalet, Copenhagen, Denmark, [2] National Bioinformatics Infrastructure Sweden (NBIS), Science for Life Laboratory, Department of Biochemistry and Biophysics, Stockholm University, Stockholm, Sweden, [3] Personalized Medicine of Infectious Complications in Immune Deficiency (PERSIMUNE), Rigshospitalet, Copenhagen, Denmark, [4] Department of Clinical Immunology, Copenhagen University Hospital, Copenhagen, Denmark, [5] The Systems Virology Lab, Division of Clinical Microbiology, Department of Laboratory Medicine, Karolinska Institute, ANA Futura, Stockholm, Sweden, [6] Manipal Institute of Virology (MIV), Manipal Academy of Higher Education, Manipal, India

OPEN ACCESS

Edited by:
Aridaman Pandit, University Medical Center Utrecht, Netherlands

Reviewed by:
Raul German Spallanzani, Harvard Medical School, United States
Rutan Zhang, University of Washington, United States

***Correspondence:**
Ujjwal Neogi
ujjwal.neogi@ki.se

Specialty section:
This article was submitted to Systems Immunology, a section of the journal Frontiers in Immunology

Received: *16 July 2021*
Accepted: *22 December 2021*
Published: *12 January 2022*

Citation:
Olund Villumsen S, Benfeitas R, Knudsen AD, Gelpi M, Høgh J, Thomsen MT, Murray D, Ullum H, Neogi U and Nielsen SD (2022) Integrative Lipidomics and Metabolomics for System-Level Understanding of the Metabolic Syndrome in Long-Term Treated HIV-Infected Individuals. Front. Immunol. 12:742736. doi: 10.3389/fimmu.2021.742736

People living with HIV (PLWH) require life-long anti-retroviral treatment and often present with comorbidities such as metabolic syndrome (MetS). Systematic lipidomic characterization and its association with the metabolism are currently missing. We included 100 PLWH with MetS and 100 without MetS from the Copenhagen Comorbidity in HIV Infection (COCOMO) cohort to examine whether and how lipidome profiles are associated with MetS in PLWH. We combined several standard biostatistical, machine learning, and network analysis techniques to investigate the lipidome systematically and comprehensively and its association with clinical parameters. Additionally, we generated weighted lipid-metabolite networks to understand the relationship between lipidomic profiles with those metabolites associated with MetS in PLWH. The lipidomic dataset consisted of 917 lipid species including 602 glycerolipids, 228 glycerophospholipids, 61 sphingolipids, and 26 steroids. With a consensus approach using four different statistical and machine learning methods, we observed 13 differentially abundant lipids between PLWH without MetS and PLWH with MetS, which mainly belongs to diacylglyceride (DAG, n = 2) and triacylglyceride (TAG, n = 11). The comprehensive network integration of the lipidomics and metabolomics data suggested interactions between specific glycerolipids' structural composition patterns and key metabolites involved in glutamate metabolism. Further integration of the clinical data with metabolomics and lipidomics resulted in the association of visceral adipose tissue (VAT) and exposure to earlier generations of antiretroviral therapy (ART). Our integrative omics data indicated disruption of glutamate and fatty acid metabolism, suggesting their involvement in the pathogenesis of PLWH with MetS. Alterations in the lipid homeostasis and glutaminolysis need clinical interventions to prevent accelerated aging in PLWH with MetS.

Keywords: HIV-1, metabolic syndrome, antiretroviral treatment, machine learning, lipidomics

INTRODUCTION

Combination antiretroviral therapy (cART) has increased life expectancy for people living with HIV (PLWH). However, an increase in incidences of comorbidities such as obesity, type 2 diabetes (T2D), and cardiovascular disease (CVD) related to metabolic syndrome (MetS) (i.e., abdominal obesity, hypertension, elevated levels of triglycerides, dyslipidemia, and altered glucose levels), has become a growing concern in successfully treated PLWH. In chronic HIV infection, complicated interactions between effects of persistent low-grade immune activation, metabolic toxicity from cART, and non-HIV-related risk factors may increase the risk of MetS in PLWH. However, the pathophysiology of MetS in PLWH is still incompletely understood (1–3).

cART is known to be associated with changes in fat distribution (i.e., lipodystrophy and dyslipidemia) and metabolic abnormalities due to side effects (4). A few studies have investigated the association of HIV infection and cART with metabolic abnormalities related to MetS (4, 5). These studies have focused on conventional blood lipids, such as triglyceride and total cholesterol. These biomarkers may not sufficiently reflect the complex alterations of the lipid metabolism in PLWH with MetS.

Plasma lipidomics studies in the general population have identified several lipid species within the lipidome to be associated with features of MetS (6). In particular, HIV infection has been recently described by our group to be independently associated with a 1.5 fold increased risk of MetS compared to the general population (5). In addition, obesity has been shown to increase the content of almost all detectable diacylglyceride (DAG) and triacylglyceride (TAG) lipid species, along with several cholesteryl esters (CE), phosphatidylcholine (PC), phosphatidylethanolamine (PE), and lysophosphatidylcholine (LPC) in a general population (7). The pathophysiology and alterations of the lipidome in PLWH with MetS are yet to be explored and may help in the discovery of new patterns and disease markers associated with MetS in PLWH (3). In prior work from our group, we identified key metabolites, which influenced and altered the metabolome of PLWH with MetS (8).

In this study, an exploratory analysis of the lipidome comparing PLWH without MetS and PLWH with MetS was conducted to identify a set of key lipids that define the mechanism of the lipid abnormalities of MetS in the context of HIV infection. Also, we have performed advanced network analysis that revealed deeper underlying patterns within the metabolome (i.e., the polar metabolome and lipidome) of PLWH with MetS. Additionally, we investigated the influence of clinical demographic parameters on integrative metabolomics and lipidomics to provide snapshots of the biological phenotypes linked with MetS in PLWH.

MATERIALS AND METHODS

Study Designing, and Patients

We obtained data from the Copenhagen Comorbidity in HIV Infection (COCOMO) study (9), an ongoing non-interventional, observational, longitudinal cohort study to assess the burden of non-AIDS comorbidities in PLWH. Sample collections and quantifications of the COCOMO cohort have previously been described (5, 9). Of the 1099 participants in the COCOMO study, 100 PLWH ≥ 40 years old were included and matched according to age, sex, duration of cART, smoking status, and current CD4+ T-cells count to 100 PLWH without MetS (5, 9). The MetS was defined according to the International Diabetes Federation (IDF) consensus worldwide definition of the MetS as previously (8, 10). For each individual, we collected clinical data from the COCOMO database with the following 13 HIV and MetS specific variables. The MetS, sex, age, ethnicity, immunodeficiency (i.e., lowest CD4+ T-cell count < 200 cells/μl or previous AIDS condition), exposure to early-generation antiretroviral therapy (ART) (i.e., medicated with thymidine analogs, didanosine and/or indinavir), visceral adipose tissue (VAT) [cm2], subcutaneous adipose tissue (SAT) [cm2], and ART drugs including the active agents; nucleotide reverse transcriptase inhibitors (NRTIs), non-nucleotide reverse transcriptase inhibitors (NNRTIs), protease inhibitors (PIs), integrase strand transfer inhibitors (INSTIs), and other/unknown active agents). Furthermore, a lipidomics dataset (see below) and a metabolomics dataset with 11 key metabolites [i.e., 1-carboxyethylisoleucine, 4-cholesten-3-one, 4-hydroxyglutamate, α-ketoglutarate, carotene diol(2), γ-glutamylglutamate, glutamate, glycerate, isoleucine, pimeloylcarnitine/3-methyladipoylcarnitine (C7-DC) (PC/3-MAPC), and palmitoyl-sphingosinephosphoethanolamine (d18:1/16:0) (PSP)] previously identified by using a combination of standard biostatistical, machine learning and network analysis technique, were collected (8). Ethical approval was obtained by the Regional Ethics Committee of Copenhagen (COCOMO: H-15017350). Written informed consent was obtained from all participants.

Plasma Lipidomic Profiling

Untargeted lipidomic profiling was performed on plasma samples collected at baseline in COCOMO through the Complex Lipid Panel™ technique (Metabolon Inc, Morrisville, NC 27560, USA). The company is ISO 9001: 2015 certified for analytical and diagnostic testing of biological specimens. Briefly, lipids were extracted from the bio-fluid using automated BUME extraction followed by infusion-MS analysis (11). Lipid species were quantified by taking the ratio of the signal intensity of each target compound to that of its assigned internal standard, then multiplying by the concentration of internal standard added to the sample. Lipid class concentrations were calculated, and fatty acid (FA) compositions were determined by calculating the proportion of each class comprised by summation of individual FAs. All the lipid quantifications were median-centered, and missing values were minimum-imputed per lipid species. We further removed variables with zero or near-zero variance from the dataset using *nearZeroVar* (i.e., 5%, n = 46 of 963). The original scale lipidomics data can be obtained from **Supplementary Data File S1**.

Statistics and Bioinformatics Analysis

All the analyses were carried out in R 4.0.3 (12). Clinical characteristics between PLWH without MetS and PLWH with MetS were compared using the Mann–Whitney U test (continuous variables) and chi-square test (categorical variables). Dimension reduction was carried out using

principal component analysis (PCA). Structural interpretation of the lipidome was carried out through *lipidomeR*. The normality of the lipidomics data was tested through Kolmogorov-Smirnov test and density plots (13). The Mann-Whitney U test was applied to raw data and a subset of lipids with a false discovery rate (FDR)<0.001, was derived. Log-transformed data were tested for differential abundance using *limma* and significant lipids with a FDR<0.001, were derived (14). Binary classification modeling was carried out by partial least squares discriminant analysis (PLS-DA) using *ropls* (15), where a subset of variables with variables importance on projection (VIP) score >1 was derived. Random forest (RF) was carried out using *MUVR* [https://github.com/CarlBrunius/MUVR]. Variables from the optimal RF modeling performance were selected according to rank. Model performance was evaluated by using the Q2Y and area under the receiver operating characteristic (AUROC) for PLS-DA and RF, respectively. Pathway enrichment was tested from the *limma* output (FDR < 0.1) with Ingenuity Pathway Analysis (IPA) (Qiagen, US) and MetaboAnalyst (16) (limma, FDR< 0.1). The FDR was controlled for by using the Benjamin-Hochberg (BH) method (13).

Network Analysis

Network analyses were used to build a biological network consisting of lipids (n = 917) and previously identified key metabolites (n = 11) (8) after Spearman's rank correlation across all species. Edges connecting nodes (i.e., biomolecules) were weighted based on positive correlations. This network was compared against a null model attained from a random network with the same number of nodes and edges based on the Erdos-Renyi model (17). All networks were built through the Python module *igraph* (18), communities within the biological network were detected through the Leiden algorithm (19). Communities were characterized functionally and phenotypically through the lipid-specific ontology web tool, LION/web (20). LION/web was used to determine lipid ontology trends within each community, using all lipids from the network as a background list. Separate analyses on each network community with all lipids as background lists were uploaded to LIPEA to identify lipid pathway enrichment (21). The community association with clinical parameters was determined through logistic and linear regression in R. Network visualization was performed using Cytoscape 3.5.1 (22).

RESULTS

Machine Learning Highlights Differences in Key Lipids in PLWH With MetS

PLWH with MetS (n = 100) and PLWH without MetS (n = 100) were included from the COCOMO study (**Table 1**). VAT and SAT significantly differed between the two groups (p-value<0.001). We further applied several univariate and machine learning approaches to characterize the effect of MetS in HIV-infected following long-term cART treatment and to investigate the underlying biological mechanisms of MetS (**Figure 1**). The lipidomic dataset consisted of 917 unique lipid species including 602 glycerolipids, 228 glycerophospholipids, 61 sphingolipids, and 26 steroids. We observed 618 and 584 significantly differentially abundant lipids between PLWH without MetS and PLWH with MetS (**Supplementary Data File S2**, FDR <0.001), using Mann-Whitney U and limma, respectively. Moreover, PLS-DA was used to identify variations between the groups based on lipid concentrations, by exploiting its ability to handle a greater number of features compared to samples. The separation of the two groups was indicated by a score plot, where the two first orthogonal components explained half of the variance in the data with 45% and 5%, respectively. We found 516 lipids with VIP values > 1, Q2Y = 0.319 (**Supplementary Data File S2**). To obtain a better model

TABLE 1 | Clinical and demographic characteristics compared between PLWH without MetS and PLWH with MetS.

Variables	PLWH without MetS	PLWH with MetS	pvalue
Sample (n)	100	100	
Sex, Male, n (%)	90 (90.0)	90 (90.0)	1.00**
Age, mean (sd)	54.4 (9.5)	54.6 (8.5)	0.80*
Ethnicity, n (%)			0.87**
Caucasian	88 (88.0)	86 (86.0)	
Asian	3 (3.0)	2 (2.0)	
Black	4 (4.0)	6 (6.0)	
Other/unknown	5 (5.0)	6 (6.0)	
Immunodeficiency, n (%)	14 (14.0)	13 (13.0)	1.00**
Exposure to early-generation ART, n (%)	34 (34.0)	46 (46.0)	0.11**
VAT, mean (sd)	76.1 (53.6)	149.4 (71)	**< 0.001***
SAT, mean (sd)	111.1 (71.1)	150.6 (77.1)	**< 0.001***
ART_NRTI, n (%)	95 (95.0)	96 (96.0)	1.00**
ART_NNRTI, n (%)	54 (54.0)	45 (45.0)	0.26**
ART_PI, n (%)	37 (37.0)	47 (47.0)	0.20**
ART_INSTI, n (%)	16 (16.0)	21 (21.0)	0.47**
ART_other/unknown, n (%)	0 (0.0)	3 (3.0)	0.24*

**Mann-Whitney U test and **Chi-square test.*

P-values in bold indicates a significant difference in the concerned variables between the two groups. Immunodeficiency was defined as the lowest CD4+ T-cell count <200 cells/µl or previous AIDS condition and exposure to early-generation ART was defined as patients medicated with thymidine analogues, didanosine and/or indinavir.

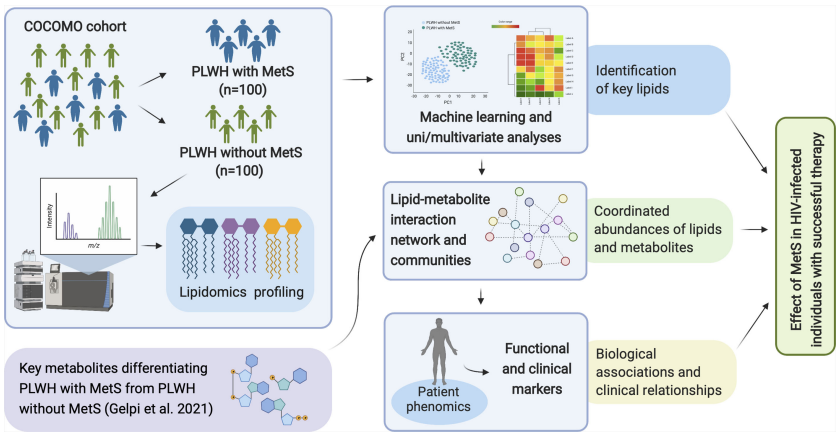


FIGURE 1 | Overview of study workflow. Analysis pipeline for characterizing the effect of MetS in HIV-infected following ART treatment and investigating the underlying biological mechanisms of PLWH with MetS (created with BioRender.com).

performance we could increase the sample size. Finally, we created three types of RF models, minimal-optimal ('Min'), geometric mean ('Mid') and all-relevant ('Max') models, which represented feature selection with minimal number of misclassifications, where we observed good performance of all models (**Figure 2A**, AUC>82.9%). Then, we identified 13 lipids as the strongest predictors of separating PLWH without MetS and PLWH with MetS (**Figure 2B**, 'Max' MUVR model, AUC = 83%), where the glycerolipid classes DAGs and TAGs were found to have the greatest significance in group separation. The number of significant lipids identified by each of the four methods varied greatly (**Supplementary Data File S2**). However, we observed 13 differentially abundant lipids between PLWH without MetS and PLWH with MetS (**Figure 2C**) which were consistently identified in all four methods (i.e., Mann-Whitney U, limma, PLS-DA, and RF). These 13 key lipids and 11 key metabolites (**Table 2**) indicated relatively good separation between PLWH without MetS and PLWH with MetS on sample clustering (**Figure 2D**). Furthermore, we observed higher abundance level of the key lipids in PLWH with MetS compared to PLWH without MetS (**Figure 2E**).

Structural Interpretation of Lipids Indicates Compositional Lipid Patterns

We then examined the structural characteristics of the lipidome by lipid class in terms of FA carbon number and saturation level (**Figure 3**). We observed an increase of ceramide (CER), DAG, dihydroceramide (DCER), lysophosphatidylethanolamine (LPE), monoacylglyceride (MAG), PE and TAG, in PLWH with MetS compared to PLWH without MetS, and a decrease in hexosylceramide (HCER) and lactosylceramide (LCER) (**Figure 3**, FDR<0.01 and Pearson's r>0.7). An increased significantly differential abundance of DAGs and TAGs was observed, indicated by the symbol and red color. The TAGs tended to display a higher abundance of polyunsaturated lipids (i.e., a double-bond content between 2-5) with long-chain fatty acids (LCFA) (i.e., C48-56) (**Figure 3**, FDR<0.01, Pearson's r>0.7). Additionally, TAGs displayed the most significant amount of lipid species. DAGs showed a tendency of increase in both saturated and unsaturated lipids (i.e., a double-bond content between 0-6) with LCFA (i.e., C30-40) (**Figure 3**, FDR<0.01, Pearson's r>0.7).

Clinical and Omics Integrated Network Identifies Biomolecular Patterns

Seeking to test whether and how any coordinated patterns of association were present throughout the samples, we generated weighted lipid-metabolite networks to understand the relationship between lipidomic profiles with those metabolites associated with MetS in PLWH. While retaining only informative metabolites, we examined the relationship between key metabolites previously identified with the entire lipidome. Briefly, we associated clinical variables with the identified communities and within the most central community, we identified associations between clinical variables and each biomolecule.

The fully connected biological network comprised 18430 edges and 917 nodes and displayed markedly distinct behavior from the null network (**Supplementary Table S1** and **Supplemetary Figure S1**). A community analysis on the biological network identified three communities of strongly interconnected lipids and metabolites (**Supplementary Table S2**). Centrality properties were evaluated identifying c1 as the most central community in the network (**Figure 4A**), which captured the most coordinated differential abundance changes. Community c1 had the largest community size (size = 339) and largest community average degree (avg. degree = 534.63) (**Supplementary Table S2**).

Structural and functional characterization of these communities (23) (**Supplementary Figure S2**) indicated that glycerolipids and especially TAGs were enriched in both c1 and c2 (**Figure 4A**, FDR<0.05). Interestingly, a coordinated structural composition pattern of the glycerolipids displayed an average lower carbon number and average lower double-bond content in c1, compared to c2 (**Supplementary Table S3**).

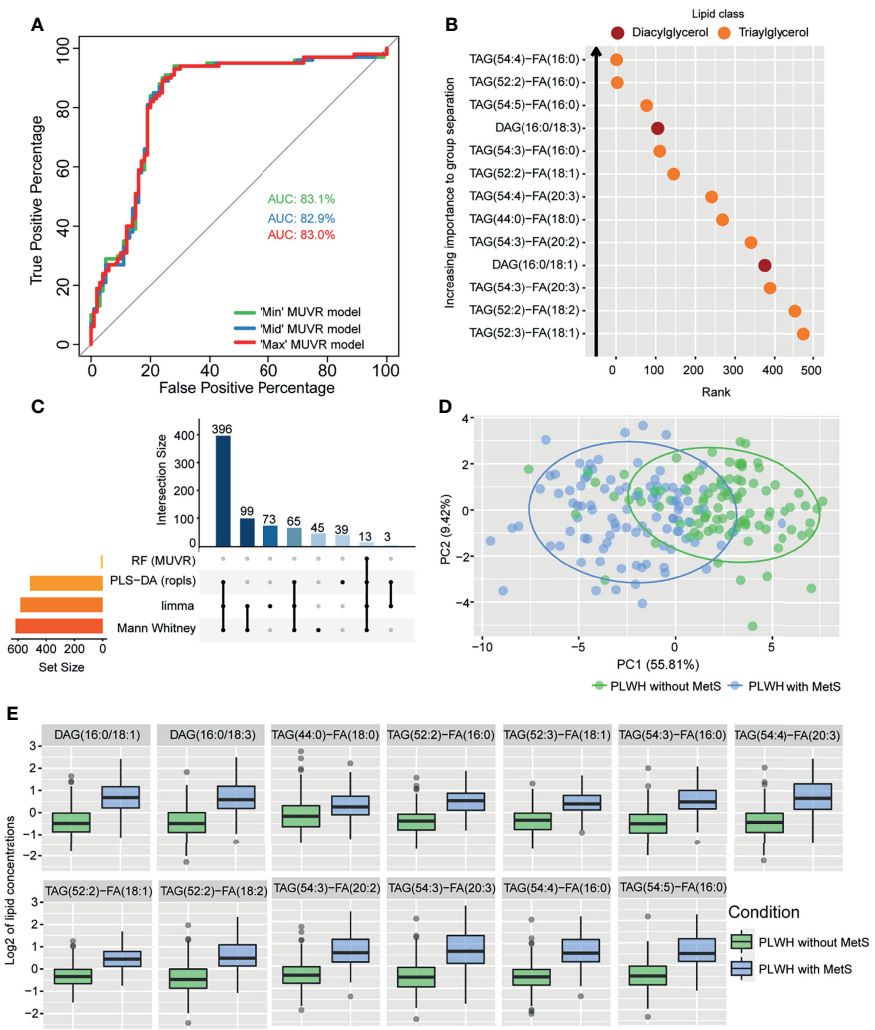


FIGURE 2 | Lipidomics analyses of PLWH without MetS vs PLWH with MetS identifying key lipids differentiating the two groups. **(A)** Performance of random forest (RF) models. Receiver operating characteristic (ROC) curve with area under the curve (AUC) values for the three *MUVR* models. **(B)** Important prediction variables separating PLWH without MetS from PLWH with MetS based on lipidomics, diacylglycerol (DAG) and triacylglycerol (TAG). Variable's importance on projection (VIP) scores plot for the 'Max' *MUVR* model, where lower rank indicates better group separation, thus better prediction variables in the model classification. **(C)** The intersection of methods identifying key lipids. UpSet plot showing number of significant lipids found *via* four statistical methods (RF, PLS-DA, limma, and Mann-Whitney U test). Note the 13 lipids (intersection size on the y-axis) are simultaneously identified by all four methods. **(D)** Separation of PLWH without MetS from PLWH with MetS based on identified key biomolecules. Principal component analysis (PCA) on key biomolecules, where lipidomics and metabolomics data were separated by the 13 identified key lipids and 11 identified key metabolites (**Table 2**). Ellipses show the 95% confidence interval of the data. **(E)** Boxplot of lipid concentration of the identified key lipids, which consist of DAGs and TAGs.

Community c3 was not further addressed, as the two other communities were interpreted to be of more importance due to their node size and average degree (**Supplementary Table S2**). We identified a positive association between community c1 with the clinical variables MetS, VAT and exposure to early-generation ART (**Supplementary Table S4**, FDR<0.12, illustrated in **Figure 4A**). In turn, community c2 was positively associated with MetS, however with a lower estimate compared to c1 (**Supplementary Table S4**, FDR<0.12). Log-fold changes indicated up-regulation of lipids in PLWH with MetS compared to PLWH in communities c1 and c2 (**Figure 4B**, limma, FDR<0.001).

Furthermore, we observed a positive association between lipids (i.e., DAGs and TAGs) and VAT (**Supplementary Table S5**, FDR<0.01). TAGs tended to consist of polyunsaturated lipids (i.e., a double-bond content ≥2) with LCFA (i.e., C48-54). Interestingly, we also observed that TAGs with LCFA (i.e., C42-48) and a low double-bond content (i.e., ≤2) were positively associated with the use of NNRTI (**Supplementary Table S5**, FDR<0.07). Additionally, four out of the 13 key lipids [i.e., TAG(52:2)-FA(16:0), TAG(52:2)-FA(18:1), DAG(16:0/18:1), and TAG(54:3)-FA(20:3)] were all found to be independently associated with VAT (**Supplementary Table S5**, FDR<0.01). Finally, one out of the 11 key metabolites

TABLE 2 | Identified key lipids and key metabolites.

Key lipids	Key metabolites
DAG(16:0/18:1)	1-carboxyethylisoleucine
DAG(16:0/18:3)	4-cholesten-3-one
TAG(44:0)-FA(18:0)	4-hydroxyglutamate
TAG(52:2)-FA(16:0)	α-ketoglutarate
TAG(52:2)-FA(18:1)	carotene diol (2)
TAG(52:2)-FA(18:2)	γ-glutamylglutamate
TAG(52:3)-FA(18:1)	glutamate
TAG(54:3)-FA(16:0)	glycerate
TAG(54:3)-FA(20:2)	isoleucine
TAG(54:3)-FA(20:3)	PC/3-MAPC*
TAG(54:4)-FA(16:0)	PSP**
TAG(54:4)-FA(20:3)	
TAG(54:5)-FA(16:0)	

*pimeloylcarnitine/3-methyladipoylcarnitine (C7-DC).
**palmitoyl-sphingosine-phosphoethanolamine (d18:1/16:0).
Overview of key lipids and key metabolites with significant differential abundance between PLWH without MetS and PLWH with MetS. Listed in alphabetical order.

(i.e., glutamate) was also found to be independently associated with VAT (**Supplementary Table S5**, FDR<0.01).

The top 10% of most interconnected biomolecules, found according to their degree, were all glycerolipids within the classes DAG and TAG (**Supplementary Table S6**). Three out of the 13 key lipids [i.e., TAG(52:2)-FA(16:0), TAG(52:2)-FA(18:1) and TAG(54:3)-FA(20:3)] were ranked among the top 10% most interconnected biomolecules in c1. Thus, these three lipids were interpreted to be among those biomolecules influencing the behavior of the global network the most. It should be noticed that the structural composition of all three lipids are polyunsaturated TAGs with LCFA and were all found to be positively associated with VAT (**Supplementary Table S5**, FDR<0.01).

The local network of community c1 included six out of the 13 key lipids (i.e., TAG(54:3)-FA(20:2), DAG(16:0/18:1), TAG (52:2)-FA(16:0), TAG(54:3)-FA(20:3), TAG(52:2)-FA(18:1), and TAG(44:0)-FA(18:0), **Figure 4C**). All of the six key lipids within c1 were glycerolipids, five TAGs and one DAG. Interestingly, we observed that all five TAGs were polyunsaturated with a doubsle-bond content between 2 and 4 with a FA carbon number between C52 and C54. Moreover, the key lipids were found to be interconnected with 6 of the 11 key metabolites within c1. Finally, all 13 key lipids and 7 metabolites within the global network were found to be interconnected with each other.

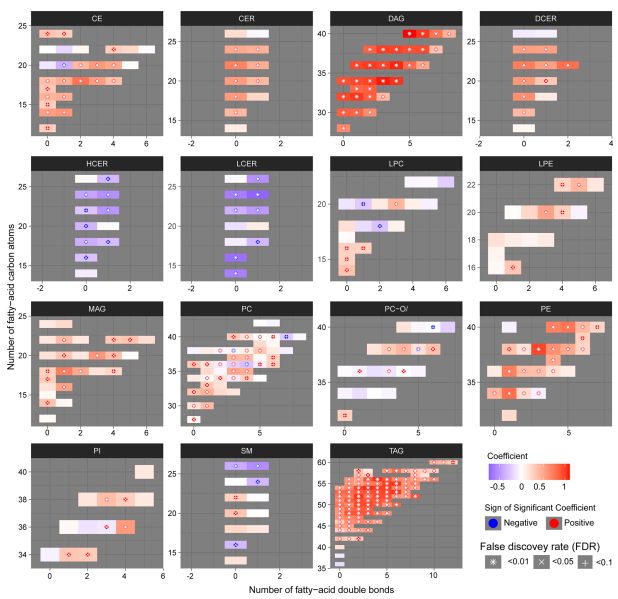


FIGURE 3 | Structural differences of the lipidomic profile of PLWH without MetS vs. PLWH with MetS. Heatmaps for each lipid class show the structural lipid composition differences between PLWH without MetS and PLWH with MetS. Each lipid species is shown as a rectangle and the color shows the abundance difference (red: higher in PLWH with MetS; white: no difference; blue: lower in PLWH with MetS), the lipids were organized by the lipid size (y-axis) and level of saturation (x-axis). Lipids with statistically significant differences between the two groups were highlighted with a symbol. P-values have been FDR adjusted.

DISCUSSION

The integrative plasma lipidomics and metabolomics analysis in a large HIV cohort of PLWH with and without MetS indicated a complex role of lipids in the link between ART and MetS in PLWH and provided a system-level understanding of MetS in PLWH. Our data indicated an increased abundance of the glycerolipids DAGs and TAGs in PLWH with MetS. The comprehensive network integration of the lipidomics and metabolomics (8) data suggested interactions between specific glycerolipids patterns and key metabolites involved in the glutamate metabolism. Finally, our data also indicated a relationship between the structural composition patterns of these specific glycerolipids with HIV and MetS-specific clinical variables, suggesting their involvement in driving the disease pathogenesis in PLWH with MetS.

In our study, we found 13 key glycerolipids from the classes DAG (n = 2) and TAG (n = 11) to be significantly altered between PLWH with and without MetS. It is worth noting the structural composition of the 13 lipids. The two DAGs [DAG(16:0/18:1) and DAG(16:0/18:3)] consist of unsaturated LCFA (i.e., C34 and 1-3 double-bonds). Additionally, 10 out of the 11 TAGs were polyunsaturated LCFA (i.e., C52-54 and a double-bond content of 2-5). The last TAG had a lower carbon number of C44, compared to the others and was saturated. These findings support previous findings of MetS in general populations that showed that lipids (especially TAGs) with lower carbon numbers (i.e., C44-54) and lower double-bond content (i.e., 1-4) were associated with an increased risk of T2D. Moreover, it had been observed that an increase in DAGs was associated with hypertension, another MetS-related factor (24). The structure of the FAs is a useful indication of the functionality of lipid metabolism. Increased accumulation of LCFA such as C(16:0), C(16:1), C(18:0), and C(18:1) suggests increased biosynthesis under MetS-conditions. Such chain compositions are observed among our 13 identified key lipids both in the DAGs and TAGs. Additionally, to the observed pattern of LCFAs, another study suggests LCFAs might cause impairment of mitochondria functions (25).

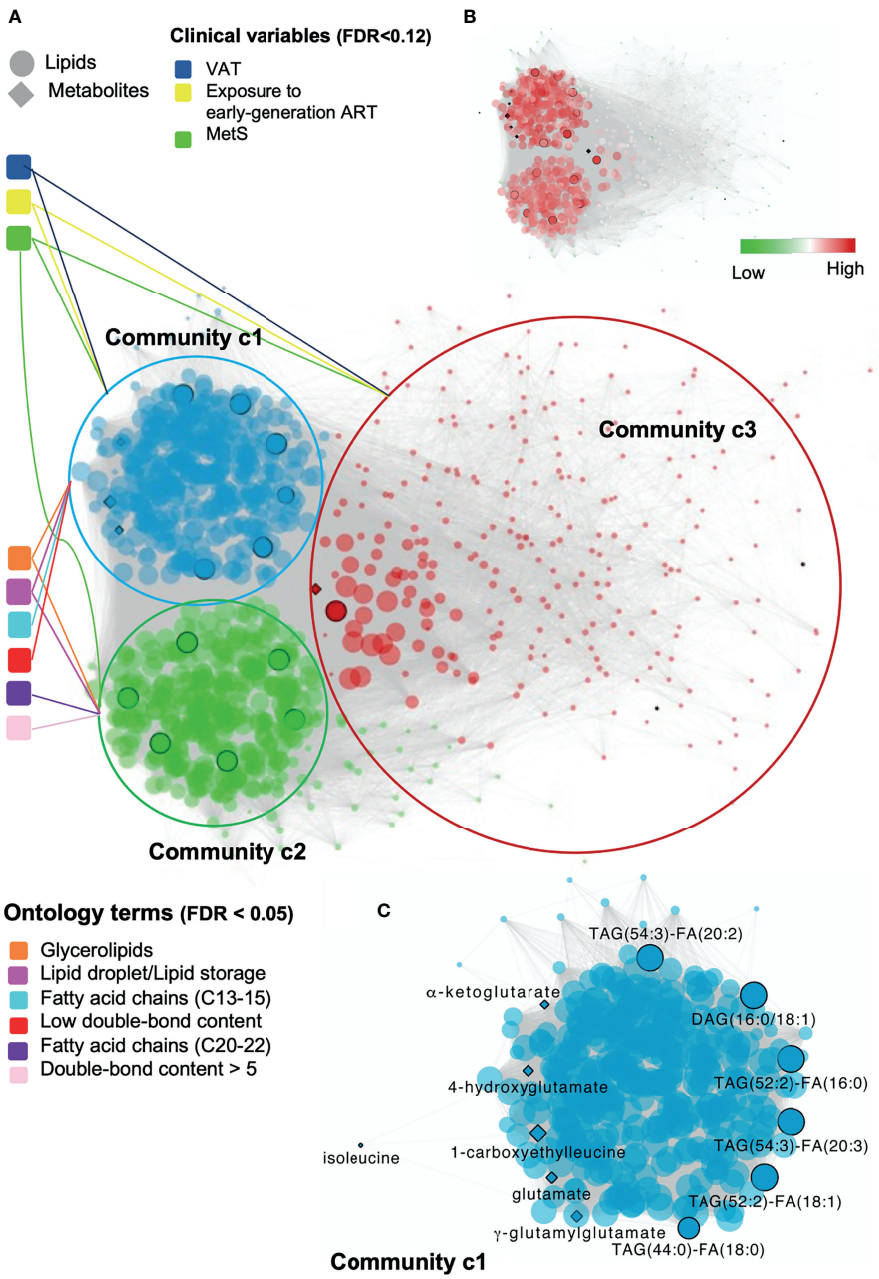


FIGURE 4 | Global and local biomolecular network of PLWH without MetS vs PLWH with MetS. **(A)** Global network illustrating the associated clinical variables and ontology terms with each community. Network of positive correlations between lipids and metabolites (FDR < 1e-07, Spearman's ρ > 0.38), colored based on the three identified communities, c1 (blue), c2 (green) and c3 (red). Communities are connected with associated clinical variables (FDR < 0.12) and ontology terms (FDR < 0.05). Black circled lipids and metabolites correspond to identified key lipids and key metabolites (**Table 2**). **(B)** Global network illustrating up and down-regulated lipids in PLWH with HIV. **(C)** Local network of community c1 highlighting key biomolecules. Biomolecular correlations within community c1 (FDR < 1e-07, Spearman's ρ > 0.38). Black circled and named biomolecules correspond to the identified key lipids and metabolites within c1.

We employed network analysis by integrating the key metabolites previously identified (8) as biomarkers in PLWH with MetS and the lipids with the clinical features (phenomics) that can provide a comprehensive view of the metabolic state related to a disease phenotype. Interestingly, we observed that community c1 contained glycerolipids with a lower carbon number and lower double-bond content compared to c2. Community c1 was further investigated and we found that c1 positively associated with the clinical variables MetS, VAT, and exposure to early-generation ART. Our findings are related to previous findings that showed TAGs with a lower carbon number and lower double-bond content play a considerable

role in MetS (23, 26). Additionally, our results suggest that exposure to early-generation ART (i.e., thymidine analogues/didanosine/indinavir) and increased VAT may also lead to a lower carbon number and double-bond content in glycerolipids, suggesting a role for polyunsaturated glycerolipids with LCFA [i.e., especially TAG (52:2)-FA(16:0), TAG(52:2)-FA(18:1) and TAG(54:3)-FA(20:3)] in the metabolic patterns in PLWH with MetS.

Previous studies have investigated the association of lipidomics profiles (211 lipids) with the progression of CVD in PLWH receiving ART treatment with carotid artery atherosclerosis, compared to HIV-negative individuals (3). The study showed elevation in lipid species with polyunsaturated LCFAs (i.e., C13-21 and double-bond content ≥2) in patients with atherosclerosis, which is also observed in PLWH with MetS in our study. Additionally, their study suggested significant alterations in lipid species such as CE, CER, LPC, LPE, PC, PIs, and PI. Other studies in both HIV and HIV-negative populations also found alterations in levels of other lipid species (i.e., different from DAG and TAG) to be associated with MetS-factors. This includes CE, CER, LPC, PC, PE, and sphingomyelin (SM) (5, 23). Some of these lipid species were also altered in our study population (i.e., CE, CER, PE, and SM); however the glycerolipids showed strongest predictive values. Our findings of coordinated abundance shift in glycerolipids may be due to our considerably larger amount of quantified lipid species than other studies (3, 23, 26).

In the same cluster (i.e., c1) we observed two trends concerning coordinated abundance shifts in glycerolipids. First, TAG species with carbon numbers between C48-54 and double-bond content ≥2, together with DAG species with carbon number between C32-36 and double-bond content ≥1 associated positively with VAT (FDR<0.01). This finding correlates with previous studies of MetS factors in HIV-negative cohorts (23, 26). Second, TAG species with carbon numbers between C42-48 and double-bond content ≤2 positively associated with the use of ART drugs containing NNRTIs (FDR<0.07). The latter trend supports previous findings suggesting that the NNRTIs drug efavirenz introduces dysfunction in the mitochondria by inducing increased levels of lipids (27). Moreover, both exposures to early-generation ART and the use of NNRTIs drugs have shown to cause disruption of the mitochondrial functions in previous studies (27, 28). To our knowledge, we present here the first evidence of association of antiretroviral treatment with specific structural composition lipid profiles.

The composition-specific glycerolipids correlated with some of the previously identified key metabolites linked to the perturbations of the glutamate metabolism in PLWH with MetS (8). This finding correlates with previous studies of MetS in HIV-negative populations, which found branched-chain amino acids (BCAAs) (i.e., leucine, isoleucine, and valine) as one of the significant metabolite groups dysregulated in obese individuals together with increased concentrations of glutamate, which is the first step of the BCAAs catabolism (29, 30). Additionally, the polar metabolite acylcarnitine (abbreviated PC/3-MAPC), an essential member of the fatty acid metabolism, was significantly down-regulated in PLWH with MetS. This molecule facilitates the transportation of LCFAs into the mitochondria for catabolism through β-oxidation (31). Besides glutamate and 4-hydroxyglutamate being a part of the glutamate metabolism, other identified key metabolites were a part of mitochondrial processes, which have critical energetic functions (e.g., regulating insulin secretion). These metabolites belonged to the isoleucine metabolism (i.e., 1-carboxyethylleucine, isoleucine) and the TCA cycle (i.e., γ-glutamylglutamate, α-ketoglutarate) (8).

The present study has limitations, such as the cross-sectional design, as no conclusions on causality could be drawn. We were only able to assess the prevalence of the diseases in the plasma samples. Finally, despite the largest study population conducted to date to type comprehensive lipid profile in PLWH, the relatively small sample size of the cohort is also considered as a limitation to this study. However, this is the first study that includes a large, well-characterized group of PLWH with or without MetS matched on MetS, sex, and age. Furthermore, the use of fully quantitative lipidomics (i.e., >900 quantified lipid species) methodology allowed us to conduct a thorough analysis of the systematic lipid profiling and its association with metabolites and clinical factors by using a combination of standard biostatistical, machine learning, and network analysis techniques.

CONCLUSIONS

In conclusion, our study suggests alterations in both fatty acid metabolism and glutamate metabolism, which depend on well-functioning mitochondria. A synergistic effect of different factors (i.e., an increased pro-inflammatory state induced by HIV, age-related pathophysiological changes, exposure to early-generation ART, and the use of ART with the active agent NNRTIs), which perturb the functions within the biological system of HIV-infected, could play a part in the alterations of the identified biological mechanisms in the phenotype PLWH with MetS. A recent large study from India and Cameroon also reported alterations in glutaminolysis as a common factor in PLWH in long-term cART (32). Alterations in the lipid homeostasis, as well as glutaminolysis in PLWH, need clinical or dietary interventions as they might drive accelerated aging in PLWH with MetS. A recent report indicated that the senescent cells depend on the glutaminolysis, and inhibition of the pathway leads to both inducing senolysis (i.e., removal of senescent cells) as well as improved serum-free fatty acids (FFAs) in the aged mice, which is a hallmark of the obesity-related disorders. We, therefore, hypothesized that clearance of the senescent cells by inhibition of the glutaminolysis and improving the lipid profile might prevent age-associated disorders and provide healthy aging in the PLWH with MetS. This further can aid in developing therapeutic targets to avoid metabolic abnormalities and accelerated aging in PLWH with MetS.

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 in the article/**Supplementary Material**.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Regional Ethics Committee of Copenhagen (COCOMO: H-15017350). The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

Conceptualization and clinical study designing: SN, SO, UN, MG, AK, and DM. Clinical data and biobank: SN, MG, AK, JH, MT, and HU. Methodology: SO, RB, and UN. Formal analysis: SO and RB. Clinical interpretation: SO, UN, RB, SN, MG, AK, and DM. Supervision: RB, AK, MG, UN, and SN. Resources: UN and SN. Writing (original draft): SO. Writing (review and editing): RB, AK, MG, HU, MT, JH, DM, UN, and SN. Visualization: SO, RB, and UN. Project administration: UN and SN. Funding acquisition: UN and SN. All authors discussed the results, commented, and approved the final version of the manuscript.

FUNDING

The study is funded by Rigshospitalet Research Council, Danish National Research Foundation (DNRF126) NovoNordisk Foundation. UN acknowledges the support received from Swedish Research Council Grants (2017-01330, 2018-06156, and 2021-01756).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2021.742736/full#supplementary-material

REFERENCES

1. Babu H, Ambikan AT, Gabriel EE, Svensson Akusjarvi S, Palaniappan AN, Sundaraj V, et al. Systemic Inflammation and the Increased Risk of Inflamm-Aging and Age-Associated Diseases in People Living With HIV on Long Term Suppressive Antiretroviral Therapy. *Front Immunol* (2019) 10:1965. doi: 10.3389/fimmu.2019.01965
2. Gelpi M, Ueland PM, Trøseid M, Mocroft A, Lebech AM, Ullum H, et al. Abdominal Adipose Tissue Is Associated With Alterations in Tryptophan-Kynurenine Metabolism and Markers of Systemic Inflammation in People With Human Immunodeficiency Virus. *J Infect Dis* (2020) 221:419–27. doi: 10.1093/infdis/jiz465
3. Chai JC, Deik AA, Hua S, Wang T, Hanna DB, Xue X, et al. Association of Lipidomic Profiles With Progression of Carotid Artery Atherosclerosis in HIV Infection. *JAMA Cardiol* (2019) 4:1239–49. doi: 10.1001/jamacardio.2019.4025
4. Freitas P, Carvalho D, Souto S, Santos AC, Xerinda S, Marques R, et al. Impact of Lipodystrophy on the Prevalence and Components of Metabolic Syndrome in HIV-Infected Patients. *BMC Infect Dis* (2011) 11:246. doi: 10.1186/1471-2334-11-246
5. Gelpi M, Afzal S, Lundgren J, Ronit A, Roen A, Mocroft A, et al. Higher Risk of Abdominal Obesity, Elevated Low-Density Lipoprotein Cholesterol, and Hypertriglyceridemia, But Not of Hypertension, in People Living With Human Immunodeficiency Virus (HIV): Results From the Copenhagen Comorbidity in HIV Infection Study. *Clin Infect Dis* (2018) 67:579–86. doi: 10.1093/cid/ciy146
6. Meikle PJ, Christopher MJ. Lipidomics Is Providing New Insight Into the Metabolic Syndrome and Its Sequelae. *Curr Opin Lipidol* (2011) 22:210–5. doi: 10.1097/MOL.0b013e3283453dbe
7. Graessler J, Schwudke D, Schwarz PE, Herzog R, Shevchenko A, Bornstein SR. Top-Down Lipidomics Reveals Ether Lipid Deficiency in Blood Plasma of Hypertensive Patients. *PLoS One* (2009) 4:e6261. doi: 10.1371/journal.pone.0006261
8. Gelpi M, Mikaeloff F, Knudsen AD, Benfeitas R, Krishnan S, Svenssson Akusjärvi S, et al. The Central Role of the Glutamate Metabolism in Long-Term Antiretroviral Treated HIV-Infected Individuals With Metabolic Syndrome. *Aging* (2021) 13:22732–51. doi: 10.18632/aging.203622
9. Ronit A, Haissman J, Kirkegaard-Klitbo DM, Kristensen TS, Lebech AM, Benfield T, et al. Copenhagen Comorbidity in HIV Infection (COCOMO) Study: A Study Protocol for a Longitudinal, Non-Interventional Assessment of non-AIDS Comorbidity in HIV Infection in Denmark. *BMC Infect Dis* (2016) 16:713. doi: 10.1186/s12879-016-2026-9
10. Alberti G, Zimmet P, Shaw J, Grundy SM. The IDF Consensus Worldwide Definition of the Metabolic Syndrome. *Brussels: Int Diabetes Fed* (2006) 23:469–80. Available at: https://www.idf.org/e-library/consensus-statements/60-idfconsensus-worldwide-definitionof-the-metabolic-syndrome.html.
11. Löfgren L, Ståhlman M, Forsberg GB, Saarinen S, Nilsson R, Hansson GI. The BUME Method: A Novel Automated Chloroform-Free 96-Well Total Lipid Extraction Method for Blood Plasma. *J Lipid Res* (2012) 53:1690–700. doi: 10.1194/jlr.D023036
12. R. Team. *RStudio: Integrated Development Environment for R*. Boston, MA (2016). Available at: http://www.rstudio.com/.
13. Checa A, Bedia C, Jaumot J. Lipidomic Data Analysis: Tutorial, Practical Guidelines and Applications. *Anal Chim Acta* (2015) 885:1–16. doi: 10.1016/j.aca.2015.02.068
14. 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:e47. doi: 10.1093/nar/gkv007
15. Thévenot EA, Roux A, Xu Y, Ezan E, Junot C. Analysis of the Human Adult Urinary Metabolome Variations With Age, Body Mass Index, and Gender by Implementing a Comprehensive Workflow for Univariate and OPLS Statistical Analyses. *J Proteome Res* (2015) 14:3322–35. doi: 10.1021/acs.jproteome.5b00354
16. Chong J, Wishart DS, Xia J. Using MetaboAnalyst 4.0 for Comprehensive and Integrative Metabolomics Data Analysis. *Curr Protoc Bioinf* (2019) 68:e86. doi: 10.1002/cpbi.86
17. Barabási AL, Oltvai ZN. Network Biology: Understanding the Cell's Functional Organization. *Nat Rev Genet* (2004) 5:101–13. doi: 10.1038/nrg1272
18. Csardi G, Nepusz T. The Igraph Software Package for Complex Network Research. *InterJ Complex Syst* (2006) 1695:1–9. doi: 10.1186/1471-2105-12-455
19. Traag VA, Waltman L, van Eck NJ. From Louvain to Leiden: Guaranteeing Well-Connected Communities. *Sci Rep* (2019) 9:5233. doi: 10.1038/s41598-019-41695-z
20. Molenaar MR, Jeucken A, Wassenaar TA, van de Lest CH, Brouwers JF, Helms JB. LION/web: A Web-Based Ontology Enrichment Tool for Lipidomic Data Analysis. *GigaScience* (2019) 8:giz061. doi: 10.1093/gigascience/giz061
21. Acevedo A, Durán C, Ciucci S, Gerl M, Cannistraci CV. LIPEA: Lipid Pathway Enrichment Analysis. *bioRxiv* (2018), 274969. doi: 10.1101/274969
22. 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
23. Stegemann C, Pechlaner R, Willeit P, Langley SR, Mangino M, Mayr U, et al. Lipidomics Profiling and Risk of Cardiovascular Disease in the Prospective Population-Based Bruneck Study. *Circulation* (2014) 129:1821–31. doi: 10.1161/CIRCULATIONAHA.113.002500

24. Hinterwirth H, Stegemann C, Mayr M. Lipidomics: Quest for Molecular Lipid Biomarkers in Cardiovascular Disease. *Circulation Cardiovasc Genet* (2014) 7:941–54. doi: 10.1161/CIRCGENETICS.114.000550
25. Hafizi Abu Bakar M, Kian Kai C, Wan Hassan WN, Sarmidi MR, Yaakob H, Zaman Huri H. Mitochondrial Dysfunction as a Central Event for Mechanisms Underlying Insulin Resistance: The Roles of Long Chain Fatty Acids. *Diabetes/metabol Res Rev* (2015) 31:453–75. doi: 10.1002/dmrr.2601
26. Rhee EP, Cheng S, Larson MG, Walford GA, Lewis GD, McCabe E, et al. Lipid Profiling Identifies a Triacylglycerol Signature of Insulin Resistance and Improves Diabetes Prediction in Humans. *J Clin Invest* (2011) 121:1402–11. doi: 10.1172/JCI44442
27. Blas-García A, Apostolova N, Ballesteros D, Monleón D, Morales JM, Rocha M, et al. Inhibition of Mitochondrial Function by Efavirenz Increases Lipid Content in Hepatic Cells. *Hepatol (Baltimore Md)* (2010) 52:115–25. doi: 10.1002/hep.23647
28. Trevillyan JM, Wong G, Puls R, Petoumenos K, Emery S, Mellett NA, et al. Changes in Plasma Lipidome Following Initiation of Antiretroviral Therapy. *PLoS One* (2018) 13:e0202944. doi: 10.1371/journal.pone.0202944
29. Wang R, Li B, Lam SM, Shui G. Integration of Lipidomics and Metabolomics for in-Depth Understanding of Cellular Mechanism and Disease Progression. *J Genet Genomics = Yi Chuan Xue Bao* (2020) 47:69–83. doi: 10.1016/j.jgg.2019.11.009
30. Rangel-Huerta OD, Pastor-Villaescusa B, Gil A. Are We Close to Defining a Metabolomic Signature of Human Obesity? A Systematic Review of Metabolomics Studies. *Metabolomics* (2019) 15:1–31. doi: 10.1007/s11306-019-1553-y
31. Kerner J, Hoppel C. Fatty Acid Import Into Mitochondria. *Biochim Biophys Acta* (2000) 1486:1–17. doi: 10.1016/S1388-1981(00)00044-5
32. Mikaeloff F, Svensson-Akusjärvi S, Ikomey GM, Krishnan S, Sperk M, Gupta S, et al. Trans Cohort Metabolic Reprograming Towards Glutaminolysis in Long-Term Successfully Treated HIV-Infection. *Commun Biol* (2021). doi: 10.1038/s42003-021-02985-3

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.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership