



## OPEN ACCESS

## EDITED BY

Ting Wang,  
University of Washington, United States

## REVIEWED BY

Paul Lyons,  
University of Cambridge, United Kingdom  
Joao Henrique Diniz Brandao Gervasio,  
Okinawa Institute of Science and Technology  
Graduate University Promotion Corp, Japan

## \*CORRESPONDENCE

Andre Franke

✉ a.franke@ikmb.uni-kiel.de

RECEIVED 15 October 2025

REVISED 06 January 2026

ACCEPTED 19 January 2026

PUBLISHED 06 February 2026

## CITATION

Mahdy AKH, Taheri Z, Høivik ML, Franke A and ElAbd H (2026) High throughput profiling of the B cell repertoire identifies systematic changes in the repertoire of individuals with Crohn's disease.

*Front. Immunol.* 17:1725813.

doi: 10.3389/fimmu.2026.1725813

## COPYRIGHT

© 2026 Mahdy, Taheri, Høivik, Franke and ElAbd. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# High throughput profiling of the B cell repertoire identifies systematic changes in the repertoire of individuals with Crohn's disease

Aya K. H. Mahdy<sup>1</sup>, Zahra Taheri<sup>1,2</sup>, Marte Lie Høivik<sup>3,4</sup>, Andre Franke<sup>1\*</sup> and Hesham ElAbd<sup>1,5</sup>

<sup>1</sup>Institute of Clinical Molecular Biology, Kiel University and University Hospital Schleswig-Holstein, Kiel, Germany, <sup>2</sup>Department of Biology and Biotechnology, University of Pavia, Pavia, Italy, <sup>3</sup>Department of Gastroenterology, Oslo University Hospital, Oslo, Norway, <sup>4</sup>Institute of Clinical Medicine, Faculty of Medicine, University of Oslo, Oslo, Norway, <sup>5</sup>Institute for Digestive Research, Lithuanian University of Health Sciences, Kaunas, Lithuania

The B cell repertoire contains the recombined sequences that encode the entire antibody repertoire of an individual. The repertoire is made from three antigenic binding chains, namely the immunoglobulin heavy chain (IGH) and two immunoglobulin light chains,  $\kappa$  (IGK) and  $\lambda$  (IGL). Compared to the T cell repertoire, the B cell repertoire is understudied in inflammatory bowel diseases (IBD) even though different antibodies such as ASCA (Anti-Saccharomyces cerevisiae) and ANCA (Anti-Neutrophil Cytoplasmic Antibodies) have been shown to be elevated in individuals with IBD. To address this limitation, we profiled the B cell repertoire of peripheral blood from 27 treatment-naïve individuals with CD and 21 age-matched symptomatic controls using bulk B cell receptor sequencing. The repertoire of individuals with CD showed a reduction in diversity and an increase in clonality. Furthermore, we observed a significant reduction in the expansion of IgM and IgD and an expansion of IgA2, and IgG2 clonotypes in individuals with CD relative to controls, suggesting an antigen-driven expansion. This was also supported by higher levels of somatic hypermutations, particularly in the complementary determining region 2 (CDR2) of immunoglobulin heavy chain, in individuals with CD relative to the control group. Thus, despite the small sample size, we identified multiple alterations in the B cell repertoire of individuals with CD, highlighting the potential of the B cell repertoire in identifying antigenic exposures implicated in the disease, demanding now larger international studies, ideally including also treatment-naïve and pre-clinical cases.

## KEYWORDS

B cell, B cell repertoire sequencing, IBD, Ig isotype, somatic hypermutation, treatment-naïve

## Introduction

Inflammatory bowel disease (IBD) is an incurable immune-mediated inflammatory disease that predominantly affects the gastrointestinal tract (GIT). It is observed clinically in two main forms: Crohn's disease (CD), which is characterized by patchy transmural inflammation of different sections of the GIT, predominantly the ileum and the colon, and ulcerative colitis (UC), which is restricted to the colon. The exact cause(s) of IBD remain unknown, however, different genetic variants have been associated with the disease, such as *ATG16L* (1, 2), and *NOD2* (3). Additionally, several human leukocyte antigen (HLA) alleles have been implicated in IBD, such as HLA-DRB1\*03:01 (4), or in a specific subset of individuals with ileal CD, such as HLA-DRB1\*07:01 (4) and HLA-DRB1\*15:01 in UC (5). Besides genetic predispositions, other risk factors have been identified, e.g., smoking, microbial dysbiosis (6), antibiotic intake (7), and previous episodes of infectious mononucleosis (8), which is mainly caused by an Epstein-Barr virus (EBV) infection.

From an immunological perspective, different alterations and dysregulated processes have been identified in individuals with IBD, including dysregulated responses toward the gut microbiome (9) and mycobiome (10). We previously observed a significant expansion of a subset of type II natural killer T cells, termed Crohn's-associated invariant T (CAIT) cells, in individuals with CD (11, 12). Elevated antibody responses against bacterial flagellins (13) and several human herpesviruses, predominantly EBV, have been reported by others (14). Recently, we also performed a large-scale analysis of the T cell repertoire of individuals with IBD in comparison to matching controls, identifying thousands of clonotypes that were significantly expanded in individuals with IBD (15, 16). These disease-associated clonotypes represent an immunological fingerprint for common antigenic exposures implicated in the disease. Nonetheless, these antigens remain to be identified as T cell repertoire sequencing enables the identification of clonotypes, i.e., V(D)J recombination sequences forming the T cell receptor, and not the exact antigens presented by HLA proteins and recognized by these T cells (17).

The dependency of T cells on HLA proteins, which are highly polymorphic (18), renders analysing the T cell repertoire a challenging task, specifically in case-control studies where thousands of samples are needed to pinpoint specific T cells involved in the disease across different HLA contexts (15, 16, 19). Thus, identifying B cells recognizing the same antigens across multiple individuals with CD might require a smaller sample size as B cell receptors (BCRs) recognise their antigenic targets directly, independent of any presenting molecules, e.g., HLA proteins. Relative to the T cell repertoire, which has been studied by others (20–23) and us (11, 15, 19), the B cell repertoire of individuals with IBD remains under-investigated.

The B cell repertoire is a composite of all immunoglobulins heavy and light chains present in a sample, such as peripheral blood. These chains are generated from somatic recombination processes termed V(D)J recombination that generate chains with enormous sequence diversity. This diversity is also augmented by a unique

process that happens in B cells, termed somatic hypermutation (SHM), where random single-nucleotide polymorphisms (SNPs) are introduced into the formed immunoglobulin chains to increase their affinity toward a particular antigen. The Ig heavy chain (IGH) has five main isotypes, namely,  $\mu$ ,  $\delta$ ,  $\epsilon$ ,  $\alpha$ , and  $\gamma$ , which form the IgM, IgD, IgE, IgA, and IgG, respectively. Furthermore, the  $\alpha$  and  $\gamma$  isotypes have different subclasses, namely,  $\alpha 1$  and  $\alpha 2$  that encode IgA1 and IgA2, and  $\gamma 1$ ,  $\gamma 2$ ,  $\gamma 3$ , and  $\gamma 4$ , which encode the IgG1, IgG2, IgG3, and IgG4 subtypes. There are two light chains,  $\kappa$  (IGK) and  $\lambda$  (IGL), but they only have one constant region and do not undergo class-switching.

Besides the seminal study by Bashford-Rogers et al. (24), which compared the BCR repertoire of six different autoimmune diseases, not many studies have investigated the immune repertoire in individuals with CD or UC. Scheid et al. (25) profiled the B cell repertoire of colonic tissues from individuals with UC and showed a bias from IgA1 and IgA2 isotype usage toward IgG2 usage in inflamed tissues. Similarly, Chen and colleagues (26) profiled the B cell repertoire of multiple tissues in individuals with IBD and showed dysregulated B cell responses in individuals with CD. Lastly, Kotagiri et al. (27) also profiled the repertoire of individuals with IBD and healthy controls and identified BCRs shared among individuals with CD.

Here, we aimed at investigating the B cell receptor of 27 treatment-naive individuals with CD and 21 matching symptomatic controls from the Norwegian inception cohort IBSEN-III (28). Symptomatic controls are individuals with symptoms of inflammatory bowel disease, but their endoscopic and radiological results excluded IBD as a potential cause for their symptoms.

## Results

### The B cell repertoire of treatment-naive individuals with CD exhibits systematic differences in clonality and diversity

We profiled the peripheral blood B cell receptor repertoire of 27 treatment-naive individuals with Crohn's disease (CD) and 21 symptomatic controls from the Norwegian inception cohort IBSEN-III (Table 1, Figure 1A). RNA was extracted from PAXgene Blood RNA tubes, and IGH, IGK and IGL transcripts were amplified by multiplex PCR and sequenced in pooled libraries. Demultiplexed reads were processed with MiXCR (29) to assemble productive clonotypes derived from the three loci (Methods). We first assessed per-sample repertoire diversity and inequality across IGH, IGK, and IGL. Shannon diversity was reduced in CD relative to SC for IGH and IGL, while IGK showed no significant difference (Figures 1B–D), indicating a selective loss of repertoire evenness rather than a uniform effect across loci. Chao1 richness did not differ significantly between groups for any locus (Figures 1E–G), arguing against a broad reduction in the number of detectable clonotypes in CD. In contrast, the Gini coefficient, a direct measure of inequality in clonotype abundance, was higher in CD for IGH and IGK but not for IGL (Figures 1H–J), consistent with increased dominance of expanded clonotypes in CD. Together, these results

TABLE 1 Clinical and demographic characteristics of the study cohort.

	Crohn's disease	Symptomatic controls
Number of individuals	27	21
Age (mean $\pm$ SD)	(27 $\pm$ 6.2)	(27.9 $\pm$ 6.57)
Percentage of females	44.4% (12 out of 27)	71.4% (15 out of 21)
Disease location (Number of individuals)	L1 (Ileum): (n=6) L2 (Colon): (n=7) L3 (Ileocolon): (n=14) L4 (Upper GI tract): (n=2)	NA
Disease modifiers (Number of individuals)	B1 (non-stricturing, non-penetrating): (n=16) B2 (Stricturing): (n=7) B3 (Penetrating): (n=2)	NA

Summary of participant metadata for individuals with Crohn's disease ( $n = 27$ ) and symptomatic controls ( $n = 21$ ), including age (mean  $\pm$  SD) and sex distribution. Symptomatic controls were individuals presenting with symptoms suggestive of inflammatory bowel disease, but with endoscopic and radiological evaluations excluding IBD as the cause of their symptoms. For Crohn's disease patients, disease location (L1-L4) and disease behavior/modifiers (B1-B3) are reported according to the Montreal classification; participants could be assigned to L4 in addition to another location category. Disease location and modifiers are not applicable (NA) for symptomatic controls.

support a model in which CD repertoires are primarily distinguished by increased clonal skewing/oligoclonality, with comparatively preserved estimated richness.

## BCR repertoires are largely private and reflect individual antigen exposure

Motivated by the increased clonal dominance observed in Crohn's disease (CD), we next asked whether repertoires share convergent clonotypes across individuals. To quantify between-sample overlap in a way that tolerates minor sequence variation, we compared repertoires using CDR3 amino-acid (CDR3-aa) similarity, defining two clonotypes as overlapping if their CDR3-aa sequences had the same length and differed by at most one amino-acid substitution (Hamming distance  $\leq 1$ ). For each locus (IGH, IGK, IGL), we computed a pairwise "fuzzy" Jaccard overlap between all samples. Across all three chains, overlap between unrelated individuals was uniformly low, with Jaccard values near zero for nearly all sample pairs, indicating highly private repertoires and limited convergent CDR3-aa sharing (Supplementary Figures S1–S3). Notably, we did not observe any consistent block structure or clustering by disease status, supporting the conclusion that any shared CDR3-aa convergence is rare and not a dominant feature distinguishing CD from symptomatic controls (SC) in peripheral blood. These results recapitulate what has been observed with functional antibody repertoire profiling using phage-immunoprecipitation sequencing (PhIP-Seq) (30). This approach aims at identifying the antigens bound by the collection of antibodies present in a sample without identifying the sequence of the BCR encoding for these antibodies. Using PhIP-Seq, we (19, 31, 32) and others (33–35) have shown that most antigenic exposures are private, *i.e.*, not shared among individuals, corroborating our observations about the private nature of humoral immunity.

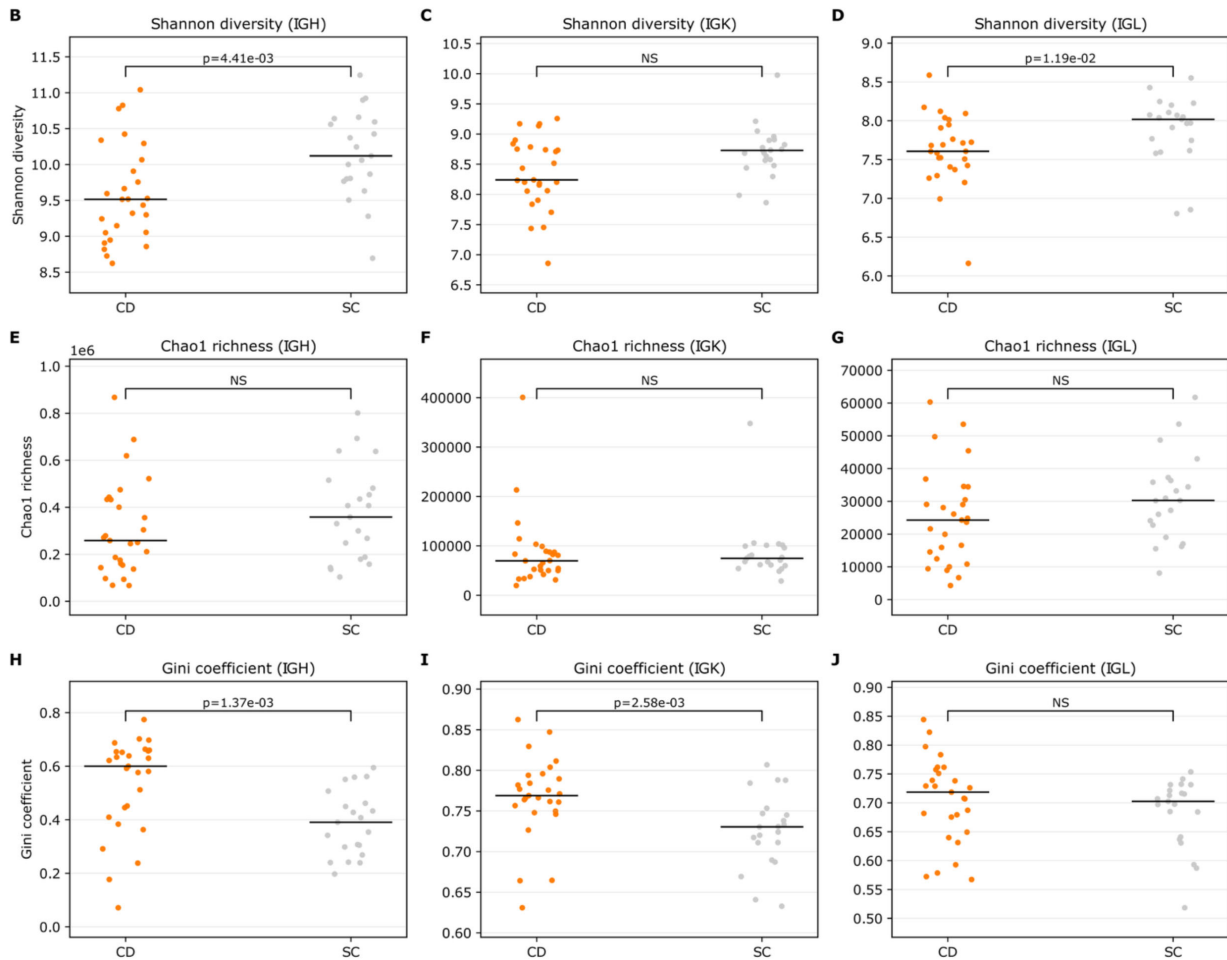
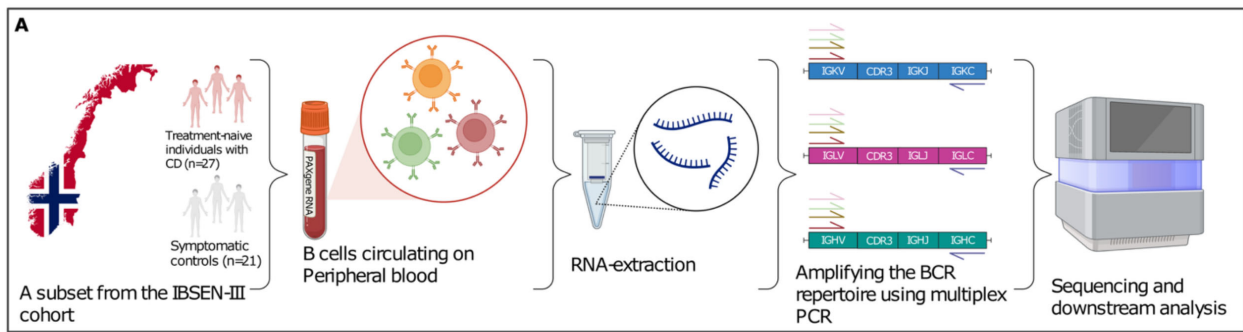
## CD does not alter the VDJ recombination landscape of the blood B cell receptor repertoire

Next, we aimed at quantifying the impact of CD on the abundance of different V and J gene combinations. Within the IGH repertoire, some VJ recombination pairs dominated the repertoire, for example, *IGHV3-23/IGHJ4* derived clonotypes (Supplementary Figure S4). The dominance of this VJ gene combination is expected, where across multiple studies, *IGHJ4* have been shown to be utilised by  $>40\%$  of clonotypes (36), similarly, the *IGHV3-23* is one of the most frequently utilised *IGHV* genes within the pool of naive and memory B cells (37). Hence, our finding corroborates previous studies showing the dominant role of the *IGHV3-23/IGHJ4* gene combination on both the naive and memory B cell compartments. Similar findings were detected at the IGK repertoire, where *IGKV3-20* paired with *IGKJ1* dominated the repertoire (Supplementary Figure S5), corroborating previous observations (38–40). Within the IGL repertoire, the *IGLV2-14* paired with *IGLJ2* formed a high percentage of the IGL repertoire (Supplementary Figure S6) in individuals with and without CD, confirming previous reports about the dominant role of these genes in shaping the IGL repertoire (38, 41). Across the dataset, we identified 75 exact overlap records involving 44 samples. All overlaps were restricted to the IGL locus ( $IGL = 75$ ;  $IGH = 0$ ;  $IGK = 0$ ) and mapped to a single PLAbDab entry (PDBmAb-24JAN2024-063013 (42)), with each positive sample contributing one or two matched clonotypes. Thus, within the "full FR1–FR4 exact overlap" definition, PLAbDab matches were detectable but rare and confined to light-chain lambda sequences in this cohort.

These findings indicate that CD induces a change in repertoire diversity and clonality, but it is not large enough to change the landscape of V(D)J recombination, at least not at a magnitude detected with our study's statistical power. To extend this analysis to other less frequent VJ combinations, we compared the expansion of each unique VJ gene combination between individuals with and without CD (Methods). Nonetheless, after correcting for multiple testing, we could not detect any significant VJ gene combination that was significantly expanded in CD relative to SC across any of the three loci (Supplementary Figure S7).

## Individuals with CD have a significant reduction in the expansion of IgM and IgD clonotypes and an expansion of IgA2 and IgG2 clonotypes

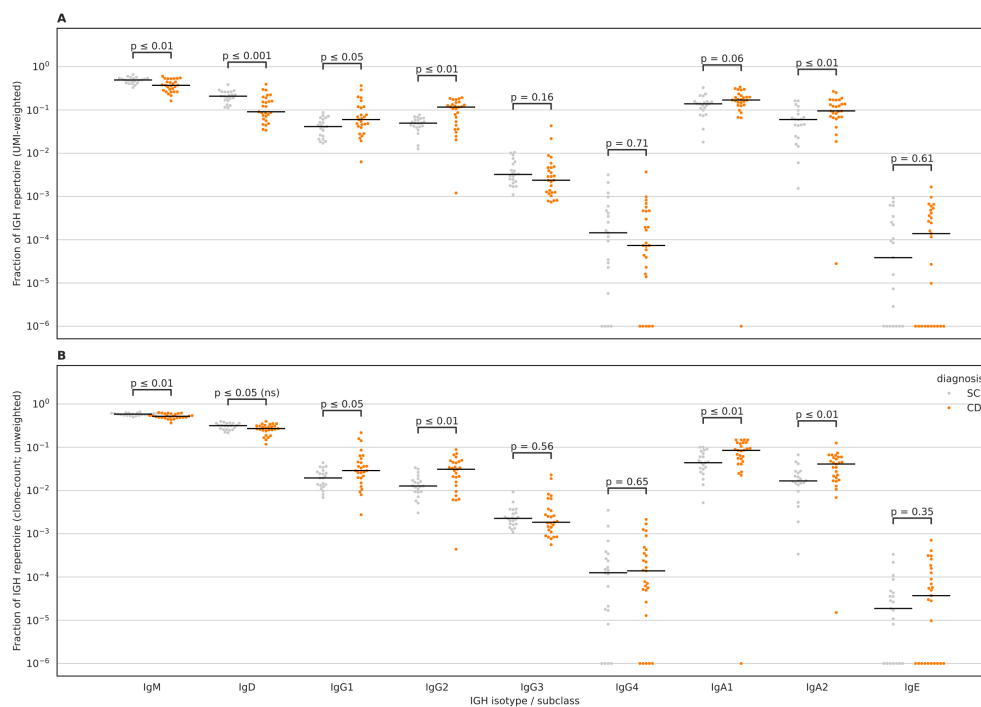
Subsequently, we aimed at investigating isotype chain usage by individuals with and without CD (Methods). We observed significantly lower levels of IgM and IgD in individuals with CD relative to SC (Figure 2A), similarly, we observed a significant increase in the levels of IgA2 and IgG2 in individuals with CD relative to SC (Figure 2A). This effect mostly persisted after removing the effect of expansion, *i.e.*, counting each clonotype as a singleton (Figure 2B). Thus, it shows that in CD there is a decrease



**FIGURE 1**  
 Repertoire diversity and inequality across BCR loci (productive clonotypes). **(A)** Schematic overview of the IBSen-III peripheral blood BCR-seq workflow. Treatment-naive individuals (Crohn's disease (CD) and symptomatic controls (SC)) were sampled; peripheral blood was collected into PAXgene tubes, RNA was extracted, and the BCR repertoire was amplified using multiplex PCR targeting IGH, IGK, and IGL rearrangements prior to sequencing and downstream analysis. Per-sample diversity metrics were calculated from productive clonotypes for the heavy chain (IGH) and light chains (IGK, IGL) and compared between Crohn's disease (CD; orange) and control subjects (SC; grey). **(B–D)** Shannon diversity index (UMI-weighted clonotype abundance) for IGH, IGK, and IGL, respectively. **(E–G)** Chao1 richness estimator for IGH, IGK, and IGL, respectively. **(H–J)** Gini coefficient of the clonotype abundance distribution (higher values indicate greater clonal inequality/expansion) for IGH, IGK, and IGL, respectively. Each dot represents one sample; horizontal black lines indicate the median. P-values are from two-sided Mann-Whitney U tests (NS, not significant). Panel **(A)** was Created in BioRender. ElAbd, H. (2026) <https://BioRender.com/cj89mj7>.

in the number and expansion levels of IgM and an increase in the number and expansion levels of IgA2 and IgG2 clonotypes. To assess whether disease status was associated with altered affinity maturation, we quantified region-resolved IGH somatic hypermutation (SHM) by comparing observed clonotype

sequences to their inferred germline counterparts after MiXCR allele reassignment. Across samples, SHM levels varied by region, with the largest values observed in the CDRs, consistent with antigen-driven selection. When comparing CD and SC, CD samples showed higher SHM levels overall, with the most



**FIGURE 2**

Differences in IGH isotype usage between Crohn's disease (CD) and symptomatic controls (SC) using MiXCR-calculated unique-molecule fractions. **(A)** UMI-weighted isotype fractions per sample, computed by summing MiXCR calculated UMI fractions across clonotypes assigned to each IGH isotype/subclass (IgM, IgD, IgG1–4, IgA1–2 and IgE). **(B)** Clonotype-count (unweighted) isotype fractions per sample, computed as the fraction of unique clonotypes assigned to each isotype/subclass. Each dot represents one sample; horizontal black lines indicate group medians. CD samples are shown in orange and SC samples in grey. Statistical comparisons were performed per isotype using two-sided Mann-Whitney U tests with Benjamini-Hochberg (BH/FDR) correction across isotypes within each panel; adjusted p-values are shown above each isotype. Because axes are log-scaled, samples with missing or zero values were displayed at a small constant ( $1 \times 10^{-6}$ ) for plotting only; raw values (including zeros) were retained for statistical testing.

pronounced separation observed in CDR2, which remained significant after FDR correction (Figure 3B). Differences in CDR1 and FR2 followed the same direction but did not reach significance after multiple testing correction (Figures 3A, C), while FR3 and FR4 exhibited broadly overlapping distributions between groups (Figures 3D, E). Together, these data indicate that disease status is associated with a modest but regionally concentrated increase in IGH SHM, strongest in CDR2.

## Discussion

By analysing peripheral blood B-cell receptor (BCR) repertoires from individuals with CD compared with symptomatic controls SC, we observed a shift toward expanded class-switched clonotypes in CD, most prominently within IgG2 and IgA2. IgG2 responses are classically enriched against carbohydrate/polysaccharide antigens and, relative to other IgG subclasses, often exhibit reduced Fc-mediated effector functions, including weaker complement activation and Fc $\gamma$ -receptor engagement (43). In parallel, IgA2 is commonly associated with mucosal immunity and has been linked to recognition of commensal bacteria, consistent with heightened antigenic exposure at the gut interface. Together, the selective enrichment of IgG2/IgA2 clonotypes in CD is compatible with increased stimulation by microbial antigens,

potentially facilitated by intestinal barrier dysfunction reported in CD. This interpretation is further supported by our SHM analyses, which indicate greater evidence of antigen-driven maturation in CD (notably in CDR2), alongside a less even repertoire. Our observations align with prior repertoire-level analyses reporting disease-associated remodelling of B-cell architecture in immune-mediated conditions including CD (24, 44).

Despite evidence of isotype-specific expansion, we detected limited sharing of exact clonotypes in the peripheral blood of the study participants. This is consistent with the strong tissue compartmentalization of B-cell responses and with recent work identifying disease-associated shared B-cell clones in lymph nodes rather than blood (27). In addition, while earlier reports suggested an IgE-associated signal in CD (24), we did not reproduce a robust IgE expansion in our cohort. Differences in the cohort size, analytical sensitivity, and genetic/environmental background may contribute to these discrepancies. Future studies with larger multi-ethnic cohorts will be important to clarify the contribution of IgE and other low-frequency isotypes to the pathogenesis of CD. Finally, we observed comparatively reduced expansion of IgM/IgD clonotypes in CD relative to SC, together with increased class-switched IgA/IgG clonotypes and elevated SHM. Collectively, these signatures support a model of increased antigenic encounter and/or immune priming in CD compared with symptomatic controls.

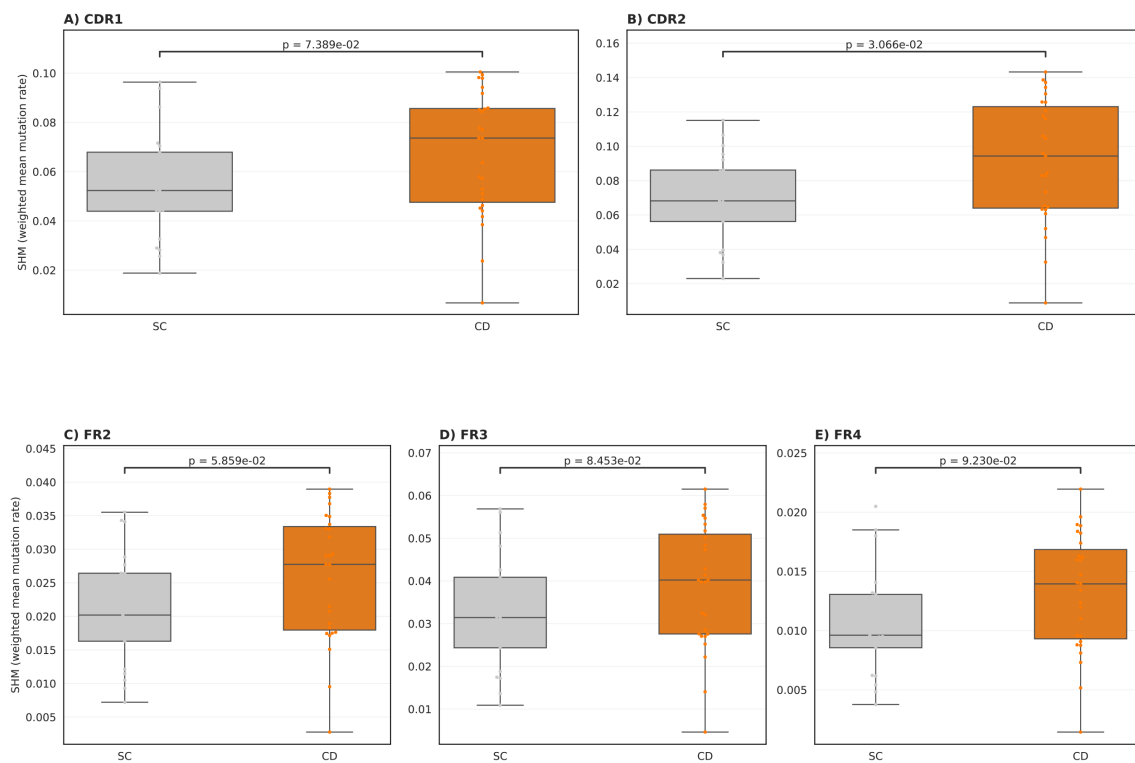


FIGURE 3

Region-resolved somatic hypermutation (SHM) in IGH differs between CD and SC. Per-sample, region-specific SHM rates were quantified from MiXCR allele-reassigned IGH clonotypes by comparing observed nucleotide sequences to their inferred germline counterparts for CDR1 (A), CDR2 (B), FR2 (C), FR3 (D), and FR4 (E). For each sample and region, we computed a UMI-weighted mean SHM rate (mismatches/compared bases), where compared bases include only positions with unambiguous A/C/G/T in both observed and germline sequences. Each dot represents one sample (CD vs SC), with boxplots summarizing the distribution across samples (median and interquartile range). Statistical differences between CD and SC were assessed using a two-sided Mann-Whitney U test per region and corrected for multiple testing across regions using Benjamini-Hochberg (BH).

## Methods

### Cohort description

The cohort contains treatment-naïve individuals with CD or symptomatic controls, *i.e.*, individuals with the symptoms of IBD but without endoscopic or radiological findings supporting their diagnosis. The study participants were recruited as a part of the Inflammatory Bowel Disease in South-Eastern Norway III (IBSEN III) study (28). The IBSEN-III study was approved by the South-Eastern Regional Committee for Medical and Health Research Ethics (Ref 2015/946-3) and performed in accordance with the Declaration of Helsinki.

### Profiling the B cell receptor

Whole blood was collected into PAXgene Blood RNA tubes (Qiagen) and total RNA was extracted from stabilized blood, that is, no cell sorting was performed and total RNA was used for profiling the repertoire. After quality controls, 300ng of RNA were used to profile the IGH, IGL and IGK repertoire of the samples using the Human IG RNA Multiplex kits from MiLaboratories according to the manufacturer's instructions. This multiplex design employs

V-gene-specific forward primers located within framework region 1 (FR1) together with constant-region primers, producing amplicons that start downstream of the 5' portion of FR1. Consequently, assembled clonotype sequences reliably cover the variable-domain segments spanning CDR1-CDR3 and framework regions FR2-FR4, while FR1 is not fully captured by the assay. After library preparation, Illumina's unique dual indexing adaptors were used to tag each sample with a unique index, then samples were pooled together and then cleaned with magnetic beads (AMPure XP, Beckman Coulter). Next, samples were sequenced using a 2×250 bp paired-end using a NextSeq 1000 at the Illumina Solutions Center in Berlin.

### Clonotype identification from sequencing reads

After sequencing and sample demultiplexing, paired-end reads from each sample were processed with MiXCR v4.7.0 (29) using the manufacturer-recommended preset for the MiLaboratories Human BCR RNA IG UMI Multiplex protocol. MiXCR performed read pre-processing, alignment to reference V(D)J gene segments, UMI-aware assembly, and clonotype calling. For each sample, clonotype tables were generated for the immunoglobulin heavy (IGH) and light

chains (IGK and IGL). Clonotype tables were then exported from MiXCR as tab-delimited files for downstream analyses, including per-clone abundance metrics (read count/fraction and UMI-derived molecule counts/fractions) and gene annotations (V and J gene calls and constant gene/isotype assignment for IGH). Unless stated otherwise, downstream analyses were restricted to productive clonotypes to focus on functional immunoglobulin rearrangements.

## Diversity and inequality calculations

Diversity metrics were computed per sample and per locus from the distribution of clonotype abundances. For a given sample, let  $R$  denote the number of clonotypes (clonal groups) and let  $n_i$  denote the abundance of clonotype  $i$  (UMI counts summed across sequences assigned to the clonotype). Abundances were converted to relative frequencies  $p_i$  by normalizing to the total abundance  $N = \sum_{i=1}^R n_i$ , i.e.,  $p_i = n_i/N$ .

Shannon diversity was calculated as:

$$\text{shannon diversity} = - \sum_{i=1}^R p_i \log(p_i)$$

Chao1 richness was used as an estimator of the total clonotype richness accounting for under sampling:

$$S_{\text{Chao1}} = S_{\text{obs}} + \frac{f_1^2}{2 f_2}$$

where  $S_{\text{obs}}$  is the observed number of clonotypes in the sample,  $f_1$  is the number of singleton-clonotypes (with  $n_i = 1$ ), and  $f_2$  is the number of doubletons clonotypes (with  $n_i = 2$ ). When  $f_2 = 0$ , we used the standard bias-corrected form:

$$S_{\text{Chao1}} = S_{\text{obs}} + \frac{f_1(f_1 - 1)}{2(f_2 + 1)}$$

To quantify repertoire unevenness (oligoclonality), we calculated the Gini coefficient per sample from clonotype abundances. Let  $n_i$  denote the abundance (UMI count) of a clonotype  $i$  and  $R$  the number of clonotypes in the sample. After sorting abundances in non-decreasing order ( $n_{(1)} \leq \dots \leq n_{(R)}$ ) and defining  $N = \sum_{i=1}^R n_{(i)}$ , the Gini coefficient was computed as:

$$\text{Gini inequality (G)} = \frac{2 \sum_{i=1}^R i n_{(i)}}{R N} - \frac{R + 1}{R}$$

where  $G \in [0, 1]$ , with  $G = 0$  indicating a perfectly even repertoire (all clonotypes have equal abundance) and higher values indicating increasing inequality consistent with oligoclonal expansion. Metrics were computed per sample and per locus using productive sequences only.

## Fuzzy CDR3 overlap analysis

To quantify inter-individual similarity while allowing for minor amino-acid variation, we computed a fuzzy repertoire overlap using CDR3 amino-acid (AA) sequences only for each locus (IGH, IGK,

IGL). For each sample, CDR3 AA strings were extracted and collapsed to unique sequences. To define fuzzy groups, all unique CDR3 AA sequences across samples were clustered under a Hamming distance threshold of  $\leq 1$  substitution, restricted to sequences of identical length. Clustering was performed separately within each length group using a wildcard indexing approach. Each sample repertoire was then represented as the set of fuzzy cluster identifiers present in that sample. Pairwise sample overlap was quantified using the Jaccard index (intersection/union) computed on these cluster sets, producing a sample-by-sample overlap matrix per locus.

## Comparing V-J recombination frequencies between conditions

To compare V-J recombination usage between individuals with Crohn's disease (CD) and symptomatic controls (SC), we first collapsed V and J calls to gene-level (alleles removed) and defined each V-J combination as V\_gene-J\_gene. For each sample, we quantified the frequency of each V-J pair as the summed UMI counts assigned to that V-J pair divided by the total UMI count in that sample for the corresponding locus. For each group (CD or SC), we computed the mean V-J frequency across samples:

$$\mu_{\text{VJ}} = \frac{1}{S} \sum_{i=1}^s f_{\text{vj},i} + \epsilon$$

where  $f_{\text{VJ},i}$  is the per-sample frequency of the V-J pair in sample  $i$ ,  $S$  is the number of samples in the group, and  $\epsilon$  is a pseudo-count ( $\epsilon = 1 \times 10^{-8}$ ) to accommodate V-J pairs not observed in a given group. We then computed the log2 fold change between groups:

$$\text{fold change}_{\text{VJ}} = \log_2 \left( \frac{\mu_{\text{VJ}}^{\text{CD}}}{\mu_{\text{VJ}}^{\text{SC}}} \right)$$

For each V-J pair, the statistical significance of differences in per-sample frequencies between CD and SC was assessed using a two-sided Mann-Whitney U test, followed by Benjamini-Hochberg correction to control the false discovery rate.

## Similarity to characterized antibodies (PLABDab)

BCR repertoires were processed with MiXCR (v4.7.0) (29), including the 'findAlleles' command, and clone tables were generated from the resulting files. For each sample and each locus (IGH, IGK, IGL), clonotypes were exported using 'mixcr exportClones' to retain clonotype identifiers and abundance metrics, namely, ('cloneId', 'readCount', 'uniqueMoleculeCount', and 'uniqueMoleculeFraction'), together with imputed amino-acid sequences across the variable domain using '-allAAFeaturesImputed FR1Begin FR4End'. For every sample in a given locus, clonotypes were ranked by UMI abundance, and the top 1000 expanded clonotypes were retained. The imputed FR1-FR4 amino-acid sequence per clonotype was defined as the concatenation of the

exported framework and CDR segments spanning FR1 through FR4 (“full FR1-FR4 overlap” string). To assess public/known antibody overlap, PLAbDab (45) paired and unpaired sequence exports were ingested, sequences were cleaned to standard amino-acid strings, and light-chain vs heavy-chain entries were assigned to the appropriate locus. Exact overlaps were then computed by strict string equality between each MiXCR clonotype’s full FR1-FR4 amino-acid sequence and the PLAbDab (45) variable-domain amino-acid sequence field. All matches were written to a consolidated overlap table including MiXCR abundance metrics and PLAbDab record metadata.

## IGH isotype/subclass utilization analysis

We quantified IGH constant-region usage per sample from MiXCR clonotype exports. Constant-region annotations (isotype) were standardized by (i) taking the first call when multiple were present, (ii) removing allele suffixes (e.g., \*00, \*01), and (iii) harmonizing IMGT-style labels to a fixed set and order: IgM, IgD, IgG1-4, IgA1-2 and IgE. Only clonotypes assigned to these categories were retained. For the UMI-weighted analysis, we used MiXCR ‘uniqueMoleculeFraction’ (per-clonotype fraction of unique molecules within a sample). For each sample and isotype, we summed uniqueMoleculeFraction across clonotypes, yielding a per-sample isotype fraction reflecting both isotype composition and clonal expansion. For the clone-count (unweighted) analysis, we counted unique clonotypes per isotype in each sample and divided by the total number of clonotypes in the sample, yielding an isotype fraction reflecting isotype diversity independent of expansion. Group comparisons (CD vs SC) were performed independently for each isotype using two-sided Mann-Whitney U tests, and p-values were adjusted using Benjamini-Hochberg FDR correction across isotypes within each panel. For visualization on log-scaled axes, missing/zero values were displayed at  $1 \times 10^{-6}$  for plotting only, while the raw values (including zeros) were used for statistical tests.

## Somatic hypermutation analysis

- I. MiXCR allele inference and clonotype export. IGH repertoires were processed with MiXCR and subjected to allele inference using ‘findAlleles’, generating allele-reassigned clonotype tables per sample. For downstream SHM quantification, clonotypes were exported with MiXCR to include: (i) clonotype identifiers and abundance, and (ii) nucleotide sequences for immunoglobulin subregions along with their corresponding inferred germline sequences.
- II. Per-sample SHM computation. SHM was computed from the exported clonotype tables using a custom Python workflow. For each sample and for each region (CDR1, CDR2, FR2, FR3, and FR4), the analysis script, compares the observed nucleotide sequence column (e.g., nSeqCDR1) to the corresponding germline column (e.g., nSeqCDR1OfGermline). Only positions where both

sequences contain unambiguous A/C/G/T bases are considered “compared bases”; positions with ambiguous characters are excluded. For each clonotype, the number of mismatches and compared bases are calculated, and then aggregated to the sample level using abundance weights:

- a. Weighted mismatches = sum (weight x mismatches).
- b. Weighted compared bases = sum (weight x compared bases).
- c. Weighted mean SHM rate = (weighted mismatches)/(weighted compared bases).

The script also reports the fraction of clonotypes with valid coverage for each region (found fraction) and the total weighted compared bases per region. The final output is a per-sample table containing region-specific weighted mean SHM values used for plotting and group comparisons. For each region, per-sample weighted mean SHM rates were compared between CD and SC using a two-sided Mann-Whitney U test. Multiple testing across regions was controlled using Benjamini-Hochberg correction. Boxplots and sample-level points were produced from these per-sample SHM summaries, and BH adjusted p-values were displayed in the figure.

## Data availability statement

The datasets presented in this article are not readily available because of data privacy regulations in Norway and our institution. However, data are available upon request, if the aims of the planned analyses are covered by the written informed consent signed by the participants, pending an amendment to the ethical approvals and a material & data transfer agreement between the institutions. Requests to access the datasets should be directed to Marte Lie Høivik, [m.l.hoivik@medisin.uio.no](mailto:m.l.hoivik@medisin.uio.no).

## Ethics statement

The studies involving humans were approved by the South-Eastern Regional Committee for Medical and Health Research Ethics (Ref 2015/946-3) and performed in accordance with the Declaration of Helsinki. All patients signed an informed consent form prior inclusion in this study, and the data were stored in services for sensitive data (TSD) at the University of Oslo. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

## Author contributions

AM: Conceptualization, Data curation, Formal Analysis, Visualization, Writing – original draft, Writing – review & editing. ZT: Formal Analysis, Methodology, Writing – original draft, Writing –

review & editing. MH: Data curation, Writing – original draft, Writing – review & editing. AF: Conceptualization, Funding acquisition, Methodology, Resources, Supervision, Writing – original draft, Writing – review & editing. HE: Conceptualization, Formal Analysis, Investigation, Software, Supervision, Visualization, Writing – original draft, Writing – review & editing.

## Funding

The author(s) declared that financial support was received for this work and/or its publication. The project was funded by the EU Horizon Europe Program grant miGut-Health: Personalized Blueprint of Intestinal Health (ID: 101095470). Additionally, the project received funding from the German Research Foundation (DFG) Research Unit 5042: miTarget (The Microbiome as a Therapeutic Target in Inflammatory Bowel Diseases), along with funding from the DFG Cluster of Excellence 2167 “Precision Medicine in Chronic Inflammation (PMI)”. AM is funded by the DFG collaborative research center CRC 1526 “Pathomechanisms of Antibody-mediated Autoimmunity (PANTAU) -Insights from Pemphigoid Diseases” and the DFG Cluster of Excellence 2167/2 “Precision Medicine in Chronic Inflammation (PMI 2.0). The IBSEN III study has received funding from the South-Eastern Health Authorities in Norway, the Dam Foundation, as well as investigator-initiated research grants from Pfizer, Ferring Pharmaceuticals, TillottsPharma. The IBSEN III study is investigator-initiated, and the sponsors did not contribute to the study design, analysis, interpretation of the data or publication. The funding agencies had neither a role in the design, collection, analysis, and interpretation of data nor in writing the manuscript.

## Conflict of interest

MH received investigator-initiated research grants from Takeda, Pfizer, Tillotts, Ferring and Janssen. Speaker honoraria

## References

- Hampe J, Franke A, Rosenstiel P, Till A, Teuber M, Huse K, et al. A genome-wide association scan of nonsynonymous SNPs identifies a susceptibility variant for Crohn disease in ATG16L1. *Nat Genet.* (2007) 39:207–11. doi: 10.1038/ng1954
- Lavoie S, Conway KL, Lassen KG, Jijon HB, Pan H, Chun E, et al. The Crohn’s disease polymorphism, ATG16L1 T300A, alters the gut microbiota and enhances the local Th1/Th17 response. *Elife.* (2019) 8:e39982. doi: 10.7554/eLife.39982
- Hampe J, Cuthbert A, Croucher PJP, Mirza MM, Mascheretti S, Fisher S, et al. Association between insertion mutation in NOD2 gene and Crohn’s disease in German and British populations. *Lancet.* (2001) 357:1925–8. doi: 10.1016/S0140-6736(00)05063-7
- 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
- Degenhardt F, Mayr G, Wendorf 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) 30:356–69. doi: 10.1093/hmg/ddab017
- Shan Y, Lee M, Chang EB. The gut microbiome and inflammatory bowel diseases. *Annu Rev Med.* (2022) 73:455–68. doi: 10.1146/annurev-med-042320-021020
- Faye AS, Allin KH, Iversen AT, Agrawal M, Faith J, Colombel J-F, et al. Antibiotic use as a risk factor for inflammatory bowel disease across the ages: a population-based cohort study. *Gut.* (2023) 72:663. doi: 10.1136/gutjnl-2022-327845
- Ebert AC, Harper S, Vestergaard MV, Mitchell W, Jess T, Elmahdi R. Risk of inflammatory bowel disease following hospitalisation with infectious mononucleosis: nationwide cohort study from Denmark. *Nat Commun.* (2024) 15:8383. doi: 10.1038/s41467-024-52195-8
- Uchida AM, Boden EK, James EA, Shows DM, Konecny AJ, Lord JD. Escherichia coli-specific CD4+ T cells have public T-cell receptors and low interleukin 10 production in Crohn’s disease. *Cell Mol Gastroenterol Hepatol.* (2020) 10:507–26. doi: 10.1016/j.jcmgh.2020.04.013
- Martini GR, Tikhonova E, Rosati E, DeCelle MB, Sievers LK, Tran F, et al. Selection of cross-reactive T cells by commensal and food-derived yeasts drives cytotoxic TH1 cell responses in Crohn’s disease. *Nat Med.* (2023) 29:2602–14. doi: 10.1038/s41591-023-02556-5

from Takeda, Tillotts, Ferring, AbbVie, Galapagos and Meda. She is also on the advisory board of Takeda, Galapagos, MSD, Lilly and AbbVie.

The remaining author(s) declared that this work 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) declared that generative AI was used in the creation of this manuscript. Generative AI was used to improve the clarity and readability of the manuscript and to assist in drafting and debugging analysis scripts. All analyses, results, and interpretations were verified by the authors.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

## 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.2026.1725813/full#supplementary-material>

11. 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
12. Mahdy A, ElAbd H, Kriukova V, Olbjorn 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/jjaf190.0299
13. Bourgonje AR, Andreu-Sánchez S, Vogl T, Hu S, Vich Vila A, Gacesa R, et al. Phage-display immunoprecipitation sequencing of the antibody epitope repertoire in inflammatory bowel disease reveals distinct antibody signatures. *Immunity*. (2023) 56:1393–1409.e6. doi: 10.1016/j.immuni.2023.04.017
14. 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) 169:150–3. doi: 10.1053/j.gastro.2025.01.247
15. 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) 19:jjaf048. doi: 10.1093/ecco-jcc/jjaf048
16. Elabd H, Mahdy A, Franke A. OP11 Analysing the T cell receptor beta chain repertoire of 2,800 Inflammatory Bowel Disease patients identifies public T cell responses involved in the pathogenesis of Crohn's Disease and Ulcerative Colitis and quantifies the impact of surgery and therapy on the immune repertoire of IBD patients. *J Crohns Colitis*. (2025) 19:i21–3. doi: 10.1093/ecco-jcc/jjaf190.0011
17. 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
18. ElAbd H, Mahdy A, Wacker EM, Gretsova M, Ellinghaus D, Franke A. T cell receptor clonotypes predict human leukocyte antigen allele carriage and antigen exposure history. *Commun Biol*. (2026) 9:50. doi: 10.1038/s42003-025-09140-2
19. ElAbd H, Pesesky M, Innocenti G, Chung BK, Mahdy AKH, Kriukova V, et al. T and B cell responses against Epstein-Barr virus in primary sclerosing cholangitis. *Nat Med*. (2025) 31:2306–16. doi: 10.1038/s41591-025-03692-w
20. Allez M, Auzolle C, Ngollo M, Bottois H, Chardin V, Corraliza AM, et al. T cell clonal expansions in ileal Crohn's disease are associated with smoking behaviour and postoperative recurrence. *Gut*. (2019) 68:1961–70. doi: 10.1136/gutjnl-2018-317878
21. Pesesky M, Carty CL, Singh N, Le Bourhis L, Rosati E, Bokemeyer B, et al. DOP47 Identification and characterization of T-cell receptor sequences associated with Crohn's Disease. *J Crohns Colitis*. (2022) 16:i096–7. doi: 10.1093/ecco-jcc/jjab232.086
22. Chapman CG, Yamaguchi R, Tamura K, Weidner J, Imoto S, Kwon J, et al. Characterization of T-cell receptor repertoire in inflamed tissues of patients with crohn's disease through deep sequencing. *Inflammation Bowel Dis*. (2016) 22:1275–85. doi: 10.1097/MIB.0000000000000752
23. Doorenspleet ME, Westera L, Peters CP, Hakvoort TBM, Esveldt RE, Vogels E, et al. Profoundly expanded T-cell clones in the inflamed and uninfamed intestine of patients with crohn's disease. *J Crohns Colitis*. (2017) 11:831–9. doi: 10.1093/ecco-jcc/jjx012
24. Bashford-Rogers RJM, Bergamaschi L, McKinney EF, Pombal DC, Mescia F, Lee JC, et al. Analysis of the B cell receptor repertoire in six immune-mediated diseases. *Nature*. (2019) 574:122–6. doi: 10.1038/s41586-019-1595-3
25. Scheid JF, Eraslan B, Hudak A, Brown EM, Sergio D, Delorey TM, et al. Remodeling of colon plasma cell repertoire within ulcerative colitis patients. *J Exp Med*. (2023) 220:e20220538. doi: 10.1084/jem.20220538
26. Chen D, Xu S, Li S, Wang Q, Li H, He D, et al. The multi-organ landscape of B cells highlights dysregulated memory B cell responses in Crohn's disease. *Natl Sci Rev*. (2025) 12:nwaf009. doi: 10.1093/nsr/nwaf009
27. Kotagiri P, Rae WM, Bergamaschi L, Pombal D, Lee J-Y, Noor NM, et al. Disease-specific B cell clones are shared between patients with Crohn's disease. *Nat Commun*. (2025) 16:3689. doi: 10.1038/s41467-025-58977-y
28. Kristensen VA, Opheim R, Perminow G, Huppertz-Hauss G, Detlie TE, Lund C, et al. Inflammatory bowel disease in South-Eastern Norway III (IBSEN III): a new population-based inception cohort study from South-Eastern Norway. *Scand J Gastroenterol*. (2021) 56:899–905. doi: 10.1080/00365521.2021.1922746
29. Bolotin DA, Poslavsky S, Mitrophanov I, Shugay M, Mamedov IZ, Putintseva EV, et al. MiXCR: software for comprehensive adaptive immunity profiling. *Nat Methods*. (2015) 12:380–1. doi: 10.1038/nmeth.3364
30. 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
31. Innocenti G, Andreu-Sánchez S, Hörstke NV, Elabd H, Barozzi I, Franke A, et al. Associations between HLA-II variation and antibody specificity are predicted by antigen properties. *Genome Med*. (2025) 17:65. doi: 10.1186/s13073-025-01486-w
32. Mahdy AKH, ElAbd H, Prinzensteiner M, Jebens H, Sivickis K, Bacher P, et al. Mapping adaptive immune responses toward fungal antigens in inflammatory bowel disease using T cell repertoire sequencing and phage-immunoprecipitation sequencing. *bioRxiv*. (2025), 2025.05.13.653535. doi: 10.1101/2025.05.13.653535
33. Vogl T, Kalka IN, Klompus S, Leviatan S, Weinberger A, Segal E. Systemic antibody responses against human microbiota flagellins are overrepresented in chronic fatigue syndrome patients. *Sci Adv*. (2024) 8:eabq2422. doi: 10.1126/sciadv.abq2422
34. Bourgonje AR, Hörstke NV, Fehringer M, Innocenti G, Vogl T. Systemic antibody responses against gut microbiota flagellins implicate shared and divergent immune reactivity in Crohn's disease and chronic fatigue syndrome. *Microbiome*. (2024) 12:141. doi: 10.1186/s40168-024-01858-1
35. 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
36. Shi B, Dong X, Ma Q, Sun S, Ma L, Yu J, et al. The usage of human IGHJ genes follows a particular non-random selection: the recombination signal sequence may affect the usage of human IGHJ genes. *Front Genet*. (2020) 11:524413. doi: 10.3389/fgene.2020.524413
37. Wu Y-C, Kipling D, Leong HS, Martin V, Ademokun AA, Dunn-Walters DK. High-throughput immunoglobulin repertoire analysis distinguishes between human IgM memory and switched memory B-cell populations. *Blood*. (2010) 116:1070–8. doi: 10.1182/blood-2010-03-275859
38. Stamatoopoulos K, Belessi C, Hadzidimitriou A, Smilevska T, Kalagiakou E, Hatzi K, et al. Immunoglobulin light chain repertoire in chronic lymphocytic leukemia. *Blood*. (2005) 106:3575–83. doi: 10.1182/blood-2005-04-1511
39. Xia M, Wu L, Sun X, Han X, Yan H, Huang J, et al. Next-generation sequencing revealed a distinct immunoglobulin repertoire with specific mutation hotspots in acute myeloid leukemia. *Biol (Basel)*. (2022) 11. doi: 10.3390/biology11020161
40. He B, Liu S, Xu M, Hu Y, Lv K, Wang Y, et al. Comparative global B cell receptor repertoire difference induced by SARS-CoV-2 infection or vaccination via single-cell V (D)J sequencing. *Emerg Microbes Infect*. (2022) 11:2007–20. doi: 10.1080/22221751.2022.2105261
41. Ghiotto F, Fais F, Albesiano E, Sison C, Valetto A, Gaidano G, et al. Similarities and differences between the light and heavy chain ig variable region gene repertoires in chronic lymphocytic leukemia. *Mol Med*. (2006) 12:300–8. doi: 10.2119/2006-00080.Ghiotto
42. Foster MH, Buckley ES, Chen BJ, Hwang K-K, Clark AG. Uncommon structural motifs dominate the antigen binding site in human autoantibodies reactive with basement membrane collagen. *Mol Immunol*. (2016) 76:123–33. doi: 10.1016/j.molimm.2016.07.004
43. Vidarsson G, Dekkers G, Rispen T. IgG subclasses and allotypes: from structure to effector functions. *Front Immunol*. (2014) 5:520. doi: 10.3389/fimmu.2014.00520
44. Timmermans WMC, van Laar JAM, van der Houwen TB, Kamphuis LSJ, Bartol SJW, Lam KH, et al. B-cell dysregulation in crohn's disease is partially restored with infliximab therapy. *PLoS One*. (2016) 11:e0160103–. doi: 10.1371/journal.pone.0160103
45. Abanades B, Olsen TH, Raybould MIJ, Aguilar-Sanjuan B, Wong WK, Georges G, et al. The Patent and Literature Antibody Database (PLAbDab): an evolving reference set of functionally diverse, literature-annotated antibody sequences and structures. *Nucleic Acids Res*. (2024) 52:D545–51. doi: 10.1093/nar/gkad1056