Convergent Evolution and B-Cell Recirculation in Germinal Centers in a Human Lymph Node

Aurelien Pelissier[1,2,*], Maria Stratigopoulou[3,*], Naomi Donner[3], Evangelos Dimitriadis[4], Richard J Bende[3], Jeroen E Guikema[3,*,†], Maria Rodriguez Martinez[1,*,†] and Carel J M van Noesel[3,*,†]

[*] *Equal contribution*
[1] *IBM Research Europe, 8803 Rüschlikon, Switzerland*
[2] *Department of Biosystems Science and Engineering, ETH Zurich, 4058 Basel, Switzerland*
[3] *Lymphoma and Myeloma Center Amsterdam, 1105 AZ Amsterdam, Netherlands*
[4] *MonetDB Solutions, 1098 XG Amsterdam, Netherlands*
[†] *Corresponding authors. j.e.guikema@amsterdamumc.nl, mrm@zurich.ibm.com, c.j.vannoesel@amsterdamumc.nl*

Abstract

Germinal centers (GCs) are specialized compartments within the secondary lymphoid organs where B cells proliferate, differentiate, and mutate their antibody genes in response to the presence of foreign antigens. They play a central role in generating an effective immune response against infectious pathogens, and failures in their regulating mechanisms can lead to the development of autoimmune diseases and cancer. While previous works study experimental systems of the immune response with mouse models that are immunized with specific antigens, our study focuses on a real life situation, with an ongoing GC response in a human lymph node (LN) involving multiple asynchronized GCs reacting simultaneously to unknown antigens. We combined laser capture microdissection (LCM) of individual GCs from human LN with next-generation repertoire sequencing (Rep-seq) to characterize individual GCs as distinct evolutionary spaces. In line with well-characterized GC responses in mice, elicited by immunization with model antigens such as NP-CGG, we observe a relatively low sequence similarity, as well as heterogeneous clonal diversity across individual GCs from the same human LN. Still, we identify shared clones in several individual GCs, and phylogenetic tree analysis combined with paratope modeling suggest the re-engagement and rediversification of B-cell clones across GCs as well as expanded clones exhibiting shared antigen responses across distinct GCs, indicating convergent evolution of the GCs. Finally, our study allows for the characterization of non-functional clones, where frequencies of V(D)J or SHM induced stop codons are quantified.

Germinal centers (GCs) are specialized microanatomical structures within the secondary lymphoid organs where B cells proliferate, differentiate, and mutate their antibody genes in response to the presence of foreign antigens [1, 2]. Through the GC lifespan, interclonal competition between B cells leads to an increased affinity of the B-cell receptors (BCRs) for antigens accompanied by a loss of clonal diversity in rodents [3] and humans [4]. Throughout the GC reaction, B cells with improved affinity for antigens as a result of these mutations are continuously selected. By competing for antibody-mediated antigen capture and subsequent acquisition of T-cell help the mutated B cells gain affinity, and the selected B cells with improved affinity for the antigen differentiate into memory B cells and plasma cells [1, 5]. GCs thus play a central role in generating an effective immune response against infectious pathogens, and failures within their tightly regulated environment can lead to the development of autoimmune diseases [6] and cancer [7].

In vivo mouse experiments have shown that individual GCs exhibit variable degrees of clonal imbalance and dominance, even when induced synchronously by immunization with various antigens (CGG, chicken gamma globulin; OVA, chicken ovalbumin; HA, influenza hemagglutinin (H3); NP-OVA, 4-hydroxy-3-nitrophenylacetyl-OVA) [3]. In particular, a subset of GCs underwent a massive expansion of higher-affinity B-cell variants (clonal bursts), leading to a loss of clonal diversity at a significantly faster rate than other GCs. These differences in clonal dynamics among GCs could partially be explained by the differences in affinity between competing B cells but also by other factors unrelated to affinity [3]. Moreover, the temporal resolution of GC reactions through *intravital* microscopy revealed substantial heterogeneity in the evolution of both foreign-antigen induced and autoreactive GCs over time [8]. Interestingly, initially dominant

clones were observed to suddenly lose competitive momentum allowing for the emergence of other clones (inversion event) [8]. GC computational models suggest that these chaotic dynamics were likely the consequences of initially small stochastic advantages in the affinity to antigens [9], amplified through the selection and proliferation of higher affinity clones. A research question of interest is to which extent these observations are applicable to a typical ongoing GC response in human LN, which involves multiple asynchronized GCs to unknown antigens.

During lymphocyte development, the antigen receptors expressed by B and T lymphocytes are assembled in an antigen-independent fashion by ordered variable gene segment recombination (V(D)J recombination) [10]. The antigen receptors of B cells are then further diversified in the GCs through somatic hyper mutation (SHM) [11] which is induced by activation-induced cytidine deaminase (AID), both Base Excision Repair (BER) and Mismatch Repair (MMR) are necessary for those processes [12, 13]. Affinity maturation through SHM is critical for antibodies to reach high affinity for the target antigen [11]. During V(D)J recombination and SHM, stop codons or frameshifts of the B-cell receptor (BCR) sequences may occur, leading to non-functional BCR sequences. GC B cells that concur SHM-derived crippling mutations, resulting in failure to express a functional BCR, undergo apoptosis [14]. On the other hand, non-functional BCR sequences derived from V(D)J recombination may be propagated in the GC but are not subjected to selection [1]. A recent study focusing on mice Peyer patches [15] sequenced both functional and non-functional V(D)J rearrangements from genomic DNA, which they used to identify positively selected mutations by comparing SHM mutation rates between functional and non-functional sequences. SHM selection was identified for some shared clonotypes among mice under different gnotobiotic conditions, but not for all of them [15].

In vivo studies in mice have revealed that recently activated B and T cells could constantly enter ongoing GC reactions [16, 17, 18], thus indicating that GC seeding is an ongoing process governed by a competitive advantage in antigen-binding affinity of naive B cells [16, 2]. In addition, *in vivo* time resolution of single GCs in mouse models demonstrated a possible reentry of output memory B-cell clones to ongoing GCs [8]. In line with these results in mice, we demonstrated reoccurring B-cell clones in multiple GCs within the human LNs [19], suggesting the migration of antigen-responsive B cells in human lymphoid tissue. Those findings support the hypothesis that high affinity BCRs do not necessarily arise during a single GC passage, but potentially after successive GC responses [19]. Nevertheless, such phenomena were shown to be infrequent in mouse models under typical boost regimens [20], with less than 10% of the clones found in multiple GCs, and secondary germinal centers consisting predominantly of B cells without prior GC experience.

Importantly, while mouse immunization models represent controlled environments in which B cells from all GCs react to the same antigen, a major question is to which extent observations in such experimental models are applicable to a typical ongoing GC response in a human LN that involves multiple asynchronized GCs reacting simultaneously to unknown antigens.

We have developed a method to study the BCR repertoire at the GC scale by combining LCM of individual GCs from human LNs with next generation sequencing (NGS)-based immunoglobulin heavy chain variable region (IGHV) repertoire analysis (Rep-seq) from genomic DNA. Our study goes beyond most of the previous GC mouse studies [8, 20] in several aspects. Firstly, the use of genomic DNA, as opposed to more commonly used RNA, gives a direct and unbiased reflection of clonal frequency, since one V(D)J rearrangement equals one B-cell, which is not the case with mRNA analysis. Secondly, it allows the analysis of the source and the fate of the non-functional BCR sequences. Furthermore, it enables the analysis of the effects of affinity-based selection on the specific mutational spectrum during SHM, at the individual GC scale. Such analyse were performed previously in mouse models in bulk naive and GC B cells from the spleen or Peyer's patches [15], but not in individual GCs. Importantly, the use of NGS-based Rep-seq of individual GCs enables the sequencing of more than 10^4 B cells in each GC, which considerably strengthens the quantitative and statistical analysis of the GC repertoires, with regards to clonal representation and diversity. Finally, phylogenetic analysis provide insights into the functional convergence of dominant clones across GCs by identifying shared clones, which represent reactivated B cells as well as expanded clones across distinct GCs.

Results

Heterogeneous Clonal Diversity Across GCs. We performed LCM on 10 individual GCs from a human LN (Supplementary Figure S9) and we applied Rep-seq in order to obtain sequencing data of the IGHV of the BCR, obtaining approximately 100.000 sequences per sample (Figure 1A and E). Two replicate PCR amplifications were performed for each GC and analyzed in order to verify the reliability and reproducibility of our PCR amplification and NGS approach, and to quantify the variance intrinsic to our measurement protocol. Mutation analysis revealed that all samples were very similar in terms of mutation rate and the nature of these mutations (Supplementary Section 1.1). Following previous conventions [3, 15, 21], sequences were grouped together into clones by shared V, J gene segments and CDR3 length, as well as more than 84% junction nucleotide sequence identity, as optimized from the distance to the nearest distribution model [22] (Supplementary Section 2). We analyzed the similarity between the clonal repertoire of each sample and observed a high similarity between samples from the same GC (Figure 1B and C), confirming the reliability of our sequencing protocol, showing a low overlap between IGHV sequences obtained from different GCs (< 0.1), suggesting that GCs are distinct evolutionary environments and thus relatively independent from each other.

We then classified sequences into three categories according to their frequency, corresponding to the dominant clone (i.e. most abundant clone), an expanded clone (frequency $> 1\%$) or a non-expanded clone (frequency $< 1\%$, typically involving less than 1000 sequences belonging to that clone in our data). As highlighted in Figure 1D, the clonal abundance within different GCs is heterogeneous.

To quantify the heterogeneity between different GCs, we computed the clonal diversity in terms of dominance (proportion of the most abundant clone), evenness (homogeneity of the clonal abundance), richness (number of clones) and Shannon entropy, an alternative measure of diversity that is less sensitive to singletons (clonal group with a unique sequence). In agreement with previous mouse models [3, 8] that identified heterogeneous clonal dominance across GCs,

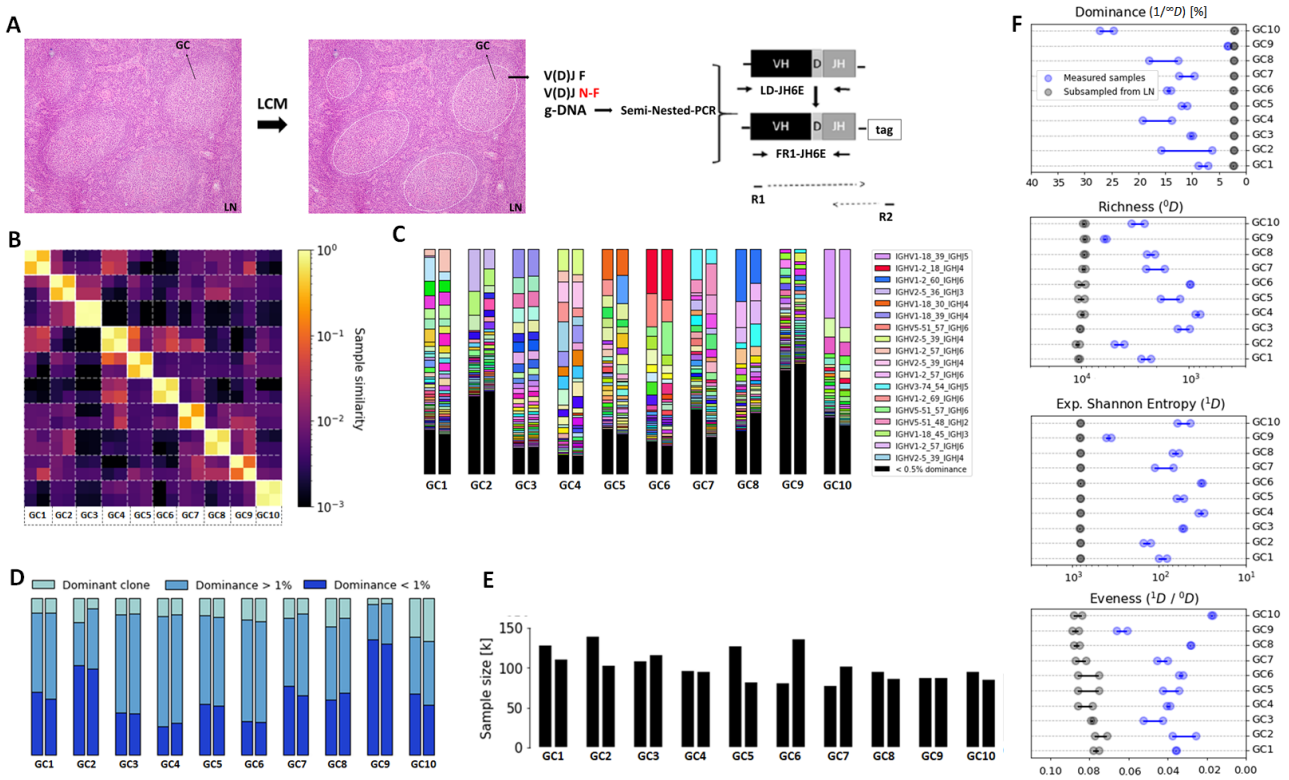


Figure 1: Sample diversity analysis. (A) Experimental framework combining laser capture microdissection (LCM) and Rep-seq from Human Lymph Node (LN) GC's genomic DNA to analyze F (functional) and NF (non-functional) rearrangements via Semi-Nested PCR amplifying the Leader (LD) and Frame work region 1 (FR1). Details of the experimental approach are provided in the methods section. The area within the dashed lines corresponds to the area that was isolated by LCM and used for genomic DNA extraction. H & E staining performed in parallel with LCM is shown (10X magnification) (B) Sørensen–Dice similarity between each sample in terms of clonal abundance. (C) Clone abundance across samples, denoted as `Vgene_JunctionLength_Jgene` (only the 20 most abundant clones are shown in the legend for visual clarity). (D) Proportion of sequences belonging to the dominant clone, expanded clones and non-expanded clones in each sample. (E) Number of sequencing reads in each sample (in units of 1000). (F) diversity analysis across samples in terms of dominance, richness, Shannon entropy and evenness, where ${}^{q}D$ corresponds to Hill's unified notation. To highlight the relevance of studying GCs individually, each sample (blue) is compared to an artificial sample of equivalent size sub-sampled from combining all the obtained sequences (grey).

we show that the clonal diversity within each GC takes a wide range of values (Figure 1C and F). As an example in our setting, the clonal dominance ranged from 5% (GC9) to 30% (GC10). In order to study the role of sample size in the GC heterogeneity quantification, we paired each sample with an artificial sample of equivalent size, obtained from randomly selecting sequences from all the GCs combined. As highlighted in Figure 1F, differences related to sample size are not statistically significant compared to the diversity variation across samples. These findings highlight the importance of studying individual GCs and prove that the heterogeneity is not a result of sub-sampling. In Supplementary Figure S10 we show that the variability in the diversity metrics between samples is consistent across different clonal identification methods, where the same conclusion can be obtained when quantifying diversity in terms of CDR3 abundance only.

Heterogeneous V Gene Repertoire Usage Across GCs. Analyzing the abundance of V genes in each sample reveals that GCs differ significantly in terms of their V gene usage (Figure 2C and Supplementary Section 3). Interestingly, three genes in particular stand out with regards to their frequency across all samples: IGHV1-2, IGHV2-5 and IGHV1-18 (Figure 2E). While these genes were shown to be the most frequently found in the peripheral blood aof adult patients suffering from end-stage renal disease [24], they were still found in a higher abundance in our data than in other LNs and the bone marrow of healthy donors [23] (Figure 2E), suggesting that these V genes may have been positively selected during the GC reaction in the reactive lymph node used in our study. This is further supported by the fact that these genes are found at a higher frequency than the most abundant V genes in non-functional BCR sequences (Figure 2F), which are not subjected to selection. The same figure with functional and non-functional V gene abundances for the same V gene label is provided Supplementary Figure S11.

Next, we analyzed the V gene usage in the top 15 dominant clones from each individual GC. We found that the frequencies of the IGHV1-2, IGHV2-5 and IGHV1-18 genes varied significantly between the GCs (Figure 2D). As an example, in GC6, the IGHV1-18 or IGHV2-5 genes were not found in the top 15 clones, while they were present in 7 and 5 of the top 15 clones from GC9, respectively. Similar results were obtained when using different thresholds for

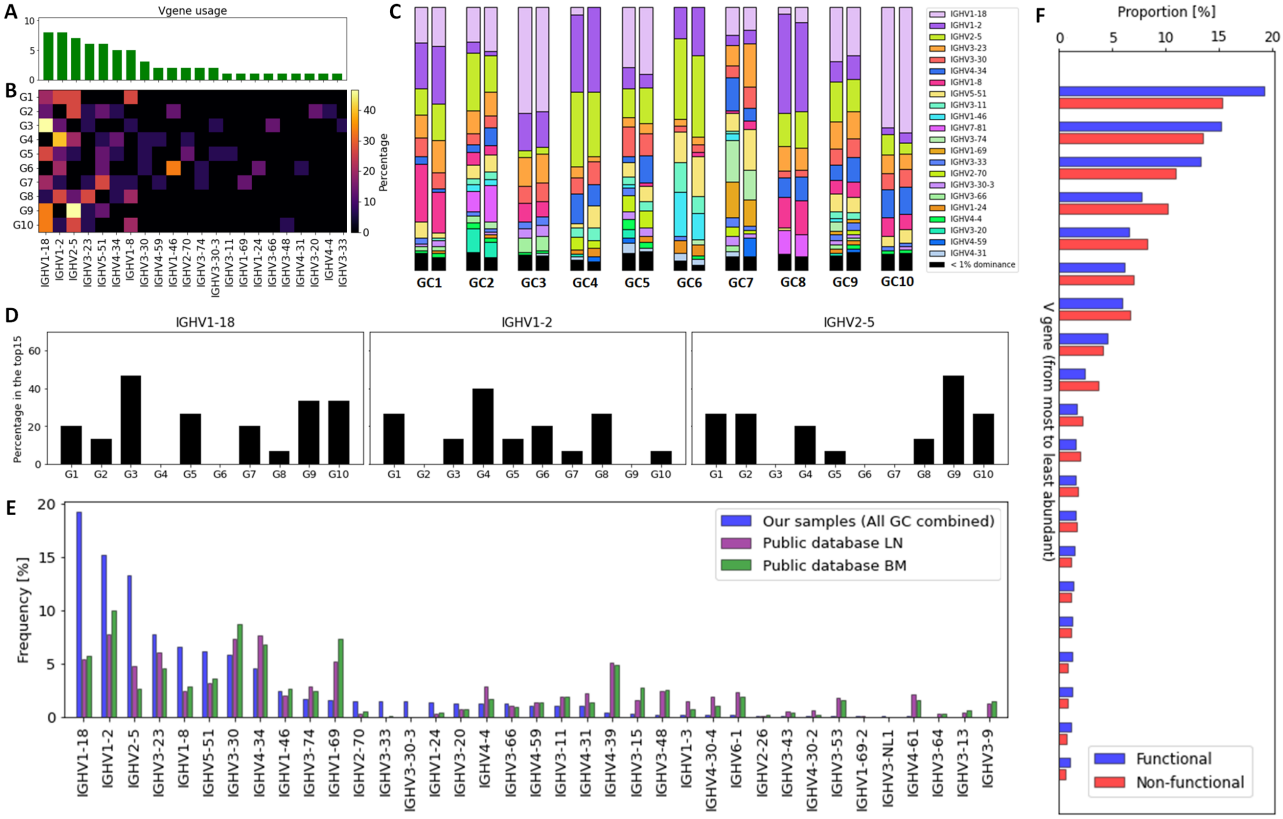


Figure 2: V gene repertoire usage across samples. (A) Histogram expressing the number of GCs for which a given V gene is used by the 15 most abundant clones of that GC, ranked from most to least abundant. (B) Heatmap where a pixel represents the frequency (in %) with which a given V gene has been observed in a given GC's 15 most abundant clones. (C) V gene abundance across samples. (D) Histogram showing the V gene usage by the 15 most abundant clones in each GC. The 3 most abundant V genes (IGHV1-18, IGHV1-2, IGHV2-5) are depicted. (E) Comparison of V genes frequency in our samples (all GCs combined) and in B-cells from a public dataset [23]. (F) V genes frequencies in functional and non-functional B-cells, ranked from the most to the least abundant. The V gene labels are not the same for non-functional and functional sequences, as we are only interested in comparing the shape of the V gene distribution. The same figure with common labels is provided as Supplementary Figure S11.

dominance (Supplementary Section 3). Moreover, we analyzed the most abundant clones and their features with two different sequence-grouping algorithms. We used the TRIP tool [25] to group the sequences with identical CDR3 amino acid sequences into clones. In parallel, we used the Hierarchical Agglomerative Clustering (HAC) algorithm to cluster the junctional sequences [26], which was previously proven to be effective at identifying clones [27]. This approach requires identical V and J gene segments, identical CDR3 lengths, and more than 84% junction nucleotide sequence overlap to assign two sequences to the same clone [27] (see Methods Section 'Grouping clone sequences' for more details). Both algorithms result in a similar identification of the most abundant clones. We observed that the most abundant clones exhibit common features in terms of CDR3 length and V, D and J gene usage (Supplementary Section 4). These data reveal different V genes expansions in individual GCs and convergent evolution of the expanded and most abundant clones.

Heterogeneous Clone Functionality Across GCs And SHM Induced Crippling Mutations. We classified the sequences into four categories based on the identification of a frameshift or a stop codon. The categories are the following: (i) functional sequences, (ii) out of frame non-functional, and in frame non-functional due to an early stop codon induced by (iii) V(D)J recombination or (iv) SHM. Out of frame sequences were defined as having an out of frame junction, with a frameshift detected by IMGT-V-Quest [28]. In frame non-functional sequences were assumed to arise either from V(D)J recombination when having at least one stop codon in the IMGT N-region or from SHM otherwise. More specifically, the BCR junction is formed when the germline V, D and J genes are associated during V(D)J recombination. During this process, additional nucleotides are inserted between the V,D and D,J genes, which are referred to as the N-regions [29]. Consequently, we assumed that stop codons in N-regions are mostly derived from V(D)J recombination errors.

We observe that the majority of sequences is functional ($\approx$ 90 %), and 78 % of the non-functional sequences result from a frame-shift. The remaining sequences are roughly equally divided between SHM-induced and V(D)J recombination-related stop codons. The distribution appears consistent between GCs, with the proportion of non-functional sequences ranging between 10 % and 25 % across samples (Figure 3B). The frequency of non-functional sequences was found to be relatively similar across samples from the same GC, except

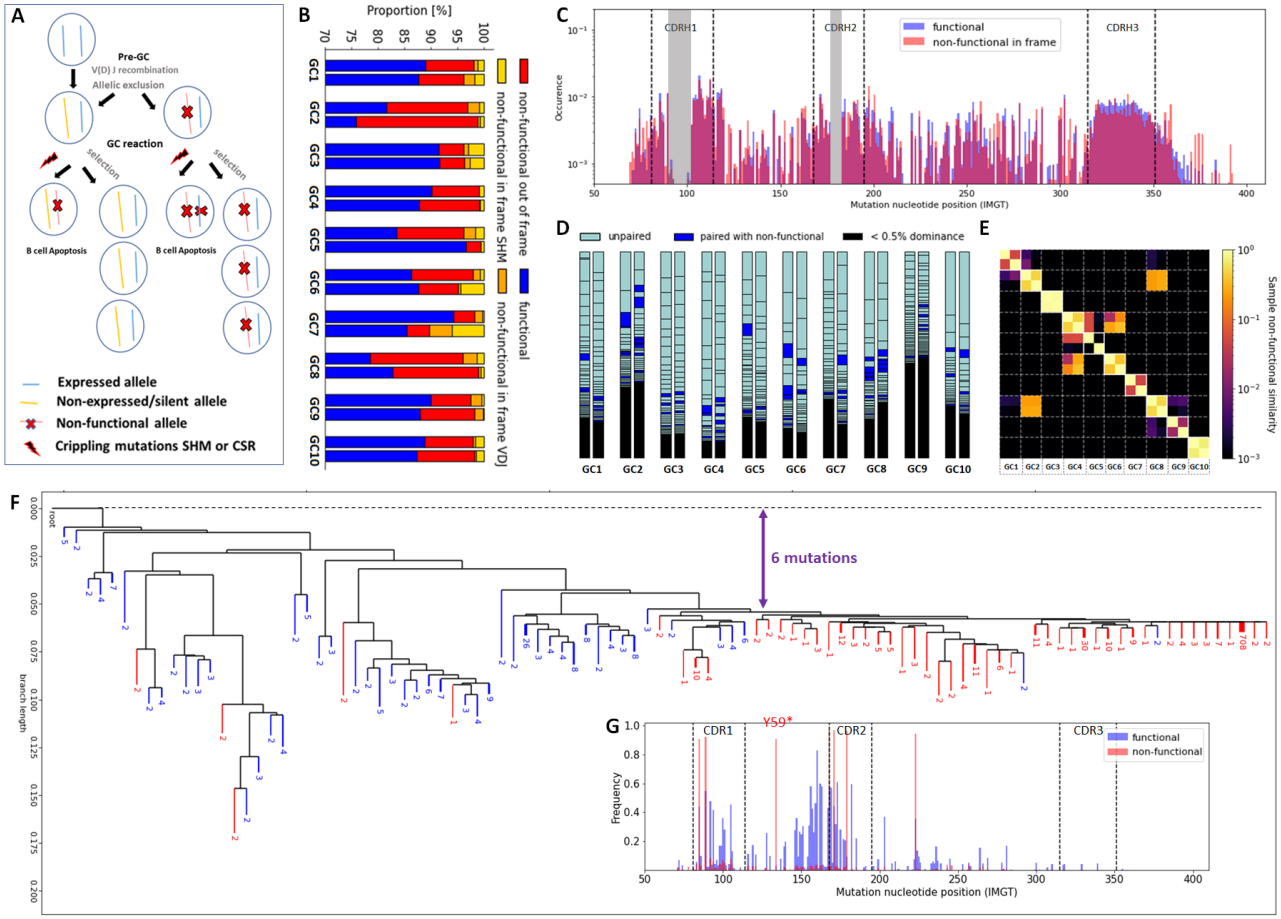


Figure 3: Analysis of functional and non-functional BCR sequences in GCs. (A) Diagram showing the possible fates of functional and non-functional B-cell alleles prior to and during a GC reaction, (B) proportion of functional and non-functional in frame and out of frame sequences in each sample. (C) Mutational spectrum regrouping the position-wise mutation frequency across all sequences. The near zero mutation frequency zones in the middle of CDR1 and CDR2, shown in gray, are due to gaps related to the IMGT numbering scheme that varies across sequences. Mutations were inferred by comparison to the most common sequence segments among all sequences within a given clone. (D) Mapping of functional and non-functional clones across samples. Each rectangle represents a functional clone. Dark blue rectangles represent functional alleles paired with non-functional alleles, while light blue rectangles represent unpaired functional alleles. (E) Sample similarity in terms of frequency of non-functional clones. (F) Phylogeny reconstruction and (G) mutational spectrum of a representative clone including both functional and non-functional alleles (`IGHV3-74_60_IGHJ5`). Mutations associated with functional sequences are depicted in blue, while non-functional ones are shown in red. The crippling mutation Y59* was probably caused by SHM and made the BCRs non-functional. The distance between the majority of non-functional BCR sequences and the root can be quantified in terms of number of mutations, highlighted in purple. The analysis of other F&NF clones is provided in Supplementary Figure S12.

for GC5, where non-functional sequences were found in a low abundance in one of the two NGS samples. The high degree of similarity of the non-functional sequences in GCs 4-6 and GCs 2-8(Figure 3E) was related to shared clones in these GCs, which also carried the non-functional allele, supporting the notion that these are due to V(D)J recombination errors. Studies in mouse models focusing on the role of crippling mutations during SHM identified a limited number of stop codons in the GC with a range of 2-5%, where the crippling mutations were strikingly enriched in B-cells expressing low surface BCR levels, being 7 times more abundant (17.2%) [14] compared to our current analysis of a human lymph node.

We computed the SHM mutational spectrum by regrouping the position-wise mutation frequencies across all sequences. Our data highlighted an expected increase in the mutation frequency in the CDR regions, but no significant differences between functional and non-functional sequences were found in this context (Figure 3C). The same comparison cannot be performed at the clonal level because each clone has different V or J genes and thus the mutation positions cannot be compared across different clones. Furthermore, we also studied the SHM spectrum with regards to the type of mutations induced by AID in each GC for the functional and the non-functional clones. Mutation analysis revealed that all samples were very similar in terms of mutation rate and the nature of these mutations.(Supplementary Section 1.1) Interestingly, no difference between functional and non-functional clones regarding the SHM spectrum was observed suggesting that, the SHM process seems to be stochastic. However, the selection pressure of the functional dominant clones was assessed by studying the replacement to silent mutation (R/S) ratio in

CDR for the functional expanded clones versus the singletons (Supplementary Section 1.2). The R/S ratio indicates that replacement mutations are being selected over silent mutations and it positively correlates to the selection pressure undergoing in a GC environment. We observed increased R/S ratio for the functional dominant clones compared to the 'singleton' sequences, verifying the difference of selection pressure between the expanded and the non expanded clones (Supplementary Section 1.2).

Pairing between functional and non-functional alleles. The expansion of non-functional IGH alleles in the GC may be derived from selected B-cells that harbor both a functional and a non-functional allele due to V(D)J recombination errors [30, 31] (Figure3A). Based on this notion, we paired each non-functional IGH allele with a functional IGH allele by assuming that equal/similar abundance indicates that they come from the same B-cell. We show in the obtained mapping that, the paired functional and non-functional alleles were observed in all the GCs, consistently in both replicates. Furthermore, the frequency of the functional alleles that were paired with non-functional ones varies between individual GCs (Figure 3D).

Identifying non-functional clones derived from SHM. We performed functionality analysis also at clonal level across GCs, and separated the clones into 3 categories. We defined a clone as functional if it contained more than 95% functional sequences (F), non-functional (NF) if it contained less than 5%, and F&NF otherwise, i.e. clones consisting of both functional and non-functional sequences. The F&NF clones corresponded to the 4% of all the clones studied, without significant correlation with their abundance (i.e. being F&NF seems to be independent of being selected). All of them were in frame and likely caused by SHM since the stop codons were outside of the junction region. We found at least 5 F&NF clones with 1000 different unique sequences and investigated these in more detail. After inference of their phylogenetic trees, we can observe a separation between functional and non-functional branches (Figure 3F, Supplementary Figure S12). Thus we can check for mutations that were selected through affinity maturation, as performed previously in mouse models [15]. For example, we observe several selected mutations of the functional sequences in the FWR2 region of the studied F&NF clone (Figure 3G). Still, we did not identify specific mutations consistently selected across clones (Supplementary Figure S12). On the other hand, our analysis revealed that specific crippling mutations occurred independently in different F&NF clones, when they shared the same V genes (Supplementary Table S1). As an example, the mutation `Y59*` occurred consistently in 5 independent clones with the IGHV3 gene.

Regarding the timing of these crippling mutations, the distance between the non-functional sequences and the root in the inferred tree can inform us about the number of mutations the BCR sequences underwent before the crippling mutation occurred, which correlates with the number of cell divisions (roughly one mutation per two cell divisions [9]). It was estimated from our generated trees that an average of 3.6 mutations occurred prior to the crippling mutation (Supplementary Figure S12).

Expansions of specific V genes are frequently involved in stereotypic rearrangements found in B-cell malignancies (IGHV1-8, IGHV1-2, IGHV3-23, IGHV4-34 [32, 33]), autoimmune diseases and infectious diseases (IGHV4-34, IGHV5-51, IGHV1-69, IGHV1-46 [34]). We observed that the majority of F&NF clones with crippling mutations (22/35) use some of those genes. More specifically, the genes IGHV1-2, IGHV3-23, IGHV4-34 and IGHV2-5 were used by the F&NF clones. The SHM can act as a double-edged sword for the organism since it is necessary for an effective immune response but at the same time it introduces mutations that can induce the recognition of self antigens and consequently can lead to autoimmune disease. The GCs not only rely on the selection of the antibodies with the highest affinity but also on autoreactivity *checkpoints* that are needed for the counter-selection of B-cells that can bind to self antigens [1]. As studies in mouse models demonstrated that self-reactive GC B-cells are counter-selected or inactivated by SHM [35], our finding support this hypothesis of counter-selection mechanism of self-reactive B-cells in the human GC. Still, the role of negative selection in the GC is not yet completely clear, despite the deepened knowledge obtained from recent studies using mouse models [36, 14].

B-Cell Reactivation In The GC. Several studies in both rodents and humans by us and others have shown various frequencies of shared clones between different GCs [19, 3, 20], which might be indicative of B-cell reactivation in ongoing GC reactions [8]. We investigated the number of shared clones in a human LN. To increase the robustness of our analysis we considered a clone to belong to a GC only if it was consistently identified in both NGS replicates. We found 10.8 % (396/3650) of functional clones to be present in at least two GCs (Figure 4B). This is in line with a recent study in the mouse, where approximately 10% of the clones were found to be shared between individual GCs induced by CGG immunization [20]. Moreover, some of the shared clones exhibit common features with other shared clones across GCs, regarding V, D and J gene usage and CDR3 length (Supplementary Section 4). Shared clones between different GCs indicate a frequent B-cell recirculation from one GC to another. (Figure 4A). Nevertheless, clones that were expanded in more than two GCs were rare and the majority of the shared clones were expanded only in one GC, with only 5 % of clones found with dominance $> 0.1\,\%$ in more than one GC (Figure 4B). The fact that shared clones exhibit different degrees of dominance in different GCs fits with the notion that GCs are distinct competitive environments, leading to clonal expansions varying by several orders of magnitude across GCs. In fact, a recirculating B-cell enters different GCs at a disadvantage compared to other already established dominant clones, where its initially low abundance makes it more difficult to compete for T cell help [9].

Interestingly, the distribution of shared clones among GCs fits a Poisson distribution well (Figure 4B), suggesting that the reactivation of a B-cell in a different GC is a memoryless stochastic process where the past evolutionary history of the B-cell does not play any role. According to the Poisson distribution and denoting as λ the occurrence rate (in our case the rate of B-cell reactivation), the number of occurrences N at the exposure time t is a random variable of mean $\mu = \lambda t$ and distribution:

$$P[N = n] = \frac{\mu^n e^{-\mu}}{n!} \,. \tag{1}$$

Fitting Eq. 1 to the observed distribution gives $\mu = 0.25$.

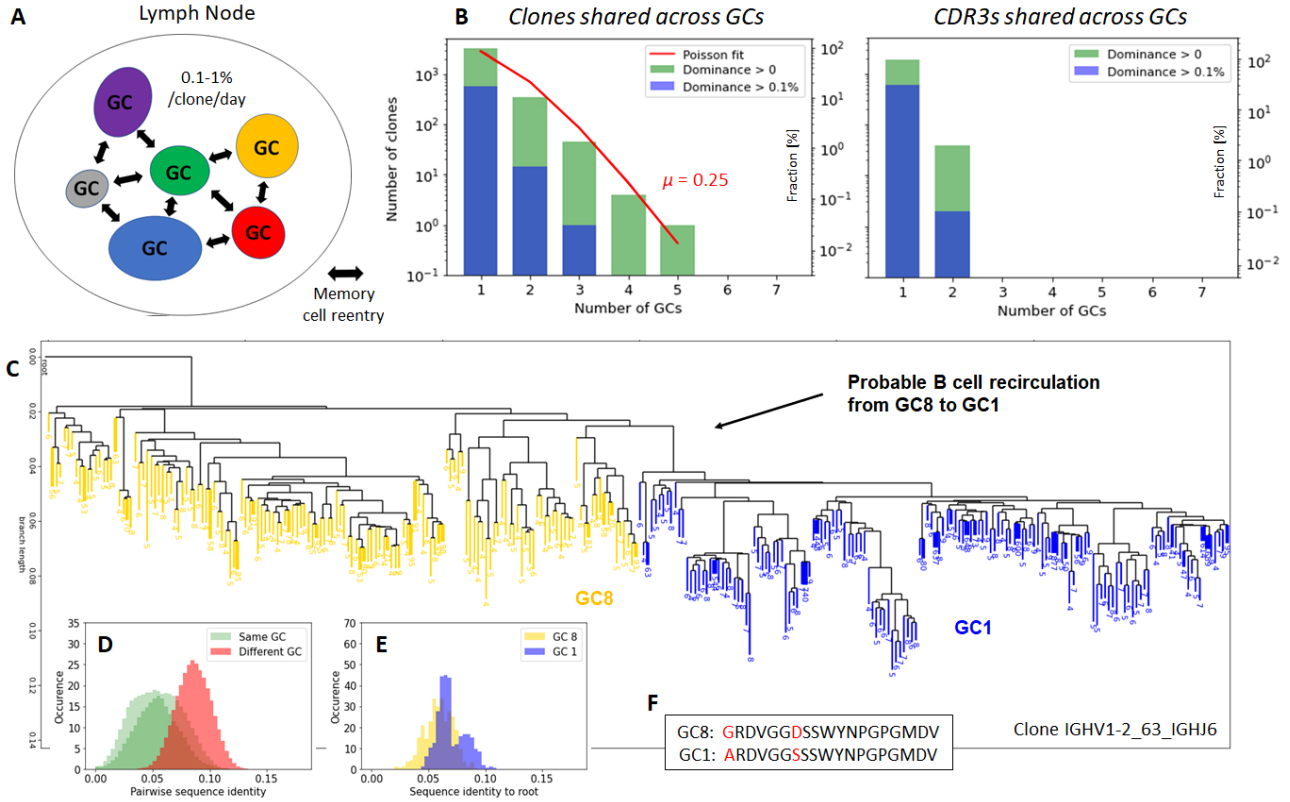


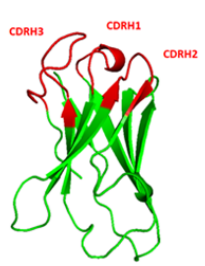
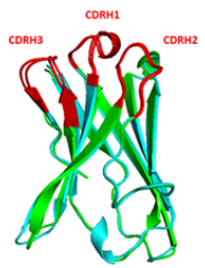
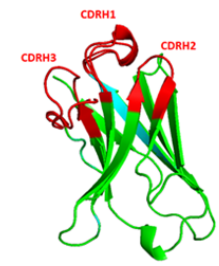
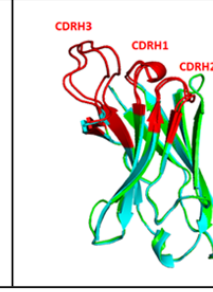
Figure 4: (A) Cartoon representing the probability of clones being shared between different GCs of the same LN. (B) Number of clones and number of identical CDR3 sequences as a function of the number of GCs in which they are found (in green), fitted to a Poisson process (in red) of parameter rate μ. The blue histogram depicts the clones that are found with at least 0.1 % dominance. (C) Phylogenetic reconstruction of clone `IGHV1-2_63_IGHJ6`, where each colored branch corresponds to an unique BCR sequence. The number of replicates of each BCR sequence is indicated at the bottom of the branch. Only the 150 most abundant sequences from both GC8 and GC1 are shown for visualization clarity. (D) Histograms of the pairwise sequence similarity between BCRs of the same GC (green) and different GCs (red). (E) Histograms of the pairwise sequence similarity between BCRs and the root of the joint tree including both GC1 and GC8. (F) Most common CDR3 sequences among BCRs from both GCs.

From this, we can compute the probability that a given clone seeds at least one other GC during the whole GC reaction $P(N \geq 1) = P(N \geq 0) - P(N = 0) = 1 - e^{-\mu} = 1 - e^{-0.25} \approx 22\,\%$.

We can also estimate the probably of seeding per clone and day. When using 20 days as a typical lifetime of a GC, generated in response to immunization with a simple hapten-protein conjugate sucha as NP-CGG [3], we can estimate the GC B-cell reactivation rate to be $\lambda \approx \frac{0.25}{20} = 0.0125$ seeding per day per clone. However, it is worth keeping in mind that the calculated reactivation rate is only an upper limit, as the GC lifetime most likely is much clonger in chronically activated LNs. For instance, lifetimes of up to 100 days have been observed for some viral infections [37], and GCs formed in response to synthetic antigens can persist for up to 1.5 years [38]. Such extended GC lifetimes would result in much smaller reactivation rates. As an example, if the data would have been collected at day 200 of the GC reaction, the estimated reactivation rate would decrease to 0.1 % (Figure 4A).

We are interested not only in how many clones (sequences sharing the same V, J gene segments and CDR3 length, as well as more than 84% junction nucleotide sequence identity) are shared across GCs but also in how many identical CDR3s are observed between different GCs. We observed that only 2 % (38/1885) of the CDR3s are shared among GCs, which is considerably less than for clones. This indicates that the CDR3 typically continues to mutate after B-cell recirculation in the new GC environment (Figure 4F). To gain insight into the parallel evolution of a given clone in different GCs, we inferred phylogenetic trees from clones found in multiple GCs. In Figure 4C, we show the phylogenetic tree inferred from clone `IGHV1-2_63_IGHJ6`, which was found expanded with dominance $> 1\%$) in two different GCs. The hierarchical reconstruction of the tree clearly separated the two GCs, where sequences in the same GC were more similar to each other than sequences in different GCs (Figure 4D). In this example, sequences in GC8 are generally closer to the root (unmutated V gene, J gene and consensus CDR3) than GC1 in terms of sequence similarity (Figure 4E), suggesting that GC8 yielded a clone that initiated GC1. This apparent temporal separation (Figure 4D&E) was also observed in the Brainbow mouse model after 20 days of subcutaneous immunization with alum-adsorbed CGG [39, 3], suggesting that mutations may be selected through affinity maturation at different rates and through different evolutionary processes in different GCs. Such a clonal tree pattern was found to different extents in other shared clones and is thus not anecdotal (other shared clones are provided in Supplementary Figure S13).

A

GC	GC3 GC7	GC3 GC5	GC4 GC6	GC1 GC3
Dominance	4.02% 1.12%	0.55% 2.22%	0.12% 0.09%	0.26% 1.01%
Vgene	IGHV1-8 IGHV1-8	IGHV3-30 IGHV3-30	IGHV4-31 IGHV4-30	IGHV3-23 IGHV3-23
Jgene	IGHJ6 IGHJ6	IGHJ4 IGHJ4	IGHJ4 IGHJ4	IGHJ3 IGHJ2
CDRH1	GYTFT**SYD** GYTFT**SYD**	GFTF**SSY**G GFTF**SSY**G	GGSISS**GGYY** GGSISS**GDYY**	GFTF**SSYA** GFTF**SSYA**
CDRH2	MNPNS**G**NT MNPNS**G**NT	I**SYD**GSNK I**SYD**GSNK	I**YYS**G**S**T I**YYS**G**S**T	I**SGSG**G**S**T I**SGS**GG**S**T
CDRH3	ASYS**RGGGYYY**GMDV ARI**GAAGFYYY**GMDV	AR**DLGGN**FDY AK**DLGANR**DY	AR**GGGYD**KAFDY AR**GRGYDFT**LDY	AKEN**QYYDFWSGYYTN**AFDI AKGS**THYDFWSGYPYWY**FDL
V region difference	1 nucleotide	3 nucleotides	4 nucleotides	5 nucleotides
J region difference	0 nucleotide	0 nucleotide	0 nucleotide	7 nucleotides
CDRH3 similarity	0.67	0.74	0.71	0.63
CDRs score	0.89	0.91	0.87	0.88
Paratype score	0.76	0,92	0.77	0.82
Ab Ligity score	0.86	0.35	0.32	0.45
Predicted structures				

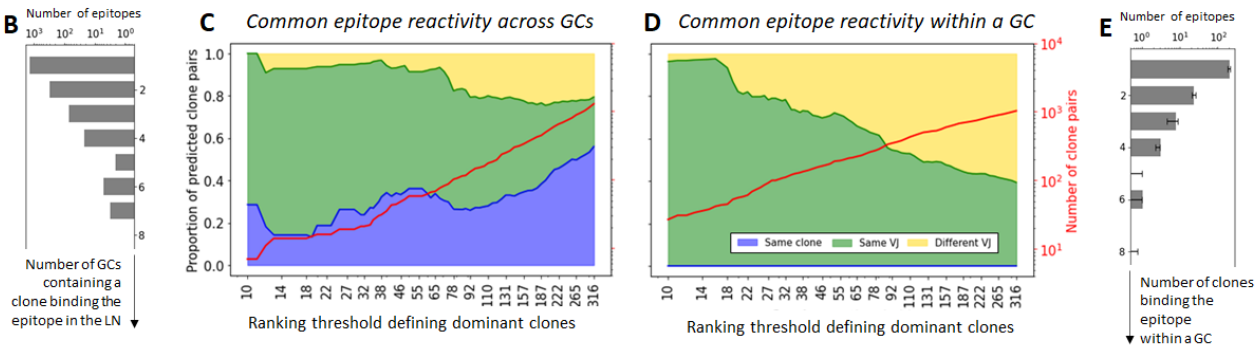


Figure 5: Common epitope reactivity across GCs. (A) Example of predicted dominant clones pairs that bind to similar epitopes. The paratopes residues are emphasised in bold, and the differences between the two clones highlighted in red. (B) Number of GCs for which at least one clone is within each structural (epitope binder) group for a dominance threshold of 300. (C) Number of predicted clone pairs with common epitope reactivity across different GCs and (D) within the same GC as a function of the dominance ranking threshold. The pairs are separated into three categories: (blue) same clone found in two different GCs, (green) clone with the same V and J gene, but with CDRs similarity < 0.84, and (yellow) clone with different V or J genes. (E) Median of the number of clones within each structural (epitope binder) group within each GC for a dominance threshold of 300. The analysis was done independently for each GC and the error bar represents the first and third quartile.

Convergent Epitope Reactivity Across GCs. We studied the structural similarity of dominant clones by considering their paratope structure via paratope modeling. In order to study the paratope similarity of dominant clones, we assigned to each clone in each GC a representative sequence, defined as the most abundant sequence within that clone. Then, we combined three metrics, the *CDRs similarity*, *Paratype* [40], and *Ab-Ligity* [41], to identify clones that are likely to bind similar epitopes (Supplementary Section 5). We refer to those predicted clones as clone pairs. If the similarity of the clone pairs is above the optimized threshold for two of the three aforementioned metrics (see Supplementary Information for details), the clone pair is predicted to have evolved towards binding the same epitope. We show a few illustrative examples of predicted clone pairs in Figure 5A. Although all pairs have a relatively low CDR3 sequence similarity (ranging from 0.63 to 0.74), the pairs exhibit identical (or nearly identical) CDR1 and CDR2, as well as similar paratope

(> 0.76) and high structure similarity. The first two clone pairs in the Figure 5A (1st and 2nd column starting from the left) could be the same clone misclassified as a different clone because of too many newly accumulated mutations in the CDR3. Alternatively, they could have originated from highly similar naive B-cells. The fact that they exhibit differences in the nucleotide sequence of the N region of the V and J genes suggests that they are different clones exhibiting common epitope reactivity. The last two clone pairs (3rd and 4th column starting from the left) have different V and J genes, respectively, with a 4(7) nucleotides difference between the two V(J) regions and considerable CDR3 dissimilarity, but their paratope and CDR3 loop structure were predicted to be highly similar (Figure 5A).

We classified the predicted clone pairs with common epitope reactivity within one GC and across GCs into the following categories:

- *Same clone present in different GCs:* This is the most straight forward way to study GC convergent evolution (blue curve in Figure 5C).
- *Same V and J gene, but CDR3 similarity* < 0.84*:* This category needs to be carefully analyzed since some clones could be miss-classified as different because they have accumulated too many mutations in the CDR3 and their CDR3 similarity is thus below 84%, the threshold we used for clone identification in our study (green curve on Figure 5C&D).
- *Different V or J gene:* This category is based on literature showing that different gene rearrangements can yield convergent epitope reactivity [42, 40]. For this category, however, there is a chance that some pairs are miss-classified due to a misalignment in the V or J genes, which is especially likely to happen when their CDR3s are nearly identical (yellow curve on Figure 5C&D).

We also investigated the number of predicted clone pairs within and across GCs based on each of these categories. We considered multiple dominance ranking thresholds, ranging from the top 10 to the top 300 clones in each GC, and showed the predicted number of identified clone pairs in each of the 3 categories with an increased number of dominant clones (Figure 5C&D). We observe an increased number of predicted clone pairs across and within GCs for all three categories, with the number of identified clone pairs approaching 1000 when the top 300 dominant clones per GC are considered (red line in Figure 5C&D). Most of the predicted clone pairs come from the second category (same V and J genes and CDR3 similarity < 0.84, depicted by the green curve in Figure 5C&D).

Regarding the paratope commonalities between dominant clones across and within GCs, we characterized paratope similarity by studying the paratope distance between clone pairs, defined as (paratype distance + CDRs distance)/2. We show that the distribution of paratope distance is very similar across and within GCs, with only a slightly lower average paratope distance within GCs than across GCs (Supplementary Section 6). This finding suggests that different evolutionary forces can lead to convergent immune responses.

Discussion

In this study we developed a method to analyze the IGHV repertoire in single GCs by combining LCM of individual GCs from human LNs with NGS Rep-seq from gDNA. Our study revealed the heterogeneity of an ongoing immune response in single GCs, within a LN of a 46 years old woman, but also the convergent evolution of different evolutionary spaces, the single GCs. The reproducibility of our method was demonstrated by performing replicate IGHV amplification and NGS analyses. Moreover, the use of genomic DNA in our method offers a better alternative to typical mRNA-based studies, which introduce bias due to differential expression of mRNA between different B-cell types, and at the same time it allows for the study of non-functional clones.

Similarity and diversity analyzes of samples showed that human GCs are distinct evolutionary spaces governed by the evolutionary pressure of antigen capture and selection driving the output of the GC reaction. In line with previous mouse studies [3, 8, 15], we observed a relatively low sequence similarity between the GCs, as well as heterogeneous clonal diversity in terms of dominance, evenness, richness and Shannon entropy between individual GCs. The heterogeneous clonal diversity was independent of the clone identification methods studied and was not a result of sub-sampling, since the differences related to sample size were not significant compared to the diversity variation across GCs. The expanded clones showed commonalities in terms of V, D, J gene usage as well as CDR3 length. This finding suggests convergent evolution of the expanded clones in different individual GCs.

Analysis of the VH gene repertoire revealed that a few VH genes were found more frequently than in public repertoire data, suggesting that they were positively selected during the GC reaction in the studied human LN. Acccordingly, The VH gene usage was found to be heterogeneous across GCs.

With the help of phylogenetic tree inference, we showed that there is convergent evolution of different GCs of the same LN by identifying the seeding of GC-experienced (potentially memory B cells) B cells in different preexisting GCs, which is in line with previous studies in rodents and humans [19, 3, 20]. While earlier studies identified the frequent presence of clones that were shared between individual GCs in the same LN, both in mice and humans [19, 3], those studies have sequenced limited numbers of B cells with limited statistical analysis/evaluation. Nevertheless, a more recent study in a mouse model argues that this is a restricted phenomenon under typical immunization-boost regimens, with less than 10 % clones found in multiple GCs, and secondary germinal centers consisting predominantly of B cells without prior GC experience [20]. Although coming from on ongoing human GC response in an uncontrolled environment, our results are in line with the most recent mouse model studies, since we identify approximately 10% of clones as being shared between GCs.

Using antibody modeling and paratope prediction tools, we quantified the functional convergence of clones in different GCs by predicting candidates of clone pairs that potentially bind the same epitope, suggesting a shared pool of antigens across different GCs. Interestingly, these cases were relatively rare, which suggests that the number of epitopes involved in the LN reaction is high (estimated to range from 500 to 5000 epitopes). In addition, the prediction of common epitope reactivity can be used to predict the number of epitopes in the GC and/or the LN. We clustered all clones within structural groups based on the predicted epitope binding, such that each clone in a given group is predicted to bind the same epitope as at least one other clone of the same group. Overall, only a minority of clones were found to bind the same epitope across GCs (Figure 5B&E), thus we expect the

number of epitopes present in GCs and LNs to be quite high. With a simple statistical model (Supplementary Section 7), we derived the number of epitopes targeted in the lymph node to be around 5000, and each GC to be specialized to ∼1000 epitopes on average. These estimations should be taken as a higher bound, as we expect to have missed a large fraction of antibodies that react to the same epitope (recall estimated to be as low as 10%, see Supplementary Section 5). Assuming that 90% of the pairs were missed, our estimate would be significantly reduced to roughly ∼500 epitopes in the lymph node and 100 epitopes in each GC. Given that epitopes are around 15 AAs long [43], and that antigen chains typically consist of hundreds of AAs, many of these epitopes could belong to a single antigen. As an example, insulin was reported to have more than 100 epitopes [44].

Finally, the use of genomic DNA, as opposed to RNA, which is common practice, allowed us to study and quantify the role of selection in the SHM spectrum by studying the origin and fate of non-functional alleles at the individual GC scale. We showed that the SHM pattern in terms of both position and spectrum was independent of GC B-cell selection. For the first time, we analyzed the non-functional alleles stemming from the SHM or V(D)J recombination process in individual GCs and we observed that the source of the non-functional alleles was approximately equally divided between these two processes. The distribution was roughly consistent across GCs. We identified several mutations that were introduced by SHM and were either selected through affinity maturation or crippled the BCR. While this type of analysis was previously performed in mouse models in naive and GC B cells from the spleen and Peyer's patches at the B-cell population level, it was not in individual GCs [15]. Interestingly, the majority of clones with crippling mutations used V genes that are known to be implicated in stereotypic rearrangements in autoimmunity and B cells malignancies [33, 34], such as IGHV4-34, suggesting counter-selection of those B cells in the GCs. Our result are in line with studies in mouse models, which showed that self-reactive GC B cells are counter selected or inactivated by SHM [35].

Our study underscores the stochastic nature of SHM, which was found to be selection-independent. It also quantifies its role in producing non-functional clones by analyzing the frequency of non-functional sequences derived from SHM. It shows that the heterogeneity of the individual GCs in terms of diversity and clonal expansions is a conserved phenomenon between mouse and human. Additionally, even though the individual GCs are heterogeneous and separate from each other, at the same time they exhibit convergent features both in mice [3, 8] and humans.

Our results will help to improve existing computational models of the GC reaction, as well as aid in the development of novel computational models, by providing more accurate parameters stemming from a real life situation in a human LN. As GCs and LNs are stochastic systems that display a high level of variability even within the same lymph node of the same individual [45], mathematical models have been widely used to deepen our understanding of the cellular and molecular processes characterizing these complex dynamic systems [46, 9, 47]. Nevertheless, these models are limited due to the lack of available data to correctly estimate the numerous unknown parameters describing cellular and intracellular interactions. Since our study characterizes clonal abundance, clonal diversity and clonal functionality data in several GCs of the same lymph node, it represents an important step towards the correct parametrization of GC models. Additionally, it could also be used to compare the normal human BCR repertoire with data from disease models, with the final goal of potentially identifying shared BCR structures that could originate from similar antigenic stimuli.

Finally, regarding further research, one question of interest would be to which extent our findings in lymph nodes are applicable to different human tissues, such as Peyer's patches and tonsils. If the comparison is relevant, our study conclusions could be extrapolated to investigate the GC reaction mechanisms in other similar environments. Additionally, the use of mouse models in combination with gDNA could enable the characterization of the origin and fate of non-functional clones in more controlled environments. Indeed, controlled mouse immunization experiments enable the in-vitro investigation of the antibodies generated as a response to known antigens, which is not possible in human studies, where GC reactions arise in response to unknown antigens.

Materials and Methods

IGH Next Generation Sequencing / Rep-seq. In brief: LCM of the individual GCs from human LN was combined with the extraction of gDNA and an NGS approach in order to study both the functional and the nonfunctional clones.

Tissue Preparation. Patient frozen LN tissue from the pathology department of the Amsterdam UMC hospital was used in this study. (LN studied was a cervical LN resected out of a 46-yr-old woman suffering from chronic sialadenitis.) The LN was fresh-frozen in liquid nitrogen shortly after surgical resection. Five serial cryosections (10 µm per cryosection, Cryostat NX-70, Thermo Scientific) of the LN tissue on PEN slides (1mm, Zeiss) were used in order to obtain higher DNA concentrations. Immunostaining was performed in order to visualize the GCs: The PEN slides were incubated with drops of Hematoxylin (KLINPATH, VWR Life Science) for 1 minute at Room Temperature (RT), after a quick and gentle wash with demi water and tap water they dried overnight at RT. In parallel, H&E staining of the frozen tissue was performed on normal (TOMO) slides at Amsterdam UMC diagnostics.

LCM Of Individual GCs. After tissue preparation, we used the objective 5x of the Leica PALM-LMD6 to draw and laser capture individual GCs from 5 serial cryosections of the human LN studied. The laser capture was performed 5 times for each GC in the same 0.2ml Eppendorf tube.

gDNA Extraction From Individual GCs And DNA Concentration. 2 µl 1x proteinase K (recombinant PCR Grade, 25 mg - Roche) dissolved in 2,5 ml TE (10 mg/ml) and 20 µl 1x proteinase K lysis buffer (50 mM TRIS-HCl (PH8,5-ENZO), 100 mM NaCl (Sigma), 1 mM EDTA (Sigma), 0,5 Tween-20 (Biorad), 0,5 NP-40 (Sigma Aldrich) were added in the tubes used for the LCM and the samples are incubated at 56°C, overnight. The proteinase K was inactivated by incubation for 5 min at 95°C. The Qubit dsDNA HS Assay kit (Invitrogen) was used in order to quantify the extracted g-DNA, according to manufacturer's protocol.

Multiplex-Nested-PCR protocol. The gDNA extracted from individual GCs was used for the Multiplex-Nested PCR, consisting of two rounds. Round 1: Amplification of the leader genes region (LD). Round 2: Amplification of the FR1 region with primers containing Illumina adapter overhang nucleotide sequences. The input used for Round 2 was the Round 1 PCR product. We used technical replicates to check the method reproducibility (duplicates of PCRs for the same DNA/GC).

For Round 1: PCR amplification of gDNA from individual GCs (2 ng input) was performed with1 ul of JH reverse primer (10 uM, provided by MERCK) and 1.8 ul of LD forward primer set pools (10 uM per primer, provided by MERCK) using 25 ul of Q5 Hot Start Master Mix (2X) (New England Biolabs) for a total volume reaction of 50 ul.

For Round 2: PCR amplification of the Round 1 PCR product (1 ul) with 1 ul of JH reverse primer (10 uM, provided by MERCK) and 1 ul of FR1 forward primer set pools (10 uM per primer, provided by MERCK) using 25 ul of Q5 Hot Start Master Mix (2X) (New England Biolabs) for a total volume reaction of 50 ul.

The following PCR program was used for both Round 1 and Round 2: 5 min at 95 °C; three cycles of 5 s at 98 °C and 2 min at 72 °C; three cycles of 5 s at 98 °C, 10 s at 65 °C, and 2 min at 72 °C; and 25 cycles of 20 s at 98 °C, 30 s at 60°C, and 2 min at 72 °C; with a final extension cycle of 7 min at 72 °C on a Biometra T-ADVANCED PCR machine. Primers and adaptors used in Round 1 and Round 2 are provided in Supplementary Figure S9.

Round 2 PCR Product Purification. Electrophoresis of the Round 2 Multiplex-Nested-PCR product was performed on a 1% agarose (Molecular grade, Bioline) gel, followed by Gel excision of the 400bp PCR product with Macherey-Nagel PCR clean up and gel extraction kit, according to manufacturer's protocol.

NGS Library Preparation And Sequencing. The concentration of the gel excised Round 2 Multiplex-Nested-PCR product was quantified by using the Qubit dsDNA HS Assay kit (Invitrogen) according to manufacturer's protocol. 100ng of this PCR product were used for the preparation of the 16S Metagenomic Sequencing Library for the Illumina MiSeq System, according to manufacturer's protocol. More specifically, for the library construction the Illumina indexes were attached to the amplicon with 8 PCR cycles using Nextera Indexted Primer. The final purified (AMPure XP beads) product was then quantified using qPCR according to the qPCR Quantification Protocol Guide (KAPA Library Quantification kits for Illumina Sequecing platforms) and qualified using the TapeStation D1000 ScreenTape (Agilent Technologies, Waldbronn, Germany). The paired-end (2×300 bp) sequencing was then performed by Macrogen using the MiSeq™ platform (Illumina, San Diego, USA)

Sample Reads Preprocessing. Two samples were acquired for each of the 10 GCs. For each sample, the reads were merged with BBmerge [48] (using the recommended command for optimal accuracy `bbmerge-auto.sh`). We compared the reads obtained from BBmerge with the ones from PEAR [49] (using a quality index of 30) and no significant differences were found. Then, the primers and adaptors sequences were removed by using a customized Cutadapt script in order to remove 7 primers from the merged reads [50]. The quality of the preprocessed reads was tested by the FastQC tool available in the Galaxy platform (`https://usegalaxy.org`). The sequences were transformed to FASTA format. Finally, the alignments to V and J germline sequences, as well as the identification of CDRs/FWRs and frameshifts were obtained by submitting the reads to the IMGT-High-V-Quest web portal [28] (`http://www.imgt.org/HighV-QUEST/search.action`).

Grouping Sequences Into Clones. Following previous conventions [3, 15, 51, 52, 23, 21], sequences were grouped together into clones if they shared the same V, J gene segments and CDR3 length, as well as more than 84% junction nucleotide sequence identity. The threshold was optimized using the distance to nearest distribution model [22] as well as a negation table approach [53], as detailed in the Supplementary Section 1. The junction clustering was performed with a Hierarchical Agglomerative Clustering (HAC) algorithm [26], previously proven to be effective at identifying clones [27]. As our data is especially susceptible to incorrect V-gene assignment, due to part of FRW1 missing in our sequencing data (the first 15 to 20 residues), clones with different V genes from the same subgroup were merged together when the CDR3 of their most common sequence were identical.

As a way of checking for the consistency of our analysis, we alternately grouped sequences into CDR3 groups irrespectively of their V and J genes, which can be directly performed with the TRIP tool [25], a software framework implemented in R shiny.

Similarity Between Samples. To quantify the similarity between samples X and Y, we first normalized the number of sequences belonging to each clone by the total number of sequences in that sample such that $|X| = |Y| = 1$. Then, the similarity was defined as the Sørensen–Dice coefficient, known to be more robust than the Jaccard index when some information is missing from the dataset [54]

$$Similarity = \frac{2|X \cap Y|}{|X| + |Y|} = |X \cap Y| = 1 - \sum_{i \in \text{clones}} \frac{|x_i - y_i|}{2} \quad (2)$$

where x_i and y_i are the abundance of clone i in sample X and Y, respectively. A similarity of one indicates perfectly identical samples and one of zero that they do not share any clone.

Quantifying Clonal Diversity. To quantify the diversity of GCs, we used different statistical representations of GC clonal composition: richness, dominance, entropy and evenness. The diversity indices are denoted with the unified framework of Hill (${}^{q}D$) [55], where values of q that are smaller than unity disproportionately favor rare species (thus more sensitive to singletons), while values greater than unity disproportionately favor the most common species.

- **Dominance ($1/{}^{\infty}D$).** The clonal dominance is defined as the number of cells belonging to the most abundant clone divided by the total number of observed cells in the GC.
- **Richness (${}^{0}D$).** The clonal richness is defined as the number of clones in the GC. To provide an estimate

of the total number of clones in a GC based on an incomplete sample, we use the bias-corrected Chao1 formula [56, 57]:

$$N_{\text{Chao}} = N_{\text{obs}} + \frac{f_1\,(f_1 - 1)}{2\,(f_2 + 1)} \tag{3}$$

where N_{obs} is the number of observed clones in the GC, f_1 is the number of clones detected exactly once, and f_2 the number of clones detected exactly twice. In our setup, the second term of Eq.3 typically equals between 50 % and 100 % of the number of observed clones.

- **Shannon entropy ($\log[^1D]$).** The Shannon entropy has been a popular diversity index in the ecological literature. It is a measure of diversity that accounts "*fairly*" for the abundance of both rare and frequent clones [58] and is thus less affected by the presence of singletons. Defining the sample size n, and $p_i = \frac{X_i}{n}$ as the normalized occurrence of clone i in the sample, the Shannon entropy is defined as:

$$H = -\sum_{i \in \text{clones}} p_i \ln(p_i) \tag{4}$$

Similar to the Chao richness, we also use an estimator to infer Shannon entropy with incomplete sample information [59]. Defining the sample coverage $C = 1 - \frac{f_1}{n}$, we adjust the relative species abundances with $\tilde{p}_i = p_i C$,

$$H_{\text{Chao}} = -\sum_{i=1}^{N_{\text{obs}}} \frac{\tilde{p}_i \log(\tilde{p}_i)}{1 - (1 - \tilde{p}_i)^n} \tag{5}$$

- **Evenness ($^1D/^0D$).** The evenness quantifies the homogeneity of clonal abundances in a GC. It is defined as the exponential of the Shannon entropy normalized by the richness of the GC [55]. The Chao corrected richness and Shannon entropy were used to compute this indicator:

$$Evenness = \frac{\exp(H_{\text{Chao}})}{N_{\text{Chao}}} \tag{6}$$

The Evenness is bounded between 0 and 1, where an evenness of one corresponds to a perfectly homogeneous sample, i.e. all clones have the same occurrence.

Classifying Non-Functional And Functional Clones. Each sequence was labeled either as non-functional when having a frameshift or a stop codon, or functional otherwise. For each clone, a *productivity index*, defined as the number of productive sequences divided by the total number of sequences in that clone (thus between 0 and 1), was computed.

We further classified non-functional sequences into three categories: (i) out of frame (ii) early stop codon induced by VDJ recombination and (iii) early stop codon induced by SHM. Out of frame sequences were defined as having an out of frame junction, with a frameshift detected by IMGT. In frame non-functional sequences were assumed to arise either from V(D)J recombination when having at least one stop codon in the IMGT N region, or from SHM otherwise.

Phylogenetic Tree Representation. We used phylogenetic trees to visualize the evolution of the antibody receptor sequences. In such representation, each founder cell defines the unmutated germline of a new tree, and newly acquired mutations are ideally represented as downstream nodes. We defined the root by taking the unmutated V,J germline and filling the remaining junction region with the consensus sequence of all available sequences within the clone. To compute the trees, the grouped sequences from each clone were first aligned with ClustalW [60]. This alignment was necessary to infer a mutation matrix, as tree inference algorithms typically require sequences to be aligned and of the same length, while experimentally determined sequences have different lengths, and include insertions and deletions. To infer trees from a large amount of sequences, we used a hierarchical clustering approach, where the most similar sequences were grouped together and progressively aggregated with other groups until the root node (maximum distance) was reached (Neighbour Joining method in ClustalW). Such a method was preferred over typical Bayesian Maximum likelihood estimation [39] or Markov chain Monte Carlo sampling [61], because the latter were intractable for the amount of sequences per clone in our data ($>$ 5000 sequences), and would thus require significant subsampling leading to an important loss of information in our phylogeny analysis.

Sequence Similarity. The Levenshtein distance [62], defined as the minimum number of edits required to transition form one sequence to the other, is a common metric to quantify sequence similarity. To reduce the bias caused by length differences, we used a normalized Levenshtein distance [63] that incorporates the length of both sequences and satisfies the triangle inequality. Given two strings s_1, s_2 and $Lev(s_1, s_2)$ the Levenshtein distance between s_1 and s_2, the normalized Levenshtein distance Lev_{norm} is defined as:

$$Lev_{\text{norm}}(s_1, s_2) = \frac{2 \cdot Lev(s_1, s_2)}{|s_1| + |s_2| + Lev(s_1, s_2)} \tag{7}$$

Paratope And Antibody Modeling. The paratope residues were predicted by submitting the antibody sequences to Parapred [64] (`https://github.com/eliberis/parapred`). To convert the output of Parapred, binding probabilities, into a binary label, we selected a threshold of 0.67, shown to be optimal by the authors of the original paper [64]. Antibody sequences were submitted to repertoire Builder [65] (`https://sysimm.org/rep_builder/`) to construct the antibody structure from a homology model. As part of the FRW1 region was missing from our sequencing data (the first 15 to 20 residues), missing nucleotides were filled with the unmutated sequence of their respective IGHV. The obtained structures were visualized with Pymol [66].

Functional Convergence. While antibodies within the same clones are likely to target a common epitope [40], it has also been observed that clones from different V and J gene backgrounds express common epitope reactivity [67]. To identify different clones that are potentially binding to the same target epitope, we combine three approaches.

- **CDR similarity:** As the majority of the paratope residues are localized on the CDR loops of the antibody ($\approx$ 90% according to [68]), antibodies that exhibit high CDR similarity are likely to express common reactivity (although expressing different V or J genes). Given two sequences s_1 and s_2 with CDRs denoted as H1,H2, and H3, we define the CDR similarity as the average

Levenshtein distance between their respective CDRs.

$$\text{CDR similarity}(s_1, s_2) = \frac{1}{3} \sum_{i=[1,2,3]} Lev_{\text{norm}}(\text{H}i_1, \text{H}i_2) \tag{8}$$

The threshold for common reactivity prediction was set to 84% for this metric, as optimized in Supplementary Section 5.

- **Paratype** [40] simplifies the complex phenomenon of antibody–antigen interaction into sets of shared paratope residues. It was shown to help in identifying antibodies from different clones binding to the same target epitope. Two antibodies are said to be *paratype* if they have a sequence identity across the predicted paratope regions greater than a given threshold. Paratope identity is defined as the number of identical paratope residues (residues that are predicted to be in the paratope in both cases) divided by the smallest number of paratope residues of either sequence being compared. The threshold for common reactivity prediction was set to 76% for this metric, as optimized in Supplementary Section 5.
- **AbLigity:** [41] is a structure-based similarity measure tailored to the antibody-antigen interfaces. Using predicted paratopes on modeled antibody structures, it allows for the identification of sequence-dissimilar antibodies that target highly similar epitopes. In short, the Ab-Ligity framework enumerates all sets of 3 residues in the paratope structures, and tokenizes them in terms of both the distance between the residues as well as their intrinsic chemical properties (aliphatic, hydroxyl, sulphur, aromatic, acidic, amine, basic). The obtained set is then compared among other antibodies to obtain a similarity score, and predicted as common epitope if above 0.26, as optimized in Supplementary Section 5.

To find pairs of antibodies that are likely to bind the same epitope, we combine these three metrics. An antibody pair is predicted to bind to a common epitope if it has a similarity above the given threshold for two of the three metrics. Additionally, since a different CDR3 length generally implies a different binding mode, we add the condition that the two antibodies should have the same CDR3 length in order to be predicted as binding the same epitope. Although some cases of antibodies with different CDR3 length expressing common epitope reactivity were observed [41], such a situation is anecdotal. In fact, removing this constrain considerably reduces the accuracy of the three metrics described above, and existing tools are currently unable to accurately identify such cases, mainly due to the difficulty of modeling the CDR3 loop accurately [69].

Comparison To Public V Gene Database. We compared the V genes frequencies with a public database of 6 human donors in Lymph node [23], downloaded from http://immunedb.com/tissue-atlas/. LN samples were retrived form tissues labeled as submandibular lymph nodes (MLN) and inguinal lymph nodes (ILN).

Supplementary Materials

The supplementary files attached to this article contains: Supplementary Section 1 and Figure S1: Mutations analysis across samples. Supplementary Section 2 and Figure S2: Distance to nearest distribution to both negation sequences and sequences within the same sample. Supplementary Section 3 and Figure S3: VJ gene usage combination heatmap. Supplementary Section 4 and Figure S4: Common features of the most abundant and shared clones. Supplementary Section 5 and Figure S5: Threshold optimization for the Paratype, CDRsim and AbLigity similarity frameworks. Supplementary Section 6 and Figure S6: Distribution of paratope distances for all dominant clones pairs in the same GC and in different GCs. Supplementary Section 7 and Figure S7 and S8: Epitope convergence model and epitope diversity in germinal centers. Figure S9: Extensive Experimental Methods. Figure S10: Samples diversity analysis with CDR3. Figure S11: V gene frequency in functional and nonfunctional clones. Figure S12: Phylogeny reconstruction of some representative F&NF clones. Figure S13: Phylogeny reconstruction of some representative shared clones across GCs. Supplementary Table S1: Crippling mutations and stereotyped V genes (excel table).

Data Availability

Processed Sequencing data were deposited on the VDJ server under UUID 8899006209436478995-242ac118-0001-012, publicly accessible at https://vdjserver.org/community/8899006209436478995-242ac118-0001-012.

Competing Interests

There is NO Competing Interest.

Author Contributions Statement

C.N. and J.G. conceived the experiments, M.S conducted the experiments. E.D and M.S performed the samples reads preprocessing. M.S also analyzed the data via the TRIP-tool, under the supervision of J.G. A.P. wrote the code and analyzed the experimental data under the supervision of M.R.M. A.P. and M.S. wrote the manuscript. All authors have read and agreed to the published version of the manuscript.

Acknowledgments And Funding

The authors thank Kostas Stamatopoulos lab for helping with the TRIP tool and for the fruitful discussions. We also thank Thera A M Wormhoudt for assisting in the experimental method development. We thank Nike Claessen for helping with the tissue preparation. This research was supported by the COSMIC European Training Network, funded from the European Union's Horizon 2020 research and innovation program under grant agreement No 765158.

Bibliography

[1] Luka Mesin, Jonatan Ersching, and Gabriel D Victora. "Germinal center B cell dynamics". In: *Immunity* 45.3 (2016), pp. 471–482.

[2] Nilushi S De Silva and Ulf Klein. "Dynamics of B cells in germinal centres". In: *Nature reviews immunology* 15.3 (2015), pp. 137–148.

[3] Jeroen MJ Tas et al. "Visualizing antibody affinity maturation in germinal centers". In: *Science* 351.6277 (2016), pp. 1048–1054.

[4] R Küppers et al. "Tracing B cell development in human germinal centres by molecular analysis of single cells picked from histological sections." In: *The EMBO journal* 12.13 (1993), pp. 4955–4967.

[5] Tanja A Schwickert et al. "A dynamic T cell–limited checkpoint regulates affinity-dependent B cell entry into the germinal center". In: *Journal of Experimental Medicine* 208.6 (2011), pp. 1243–1252.

[6] Carola G Vinuesa, Iñaki Sanz, and Matthew C Cook. "Dysregulation of germinal centres in autoimmune disease". In: *Nature Reviews Immunology* 9.12 (2009), pp. 845–857.

[7] Laura Pasqualucci. "Molecular pathogenesis of germinal center-derived B cell lymphomas". In: *Immunological reviews* 288.1 (2019), pp. 240–261.

[8] Daniel J Firl et al. "Capturing change in clonal composition amongst single mouse germinal centers". In: *Elife* 7 (2018), e33051.

[9] Aurélien Pélissier et al. "Computational Model Reveals a Stochastic Mechanism behind Germinal Center Clonal Bursts". In: *Cells* 9.6 (2020), p. 1448.

[10] Susumu Tonegawa. "Somatic generation of antibody diversity". In: *Nature* 302.5909 (1983), pp. 575–581.

[11] Javier M Di Noia and Michael S Neuberger. "Molecular mechanisms of antibody somatic hypermutation". In: *Annu. Rev. Biochem.* 76 (2007), pp. 1–22.

[12] Mahnoush Bahjat and Jeroen EJ Guikema. "The complex interplay between DNA injury and repair in enzymatically induced mutagenesis and DNA damage in B lymphocytes". In: *International journal of molecular sciences* 18.9 (2017), p. 1876.

[13] Maria Stratigopoulou, Tijmen P van Dam, and Jeroen EJ Guikema. "Base excision repair in the immune system: small DNA lesions with big consequences". In: *Frontiers in immunology* 11 (2020), p. 1084.

[14] Isabelle Stewart et al. "Germinal center B cells replace their antigen receptors in dark zones and fail light zone entry when immunoglobulin gene mutations are damaging". In: *Immunity* 49.3 (2018), pp. 477–489.

[15] Huan et al. "BCR Selection and Affinity Maturation in Payer's Patches Germinal Centers". In: *Nature* 528.7812 (2020), pp. 421–425.

[16] Tanja A Schwickert et al. "In vivo imaging of germinal centres reveals a dynamic open structure". In: *Nature* 446.7131 (2007), pp. 83–87.

[17] Tanja A Schwickert et al. "Germinal center reutilization by newly activated B cells". In: *Journal of Experimental Medicine* 206.13 (2009), pp. 2907–2914.

[18] Ziv Shulman et al. "T follicular helper cell dynamics in germinal centers". In: *Science* 341.6146 (2013), pp. 673–677.

[19] Richard J Bende et al. "Germinal centers in human lymph nodes contain reactivated memory B cells". In: *Journal of Experimental Medicine* 204.11 (2007), pp. 2655–2665.

[20] Luka Mesin et al. "Restricted clonality and limited germinal center reentry characterize memory B cell reactivation by boosting". In: *Cell* 180.1 (2020), pp. 92–106.

[21] Aurelien Pelissier et al. "Quantifying B-cell Clonal Diversity In Repertoire Data". In: *bioRxiv* (2022).

[22] Nima Nouri and Steven H Kleinstein. "Optimized threshold inference for partitioning of clones from high-throughput B cell repertoire sequencing data". In: *Frontiers in immunology* 9 (2018), p. 1687.

[23] Wenzhao Meng et al. "An atlas of B-cell clonal distribution in the human body". In: *Nature biotechnology* 35.9 (2017), pp. 879–884.

[24] Lei Wang et al. "Assessment of variation in B-cell receptor heavy chain repertoire in patients with end-stage renal disease by high-throughput sequencing". In: *Renal failure* 41.1 (2019), pp. 1–13.

[25] Maria Th Kotouza et al. "TRIP-T cell receptor/immunoglobulin profiler". In: *BMC bioinformatics* 21.1 (2020), pp. 1–21.

[26] Daniel Müllner. "Modern hierarchical, agglomerative clustering algorithms". In: *arXiv preprint arXiv:1109.2378* (2011).

[27] Namita T Gupta et al. "Hierarchical clustering can identify B cell clones with high confidence in Ig repertoire sequencing data". In: *The Journal of Immunology* 198.6 (2017), pp. 2489–2499.

[28] E Alamyar et al. "IMGT/HighV-QUEST: A high-throughput system and Web portal for the analysis of rearranged nucleotide sequences of antigen receptors-High-throughput version of IMGT/V-QUEST". In: *V-QUEST 11èmes Journées Ouvertes en Biologie, Informatique et Mathématiques (JOBIM)* (2010), pp. 7–9.

[29] David G Schatz and Yanhong Ji. "Recombination centres and the orchestration of V (D) J recombination". In: *Nature Reviews Immunology* 11.4 (2011), pp. 251–263.

[30] Raul Mostoslavsky, Frederick W Alt, and Klaus Rajewsky. "The lingering enigma of the allelic exclusion mechanism". In: *Cell* 118.5 (2004), pp. 539–544.

[31] Joyce K Hwang, Frederick W Alt, and Leng-Siew Yeap. "Related mechanisms of antibody somatic hypermutation and class switch recombination". In: *Mobile DNA III* (2015), pp. 325–348.

[32] Andreas Agathangelidis, Fotis Psomopoulos, and Kostas Stamatopoulos. "Stereotyped B cell receptor immunoglobulins in B cell lymphomas". In: *Lymphoma* (2019), pp. 139–155.

[33] Katerina Gemenetzi et al. "B cell receptor immunogenetics in B cell lymphomas: immunoglobulin genes as key to ontogeny and clinical decision making". In: *Frontiers in oncology* 10 (2020), p. 67.

[34] Carole J Henry Dunand and Patrick C Wilson. "Restricted, canonical, stereotyped and convergent immunoglobulin responses". In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 370.1676 (2015), p. 20140238.

[35] Tyani D Chan et al. "Elimination of germinal-center-derived self-reactive B cells is governed by the location and concentration of self-antigen". In: *Immunity* 37.5 (2012), pp. 893–904.

[36] Christian T Mayer et al. "The microanatomic segregation of selection by apoptosis in the germinal center". In: *Science* 358.6360 (2017).

[37] Martin F Bachmann et al. "Induction of long-lived germinal centers associated with persisting antigen after viral infection." In: *The Journal of experimental medicine* 183.5 (1996), pp. 2259–2269.

[38] Sudhir Pai Kasturi et al. "Programming the magnitude and persistence of antibody responses with innate immunity". In: *Nature* 470.7335 (2011), pp. 543–547.

[39] William S DeWitt III et al. "Using genotype abundance to improve phylogenetic inference". In: *Molecular biology and evolution* 35.5 (2018), pp. 1253–1265.

[40] Eve Richardson et al. "A computational method for immune repertoire mining that identifies novel binders from different clonotypes, demonstrated by identifying anti-Pertussis toxoid antibodies". In: *Mabs*. Vol. 13. Taylor & Francis. 2021, p. 1869406.

[41] Wing Ki Wong et al. "Ab-Ligity: Identifying sequence-dissimilar antibodies that bind to the same epitope". In: *mAbs*. Vol. 13. Taylor & Francis. 2021, p. 1873478.

[42] Sara D'Angelo et al. "Many routes to an antibody heavy-chain CDR3: necessary, yet insufficient, for specific binding". In: *Frontiers in immunology* 9 (2018), p. 395.

[43] David C Benjamin and Samuel S Perdue. "Site-directed mutagenesis in epitope mapping". In: *Methods* 9.3 (1996), pp. 508–515.

[44] Joyce A Schroer et al. "Mapping epitopes on the insulin molecule using monoclonal antibodies". In: *European journal of immunology* 13.9 (1983), pp. 693–700.

[45] Nicole Wittenbrink et al. "Is there a typical germinal center? A large-scale immunohistological study on the cellular composition of germinal centers during the hapten-carrier–driven primary immune response in mice". In: *The Journal of Immunology* 187.12 (2011), pp. 6185–6196.

[46] Marcel Jan Thomas et al. "A Probabilistic Model of the Germinal Center Reaction". In: *Frontiers in Immunology* 10 (Apr. 2019), p. 689. ISSN: 1664-3224. DOI: 10.3389/fimmu.2019.00689. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6456718/ (visited on 09/21/2021).

[47] Theinmozhi Arulraj et al. "Synchronous germinal center onset impacts the efficiency of antibody responses". In: *Frontiers in immunology* 10 (2019), p. 2116.

[48] Brian Bushnell, Jonathan Rood, and Esther Singer. "BBMerge–accurate paired shotgun read merging via overlap". In: *PloS one* 12.10 (2017), e0185056.

[49] Jiajie Zhang et al. "PEAR: a fast and accurate Illumina Paired-End reAd mergeR". In: *Bioinformatics* 30.5 (2014), pp. 614–620.

[50] Marcel Martin. "Cutadapt removes adapter sequences from high-throughput sequencing reads". In: *EMBnet. journal* 17.1 (2011), pp. 10–12.

[51] Cinque Soto et al. "High frequency of shared clonotypes in human B cell receptor repertoires". In: *Nature* 566.7744 (2019), pp. 398–402.

[52] Matthew IJ Raybould et al. "Public Baseline and shared response structures support the theory of antibody repertoire functional commonality". In: *PLoS computational biology* 17.3 (2021), e1008781.

[53] Ofir Lindenbaum et al. "Alignment free identification of clones in B cell receptor repertoires". In: *Nucleic acids research* 49.4 (2021), e21–e21.

[54] Alonso Gragera and Vorapong Suppakitpaisarn. "Semimetric properties of sørensen-dice and tversky indexes". In: *International Workshop on Algorithms and Computation*. Springer. 2016, pp. 339–350.

[55] Mark O Hill. "Diversity and evenness: a unifying notation and its consequences". In: *Ecology* 54.2 (1973), pp. 427–432.

[56] Anne Chao. "Nonparametric estimation of the number of classes in a population". In: *Scandinavian Journal of statistics* (1984), pp. 265–270.

[57] Nicholas J Gotelli and Robert K Colwell. "Estimating species richness". In: *Biological diversity: frontiers in measurement and assessment* 12 (2011), pp. 39–54.

[58] Lou Jost. "Entropy and diversity". In: *Oikos* 113.2 (2006), pp. 363–375.

[59] Anne Chao and Tsung-Jen Shen. "Nonparametric estimation of Shannon's index of diversity when there are unseen species in sample". In: *Environmental and ecological statistics* 10.4 (2003), pp. 429–443.

[60] Julie D Thompson, Desmond G Higgins, and Toby J Gibson. "CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice". In: *Nucleic acids research* 22.22 (1994), pp. 4673–4680.

[61] Katharina Jahn, Jack Kuipers, and Niko Beerenwinkel. "Tree inference for single-cell data". In: *Genome biology* 17.1 (2016), p. 86.

[62] Vladimir I Levenshtein. "Binary codes capable of correcting deletions, insertions, and reversals". In: *Soviet physics doklady*. Vol. 10. Soviet Union. 1966, pp. 707–710.

[63] Li Yujian and Liu Bo. "A normalized Levenshtein distance metric". In: *IEEE transactions on pattern analysis and machine intelligence* 29 (2007), pp. 1091–1095.

[64] Edgar Liberis et al. "Parapred: antibody paratope prediction using convolutional and recurrent neural networks". In: *Bioinformatics* 34.17 (2018), pp. 2944–2950.

[65] Dimitri Schritt et al. "Repertoire Builder: high-throughput structural modeling of B and T cell receptors". In: *Molecular Systems Design & Engineering* 4.4 (2019), pp. 761–768.

[66] Schrödinger, LLC. "The PyMOL Molecular Graphics System, Version 1.8". Nov. 2015.

[67] Johannes F Scheid et al. "Sequence and structural convergence of broad and potent HIV antibodies that mimic CD4 binding". In: *Science* 333.6049 (2011), pp. 1633–1637.

[68] Rahmad Akbar et al. "A compact vocabulary of paratope-epitope interactions enables predictability of antibody-antigen binding". In: *Cell Reports* 34.11 (2021), p. 108856.

[69] C Marks and CM Deane. "Antibody H3 structure prediction". In: *Computational and structural biotechnology journal* 15 (2017), pp. 222–231.