Check for updates

Decoding the etiology of immune-mediated inflammatory diseases statistically

Hesham ElAbd[1,2]* and Aya K. H. Mahdy[1]

[1]Institute of Clinical Molecular Biology, Kiel University and University Hospital Schleswig-Holstein, Kiel, Germany, [2]Institute for Digestive Research, Lithuanian University of Health Sciences, Kaunas, Lithuania

Immune-mediated inflammatory diseases (IMIDs) are incurable pathologies with an increased prevalence. Whereas different risk factors for IMIDs have been identified, such as microbial dysbiosis, diet, Epstein-Barr virus infection, the exact cause of most of these diseases remains unknown and it is thought to be a combination of environmental exposures and genetic predispositions. Despite their different clinical presentation, most IMIDs are genetically associated with variants at multiple immune-related genes, predominately with different human leukocyte antigen (HLA) alleles suggesting a strong pathological involvement of adaptive immune responses. However, antigens causing these diseases remain, in most cases, unknown. Using statistical analyses of the immune repertoire, several markers of antigenic exposures have been associated with IMIDs. Here, we discuss different approaches to identify disease-associated antigenic exposure markers and formulate a framework to test their causal role in IMIDs. We then discuss the potential contribution of risk HLA alleles to diseases development and lastly, we discuss how either antigens causing IMIDs or their signatures on the immune repertoire can be exploited therapeutically.

KEYWORDS

immune-mediated inflammatory diseases, immune repertoire, T-cell repertoire profiling, T-cell therapies, statistical analyses, etiology, immunogenetics

There is an urgent need for an etiological understanding of immune-mediated inflammatory diseases

IMIDs are a group of pathologies where chronic inflammation is evidenced, leading to tissue destruction, remodeling and eventually a loss-of-function. These diseases can be organ-specific, such as multiple sclerosis (MS), which affects the central nervous system, or systematic, impacting multiple organs simultaneously such as systemic sclerosis (SC). Several risk factors have been implicated in the pathogenesis of IMIDs such as Epstein-Barr virus (EBV) (1), for example, epidemiological and molecular studies established a strong link between infectious mononucleosis and inflammatory bowel disease (IBD) (2, 3). Furthermore, dysregulated immune responses toward EBV have been observed in other

IMIDs as well, *e.g.* MS (4–6), rheumatoid arthritis (RA) (7), systemic lupus erythematosus (8), and Sjögren's syndrome (9). Whereas a mechanistic understanding of the pathological role of EBV in these diseases remains to be identified, several mechanisms have been proposed, such as molecular mimicry between EBV and human proteins, for example, EBNA1 and GlialCAM (10) and EBNA1 and C1q (11). Beside EBV, other disease-specific alterations have also been identified, such as increased antibody responses toward citrullinated peptides in RA (12) and an expansion of a specific group of unconventional T cells in Crohn's disease (CD) (13, 14), which is a subset of IBD, among other disease-specific immune dysregulations.

As an exact cause for most of these diseases remains to be identified, treatments are mainly directed at inhibiting the inflammation, to clinically control disease symptoms and induce remission. This, arguably, partially non-specific inhibition of the immune system is achieved using different ways, such as anti-TNFs, anti-integrins and anti-cytokines antibodies, among others. Nonetheless, these therapies fail to introduce remission in all affected individuals, *i.e.* primary non-responders (15–17). Even primary responders might develop resistance to these therapies, *i.e.* secondary loss of response (15, 17), reaching what can be called a "therapeutic celling" at least in some diseases such as IBD (18). This problem is also aggravated by the lack of any approved prognostic marker for therapy response, despite ongoing efforts (19).

The prevalence of some of these IMIDs have increased significantly over the second half of the twentieth century, for example, the prevalence of IBD in the Olmsted County in the US, increased from 0.12% in 1960 to 0.63% in 2019 (20). Based on current estimates, it is projected that the prevalence of IBD will be ~1% in Canada in the upcoming decade (20–22). Besides IBD, other IMIDs are also prevalent in the population, for example, in the US alone there is between 400,000 (23) to 700,000 (24) individuals with MS and between 2 to 2.8 million individuals are living with the disease globally (25, 26). A higher prevalence is seen with RA with 17.6 million people affected globally (27) and with atopic dermatitis (AD) where more than 200 million individuals are living with the disease worldwide (28).

The combination of high prevalence, lack of accurate prognostic markers and the high cost of these medications is having a deleterious impact on the quality-of-life of affected individuals and healthcare systems. Thus, there is an urgent need for a better understanding of these diseases which could lead to more personalized therapies that induce a long-lasting remission in most patients as well as the development of preventive strategies, *e.g.* vaccines, in high-risk individuals.

Disease-associated genetic variants do not predetermine the development of IMIDs

With the rise of genome-wide association studies (GWAS) during the last couple of decades, the genetic signatures of IMIDs have been heavily investigated (29–34). For example, using the genetic data of 47,429 individuals with MS and 68,374 controls, more than 233 variants were associated with MS, 32 of them were located within the extend HLA loci (35). Similarly for RA, a recent meta-analysis spanning 35,871 individuals with RA and 240,149 controls from different ancestries identified 124 loci associated with RA (32). Similar meta-analyses were conducted in other IMIDs such as psoriasis, where a recent study has identified 109 loci that are implicated in the disease using a large cohort of 36,466 cases and 458,078 controls (36). Also, in AD 71 genetic variants were implicated in the disease by analyzing the genetic background of 65,107 individuals with AD and 1,021,287 controls (33).

A common denominator among these IMIDs was the lack of clear causative genetic mutations, as opposed to Mendelian genetic diseases, instead there was multiple associations to different common genetic variants. For most IMIDs, these associations resolved to genetic variants in different innate and adaptive immunity-related genes and loci (32, 33, 35, 36) such as the human leukocyte antigen (HLA) loci. Most of the associated HLA alleles have a moderate association odds ratio (OR) and were frequent in the study population in general. For example, in MS, the strongest genetic association is with the HLA-DRB1*15:01 with an OR of ~3 (37–39). Whereas the frequency of this alleles varies across populations and ancestries, it is frequent in European populations (frequency >10%) (39).

This implies that millions of individuals are carrying disease-associated HLA alleles and are not affected, at least symptomatically, with these diseases. This is clearly seen in celiac disease (CeD) which is strongly associated with HLA-[DQ2.2, DQ2.5 and DQ8] alleles (40, 41), nonetheless, not all carriers of these alleles are developing CeD. Thus, other environmental factors besides genetics are contributing to IMIDs such as gluten in the context of CeD.

The adaptive immune system records previous antigenic exposures as V(D)J generated sequences

HLA proteins are a central hub for communication among different parts of the immune system. They are classified into two classes, class I which presents peptides to CD8+ T cells and class II which presents peptides to CD4+ T cells. Thus, they convey critical information about potential peptide antigens between all nucleated cells and CD8+ T cells, between B cells and CD4+ T cells and between dendritic cells and T cells. A hallmark of adaptive immunity is the formation of an immunological memory after an antigenic exposure. This immune memory is composite of three main elements, first, a unique immune receptor that recognize different antigenic peptides from pathogens, *i.e.* T and B cell receptors, TCRs and BCRs, respectively. These receptors are generated via V(D)J recombination events and are engraved in the DNA encoding the TCR and the BCR of this antigen-specific T and B cells.

Nonetheless, before we continue our discussion, we need to highlight important distinctions between TCRs and BCRs, namely,

somatic hypermutation and class-switching which are exclusive to BCRs. Somatic hypermutation is a process used by B cells to enhance the affinity of their receptors, *i.e.* BCRs, toward a specific antigen, through random mutations introduced in their immune receptor chains followed by selection of mutations that increase the affinity of the BCR toward its cognate antigen, via an interaction with follicular helper T (Tfh) cells (42). This results in a family of related BCRs that bind the antigen with varying affinities, that is, a phylogenetic tree of "evolutionary-related" sequences that respond to the initial antigenic exposure (43). A second mechanism that is also specific to BCRs is class-switching where the isotype of the immunoglobin heavy chain is changing from μ, which is formed during the early phase of antigenic exposure, to other isotypes, for example, α, γ or ε which are used in IgA, IgG and IgE antibodies, respectively. Nonetheless, beside these differences between BCRs and TCRs, an antigenic exposure is associated with the formation of long-lived plasma cells and memory B cells that respond to this infection and record this exposure in the form of a DNA-encoded V(D)J recombination sequences (44).

The second part of an immune memory is a transcriptional program that shapes the behaviors of antigen-specific T and B cells and govern the phenotype of these cells, for example, T helper 1, 2, or 17. Also, naive B cells can follow different developmental trajectories after an antigenic encounter, for example, they can develop into short-lived plasma cells, into germinal center (GC) B cells or into GC independent memory B cells (44). The last part is an epigenetic memory, which enforces and constrains the formed transcriptional program of these antigen-specific T and B memory cells (Figure 1A). These formed immune memories are mostly long-lived and provide protection against repeated infections by the same pathogen.

Hence, as we age, we accumulate more antigenic exposures, either from natural infections, or vaccines, each of these exposures elicit the formation of an immune memory, resulting in the accumulation of memory cells that record this exposure history. Before we continue our discussion, we need to introduce two temporal events, first, a starting point, which will be the 1st trimester of gestation in humans where T and B cells begin to form (Figure 1B). Indeed, different compartments of the adaptive immune system develop at different stages of gestation for example, thymic development of T cells beings in the 1st trimester, however, T cells egress from the thymus at the beginning of the 2nd trimester, between the 12th and 14th week of gestation (45–47).

Second, a sampling timepoint, it is the timepoint of sampling a subset of the immune repertoire (Figure 1B). Based on these two events, the immune repertoire is defined here as the collection of immune memories, *i.e.* exposure histories, accumulated between the beginning and the sampling timepoint. For the sake of simplification, we are going to narrow down the definition of immune memories into the unique collection of V(D)J generated sequences. It is also worth mentioning that beside memory cells, the repertoire also contains V(D)J sequences from naive cells, which have not encountered their cognate antigen yet. Additionally, it contains V(D)J recombination sequences derived from effector cells responding to ongoing infections. For the sake of clarity, we focus

on the memory compartment of the immune repertoire unless stated otherwise.

A powerful method to study the collection of V(D)J recombination events encoding immunological exposure histories, is bulk immune receptor sequencing (Supplementary Figure S1A) (48). Nonetheless, it has three main limitations, first, it only provides the sequence of the generated receptor and not the antigen to which it binds. Second, the temporal order or exposure histories is almost not-captured, unless it is a very recent or an ongoing exposure that results in the expansion of some V(D)J recombination sequences (Supplementary Figure S1B). Third, it does not reveal the functional state of cells expressing these receptors, for example, Th1, Th2, Th17, among others. Furthermore, bulk immune-sequencing methods do not provide the full sequence of the immune receptor only part of it, for example, in case of TCRs, only the alpha (TRA) or the beta (TRB) chain, that is, the pairing information is lost in bulk repertoire profiling experiments.

Whereas the pairing information of TCRs and the transcriptional landscape of cells expressing these receptors can be identified via single-cell T cell receptor sequencing either using short-reads (49) or long-read sequencing (50), this method has several limitations. First, it is expensive, labor intensive and provides a shallow profiling of the repertoire where only few 1000s of clonotypes are profiled using single cell approaches, while in bulk immune sequencing 100,000s of clonotypes can be identified (48). Second, it requires access to intact cells, *e.g.* fresh or frozen PBMCs, which possess logistical problem when profiling the repertoire of thousands of samples. Lastly, if the aim is to generate pairing information without information about the cell type, then probabilistic mapping of profiled immune receptor chains using frameworks, such as pairSEQ (51) and TIRTL-Seq (52) might be a more cost-efficient approach. Hence, by integrating these different frameworks and methods, a better understanding of different aspects of immune receptor chains can be obtained, for example, using bulk TCR-Seq to profile the repertoire of thousands of individuals. Then, utilize probabilistic pairing to obtain the pairing information for candidate clonotypes across hundreds of individuals and lastly, using single cell technologies to understand the transcriptional landscape of these individuals in tens of samples.

Genetics has a fixed, robust effect on the formed immune repertoire that can be studied statistically

Before we delve deeper into how immune repertoires can be investigated to identify the etiological causes of IMIDs, we need to distinguish between two factors shaping the repertoire. First, fixed effects that is genetically predetermined regardless of antigenic exposures and second dynamic effects that result from a combination of genetic predeterminants and antigenic exposures. The fixed effects have three pillars, (i) germline encoded variation in the V and J genes, which forms the basis for V(D)J generated sequences (53–55). Second allelic variation in the HLA region (56–58) and third variation in other genomic loci (59). HLA proteins,
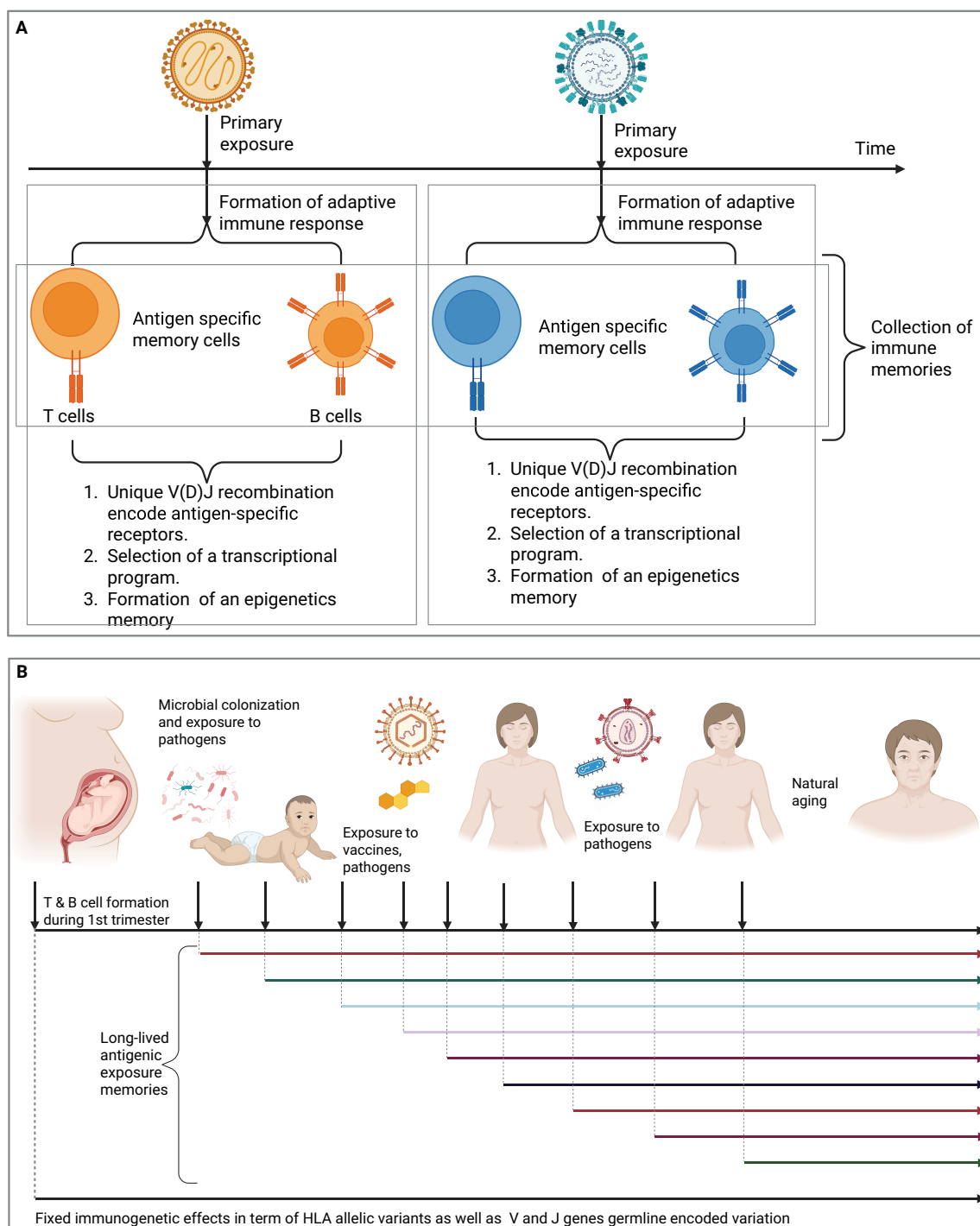
FIGURE 1

T and B cell repertoires record the antigenic exposure history of an individual. **(A)** The formation of a long-lived adaptive immune memories after the exposure to two distinct viruses, each of which will result in the formation of a distinct immune memory that records this antigenic exposure **(B)** The immune repertoire contains long-lived adaptive immune memories the records previous antigenic exposure histories. Created in BioRender. Elabd, H. (2025) https://BioRender.com/tn5de67.

regardless of any antigenic exposure, have a major impact on shaping the formed T cell immune repertoire, because of thymic selection. Different HLA proteins present different peptides to T cells, as shown previously by others and us (60–63), and during positive selection only V(D)J recombination sequences able to recognize self-peptides loaded into HLA proteins receive survival signal. Other genetic variants can have an influence by biasing the process of V(D)J recombination prior (59) to selection either by HLA proteins or by having coding variants that upon presentation by HLA alleles will shape T-cell selection. Alternatively, other

somatic genes might encode for signaling molecules that change the perception and the execution of T and B cells to an antigenic stimulus (64, 65).

From a molecular perspective, HLA exerts two effects on the immune repertoire, first, it biases the frequency of utilizing different V genes in the repertoire (57). Additionally, HLA proteins have a strong effect on the frequency of amino acids in the complementarity-determining region 3 (CDR3) (56). Thus, prior to any antigenic exposure, an interaction between the germline encoded genes, HLA allelic variants and other coding and non-coding variants will shape the formation of naive T cells primarily by shaping which V(D)J recombination is selected. Thus, forming the base to which immune memories will be formed upon antigenic exposures. Additionally, the generated TCRs can also shape the transcriptional landscape of the generated naive cells, for example, different TCR signaling intensities can module the differentiation of double-positive T cells into either CD8+ or CD4+ single-positive T cells (66). These sequence features can also play a role in the fate-determination process of regulatory T cells (67). Hence, an interaction among these different factors will have a strong effect on shaping the generated repertoire not only in terms of sequence diversity but also the functional landscape.

Using large-scale statistical analyses of the immune repertoire, the fixed effect of HLA proteins on the immune repertoire can be elucidated. For example, using >5,500 paired T cell immune repertoire and HLA genotypes, we were able to discover hundreds of thousands of clonotypes associated with tenths of HLA alleles (68). These clonotypes could accurately impute the carriership of these HLA alleles, indicating the strong impact of HLA protein on shaping the generated immune memories. Using different statistical frameworks, the impact of variable HLA sites on the frequency of amino acids in the CDR3 of the TCR beta chains (69) was studied by others (56) and us (69). Highlighting several paths by which HLA proteins exhibit an effect on the formed immune repertoires.

Identifying disease-associated shared antigenic exposures markers statistically

An immune memory is formed upon an antigenic exposure that results in the induction of a memory cell, the formation of these memory cells depends on the antigen, and fixed effects encoded genetically. With current technologies we can sequence V(D)J events, but unfortunately, we cannot, in most cases, decode their antigenic specificities, resulting in a trajectory of unknown antigenic exposures (Supplementary Figure S1B). Whereas newer technologies developed to decode the antigenic specificities of immune receptors, such as T-Scan (70), TScan-II (71) and receptor–antigen pairing by targeted retroviruses (RAPTR) (72), they require a rationally selected library of candidate TCRs, as well as peptide-HLA complexes. This represents a major hurdle for identifying the etiological causes of IMIDs for multiple reasons, first, in most cases neither the antigen nor the exact TCR(s) deriving these diseases are known. Second, the

immune repertoire is extremely diverse and personal, that is, most clonotypes are observed in one individual and are not shared among individuals. Thus, it is not feasible neither financially nor logistically to conduct these assays on all T cells of a cohort of individuals living with an IMID of interest. Third, there might be a long variable time span between the causative antigenic exposure and disease development, which is evidenced in some IMIDs, e.g. MS (5) which complicates the identification process of antigens implicated in the disease. Hence, a narrowing down of candidate T cells involved in the disease is needed.

Assuming that there is a specific antigen or a group of antigens that are causing IMIDs, then within individuals having the same IMID and sharing the fixed-repertoire effects, e.g. similar HLA background, we expect the same immune memories toward this exposure to be formed. Thus, we expect some V(D)J recombination sequences to be shared among individuals with a specific IMIDs relative to individuals without this IMID. These shared V(D)J sequences represent the exposure signature of the disease, e.g. the memory T cells associated with an antigenic exposure implicated in the disease. From large repertoire profiling studies, it was observed that most of the immune repertoire is private, that is, most responses are detected in only one individual, and that shared or public immune responses represent a small fraction of the repertoire (73). These shared responses might represent an exposure marker toward prevalent antigenic exposures, for example, common viral and bacterial infections (74, 75). Our ability to identify shared clonotypes involved in responding to a known antigenic exposure by comparing the repertoires of exposed to non-exposed (73, 76, 77), provides evidence to suggest that an IMIDs-associated antigenic exposures can be identified by comparing the repertoire of cases and controls (78).

Nonetheless, relative to disentangling the antigenic exposure of a specific infectious agent, different factors might complicate our analysis to identify exact antigenic exposures implicated in IMIDs. First, the likelihood of an antigenic exposure causing the disease, which is similar to the concept of "penetrance" in genetics. Here, some antigenic exposures might not exhibit a perfect or near perfect penetrance, where having the exposure does not guarantee a certain likelihood to develop the disease, just an increase in the odds of developing the disease. A perfect example of this is EBV, which have been implicated in multiple IMIDs as discussed above, however, it is a very prevalent infection where more than 95% of the population is affected.

Second, some antigenic exposures might generate an immune response in almost everyone in the population. For example, in phage-immunoprecipitation (PhIP-Seq) (79, 80) studies of the immune repertoire, multiple antigens have been shown to be recognized by almost every individual in the population with a prevalence of the immune response that is >95% (81–83). While these antigens might be recognized by the immune system of every individual it does not imply that they will be recognized by the same immune receptor, or that the epitope is going to be mapped to the same part of the protein in every individual. Hence, differences between individuals with and without IMIDs, might not be per se in the antigenic exposure but in the antigenic region, or the exact epitope, targeted by the immune system.

As a result, to disentangle disease signatures in IMIDs, a methods that can identify immune signatures at the epitope level is needed. This method should be able to identify antigenic exposures at the infectious agent level, at the antigenic protein of this pathogen and lastly at the epitope level, that is sub-antigenic protein level. A powerful method to identify these disease signatures is to analyze the T cell repertoire of thousands of individuals statistically to identify disease signatures. Here, the immune repertoire of T cells, which recognize short class of peptides, between 9 and 17 amino acids, presented by either HLA-I or HLA-II proteins is analyzed which provides us with an epitope-based mapping of immune responses. By statistically investigating the T cell repertoire of >5,000 individuals with IBD and >5,000 healthy controls we were able to identify >1,800 distinct V(D)J recombination sequences implicated in IBD (84).

Decoding the etiologies of IMIDs is similar to solving a temporal-credit assignment problem but with incomplete action history

While the approach described above provides an unparallel opportunity to identify disease etiologies, it has limitations primarily rated to the lack of temporal exposure order as most repertoire profiling methods provides a collection of immune memories without a temporal order. Given that most studies depend on a cross-sectional study design (Figure 2A) that includes individuals after their diagnosis with the disease, proving causation between the associated V(D)J sequences and the disease is not possible. Because these shared V(D)J sequences can be a consequence of the diseases instead of being the cause of the disease, *i.e.* a reverse-causation (Figure 2B).

A solution to this problem would require the temporal order of antigenic exposures before and after diagnosis to be resolved cohort-wide, to identify which antigenic exposure caused the disease and which exposure was caused by the disease. In essence, this will reduce the problem of identifying etiological factors into a temporal credit assignment problem, where exposures can be thought of as actions and developing the disease can be thought of as a reward. Hence, the solution to this problem becomes finding the action or series of actions, here antigenic exposures, responsible for disease development, that is, the reward in this formulation.

Nonetheless, resolving the temporal order of immune exposures from a profiled immune repertoire is still not possible with current technologies. A more practical approach will be to sample the repertoire of the study cohort across multiple timepoints ideally before disease development. This can be done using large-scale prospective or retrospective cohorts where multiple samples are collected from individuals before they develop the disease, as well as after they develop the disease (5, 85–87). By decoding the immune exposure across multiple points, a better understanding of exposures responsible for developing the disease can be obtained (Figure 2B). Given the high cost of immune profiling and the low incidence rate of most of these diseases, focusing the profiling on high-risk

individuals, *e.g.* patients' relatives (88), might provide a cost-efficient way to identify antigenic exposures causing the disease.

Which or when, the importance of exposure timing

In our discussion so far, we have focused on identifying antigens deriving the disease regardless of the timing of exposure, which might be critical for shaping the outcome of the disease. For example, infectious mononucleosis, which is predominantly caused by an EBV infection has been implicated in many chronic inflammatory diseases, *e.g.* MS (5, 6, 89), RA (7, 90), and IBD (2, 91), among others. While EBV infection is commonly associated with infectious mononucleosis in adults, this rarely happens in children (92), suggesting that the same antigenic exposure can lead to different outcomes based on the timing of exposure. Beside biological age, previous exposures can also have a strong influence on shaping the outcomes of an exposure (93), hence, the exposure trajectory plays important roles in shaping the generated outcome.

Do we need to identify etiological factors causing IMIDs to treat these diseases?

Given the high prevalence of IMIDs and their associated burden on health care systems, there is an urgent need for developing better therapeutic strategies to treat these diseases. However, do we need to know the etiology to effectively treat these diseases? By identifying V(D)J sequences associated with ankylosing spondylitis (78, 94) and depleting T cell populations containing these V(D)J sequences a novel therapy that induce remission in ankylosing spondylitis patients was developed (95). Whether this approach would generalize to other diseases with different etiologies and driver-antigens remains an open question.

Can we escape the inevitable?

In some cases, antigens causing and deriving the disease are common environmental exposures, *e.g.* gluten in CeD, making avoiding the exposure a practical approach to control the disease (Supplementary Figure S2A). Nonetheless, these avoidance approaches still have challenges, for example, a gluten-free diet is expensive with a limited set of options available in the market (96) and the risk of gluten contamination and mislabeling exists (97). As a result, different pharmacological interventions are being developed to treat CeD, for example, peptidases to digest gluten before it can elicit an immune response and intestinal barrier regulators (98).

In other cases, disease-causing antigenic exposures might be a common viral infection, such as EBV which infects >90-95% of the
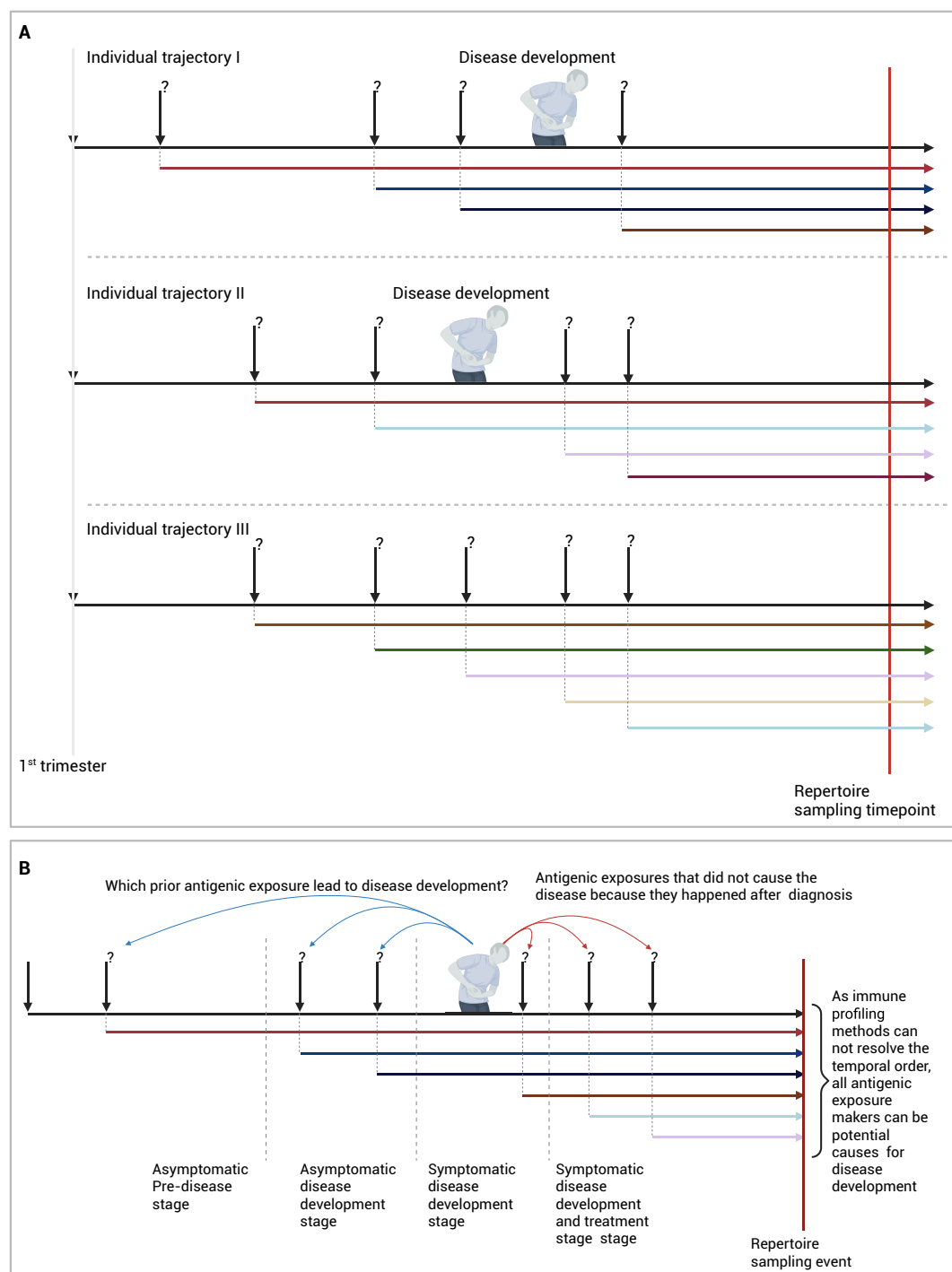
FIGURE 2
Identifying antigenic exposures deriving IMIDs. **(A)** Identifying diseases-associated antigenic exposures from cross-sectional studies. **(B)** An alternative framework to identify disease-causing antigenic exposures through longitudinal sampling of pre-clinical individuals. Created in BioRender. Elabd, H. (2025) https://BioRender.com/2cvfi4o.

population, making avoidance a much harder problem. While identifying antigenic exposures causing the disease might provide a promising strategy for therapeutic interventions, developing a preventive strategy might not be trivial, *e.g.* avoiding the exposure might not be possible. Alternatively, other sophisticated approaches, such as vaccines for either inducing tolerance (Supplementary Figure

S2B) (99), or protection against a particular disease-causing exposure (100) might be needed to prevent disease development. Lastly, by identifying disease-causing antigens and their cognate immune cells, technologies like monoclonal antibodies targeting the V(D)J recombination of these cells (Supplementary Figure S2C) (95) and chimeric autoantibody receptor (CAAR) T cells can be used to

specifically deplete immune cells driving the disease (Supplementary Figure S2D) (101).

Concluding remarks

There is an urgent, unmet need for a better understanding of IMIDs. The strong association between most of these diseases and several HLA alleles suggest an import role for adaptive immunity, specifically, T cell mediated responses in the disease. Nonetheless, the antigens causing these diseases remain to be identified, here we discussed several approaches and frameworks to identify antigenic exposure markers implicated in the disease, as well as potential experimental designs to establish causation. Despite the recent progress, many open questions remain to be addressed, for example, can we develop a more sample-efficient algorithm to identify disease-associated clonotypes? Current approaches depend on utilizing large cohorts of cases and controls; however, it is a very costly approach and not suitable for rarer diseases where assembling large cohorts is not feasible. Additionally, can we infer the antigenic specificity of a given V(D)J recombination event computationally? Can we infer the antigenic exposure trajectories, *i.e.* the order of antigenic exposures from the profiled T cell repertoire? Further, what is the contribution of private immune responses to IMIDs, relative to shared or public responses? In conclusion, the identification of disease-causing antigens or even their corresponding V(D)J sequences, will have a transformative utility on the development of therapeutic and preventive strategies for IMIDs.

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

HE: Conceptualization, Supervision, Visualization, Writing – original draft, Writing – review & editing. AM: Conceptualization, Visualization, Writing – original draft, Writing – review & editing.

Funding

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.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2025.1610662/full#supplementary-material

References

1. Robinson WH, Younis S, Love ZZ, Steinman L, Lanz TV. Epstein–Barr virus as a potentiator of autoimmune diseases. *Nat Rev Rheumatol*. (2024) 20:729–40. doi: 10.1038/s41584-024-01167-9

2. Ebert AC, Harper S, Vestergaard MV, Mitchell W, Jess T, Elmahdi R. Risk of inflammatory bowel disease following hospitalization with infectious mononucleosis: nationwide cohort study from Denmark. *Nat Commun*. (2024) 15:8383. doi: 10.1038/s41467-024-52195-8

3. Nandy A, Petralia F, Porter C, Elledge S, Anand R, Croitoru K, et al. Epstein-barr virus (EBV) exposure precedes crohn`s disease development. *Gastroenterology*. (2025). doi: 10.1053/j.gastro.2025.01.247

4. Ascherio A, Munger KL, Lennette ET, Spiegelman D, Hernán MA, Olek MJ, et al. Epstein-barr virus antibodies and risk of multiple sclerosisA prospective study. *JAMA*. (2001) 286:3083–8. doi: 10.1001/jama.286.24.3083

5. Bjornevik K, Cortese M, Healy BC, Kuhle J, Mina MJ, Leng Y, et al. Longitudinal analysis reveals high prevalence of Epstein-Barr virus associated with multiple sclerosis. *Sci (1979)*. (2022) 375:296–301. doi: 10.1126/science.abj8222

6. Bjornevik K, Münz C, Cohen JI, Ascherio A. Epstein–Barr virus as a leading cause of multiple sclerosis: mechanisms and implications. *Nat Rev Neurol*. (2023) 19:160–71. doi: 10.1038/s41582-023-00775-5

7. Fechtner S, Berens H, Bemis E, Johnson RL, Guthridge CJ, Carlson NE, et al. Antibody responses to epstein-barr virus in the preclinical period of rheumatoid arthritis suggest the presence of increased viral reactivation cycles. *Arthritis Rheumatol.* (2022) 74:597–603. doi: 10.1002/art.41994

8. Moon UY, Park SJ, Oh ST, Kim W-U, Park S-H, Lee S-H, et al. Patients with systemic lupus erythematosus have abnormally elevated Epstein–Barr virus load in blood. *Arthritis Res Ther.* (2004) 6:R295. doi: 10.1186/ar1181

9. Pasoto SG, Natalino RR, Chakkour HP, Viana V dos ST, Bueno C, Leon EP, et al. EBV reactivation serological profile in primary Sjögren's syndrome: an underlying trigger of active articular involvement? *Rheumatol Int.* (2013) 33:1149–57. doi: 10.1007/s00296-012-2504-3

10. Lanz TV, Brewer RC, Ho PP, Moon J-S, Jude KM, Fernandez D, et al. Clonally expanded B cells in multiple sclerosis bind EBV EBNA1 and GlialCAM. *Nature.* (2022) 603:321–7. doi: 10.1038/s41586-022-04432-7

11. Csorba K, Schirmbeck LA, Tuncer E, Ribi C, Roux-Lombard P, Chizzolini C, et al. Anti-C1q antibodies as occurring in systemic lupus erythematosus could be induced by an epstein-barr virus-derived antigenic site. *Front Immunol.* (2019) 10. doi: 10.3389/fimmu.2019.02619

12. Liu J, Gao J, Wu Z, Mi L, Li N, Wang Y, et al. Anti-citrullinated protein antibody generation, pathogenesis, clinical application, and prospects. *Front Med (Lausanne).* (2022) 8-2021. doi: 10.3389/fmed.2021.802934

13. Rosati E, Martini GR, Pogorelyy MV, Minervina AA, Degenhardt F, Wendorff M, et al. A novel unconventional T cell population enriched in Crohn's disease. *Gut.* (2022) 71:2194–204. doi: 10.1136/gutjnl-2021-325373

14. Mahdy A, ElAbd H, Kriukova V, Olbjørn C, Perminow G, Bengtson MB, et al. P0125 Crohn's-associated invariant T Cells are associated with disease severity and location and are not affected by medication intake. *J Crohns Colitis.* (2025) 19:i507–8. doi: 10.1093/ecco-jcc/jjae190.0299

15. Krishna M, Spartz EJ, Maas L, Cusumano V, Sharma S, Limketkai B, et al. Retrospective cohort study on the predictive value of primary non-response to initial biologic for future biologic use in patients with inflammatory bowel disease. *Dig Dis Sci.* (2025), 70. doi: 10.1007/s10620-024-08816-9

16. Papamichael K, Gils A, Rutgeerts P, Levesque BG, Vermeire S, Sandborn WJ, et al. Role for therapeutic drug monitoring during induction therapy with TNF antagonists in IBD: evolution in the definition and management of primary nonresponse. *Inflammation Bowel Dis.* (2015) 21:182–97. doi: 10.1097/MIB.0000000000000202

17. Roda G, Jharap B, Neeraj N, Colombel J-F. Loss of response to anti-TNFs: definition, epidemiology, and management. *Clin Transl Gastroenterol.* (2016) 7. doi: 10.1038/ctg.2015.63

18. Alsoud D, Verstockt B, Fiocchi C, Vermeire S. Breaking the therapeutic ceiling in drug development in ulcerative colitis. *Lancet Gastroenterol Hepatol.* (2021) 6:589–95. doi: 10.1016/S2468-1253(21)00065-0

19. Gisbert JP, Chaparro M. Predictors of primary response to biologic treatment [Anti-TNF, vedolizumab, and ustekinumab] in patients with inflammatory bowel disease: from basic science to clinical practice. *J Crohns Colitis.* (2020) 14:694–709. doi: 10.1093/ecco-jcc/jjz195

20. Hracs L, Windsor JW, Gorospe J, Cummings M, Coward S, Buie MJ, et al. Global evolution of inflammatory bowel disease across epidemiologic stages. *Nature.* (2025). doi: 10.1038/s41586-025-08940-0

21. Coward S, Clement F, Benchimol EI, Bernstein CN, Avina-Zubieta JA, Bitton A, et al. Past and future burden of inflammatory bowel diseases based on modeling of population-based data. *Gastroenterology.* (2019) 156:1345–1353.e4. doi: 10.1053/j.gastro.2019.01.002

22. Coward S, Benchimol EI, Bernstein CN, Avina-Zubieta A, Bitton A, Carroll MW, et al. Forecasting the incidence and prevalence of inflammatory bowel disease: a Canadian nationwide analysis. *Off J Am Coll Gastroenterology| ACG.* (2024) 119:1563–70. doi: 10.14309/ajg.0000000000002687

23. Zwibel HL, Smrtka J. Improving quality of life in multiple sclerosis: an unmet need. *Am J Managed Care.* (2011) 17:S139.

24. Hittle M, Culpepper WJ, Langer-Gould A, Marrie RA, Cutter GR, Kaye WE, et al. Population-based estimates for the prevalence of multiple sclerosis in the United States by race, ethnicity, age, sex, and geographic region. *JAMA Neurol.* (2023) 80:693–701. doi: 10.1001/jamaneurol.2023.1135

25. Reich DS, Lucchinetti CF, Calabresi PA. Multiple Sclerosis. *N Engl J Med.* (2018) 378:169–80. doi: 10.1056/NEJMra1401483

26. Portaccio E, Magyari M, Havrdova EK, Ruet A, Brochet B, Scalfari A, et al. Multiple sclerosis: emerging epidemiological trends and redefining the clinical course. *Lancet Regional Health - Europe.* (2024) 44:100977. doi: 10.1016/j.lanepe.2024.100977

27. Black RJ, Cross M, Haile LM, Culbreth GT, Steinmetz JD, Hagins H, et al. Global, regional, and national burden of rheumatoid arthritis, 1990–2020, and projections to 2050: a systematic analysis of the Global Burden of Disease Study 2021. *Lancet Rheumatol.* (2023) 5:e594–610. doi: 10.1016/S2665-9913(23)00211-4

28. Tian J, Zhang D, Yang Y, Huang Y, Wang L, Yao X, et al. Global epidemiology of atopic dermatitis: a comprehensive systematic analysis and modelling study. *Br J Dermatol.* (2024) 190:55–61. doi: 10.1093/bjd/ljad339

29. de Lange KM, Moutsianas L, Lee JC, Lamb CA, Luo Y, Kennedy NA, et al. Genome-wide association study implicates immune activation of multiple integrin genes in inflammatory bowel disease. *Nat Genet.* (2017) 49:256–61. doi: 10.1038/ng.3760

30. Degenhardt F, Mayr G, Wendorff M, Boucher G, Ellinghaus E, Ellinghaus D, et al. Trans-ethnic analysis of the human leukocyte antigen region for ulcerative colitis reveals shared but also ethnicity-specific disease associations. *Hum Mol Genet.* (2021). doi: 10.1093/hmg/ddab017

31. Goyette P, Boucher G, Mallon D, Ellinghaus E, Jostins L, Huang H, et al. High-density mapping of the MHC identifies a shared role for HLA-DRB1*01:03 in inflammatory bowel diseases and heterozygous advantage in ulcerative colitis. *Nat Genet.* (2015) 47:172–9. doi: 10.1038/ng.3176

32. Ishigaki K, Sakaue S, Terao C, Luo Y, Sonehara K, Yamaguchi K, et al. Multi-ancestry genome-wide association analyses identify novel genetic mechanisms in rheumatoid arthritis. *Nat Genet.* (2022) 54:1640–51. doi: 10.1038/s41588-022-01213-w

33. Budu-Aggrey A, Kilanowski A, Sobczyk MK, Shringarpure SS, Mitchell R, Reis K, et al. European and multi-ancestry genome-wide association meta-analysis of atopic dermatitis highlights importance of systemic immune regulation. *Nat Commun.* (2023) 14:6172. doi: 10.1038/s41467-023-41180-2

34. Patsopoulos NA, the Bayer Pharma MS Genetics Working Group ANZgene Consortium GeneMSA International Multiple Sclerosis Genetics Consortium the SC of SEIfn-1b and a C-A, de Bakker PIW. Genome-wide meta-analysis identifies novel multiple sclerosis susceptibility loci. *Ann Neurol.* (2011) 70:897–912. doi: 10.1002/ana.22609

35. Consortium*† IMSG, ANZgene, IIBDGC, WTCCC2. Multiple sclerosis genomic map implicates peripheral immune cells and microglia in susceptibility. *Sci (1979).* (2019) 365:eaav7188. doi: 10.1126/science.aav7188

36. Dand N, Stuart PE, Bowes J, Ellinghaus D, Nititham J, Saklatvala JR, et al. GWAS meta-analysis of psoriasis identifies new susceptibility alleles impacting disease mechanisms and therapeutic targets. *Nat Commun.* (2025) 16:2051. doi: 10.1038/s41467-025-56719-8

37. Stürner KH, Siembab I, Schön G, Stellmann J-P, Heidari N, Fehse B, et al. Is multiple sclerosis progression associated with the HLA-DR15 haplotype? *Multiple sclerosis journal–experimental Trans Clin.* (2019) 5:2055217319894615. doi: 10.1177/2055217319894615

38. Kular L, Liu Y, Ruhrmann S, Zheleznyakova G, Marabita F, Gomez-Cabrero D, et al. DNA methylation as a mediator of HLA-DRB1*15:01 and a protective variant in multiple sclerosis. *Nat Commun.* (2018) 9:2397. doi: 10.1038/s41467-018-04732-5

39. Hollenbach JA, Oksenberg JR. The immunogenetics of multiple sclerosis: A comprehensive review. *J Autoimmun.* (2015) 64:13–25. doi: 10.1016/j.jaut.2015.06.010

40. Almeida LM, Gandolfi L, Pratesi R, Uenishi RH, de Almeida FC, Selleski N, et al. Presence of DQ2.2 associated with DQ2.5 increases the risk for celiac disease. *Autoimmune Dis.* (2016) 2016:5409653. doi: 10.1155/2016/5409653

41. Vader W, Stepniak D, Kooy Y, Mearin L, Thompson A, van Rood JJ, et al. The HLA-DQ2 gene dose effect in celiac disease is directly related to the magnitude and breadth of gluten-specific T cell responses. *Proc Natl Acad Sci.* (2003) 100:12390–5. doi: 10.1073/pnas.2135229100

42. Merkenschlager J, Pyo AGT, Silva Santos GS, Schaefer-Babajew D, Cipolla M, Hartweger H, et al. Regulated somatic hypermutation enhances antibody affinity maturation. *Nature.* (2025) 641:495–502. doi: 10.1038/s41586-025-08728-2

43. Hoehn KB, Kleinstein SH. B cell phylogenetics in the single cell era. *Trends Immunol.* (2024) 45:62–74. doi: 10.1016/j.it.2023.11.004

44. Akkaya M, Kwak K, Pierce SK. B cell memory: building two walls of protection against pathogens. *Nat Rev Immunol.* (2020) 20:229–38. doi: 10.1038/s41577-019-0244-2

45. Lobach DF, Haynes BF. Ontogeny of the human thymus during fetal development. *J Clin Immunol.* (1987) 7:81–97. doi: 10.1007/BF00916002

46. Darrasse-Jeze G, Marodon G, Salomon BL, Catala M, Klatzmann D. Ontogeny of CD4+ CD25+ regulatory/suppressor T cells in human fetuses. *Blood.* (2005) 105:4715–21. doi: 10.1182/blood-2004-10-4051

47. Rackaityte E, Halkias J. Mechanisms of fetal T cell tolerance and immune regulation. *Front Immunol.* (2020) 11-2020. doi: 10.3389/fimmu.2020.00588

48. Mahdy AKH, Lokes E, Schöpfel V, Kriukova V, Britanova OV, Steiert TA, et al. Bulk T cell repertoire sequencing (TCR-Seq) is a powerful technology for understanding inflammation-mediated diseases. *J Autoimmun.* (2024) 149:103337. doi: 10.1016/j.jaut.2024.103337

49. Pai JA, Satpathy AT. High-throughput and single-cell T cell receptor sequencing technologies. *Nat Methods.* (2021) 18:881–92. doi: 10.1038/s41592-021-01201-8

50. Singh M, Al-Eryani G, Carswell S, Ferguson JM, Blackburn J, Barton K, et al. High-throughput targeted long-read single cell sequencing reveals the clonal and transcriptional landscape of lymphocytes. *Nat Commun.* (2019) 10:3120. doi: 10.1038/s41467-019-11049-4

51. Howie B, Sherwood AM, Berkebile AD, Berka J, Emerson RO, Williamson DW, et al. High-throughput pairing of T cell receptor α and β sequences. *Sci Transl Med.* (2015) 7:301ra131–301ra131. doi: 10.1126/scitranslmed.aac5624

52. Pogorelyy MV, Kirk AM, Adhikari S, Minervina AA, Sundararaman B, Vegesana K, et al. TIRTL-seq: Deep, quantitative, and affordable paired TCR repertoire sequencing. *bioRxiv.* (2024). doi: 10.1101/2024.09.16.613345

53. Corcoran M, Chernyshev M, Mandolesi M, Narang S, Kaduk M, Ye K, et al. Archaic humans have contributed to large-scale variation in modern human T cell receptor genes. *Immunity.* (2023) 56:635–652.e6. doi: 10.1016/j.immuni.2023.01.026

54. Rodriguez OL, Safonova Y, Silver CA, Shields K, Gibson WS, Kos JT, et al. Genetic variation in the immunoglobulin heavy chain locus shapes the human antibody repertoire. *Nat Commun*. (2023) 14:4419. doi: 10.1038/s41467-023-40070-x

55. Pushparaj P, Nicoletto A, Sheward DJ, Das H, Castro Dopico X, Perez Vidakovics L, et al. Immunoglobulin germline gene polymorphisms influence the function of SARS-CoV-2 neutralizing antibodies. *Immunity*. (2023) 56:193–206.e7. doi: 10.1016/j.immuni.2022.12.005

56. Ishigaki K, Lagattuta KA, Luo Y, James EA, Buckner JH, Raychaudhuri S. HLA autoimmune risk alleles restrict the hypervariable region of T cell receptors. *Nat Genet*. (2022) 54:393–402. doi: 10.1038/s41588-022-01032-z

57. Sharon E, Sibener LV, Battle A, Fraser HB, Garcia KC, Pritchard JK. Genetic variation in MHC proteins is associated with T cell receptor expression biases. *Nat Genet*. (2016) 48:995–1002. doi: 10.1038/ng.3625

58. Zahid HJ, Taniguchi R, Ebert P, Chow I-T, Gooley C, Lv J, et al. Large-scale statistical mapping of T-cell receptor β sequences to Human Leukocyte Antigens. *bioRxiv*. (2024). doi: 10.1101/2024.04.01.587617

59. Russell ML, Souquette A, Levine DM, Schattgen SA, Allen EK, Kuan G, et al. Combining genotypes and T cell receptor distributions to infer genetic loci determining V(D)J recombination probabilities. *Elife*. (2022) 11:e73475. doi: 10.7554/eLife.73475

60. ElAbd H, Degenhardt F, Koudelka T, Kamps A-K, Tholey A, Bacher P, et al. Immunopeptidomics toolkit library (IPTK): a python-based modular toolbox for analyzing immunopeptidomics data. *BMC Bioinf*. (2021) 22:405. doi: 10.1186/s12859-021-04315-0

61. Erhard F, Dölken L, Schilling B, Schlosser A. Identification of the cryptic HLA-I immunopeptidome. *Cancer Immunol Res*. (2020) 8:1018–26. doi: 10.1158/2326-6066.CIR-19-0886

62. Purcell AW, Ramarathinam SH, Ternette N. Mass spectrometry–based identification of MHC-bound peptides for immunopeptidomics. *Nat Protoc*. (2019) 14:1687–707. doi: 10.1038/s41596-019-0133-y

63. ElAbd H, Franke A, Schrader M, Fricker LD. editors. *Peptidomics: Methods Strategies*. (2024) 425–43. doi: 10.1007/978-1-0716-3646-6_23

64. Altin JA, Tian L, Liston A, Bertram EM, Goodnow CC, Cook MC. Decreased T-cell receptor signaling through CARD11 differentially compromises forkhead box protein 3x2013;positive regulatory versus T$_H$2 effector cells to cause allergy. *J Allergy Clin Immunol*. (2011) 127:1277–1285.e5. doi: 10.1016/j.jaci.2010.12.1081

65. Vallois D, Dobay MPD, Morin RD, Lemonnier F, Missiaglia E, Juilland M, et al. Activating mutations in genes related to TCR signaling in angioimmunoblastic and other follicular helper T-cell–derived lymphomas. *Blood*. (2016) 128:1490–502. doi: 10.1182/blood-2016-02-698977

66. Hernández-Hoyos G, Anderson MK, Wang C, Rothenberg EV, Alberola-Ila J. GATA-3 expression is controlled by TCR signals and regulates CD4/CD8 differentiation. *Immunity*. (2003) 19:83–94. doi: 10.1016/S1074-7613(03)00176-6

67. Lagattuta KA, Kang JB, Nathan A, Pauken KE, Jonsson AH, Rao DA, et al. Repertoire analyses reveal T cell antigen receptor sequence features that influence T cell fate. *Nat Immunol*. (2022) 23:446–57. doi: 10.1038/s41590-022-01129-x

68. ElAbd H, Mahdy A, Wacker EM, Gretsova M, Ellinghaus D, Franke A. Decoding the restriction of T cell receptors to human leukocyte antigen alleles using statistical learning. *bioRxiv*. (2025), 2022–25. doi: 10.1038/s41590-022-01129-x

69. Lokes E, Franke A, Mayr G, ElAbd H. OP26 Disentangling the role of HLA proteins in shaping the T-cell repertoire in Inflammatory Bowel Disease using CDR3-QTL analysis. *J Crohns Colitis*. (2025) 19:i51–2. doi: 10.1093/ecco-jcc/jjae190.0026

70. Kula T, Dezfulian MH, Wang CI, Abdelfattah NS, Hartman ZC, Wucherpfennig KW, et al. T-scan: A genome-wide method for the systematic discovery of T cell epitopes. *Cell*. (2019) 178:1016–1028.e13. doi: 10.1016/j.cell.2019.07.009

71. Dezfulian MH, Kula T, Pranzatelli T, Kamitaki N, Meng Q, Khatri B, et al. TScan-II: A genome-scale platform for the *de novo* identification of CD4+ T cell epitopes. *Cell*. (2023) 186:5569–5586.e21. doi: 10.1016/j.cell.2023.10.024

72. Dobson CS, Reich AN, Gaglione S, Smith BE, Kim EJ, Dong J, et al. Antigen identification and high-throughput interaction mapping by reprogramming viral entry. *Nat Methods*. (2022) 19:449–60. doi: 10.1038/s41592-022-01436-z

73. Emerson RO, DeWitt WS, Vignali M, Gravley J, Hu JK, Osborne EJ, et al. Immunosequencing identifies signatures of cytomegalovirus exposure history and HLA-mediated effects on the T cell repertoire. *Nat Genet*. (2017) 49:659–65. doi: 10.1038/ng.3822

74. May DH, Woodhouse S, Zahid HJ, Elyanow R, Doroschak K, Noakes MT, et al. Identifying immune signatures of common exposures through co-occurrence of T-cell receptors in tens of thousands of donors. *bioRxiv*. (2024). doi: 10.1101/2024.03.26.583354

75. DeWitt WSIII, Smith A, Schoch G, Hansen JA, Matsen FAIV, Bradley P. Human T cell receptor occurrence patterns encode immune history, genetic background, and receptor specificity. *Elife*. (2018) 7:e38358. doi: 10.7554/eLife.38358

76. Greissl J, Pesesky M, Dalai SC, Rebman AW, Soloski MJ, Horn EJ, et al. Immunosequencing of the T-cell receptor repertoire reveals signatures specific for identification and characterization of early lyme disease. *medRxiv*. (2022). doi: 10.1101/2021.07.30.21261353

77. Littera R, Campagna M, Deidda S, Angioni G, Cipri S, Melis M, et al. Human leukocyte antigen complex and other immunogenetic and clinical factors influence susceptibility or protection to SARS-coV-2 infection and severity of the disease course. The sardinian experience. *Front Immunol*. (2020) 11:605688. doi: 10.3389/fimmu.2020.605688

78. Faham M, Carlton V, Moorhead M, Zheng J, Klinger M, Pepin F, et al. Discovery of T cell receptor β motifs specific to HLA–B27–positive ankylosing spondylitis by deep repertoire sequence analysis. *Arthritis Rheumatol*. (2017) 69:774–84. doi: 10.1002/art.40028

79. Mohan D, Wansley DL, Sie BM, Noon MS, Baer AN, Laserson U, et al. PhIP-Seq characterization of serum antibodies using oligonucleotide-encoded peptidomes. *Nat Protoc*. (2018) 13:1958–78. doi: 10.1038/s41596-018-0025-6

80. Larman HB, Zhao Z, Laserson U, Li MZ, Ciccia A, Gakidis MAM, et al. Autoantigen discovery with a synthetic human peptidome. *Nat Biotechnol*. (2011) 29:535–41. doi: 10.1038/nbt.1856

81. Leviatan S, Vogl T, Klompus S, Kalka IN, Weinberger A, Segal E. Allergenic food protein consumption is associated with systemic IgG antibody responses in non-allergic individuals. *Immunity*. (2022) 55:2454–2469.e6. doi: 10.1016/j.immuni.2022.11.004

82. Vogl T, Klompus S, Leviatan S, Kalka IN, Weinberger A, Wijmenga C, et al. Population-wide diversity and stability of serum antibody epitope repertoires against human microbiota. *Nat Med*. (2021) 27:1442–50. doi: 10.1038/s41591-021-01409-3

83. Andreu-Sánchez S, Bourgonje AR, Vogl T, Kurilshikov A, Leviatan S, Ruiz-Moreno AJ, et al. Phage display sequencing reveals that genetic, environmental, and intrinsic factors influence variation in human antibody epitope repertoire. *Immunity*. (2023) 56:1376–1392.e8. doi: 10.1016/j.immuni.2023.04.003

84. Pesesky M, Bharanikumar R, Le Bourhis L, ElAbd H, Rosati E, Carty CL, et al. Antigen-driven expansion of public clonal T cell populations in inflammatory bowel diseases. *J Crohns Colitis*. (2025), jjaf048. doi: 10.1093/ecco-jcc/jjaf048

85. Scholtens S, Smidt N, Swertz MA, Bakker SJL, Dotinga A, Vonk JM, et al. Cohort Profile: LifeLines, a three-generation cohort study and biobank. *Int J Epidemiol*. (2015) 44:1172–80. doi: 10.1093/ije/dyu229

86. Consortium GNC. The German National Cohort: aims, study design and organization. *Eur J Epidemiol*. (2014) 29:371–82. doi: 10.1007/s10654-014-9890-7

87. Grännö O, Bergemalm D, Salomon B, Lindqvist CM, Hedin CRH, Carlson M, et al. Preclinical protein signatures of Crohn's disease and ulcerative colitis: A nested case-control study within large population-based cohorts. *Gastroenterology*. (2025) 168 (4):741–53. doi: 10.1053/j.gastro.2024.11.006

88. Rausch P, Ratjen I, Tittmann L, Enderle J, Wacker EM, Jaeger K, et al. First Insights into microbial changes within an Inflammatory Bowel Disease Family Cohort study. *medRxiv*. (2024). doi: 10.1101/2024.07.23.24310327

89. Vietzen H, Berger SM, Kühner LM, Furlano PL, Bsteh G, Berger T, et al. Ineffective control of Epstein-Barr-virus-induced autoimmunity increases the risk for multiple sclerosis. *Cell*. (2023) 186:5705–5718.e13. doi: 10.1016/j.cell.2023.11.015

90. Balandraud N, Roudier J. Epstein-Barr virus and rheumatoid arthritis. *Joint Bone Spine*. (2018) 85:165–70. doi: 10.1016/j.jbspin.2017.04.011

91. Loosen SH, Kostev K, Schöler D, Orth H-M, Freise NF, Jensen B-EO, et al. Infectious mononucleosis is associated with an increased incidence of Crohn's disease: results from a cohort study of 31–862 outpatients in Germany. *Eur J Gastroenterol Hepatol*. (2023) 35:255–60. doi: 10.1097/MEG.0000000000002505

92. Jayasooriya S, de Silva TI, Njie-jobe J, Sanyang C, Leese AM, Bell AI, et al. Early virological and immunological events in asymptomatic epstein-barr virus infection in african children. *PLoS Pathog*. (2015) 11:e1004746. doi: 10.1371/journal.ppat.1004746

93. Saggau C, Martini GR, Rosati E, Meise S, Messner B, Kamps A-K, et al. The pre-exposure SARS-CoV-2-specific T cell repertoire determines the quality of the immune response to vaccination. *Immunity*. (2022) 55:1924–1939.e5. doi: 10.1016/j.immuni.2022.08.003

94. Komech EA, Pogorelyy MV, Egorov ES, Britanova OV, Rebrikov DV, Bochkova AG, et al. CD8+ T cells with characteristic T cell receptor beta motif are detected in blood and expanded in synovial fluid of ankylosing spondylitis patients. *Rheumatology*. (2018) 57:1097–104. doi: 10.1093/rheumatology/kex517

95. Britanova OV, Lupyr KR, Staroverov DB, Shagina IA, Aleksandrov AA, Ustyugov YY, et al. Targeted depletion of TRBV9+ T cells as immunotherapy in a patient with ankylosing spondylitis. *Nat Med*. (2023) 29:2731–6. doi: 10.1038/s41591-023-02613-z

96. Aspasia S, Emmanuela-Kalliopi K, NikoLaos T, Eirini S, Ioannis S, Anastasia M. The gluten-free diet challenge in adults with coeliac disease: the Hellenic survey. *PEC Innovation*. (2022) 1:100037. doi: 10.1016/j.pecinn.2022.100037

97. Poslt Königová M, Sebalo Vňuková M, Řehořková P, Anders M, Ptáček R. The effectiveness of gluten-free dietary interventions: A systematic review. *Front Psychol*. (2023) 14-2023. doi: 10.3389/fpsyg.2023.1107022

98. Scalvini D, Scarcella C, Mantica G, Bartolotta E, Maimaris S, Fazzino E, et al. Beyond gluten-free diet: a critical perspective on phase 2 trials on non-dietary

pharmacological therapies for coeliac disease. *Front Nutr.* (2025) 11-2024. doi: 10.3389/fnut.2024.1501817

99. Krienke C, Kolb L, Diken E, Streuber M, Kirchhoff S, Bukur T, et al. A noninflammatory mRNA vaccine for treatment of experimental autoimmune encephalomyelitis. *Sci (1979).* (2021) 371:145–53. doi: 10.1126/science.aay3638

100. Zhong L, Zhao Q, Zeng M-S, Zhang X. Prophylactic vaccines against Epstein&x2013;Barr virus. *Lancet.* (2024) 404:845. doi: 10.1016/S0140-6736(24)01608-8

101. Ellebrecht CT, Bhoj VG, Nace A, Choi EJ, Mao X, Cho MJ, et al. Reengineering chimeric antigen receptor T cells for targeted therapy of autoimmune disease. *Sci (1979).* (2016) 353:179–84. doi: 10.1126/science.aaf6756