Memory persistence and differentiation into antibody-secreting cells accompanied by positive selection in longitudinal BCR repertoires

Artem I. Mikelov[1,2,3†], Evgeniia I. Alekseeva[1†], Ekaterina A. Komech[2,3], Dmitriy B. Staroverov[2], Maria A. Turchaninova[2], Mikhail Shugay[2,3], Dmitriy M. Chudakov[1,2,3], Georgii A. Bazykin[1,4], Ivan V. Zvyagin[2,3*]

[1] Skolkovo Institute of Science and Technology, Moscow, Russia;
[2] Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, Moscow, Russia;
[3] Institute of Translational Medicine, Pirogov Russian National Research Medical University, Moscow, Russia;
[4] A.A. Kharkevich Institute for Information Transmission Problems of the Russian Academy of Sciences, Moscow, Russia.

*Corresponding author: izvyagin@gmail.com (I.V.Z.)

[†]These authors contributed equally to this work.

Key words: memory B cells, plasmablasts, plasma cells, BCR repertoire, somatic hypermutation, B cell somatic evolution, natural selection, affinity maturation

Abstract

The stability and plasticity of B cell-mediated immune memory ensures the ability to respond to the repeated challenges. We have analyzed the longitudinal dynamics of immunoglobulin heavy chain repertoires from memory B cells, plasmablasts, and plasma cells from the peripheral blood of generally healthy volunteers. We reveal a high degree of clonal persistence in individual memory B cell subsets, with inter-individual convergence in memory and antibody-secreting cells (ASCs). ASC clonotypes demonstrate clonal relatedness to memory B cells, and are transient in peripheral blood. We identify two clusters of expanded clonal lineages with differing prevalence of memory B cells, isotypes, and persistence. Phylogenetic analysis revealed signs of reactivation of persisting memory B cell-enriched clonal lineages, accompanied by new rounds of affinity maturation during proliferation and differentiation into ASCs. Negative selection contributes to both persisting and reactivated lineages, preserving the functionality and specificity of BCRs to protect against current and future pathogens.

Introduction

B cells play a crucial role in protection from various pathogens and cancer cells as well as regulation of the immune response. The structural diversity of B cell receptors (BCRs) is responsible for the B cell-mediated immune system's capacity to recognize a wide variety of different antigens, and every individual harbors a large pool of naive B cell clones, each with a unique BCR. Antigenic challenge triggers the proliferation and maturation of naive B cells with cognate BCRs, and the resulting progeny comprise a number of cell subsets with differing functions and lifespans. During the affinity maturation process, the initial structure of a given BCR can change at the genomic level as a result of somatic hypermutation (SHM), a process that accompanies B cell proliferation after antigen-specific activation. Cells bearing BCRs with higher affinity to the antigen are favored during the affinity maturation process, and produce signals that stimulate further differentiation and expansion (De Silva and Klein 2015). Another process called class-switch recombination further increases the dimensionality of the BCR space. The five main classes, or isotypes, of antibodies (*i.e.,* IgA, IgD, IgE, IgG, and IgM) have different functions in the immune response (Stavnezer, Guikema, and Schrader 2008; Vidarsson, Dekkers, and Rispens 2014), and isotype switching during clonal proliferation can thereby change the functional capabilities of B cells and the antibodies they produce. As a consequence, antigen challenge yields a population of clonally related cells with different BCRs and functionalities.

Recently developed immune repertoire sequencing techniques provide valuable insights into the development and structure of B-cell immunity with clonal-level resolution (Chaudhary and Wesemann 2018). For example, the clonal relatedness of B cells in a given lineage, as well as the number and dynamics of B cell groups with distinct antigen specificities, can be studied based on BCR sequence homology. Numerous studies have generated valuable data by analyzing repertoire characteristics such as clonal diversity and tissue distribution, magnitude of clonal expansion and BCR SHM, V(D)J usage frequency and distribution of CDR3 length, and

the degree of repertoire convergence and individuality (Briney et al. 2019; Soto et al. 2019; Shah et al. 2019; Mandric et al. 2020; Yang et al. 2021). Studies of BCR repertoires of patients with different diseases have made an important contribution to the understanding of mechanisms of pathology and B-cell-mediated immunity (Bashford-Rogers et al. 2019; S. C. A. Nielsen et al. 2020; Gaebler et al. 2021; Sakharkar et al. 2021).
Longitudinal analysis of repertoires at different timepoints has made it possible to study the dynamics of B cell response following antigenic challenge or therapy (Laserson et al. 2014; Davydov et al. 2018; Horns et al. 2019; Nourmohammad et al. 2019; Hoehn et al. 2021). Reconstruction of BCR evolution in B cell clonal lineages and phylogenetic analysis can reveal which evolutionary forces predominate at different stages of clonal lineage development. De Bourcy *et al.* recently reported on age-related differences in the structure of clonal lineages, somatic hypermutagenesis and affinity maturation processes, and differences in recall response of persisting lineages upon vaccination depending on CMV seropositivity status (de Bourcy et al. 2017). Other studies have described in detail the action of positive selection in the evolution of clonal lineages in vaccination and chronic HIV infection (Bonsignori et al. 2017; Horns et al. 2019; Nourmohammad et al. 2019). Reports have also described persisting clonal lineages which are predominantly represented by cells with IgM/IgD isotypes, and which demonstrate signs of neutral evolution (Horns et al. 2019). Comparison of BCR repertoires between different cell subsets also makes it possible to investigate factors governing the functional assignment of B cells during proliferation, and thereby to understand fundamental aspects of B cell immunity. For example, recent studies have described differences in BCR repertoires of IgM and switched memory B cells as well as the complex interplay between $CD27^{high}$ and $CD27^{low}$ B-cell memory subsets, showing the complex nature of B cell immune memory (Wu et al. 2010; Grimsholm et al. 2020).

BCR repertoires of antigen-experienced B cell subsets and their dynamics are usually studied in the context of pathologic conditions and vaccination, and there is a lack of equivalent data in the absence of acute or chronic immune response. We have therefore investigated immunoglobulin heavy chain repertoires from memory B cells, plasmablasts, and plasma cells from peripheral blood collected from generally healthy volunteers at three time-points over the course of a year. In order to obtain detailed and unbiased repertoire data, we used advanced IgH repertoire profiling technology that provides full-length IgH sequences with isotype annotation. Based on comparative and phylogenetic analysis of the resulting data, we are able to describe the structure, distinctive features, clonal relations, isotype distribution and temporal dynamics of B cell subset repertoires, as well as the phylogenetic history of large clonal lineages.

Results

IGH repertoire sequencing statistics and analysis depth

We collected peripheral blood from six healthy donors at three time-points, where the second sample was collected one month after the first, and the third was collected 11 months after that (**Fig. 1A; Supplementary Table S1**). These samples were subjected to fluorescence-activated cell sorting to isolate memory B cells (Bmem: $CD19^{+}$ $CD20^{+}$ $CD27^{+}$), plasmablasts (PBL: $CD19^{low/+}$ $CD20^{-}$ $CD27^{high}$ $CD138^{-}$) and plasma cells (PL: $CD19^{low/+}$ $CD20^{-}$ $CD27^{high}$ $CD138^{+}$)

(**Supplementary Fig. S1A**). Most of the cell samples were collected and processed in two independent replicates (**Supplementary Data SD1**). For each cell sample, we obtained full-length IGH clonal repertoires were obtained using a 5'-RACE-based protocol, which allows unbiased amplification of full-length IGH variable domain cDNA while preserving isotype information, with subsequent unique molecular identifier (UMI)-based sequencing data normalization and error correction (Turchaninova et al. 2016; Shugay et al. 2014). From a total of 83 cell samples, we obtained 1.06 x 10^7 unique IGH cDNA molecules, each covered by at least three sequencing reads, representing 8.4 x 10^5 unique IGH clonotypes (**Supplementary Data SD1**). An IGH clonotype was defined as a unique nucleotide sequence spanning from the beginning of IGH V gene framework 1 to the 5' end of the C segment, sufficient to determine isotype. The number of unique clonotypes (*i.e.,* species richness) depended on the number of cells per sample (**Supplementary Fig. S1B**), even after data normalization by sampling an equal number of unique IGH cDNA sequences. To characterize the number of distinct IGH clonotypes in each cell subset, we selected the samples with the most common number of sorted cells for each sample set. The median number of clonotypes was 20,072 (14,572–32,806, n = 14) per 5 x 10^4 memory B cells, 628 (528–981, n = 8) per 1 x 10^3 plasmablasts, and 800 (623–1,183, n = 9) per 1 x 10^3 plasma cells. Rarefaction analysis in the Bmem subpopulation (**Supplementary Fig. S1B,** left) revealed an asymptotic increase of species richness that did not reach a plateau, indicating that the averaged species richness can only serve as a lower limit of sample diversity estimation. For all samples of PBL and PL subpopulations, species richness curves plateaued, meaning that we had reached sufficient sequencing depth to evaluate the clonal diversity of the sorted cell samples (**Supplementary Fig. S1B** center and right).

B cell subsets display both divergent and similar characteristics in their IGH repertoires

First, we aimed to characterize features of the IGH repertoires of the Bmem, PBL, and PL subset based on several key properties: usage of germline encoded IGHV segments, clonal distribution by isotypes, rate of SHM in CDR1-2 and FWR1-3, and features of the hypervariable CDR3 region. The proportion of overall clonal diversity occupied by the five major IGH isotypes was strikingly different between Bmem cells and antibody-secreting cells (ASCs; *i.e.,* PBL and PL). IgM represented more than half of the repertoire in Bmem, while IgA was dominant in PBL and PL (**Fig. 1B**, **Supplementary Data SD2**). The second most prevalent isotype in ASCs was IgG, which was also less abundant in Bmem compared to IgA. IgD represented a substantial part of the Bmem clonal repertoire, while < 1% clonotypes of ASCs expressed IgD. The proportion of each isotype varied between donors and time points, but IgM and IgA or IgA and IgG consistently remained the most abundant isotypes in Bmem cells or ASCs, respectively (**Supplementary Fig. S2A, Supplementary Data SD2**). In all studied subsets, the isotype distribution in terms of number of unique clonotypes roughly mirrored the isotype distribution based on the number of IGH cDNA molecules, indicating absence of large clonal expansions or differences in IGH expression level distorting abundance of isotypes. This could not be determined by sequencing of bulk PBMCs, as higher levels of IGH expression by ASCs can change the isotype proportions and thereby bias the quantitation of clonotype abundance (**Supplementary Fig. S2B**). The obtained IGH isotype distributions based on unique clonotypes

140 roughly correspond to the distribution of IGH isotypes typically detected by flow cytometry of the
141 same subsets (Perez-Andres et al. 2010).
142 The level of SHM was on average significantly higher in ASC subsets, reflecting that PBLs and
143 PLs are enriched for clones that have undergone affinity maturation (**Fig. 1C**). The average
144 number of SHMs for IgE clonotypes did not differ significantly between cell subsets, but was
145 significantly higher compared to the level of SHM detected for IgM and IgD clonotypes in Bmem.
146 Of note, the rate of SHM in PBLs was higher than that in PLs in clonotypes from the three most
147 abundant isotypes (*i.e.,* IgM, IgA and IgG). We also compared the distributions of the lengths of
148 the hypervariable CDR3 region between IGH clonotypes in different cell subsets. PBLs had
149 significantly longer CDR3 regions compared to Bmem cells on average in every isotype except
150 for IgE (**Fig. 1D**). Of note, the average CDR3 length in PL clonotypes was significantly higher
151 compared to Bmem for IgA and IgD, but not for the other isotypes.

152 IGHV gene segment usage was roughly similar between Bmem, PBL, and PL cells from all
153 donors, indicating generally equal probabilities of memory-to-ASC conversion for B cells
154 carrying BCRs encoded by distinct gene segments (**Fig. 1E**, **Supplementary Fig. S3A**). This
155 distribution differed significantly between the studied cell subsets and naive B cells (based on
156 data from Gidoni et al. 2019). The repertoire of total B cells (Btot; $CD19^+$ $CD20^+$), which
157 contained a large fraction of naive B cells, demonstrated similar IGH V gene segment usage to
158 the naive B-cell repertoire (**Supplementary Fig. S3A**). We observed statistically significant
159 Pearson correlations in terms of IGHV gene frequencies for all pairs between Bmem, PBL or PL
160 (> 0.95 correlation, $p < 0.01$), and for naive vs Btot (0.79 correlation, $p < 0 .01$). We observed
161 high concordance in terms of under- or overrepresentation of specific IGHV gene segments in
162 repertoires of all antigen-experienced B-cell subsets compared to naive B-cells; Pearson
163 correlation coefficients for the fold-change of IGHV gene segment usage frequencies were 0.95
164 for Bmem and PBL, 0.96 for Bmem and PL, and 0.98 for PBL and PL ($p < 0.01$ for all pairs).
165 Moreover, IGHV gene segment under- or overrepresentation clearly depended on the given
166 gene sequence. We clustered IGHV genes based on their sequence similarity, and observed
167 that most IGHV segments in each of the four major clusters behaved concordantly with other
168 segments in that cluster (**Fig. 1E**). This effect was also observed at the level of individual
169 repertoires (**Supplementary Fig. S3B**) with discrepancies that could probably be attributed to
170 genetic polymorphism of the IGH loci of particular donors.

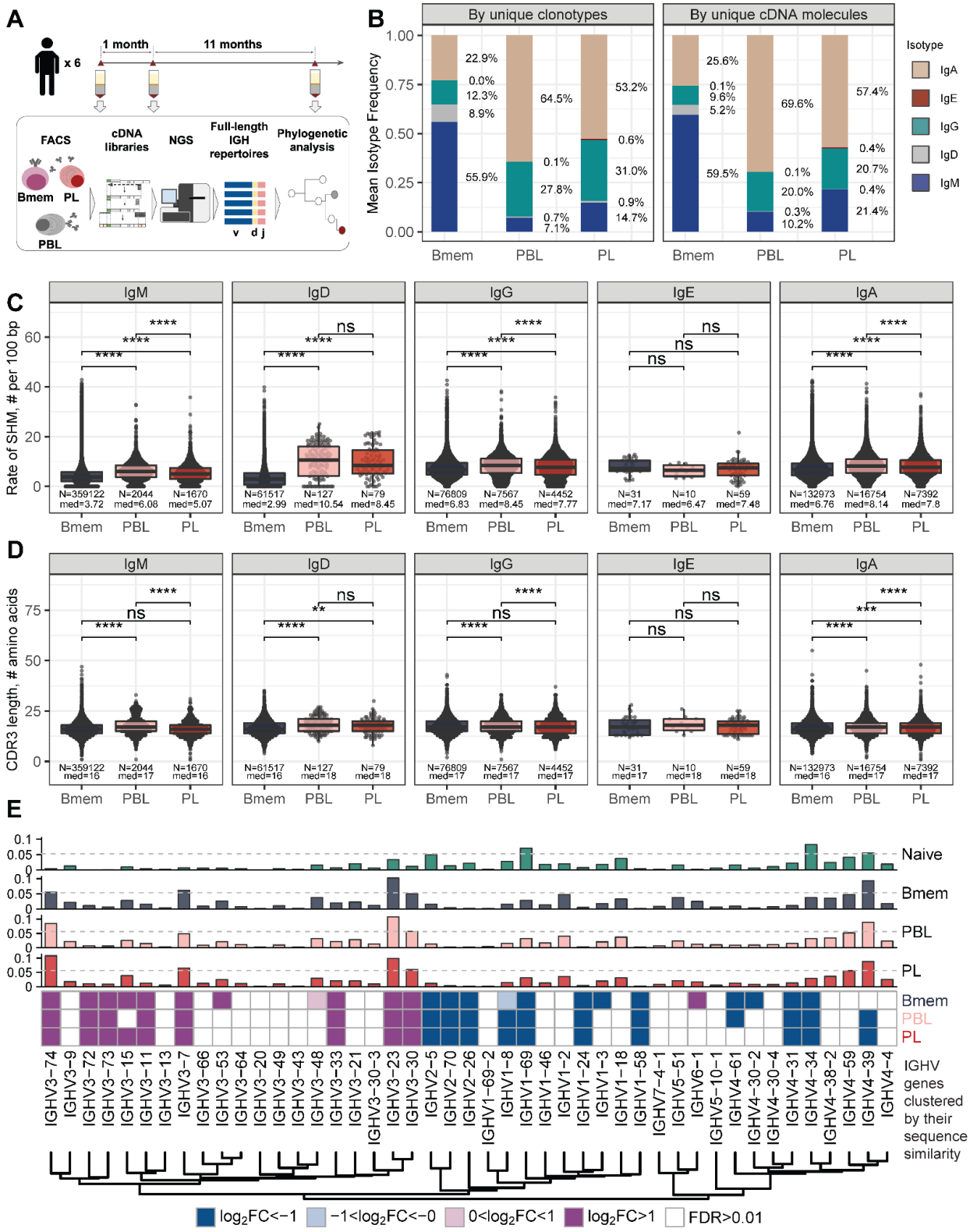
A
x 6
1 month
11 months
FACS
cDNA libraries
NGS
Full-length IGH repertoires
Phylogenetic analysis
Bmem
PL
PBL
v
d
j
B
By unique clonotypes
By unique cDNA molecules
Mean Isotype Frequency
Isotype
IgA
IgE
IgG
IgD
IgM
Bmem
PBL
PL
C
Rate of SHM, # per 100 bp
IgM
IgD
IgG
IgE
IgA
N=359122 med=3.72
N=2044 med=6.08
N=1670 med=5.07
N=61517 med=2.99
N=127 med=10.54
N=79 med=8.45
N=76809 med=6.83
N=7567 med=8.45
N=4452 med=7.77
N=31 med=7.17
N=10 med=6.47
N=59 med=7.48
N=132973 med=6.76
N=16754 med=8.14
N=7392 med=7.8
D
CDR3 length, # amino acids
N=359122 med=16
N=2044 med=17
N=1670 med=16
N=61517 med=16
N=127 med=18
N=79 med=18
N=76809 med=17
N=7567 med=17
N=4452 med=17
N=31 med=17
N=10 med=18
N=59 med=18
N=132973 med=16
N=16754 med=17
N=7392 med=17
E
Naive
Bmem
PBL
PL
IGHV genes clustered by their sequence similarity
log2FC<−1
−1<log2FC<−0
0<log2FC<1
log2FC>1
FDR>0.01

Figure 1. General characteristics of IGH repertoires in differentiated B cell lineage subsets. A: Study design. Peripheral blood from six donors was sampled at three time points: T1 - initial time point, T2 - 1 month and T3 - 12 months after the start of the study. At each time point, we isolated PBMCs and sorted memory B cells (Bmem: $CD19^{+}$ $CD20^{+}$ $CD27^{+}$), plasmablasts (PBL: $CD19^{low/+}$ $CD20^{-}$ $CD27^{high}$ $CD138^{-}$) and plasma cells (PL: $CD19^{low/+}$ $CD20^{-}$ $CD27^{high}$ $CD138^{+}$) in two replicates using FACS. For each cell sample, we obtained clonal repertoires of full-length IGH by sequencing respective cDNA libraries. **B**: Proportion of isotypes in studied cell subsets averaged across all obtained repertoires. Left, frequency of unique IGH clonotypes with each particular isotype. Right, frequency of each isotype based on IGH cDNA molecules detected in a sample. **C**: Distribution of the number of somatic hypermutations identified per 100 bp length of IGHV segment for clonotypes within each particular isotype. **D**: Distribution of CDR3 length of clonotypes in each cell subset by isotype; **E**: Distributions of average IGHV gene frequencies based on number of clonotypes in naive B cells (data from Gidoni et al. 2019), Bmem, PBL, and PL repertoires are shown at the top. Colored squares on heatmap indicate significantly different (false discovery rate, FDR < 0.01) frequencies for IGHV gene segments in corresponding B cell subsets compared to naive B cell repertoires. Color intensity reflects the magnitude of the difference (FC = fold change). Only V genes represented by more than two clonotypes on average are shown. IGHV gene segments are clustered based on the similarity of their amino acid sequence, as indicated by the dendrogram at the bottom. In **C** and **D**, the numbers at the bottom of the plots represent the number of clonotypes in the corresponding group, pooled from all donors, and the median measurements from each cell type. Comparisons between subsets were performed with two-sided Mann-Whitney U test. * = $p \leq 0.05$, ** = $p \leq 0.01$, *** = $p \leq 10^{-3}$, **** = $p \leq 10^{-4}$.

Memory B cell repertoires are stable over time and contain a large number of public clonotypes

We further studied the similarity of IGH clonal repertoires of B cell subsets across time-points and between individuals, evaluating repertoire stability (*i.e.,* distance between different time-points) and degree of individuality (*i.e.,* distance between repertoires from different donors). We evaluated repertoire similarity at two levels of IGH sequence identity: frequency of clonotypes with identical nucleotide sequence-defined variable regions (FR1–4), and number of clonotypes with identical CDR3 amino acid sequences, IGHV gene segments, and isotypes. Both metrics showed significantly higher inter-individual differences compared to the divergence of repertoires derived from the same donor, reflecting the fact that IGH repertoires of Bmem, PBL, and PL subsets are private to a large degree (**Fig. 2A,B**). We observed identical clonotypes in the repertoires of PBL and PL collected at different time-points, whereas the repertoire similarity was much lower compared to that between replicate samples, reflecting the transient nature of PBL and PL populations in peripheral blood. Notably, we observed lower clonal overlap in PBL and PL for more distant time-points (separated by 11 or 12 months) than those that are closer together (1 month) (**Supplementary Fig. S4A**). The dissimilarity between samples collected on the same day versus 1 month or even 1 year later was much lower for Bmem, demonstrating the high stability of the clonal repertoire and long-term persistence of IGH clonotypes in these cells (**Fig. 2B**, **Supplementary Fig. S4A**).

To better describe the inter-individual IGH repertoire convergence, we analyzed the number of IGH amino acid clonotypes shared between different donors (*i.e.,* public clonotypes) among 5000 most expanded clonotypes in each Bmem repertoire, assuming that functional convergence could be detected amongst the most abundant clonotypes due to clonal expansions in response to common pathogens. Indeed, the average number of shared clonotypes in Bmem was significantly higher between fractions of the most abundant clonotypes

217 compared to randomly-sampled clonotypes (**Fig. 2C**), as well as when compared to the most
218 abundant clonotypes shared by two naive repertoires (from Gidoni et al. 2019) or to pre-immune
219 IGH repertoires obtained by *in silico* generation using OLGA software (Sethna et al. 2019) (**Fig.**
220 **2C**). We also noted that there were no shared clonotypes defined by their full-length nucleotide
221 sequence (**Supplementary Fig. S4B**). Public clonotypes were also hypermutated, although the
222 rate of SHM was slightly lower compared to that in clonotypes specific to one donor (private)
223 (**Supplementary Fig. S4C**). These observations indicate functional convergence in Bmem
224 repertoires, which is presumably driven by exposure to common pathogens. Of note, the extent
225 of clonal overlap was significantly higher between naive repertoires than for *in silico*-generated
226 repertoires, indicating functional convergence even in pre-immune repertoires. Furthermore, the
227 distance between V segment usage distributions in Bmem repertoires was not significantly
228 different compared to that in naive B cells repertoires. That indicates that the higher clonotype
229 sharing seen in Bmem cells cannot be attributed to lower diversity in IGHV germline usage (**Fig.**
230 **2D**). The same analysis in PBL and PL subpopulations for the 600 and 200 most abundant
231 clonotypes respectively yielded no shared clonotypes between repertoires of different donors,
232 demonstrating no detectable convergence at this sampling depth. Finally, we found that public
233 clonotypes were more likely to be detected than private ones in samples collected at different
234 time-points (**Fig. 2E**), again suggesting persistent memory to common antigens.

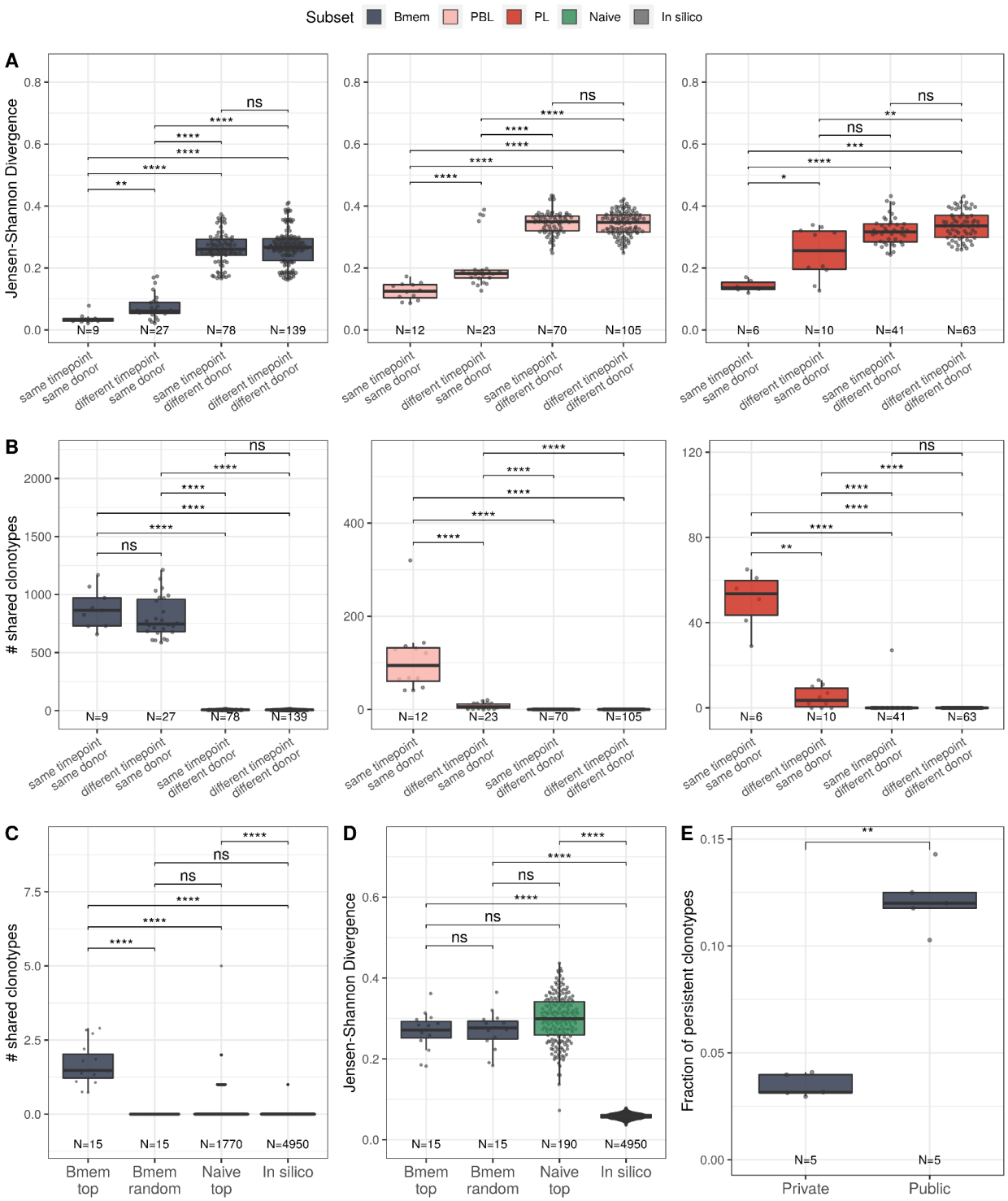
Subset
Bmem
PBL
PL
Naive
In silico
A
Jensen-Shannon Divergence
N=9
N=27
N=78
N=139
N=12
N=23
N=70
N=105
N=6
N=10
N=41
N=63
same timepoint same donor
different timepoint same donor
same timepoint different donor
different timepoint different donor
B
shared clonotypes
C
Bmem top
Bmem random
Naive top
In silico
N=15
N=15
N=1770
N=4950
D
N=15
N=15
N=190
N=4950
E
Fraction of persistent clonotypes
N=5
N=5
Private
Public

Figure 2. Bmem, PBL, and PL IGH repertoire stability over time and similarity between individuals. **A**: Distance between repertoires obtained at different time-points from the same or different donors as calculated by Jensen-Shannon divergence index for IGHV gene frequency distribution. **B**: Number of shared clonotypes between pairs of repertoires from the same or different donors and time-points. For data normalization, we assessed the most abundant 14,000 Bmem, 600 PBL, and 300 PL clonotypes. **C:** The average number of shared clonotypes between repertoires from pairs of unrelated donors for the 5,000 most abundant Bmem clonotypes, randomly-selected Bmem clonotypes, most abundant clonotypes from naive repertoires of unrelated donors (from Gidoni et al. 2019), or from synthetic repertoires generated with OLGA software. We calculated the number of shared clonotypes between these repertoires, each of which contained 5,000 clonotypes. **D**: Inter-individual distance between distributions of V genes in repertoires, calculated as Jensen-Shannon divergence indices for the pairs of repertoires depicted in **C**. **E**: Fraction of persistent clonotypes detected at more than one time-point among clonotypes detected in repertoires from only one donor (private) or in at least two donors (public). Each dot represents the fraction of persistent clonotypes from one donor. In all plots, clonotypes are defined as having identical CDR3 amino acid sequences and the same IGHV gene segment and isotype. For **A**–**D**, each dot represents a pair of repertoires of the corresponding type; N indicates the number of pairs of repertoires in the group. Comparisons in all panels were performed with two-sided Mann-Whitney U test. * = $p \leq 0.05$, ** = $p \leq 0.01$, *** = $p \leq 10^{-3}$, **** = $p \leq 10^{-4}$.

Temporal dynamics of clonal lineages are associated with cell subset composition

SHM during BCR affinity maturation leads to the formation of clonal lineages — *i.e.*, BCR clonotypes evolved from a single ancestor after B cell activation. To study the structure and dynamics of clonal lineages originating from a single BCR ancestor, we grouped clonotypes from each individual based on their sequence similarity (see Materials and Methods for details). We focused on larger clonal lineages consisting of at least 20 unique clonotypes from the corresponding donor. On average, these clonal lineages covered 3.4% of a given donor's repertoire, and we identified 190 such lineages across the four donors from whom samples were collected at each of the three time-points (**Supplementary Fig. S5A**).

After performing principal component analysis (PCA), we determined that these clonal lineages could be divided into two large clusters, HBmem and LBmem, based on the proportion of cell subsets and BCR isotypes represented (**Fig. 3A**, **Supplementary Fig. S5B**). The more abundant HBmem cluster included 138 clonal lineages, and was mostly composed of memory B-cell clonotypes of non-switched isotype IgM. Conversely, the smaller LBmem cluster (52 clonal lineages) was more diverse and largely composed of ASCs, and enriched in IgG and IgA clonotypes (**Fig. 3B**). The average size of clonal lineages (*i.e.*, the number of unique clonotypes per lineage) did not differ between the two clusters (**Supplementary Fig. S5C**). Clonal lineages belonging to both clusters were observed in repertoires of all donors, and the HBmem cluster was more prevalent in each donor (**Supplementary Fig. S5D**).

To better understand the differences between HBmem and LBmem clonal lineages, we tracked the abundance of each clonal lineage in the repertoire across each time-point. The two clusters of lineages demonstrated different temporal behavior; while HBmem groups were quite stable over time, LBmem lineages had a burst of increased frequency at one of the time points (**Fig. 3C**). To compare the temporal stability of clonal lineages, we defined the lineage persistence metric, which equals 1 when a clonal lineage was equally frequent at all three time-points and is close to 0 when it was detected at just one time-point (**Fig. 3D**). Persistence of a clonal lineage

280 was strongly associated with its composition (**Fig. 3E, F**). Clonal lineages enriched with
281 clonotypes or with the IgM isotype — including all HBmem lineages — were more likely to
282 persist through time. Conversely, lineages with larger proportions of ASCs or IgG/IgA isotypes,
283 including most LBmem lineages, tended to have lower persistence, with a burst of increased
284 frequency at one particular time-point. The persistence of a clonal lineage was not associated
285 with its size, and the frequencies of clonal lineages were highly correlated among replicate
286 samples (**Supplementary Fig. S5E, F**), indicating that differences in persistence cannot be
287 attributed to clonotype sampling noise.

288 Besides their higher persistence, the HBmem lineages were enriched in clonotypes detected at
289 multiple time-points (**Supplementary Fig. S5G**), indicating that persistent clonal lineages are
290 supported by persistent clonotypes. Furthermore, 29.7% of the HBmem cluster was represented
291 by public clonal lineages shared between at least two donors, compared to 3.8% for the LBmem
292 cluster. The only two shared LBmem lineages had atypically high persistence, which made them
293 more similar to HBmem (**Fig. 3G**).

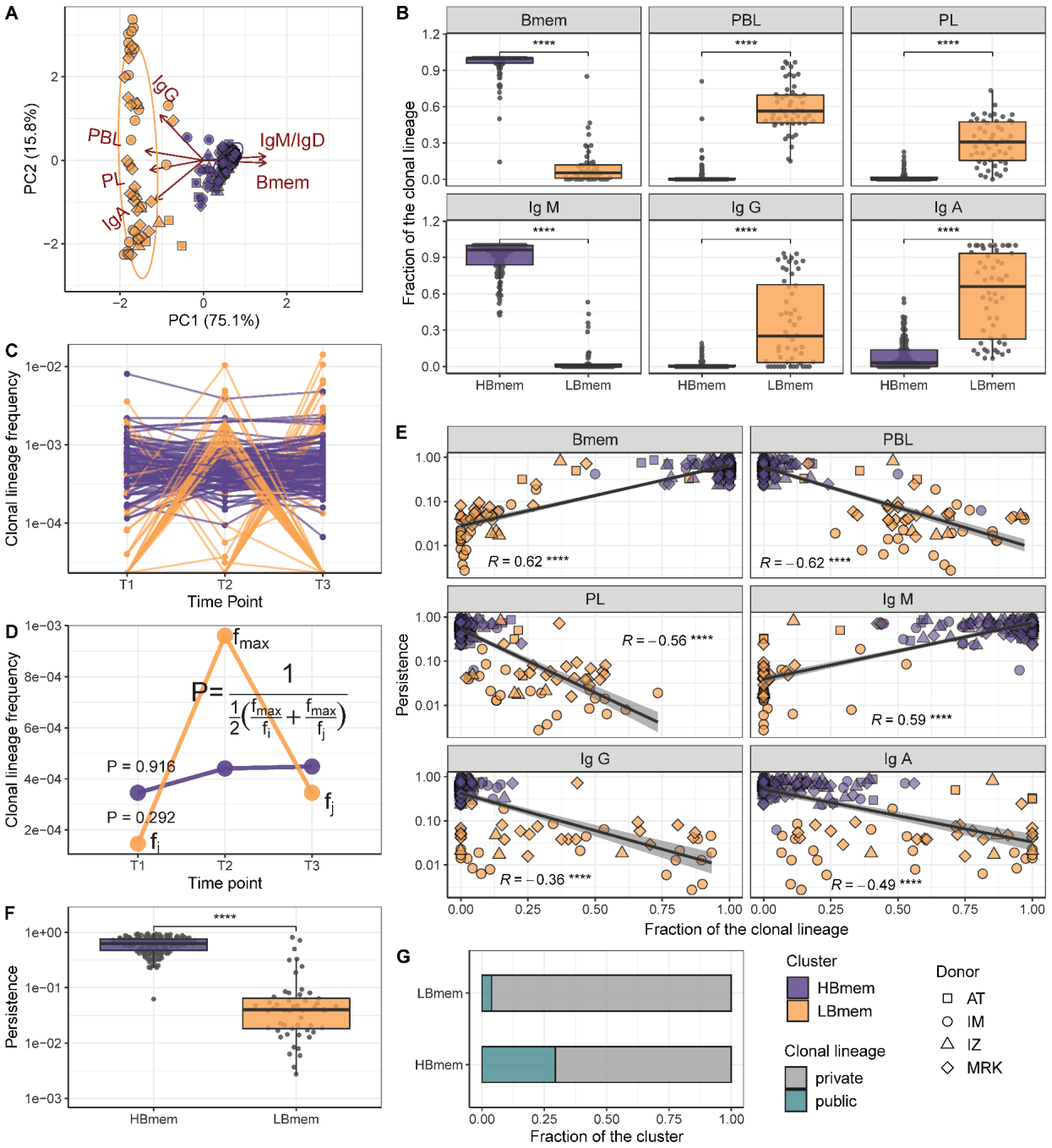
A
PC2 (15.8%)
PC1 (75.1%)
IgG
PBL
PL
IgA
IgM/IgD
Bmem
B
Bmem
PBL
PL
Ig M
Ig G
Ig A
Fraction of the clonal lineage
HBmem
LBmem
C
Clonal lineage frequency
Time Point
T1
T2
T3
D
Clonal lineage frequency
Time point
P = 0.916
P = 0.292
E
Persistence
Fraction of the clonal lineage
R = 0.62 ****
R = −0.62 ****
R = −0.56 ****
R = 0.59 ****
R = −0.36 ****
R = −0.49 ****
F
Persistence
HBmem
LBmem
G
Fraction of the cluster
Cluster
HBmem
LBmem
Clonal lineage
private
public
Donor
AT
IM
IZ
MRK

Figure 3. Temporal dynamics and composition of clonal lineages. A: Principal component analysis (PCA) of clonal lineage composition: proportions of Bmem, PBL, and PL as well as proportions of isotypes. Arrows represent projections of the corresponding variables onto the two-dimensional PCA plane, with lengths reflecting how well the variable explains the variance of the data. The two principal components (PC1 and PC2) cumulatively explain 90.9% of the variance. **B**: Proportion of clonotypes from the various cell subset or isotypes in clonal lineages falling into the HBmem or LBmem clusters; **C**: Dynamics of clonal lineage frequency, defined as the number of clonotypes in a lineage divided by the total number of clonotypes detected at a given time-point. Each line connects points representing a unique clonal lineage (N=190). **D**: Schematic representation of how we calculated clonal lineage persistence. f_{max} is the maximum clonal lineage frequency among the three time-points, and $f_{i,j}$ are the frequencies at the remaining two time-points. **E**: Spearman's correlation between persistence of a clonal lineage and proportions of its clonotypes associated with a given B cell subset or isotype. **F**: Comparison of persistence between HBmem and LBmem. **G:** Fraction of public clonal lineages in the two clusters. Statistical significance for **B**, **F**, and **G** is calculated by the two-sided Mann-Whitney test. * = $p \leq 0.05$, ** = $p \leq 0.01$, *** = $p \leq 10^{-3}$, **** = $p \leq 10^{-4}$.

LBmem clonal lineages could arise from HBmem clonal lineages

The evolutionary past of a clonal lineage can be described by inferring the history of accumulation of SHMs leading to individual clonotypes—*i.e.*, by reconstructing the phylogenetic tree of the clonal lineage. The initial germline sequence of each clonal lineage partially matches the germline VDJ segments, and can be reconstructed in a manner corresponding to the root of the phylogenetic tree of this lineage (see Methods). However, the first node of the phylogenetic tree (green diamond in **Fig. 4A**), the most recent common ancestor (MRCA) of the sampled part of the lineage, can be different from the inferred germline sequence. These differences, referred to as the G-MRCA distance, correspond to SHMs accumulated during the evolution of the clonal lineage prior to divergence of the observed clonotypes. The G-MRCA distance depends on how clonotypes of the tree were sampled. Sampling of clonotypes regardless of their position on the tree results in a low G-MRCA distance (**Fig. 4A**, top panel), while sampling just those clonotypes belonging to a particular clade can conceal early stages of lineage evolution and thus result in a large G-MRCA distance (**Fig. 4A**, bottom panel).
The G-MRCA distance was on average five-fold higher in LBmem clonal lineages compared to HBmem (median = 0.044 vs. 0.008, **Fig. 4B**). This means that even though nearly all the evolution of an HBmem clonal lineage leaves a trace in the observed diversity of that lineage (**Fig. 4D,G**), the sequence variants of an LBmem lineage typically result from divergence of an already-hypermutated clonotype (**Fig. 4E,H**). In most (38 out of 52) LBmem lineages, some Bmem clonotypes were observed at the time-point preceding expansion. Moreover, clonotypes of LBmem lineages are typically characterized by lower pairwise divergence compared to that in HBmem lineages (median = 0.11 vs 0.13, **Fig. 4C**). Together with the burst-like dynamics characteristic of LBmem lineages (**Fig. 3F**), this implies that LBmem lineages may represent recent, rapid clonal expansion of preexisting memory.

Based on these results and the compositional features of the two clusters, we further hypothesized that LBmem clonal lineages may arise from reactivation of pre-existing memory cells belonging to the HBmem cluster. In search of examples of such a transition, we examined all clonal lineages that were persistent but included ASC clonotypes. We found one clear example of a transition from HBmem to LBmem state in the evolutionary history of a clonal lineage (**Fig. 4F, I**). While the MRCA of this lineage nearly matched the germline sequence, all

339 ASC clonotypes were grouped in a single monophyletic clade (sublineage), such that its
340 ancestral node was remote from the MRCA. The ASC sublineage demonstrated all features
341 characteristic of LBmem, including predominance of IgG and IgA isotypes, low persistence, and
342 low clonotype divergence. Conversely, the remainder of the clonal lineage had features of
343 HBmem: predominance of IgM, high persistence, and high levels of clonotype divergence.

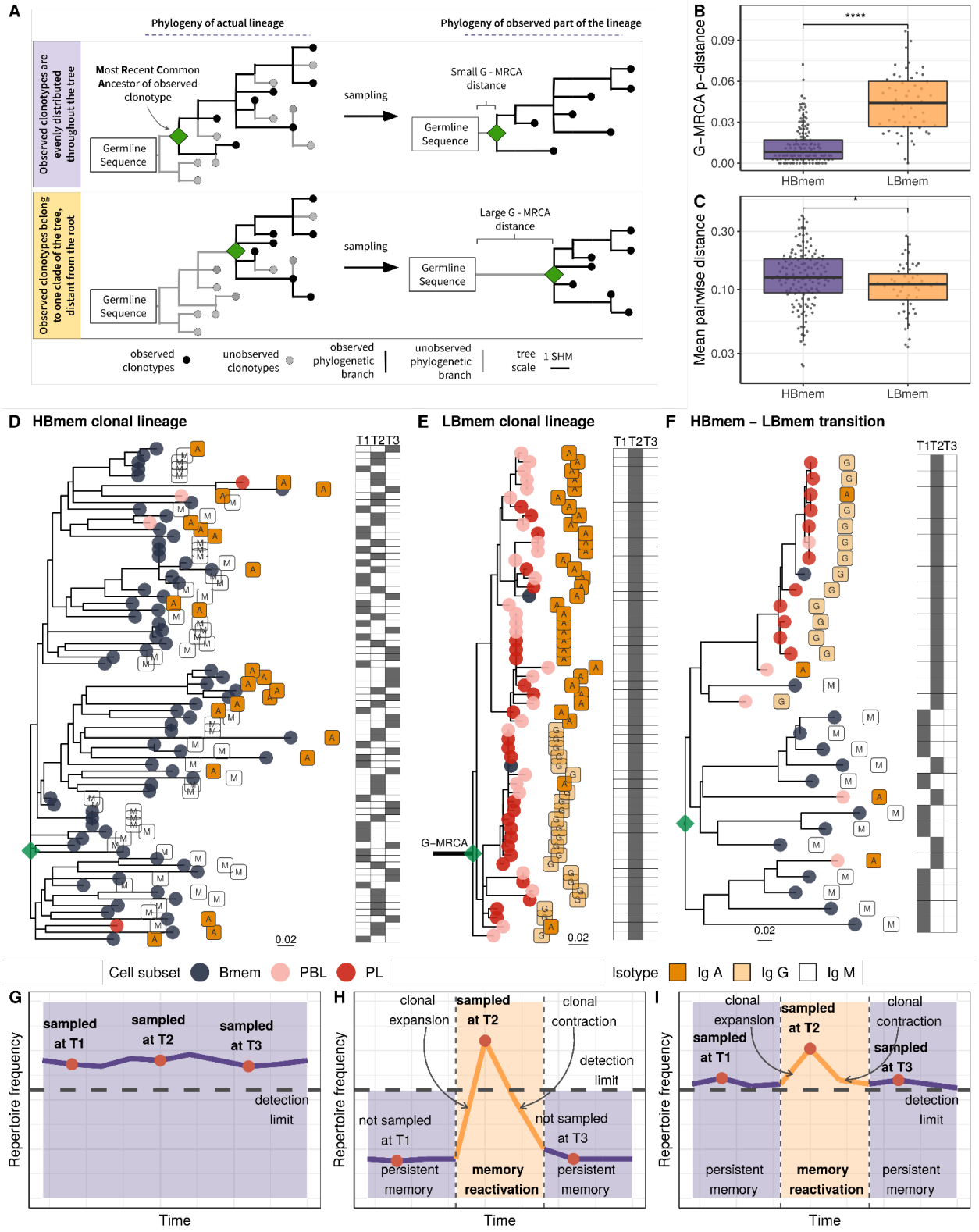
A
Phylogeny of actual lineage
Phylogeny of observed part of the lineage
Observed clonotypes are evenly distributed throughout the tree
Most Recent Common Ancestor of observed clonotype
Germline Sequence
sampling
Small G - MRCA distance
Observed clonotypes belong to one clade of the tree, distant from the root
Large G - MRCA distance
observed clonotypes
unobserved clonotypes
observed phylogenetic branch
unobserved phylogenetic branch
tree scale
1 SHM
B
G–MRCA p–distance
HBmem
LBmem
C
Mean pairwise distance
D HBmem clonal lineage
E LBmem clonal lineage
F HBmem – LBmem transition
T1 T2 T3
G–MRCA
0.02
Cell subset
Bmem
PBL
PL
Isotype
Ig A
Ig G
Ig M
G
H
I
Repertoire frequency
sampled at T1
sampled at T2
sampled at T3
detection limit
clonal expansion
clonal contraction
not sampled at T1
not sampled at T3
persistent memory
memory reactivation
Time

Figure 4. Phylogenetic history of HBmem and LBmem clonal lineages. A: A schematic illustration of how the distances between the germline sequence and the MRCA of a clonal lineage (G-MRCA distance) vary depending on which subset of clonotypes is sampled: a sample uniform with regard to the position on the tree (top panel), or only those belonging to a particular clade of the tree (bottom panel). **B**: Comparison of G-MRCA p-distance (*i.e.*, the fraction of differing nucleotides) for HBmem and LBmem lineages. **C**: Mean pairwise phylogenetic distance (*i.e.*, the distance along the tree) between clonotypes of the same lineage for HBmem and LBmem clusters. **D–F**: Representative phylogenetic trees for clonal lineages belonging to HBmem (**D**), LBmem (**E**), and an example of HBmem-LBmem transition (**F**). The LBmem sublineage in **F** is nested deep in the phylogeny of the memory clonotypes, and is not characterized by a particularly long ancestral branch, indicating that it is not an artifact of clonal lineage assignment. Circles correspond to individual clonotypes, with the cellular subset indicated by color, and the isotype by label. Tables at right indicate the presence or absence of the corresponding clonotype at each time-point. The G-MRCA distance is indicated with a thick line. **G–I:** Schematic representation of the hypothetical dynamics of relative size for clonal lineages represented in **D**, **E,** and **F**, respectively. Significance for B and C was obtained by the two-sided Mann-Whitney test. * = $p \leq 0.05$, ** = $p \leq 0.01$, *** = $p \leq 10^{-3}$, **** = $p \leq 10^{-4}$.

Reactivation of LBmem clonal lineages is driven by positive selection

Having shown that the LBmem lineages likely originate from clonal expansion of pre-existing memory, we further compared the contribution of positive (favoring new beneficial SHMs) and negative (preserving the current variant) selection between the LBmem and HBmem clusters. Since we observed only one clear example of an HBmem-LBmem transition (**Fig. 4F**), we could not claim with certainty that LBmem lineages always emerge from preexisting HBmem lineages rather than from some other memory type. Still, we were able to study LBmem reactivation by comparing differences in substitution patterns at the origin of HBmem and LBmem clusters. We reasoned that the G-MRCA distance of an HBmem lineage contains mutations fixed by primary affinity maturation after the initial lineage activation. In contrast, the G-MRCA distance of an LBmem lineage contains both mutations arising during primary affinity maturation and subsequent changes occurring later in the evolution of the lineage. Differences in the characteristics of the G-MRCA mutations between clusters are therefore informative of the process prior to the observed expansion of LBmem lineages.

To assess selection at the origin of the HBmem and LBmem lineages, we measured the divergence of nonsynonymous sites relative to synonymous sites (*i.e.*, the DnDs ratio). Classically, DnDs > 1 is interpreted as evidence for positive selection. However, DnDs > 1 is rare, because the signal of positive selection is usually swamped by that of negative selection. In the McDonald-Kreitman (MK) framework, positive selection is instead revealed by excessive nonsynonymous divergence relative to nonsynonymous polymorphism (*i.e.,* DnDs > PnPs; see Methods and **Supplementary Table S2** for examples), under the logic that advantageous changes contribute more to divergence than to polymorphism (McDonald and Kreitman 1991). The fraction of adaptive nonsynonymous substitutions (α) can then be estimated from this excess. We designed an MK-like analysis, comparing the relative frequencies of nonsynonymous and synonymous SHMs at the G-MRCA branch (equivalent to divergence in the MK test) to those in subsequent evolution of clonal lineages (equivalent to polymorphism in the MK test; **Fig. 5A**, see Methods).

In both the HBmem and LBmem clonal lineages, we observed a higher ratio of nonsynonymous to synonymous SHMs in the G-MRCA branches compared to subsequent tree branches, meaning that a fraction of SHMs acquired by MRCA was further fixed by positive selection. However, this fraction was higher in LBmem lineages (Fisher's exact test: $\alpha = 0.58$ and 0.65, $p < 10^{-6}$ and $< 10^{-15}$ in HBmem and LBmem, respectively). α of distinct clonal lineages was also generally higher in LBmem than in HBmem (median $\alpha = 0.57$ vs $\alpha = 0.18$, **Fig. 5B**), showing that positive selection more frequently preceded expansion of LBmem than HBmem lineages. The observation of excess α in the LBmem cluster compared to HBmem was robust to the peculiarities of the MK analysis (**Supplementary Table S3**). The higher α for LBmem compared to HBmem implies that a fraction of SHMs was positively selected in LBmem clonal lineages after primary affinity maturation.

Subsequent evolution of LBmem clonal lineages is affected by negative and positive selection

Next, we considered the effects of selection on HBmem and LBmem clusters following their divergence from their MRCAs — *i.e.*, in the subsequent evolution of a clonal lineage leading to the diversity of the observed clonotypes. We calculated the per-site ratio of nonsynonymous and synonymous SHMs ($\pi N \pi S$) among those that originated after the MRCA. The $\pi N \pi S$ of both clusters was < 1 (**Fig. 5C**). This deficit of nonsynonymous SHMs indicates negative selection in the observed part of clonal lineage evolution. The $\pi N \pi S$ ratio was lower in the LBmem cluster, indicating stronger negative selection.

To examine the selection affecting these post-MRCA SHMs in more detail, we studied the frequency distribution of SHMs in individual lineages, or their site frequency spectra (SFS) (R. Nielsen 2005; Neher and Hallatschek 2013; Nei and Kumar 2000; Horns et al. 2019; Nourmohammad et al. 2019) (**Fig. 5A**). SFS reflects the effect of selection on these SHMs. Deleterious SHMs are held back by negative selection, so that their frequency in the lineage remains low. By contrast, positive selection favors the spread of adaptive SHMs, increasing their frequency. Therefore, negative selection biases the SFS towards low frequencies, and positive selection, towards high frequencies.For each clonal lineage, we reconstructed the SFS of the SHMs accumulated since divergence from MRCA (**Fig. 5A**), and then averaged these SFSs within the HBmem and LBmem clusters. A larger proportion of the LBmem SFS corresponds to high frequencies compared to HBmem (**Fig. 5D**), indicating weaker negative and/or stronger positive selection in LBmem SFS.

To distinguish between these selection types, we calculated the proportion of the SFS distribution falling into each frequency bin for nonsynonymous SHMs, and divided it by the same value for synonymous SHMs (normalized $\pi N \pi S$; see Methods, **Fig. 5E**). The inter-cluster differences in the normalized $\pi N \pi S$ in low-frequency bins were generally reflective of negative selection, while the differences in the high-frequency bins were reflective of positive selection. Normalized $\pi N \pi S$ was significantly higher in the high-frequency (>60%) bins of SHMs in LBmem clonal lineages. This indicates that for LBmem, those nonsynonymous changes that were not removed by negative selection reached high frequencies more often than in HBmem. In total, these data indicate that a fraction of nonsynonymous mutations accumulated by LBmem

429 lineages were adaptive. We thus observed that reactivation of LBmem lineages is coupled with
430 strengthening of both types of selection: positive on the G-MRCA branch, and both positive and
431 negative during subsequent clonal lineage expansion. This pattern is most likely evidence of
432 new rounds of affinity maturation, which result in the acquisition of new advantageous changes
433 and preserve the resulting BCRs from deleterious ones. HBmem, in contrast, evolved more
434 neutrally under weaker negative selection, suggesting absence of antigen challenge during the
435 observation period (**Fig. 5F**).

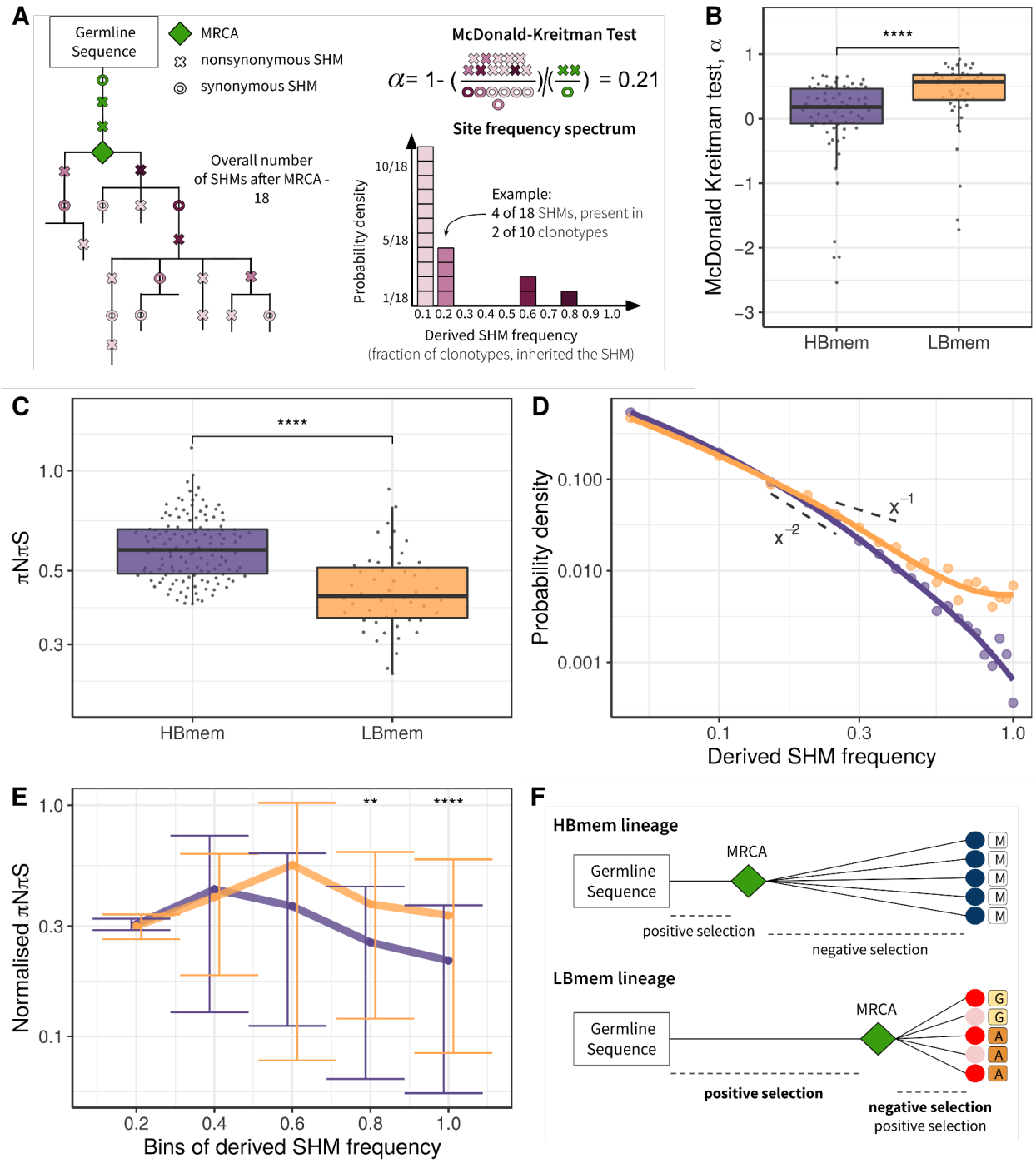
A
Germline Sequence
MRCA
nonsynonymous SHM
synonymous SHM
Overall number of SHMs after MRCA - 18
McDonald-Kreitman Test
α= 1- ( )/( ) = 0.21
Site frequency spectrum
Probability density
10/18
5/18
1/18
Example:
4 of 18 SHMs, present in 2 of 10 clonotypes
0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
Derived SHM frequency
(fraction of clonotypes, inherited the SHM)
B
McDonald Kreitman test, α
****
HBmem
LBmem
C
πNπS
****
HBmem
LBmem
D
Probability density
Derived SHM frequency
E
Normalised πNπS
**
****
Bins of derived SHM frequency
F
HBmem lineage
Germline Sequence
MRCA
positive selection
negative selection
LBmem lineage
Germline Sequence
MRCA
positive selection
negative selection
positive selection

Figure 5. Signatures of positive and negative selection in HBmem and LBmem clusters. A: Schematic of the McDonald-Kreitman (MK) test and site frequency spectrum (SFS) concept. **B**: MK estimate of the fraction of adaptive non-synonymous changes (α) between germline and MRCA in HBmem and LBmem clonal lineages. Only lineages with nonzero G-MRCA distance are included. N = 68 for HBmem, 49 for LBmem, see **Supplementary Table S3**); **C**: Comparison of mean pairwise πNπS of HBmem and LBmem lineages. **D**: Averaged SFS for HBmem and LBmem clonal lineages. The two dashed lines correspond to $f(x)=x^{-1}$, which is the expected neutral SFS under Kingman's coalescent model (Kingman 1982), and $f(x)=x^{-2}$. **E**: Comparison of normalized πNπS for HBmem and LBmem clonal lineages in various SHM frequency bins. The number of polymorphisms in each bin is normalized to the overall number of polymorphisms in a corresponding clonal lineage. **F**: Scheme summarizing features of HBmem and LBmem clonal lineages. Comparisons in **B**, **C**, and **E** were performed by two-sided Mann-Whitney test, with Bonferroni-Holm multiple testing correction in **E**. * = $p \leq 0.05$, ** = $p \leq 0.01$, *** = $p \leq 10^{-3}$, **** = $p \leq 10^{-4}$.

Discussion

Using advanced library preparation technology, we performed a longitudinal study of full-length BCR repertoires of the three main antigen-experienced B cell subsets — memory B cells, plasmablasts, and plasma cells — from peripheral blood of six donors, sampled three times over the course of a year. We analyzed these repertoires from two conceptually different but complementary points of view. First, we compared various repertoire features between the cell subsets, including clonotype stability in time and convergence between individuals. Second, we tracked the most abundant B cell clonal lineages in time and analyzed their cell subset and isotype composition, phylogenetic history, and mode of selection.

Comparative analysis of the cell subsets revealed significant differences in IGH isotype distribution, rate of SHM, and CDR3 length. IgM clonotypes predominated in the Bmem subset, whereas in ASCs the switched isotypes IgA and IgG together represented >80% of repertoire diversity on average. As expected, classical switched isotypes have higher rates of SHM, and the rate of SHM in ASCs is in general higher than in Bmem. The IgD isotype in Bmem cells showed similarities to IgM, where most IgD clonotypes had low levels of SHM, although there was a fraction of heavily-mutated clonotypes. On average, IgD-switched PL and PBL had a comparable level of SHM with IgG- and IgA-expressing ASC clonotypes. Notably, the level of SHM and CDR3 length in PBL on average exceeded that of PL in IgM, IgA, and IgG isotypes. We hypothesize that such PBLs with heavily hypermutated BCRs could be the subset of B cell progeny that continue to acquire mutations after optimal affinity has been achieved, while another part of the clonal progeny is committed to a long-lived PL fate and acquires the CD138 marker characteristic of this cell subset (Garimilla et al. 2019).

While different in many aspects, immune-experienced B cell subsets are similar — and concordantly distinct from naive B cells — in terms of IGHV gene segment usage. Moreover, we observed that the correlated enrichment or depletion in V segment usage frequency generally coincides with the level of sequence similarity of the V segments. Most IGHV-3 family members were observed more frequently in antigen-experienced B cells compared to naive subsets in all donors and time-points, while most of the other V genes that are well-represented in the naive subset decreased in frequency. These differences in V usage frequencies between naive and

478 antigen-experienced B cell subsets have also been reported in several previous studies, even
479 though different FACS gating strategies were used (Mitsunaga and Snyder 2020; Ghraichy et al.
480 2021). Our findings further support the idea that initial recruitment of B cells to the immune
481 response is in many cases determined by the germline-encoded parts of the BCR, presumably
482 CDR1 and CDR2. Previous studies have shown high levels of convergence in IGHV usage
483 between B cell clonotypes specific for particular pathogens or self-antigens (Peng et al. 2019;
484 Galson et al. 2015; Bashford-Rogers et al. 2019).

485 We further analyzed the repertoire similarity of cell subsets over time and between individuals.
486 Intuitively, the Bmem subset is the most stable over time, showing less repertoire divergence
487 and a greater number of shared clonotypes between sampling time-points in the same
488 individuals. Compared to between-individual sharing, we detected a very small number of
489 common clonotypes in Bmem cells. Those clonotypes have comparable levels of SHM to
490 private ones, assuming germinal center-dependent origin. Two recent studies on extra-deep
491 repertoires of bulk peripheral blood B cells reported 1–6% (Soto et al. 2019) or ~1% (Briney et
492 al. 2019) shared V-CDR3aa-J clonotypes between pairs of unrelated donors, with lower
493 repertoire convergence for class-switched clonotypes shown in the latter study. Using the same
494 method, we similarly measured 0.06% repertoire overlap in the Bmem subset (**Supplementary**
495 **Figure S4D**). Complementing the model proposed by Briney *et al.* — wherein IGH repertoires
496 are initially dissimilar and then homogenize during B cell development before finally becoming
497 highly individualized after immunological exposure — we found a significantly higher number of
498 shared clonotypes between IGH repertoires among the most abundant Bmem clonotypes,
499 indicating functional convergence presumably due to exposure to common environmental
500 antigens. The latter is further supported by the higher number of persisting Bmem clonotypes
501 observed among public clonotypes compared to private ones.

502 Next, we focused on the most abundant B cell clonal lineages, which are large enough to study
503 the interconnection between cell subsets and phylogenetic features of lineages. In all
504 individuals, the observed clonal lineages clearly fell into two clusters. HBmem represents
505 persistent memory with a predominant IgM isotype; such clonal lineages were equally sampled
506 from all time-points and rarely included ASC clonotypes. The MRCA of observed clonotypes in
507 HBmem lineages almost matched the predicted germline sequence — and in 14.5% of the
508 lineages, matched completely — indicating that the probability of observing a clonotype from
509 these lineages has no association with the position in that lineage's phylogeny. Horns and
510 colleagues observed lineages with very similar features to HBmem, which also possessed
511 persistent dynamics against a background of vaccine-responsive lineages and were
512 predominantly composed of the IgM isotype (Horns et al. 2019). However, their study was
513 performed on bulk B cells, so there was no possibility to track their relatedness to the Bmem
514 subset. In contrast, the LBmem cluster demonstrates completely different features, with
515 lineages largely composed of ASC clonotypes with switched IgA or IgG isotypes, showing active
516 involvement in ongoing immune response. The MRCA of LBmem lineages differed from the
517 germline sequence by some number of SHMs, and only 1.9% of LBmem lineages had a
518 complete match between the MRCA and the germline sequence. A large G-MRCA distance
519 implies that the observed clonotypes originated from an already-hypermutated ancestor, and
520 that we had therefore sampled clonotypes from a single clade of the lineage phylogeny. Such
521 an effect can be caused both by rapid expansion of the clade and migration of the clade's

522 clonotypes within the peripheral blood. We also observed that most LBmem lineages expanded
523 at T2 or T3 (38 out of 45, > 80%) had at least one clonotype detected in the Bmem subset at the
524 previous time-point, leading us to conclude that LBmems represent the progeny of reactivated
525 persistent B cells. We found one lineage that possesses all features of the HBmem cluster
526 except for one monophyletic clade, typical for LBmem lineage; thus, at least some LBmem
527 clonal lineages represent the progeny of reactivated persistent Bmem cells that are represented
528 among HBmem lineages. This example of HBmem-LBmem transition is very similar to
529 reactivated persistent memory, as observed by Hoehn *et al.* in response to seasonal flu
530 vaccination (Hoehn et al. 2021).

531 Our analysis of the selection mode in HBmem and LBmem lineages supported our
532 assumptions. We showed that both lineages experienced positive selection from the germline
533 sequence to the MRCA of the observed clonotypes — as expected, assuming that primary B
534 cell activation is followed by affinity maturation associated with clonal lineage expansion.
535 However, the pressure of positive selection was stronger in LBmem lineages than in HBmem. In
536 addition, we detected an excess of sites under positive selection in LBmem lineages that
537 underwent evolution after the MRCA as well. This leads us to the conclusion that LBmem cells
538 underwent additional rounds of affinity maturation after reactivation. *Hoehn et al.* did not study
539 the mode of selection in their reactivated lineages, but some clonotypes were sampled from
540 germinal centers, supporting the involvement of affinity maturation. In subsequent evolution
541 after the MRCA, we detected negative selection in both lineages — but again, stronger in
542 LBmem. We consider this excess negative selection in LBmem as an additional signature of
543 affinity maturation, purifying the lineage from deleterious BCR variants.

544 In this work, we performed a detailed longitudinal analysis of BCR repertoires from immune-
545 experienced B cell subsets from donors without severe pathologies, and from these data, we
546 have produced a framework for the comprehensive analysis of selection in BCR clonal lineages.
547 Our results demonstrate the interconnection of B cell subsets at a clonal level, B cell memory
548 convergence in unrelated donors, and the long-term persistence of memory-enriched clonal
549 lineages in peripheral blood. Signs of positive selection were detected in both memory- and
550 ASC-dominated B cell lineages. Together, the results of our evolutionary analysis of B cell
551 clonal lineages coupled with B cell subset annotation suggest that the reactivation of pre-
552 existing memory B cells is accompanied by new rounds of affinity maturation.

553 Methods

554 Donors, cells, and timepoints

555 Blood samples from six young and middle-aged (**Supplementary Table S1**) donors without
556 severe inflammatory diseases, chronic or recent acute infectious diseases, or vaccinations were
557 collected at three time-points (T1 - 0, T2 - 1 month, T3 - 12 months). Four donors suffered
558 allergic rhinitis to pollen, and two also suffered from food allergy. Informed consent was
559 obtained from each donor. The study was approved by the Ethical Committee of Pirogov
560 Russian National Research Medical University, Moscow, Russia. At each time point, 18–22 mL
561 of peripheral blood was collected in BD Vacuette tubes with EDTA. Peripheral blood

mononuclear cells were isolated using Ficoll gradient density centrifugation. To isolate subpopulations of interest, cells were stained with anti-CD19-APC, anti-CD20-VioBlue, anti-CD27-VioBright FITC, and anti-CD138-PE-Vio770 (all Miltenyi Biotec) in the presence of FcR Blocking Reagent (Miltenyi Biotec) according to the manufacturer's protocol, and then sorted using fluorescence-activated cell sorting (FACS; BD FacsAria III, BD Biosciences) into the following populations: memory B cells (Bmem; $CD19^{+}$ $CD20^{+}$ $CD27^{+}$ $CD138^{-}$, plasmablasts (PBL; $CD20^{-}$ $CD19^{Low/+}$ $CD27^{++}$ $CD138^{-}$), plasma cells (PL; $CD20^{-}$ $CD19^{Low/+}$ $CD27^{++}$ $CD138^{+}$). For each donor at T1, one replicate sample of each cell subpopulation was collected. At T2 and T3, two replicate samples were collected (50×10^3 to 100×10^3 Bmem, 1×10^3 to 2×10^3 PBL, 0.5×10^3 to 1×10^3 PL per sample).

Full-length IGH cDNA libraries and sequencing

IGH cDNA libraries were prepared as described previously (Turchaninova et al. 2016) with several modifications. Briefly, we used a rapid amplification of cDNA ends (RACE) approach with a template-switch effect to introduce 5' adaptors during cDNA synthesis. These adaptors contained both unique molecular identifiers (UMIs), allowing error-correction, and sample barcodes (described in Zvyagin et al. 2017), allowing us to rule out potential cross-sample contaminations. In addition to a universal sequence for annealing the forward PCR primer, we also introduced a 5' adaptor during the reverse transcription (RT) reaction, which allowed us to avoid using multiplexed forward primers specific for V segments, thereby reducing PCR amplification biases. Multiplexed C-segment-specific primers were used for RT and PCR, allowing us to preserve isotype information. Prepared libraries were then sequenced with an Illumina HiSeq 2000/2500, (paired-end, 2 x 310 bp).

Sequencing data pre-processing and repertoire reconstruction

Sample demultiplexing by sample-barcodes introduced in the 5' adapter and UMI-based error-correction were performed using MIGEC v1.2.7 software (Shugay et al. 2014). For further analysis, we used sequences covered by at least two sequencing reads. Alignment of sequences, V-, D-, J-, and C-segment annotation, and reconstruction of clonal repertoires were accomplished using MiXCR v3.0.10 (Bolotin et al. 2015) with prior removal of the primer-originated component of the C-segment. We defined clonotypes as a unique IGH nucleotide sequence starting from the framework 1 region of the V segment to the end of the J segment, and taking into account isotype. Using TIgGER (Gadala-Maria et al. 2015) software, we derived an individual database of V gene alleles for each donor and realigned all sequences for precise detection of hypermutations. For analysis of general repertoire characteristics (isotype frequencies, SHM levels, CDR3 length, IGHV gene usage, and repertoire similarity metrics) we used samples covered by at least 0.1 cDNA molecules per cell for Bmem, and at least 5 cDNA per cell for PBL and PL.

Repertoire characteristics analysis

Isotype frequencies, rate of SHM, and CDR3 lengths were determined using MiXCR v3.0.10 (Bolotin et al. 2015). For calculation of background IGHV gene segment usage and number of

shared clonotypes, we utilized data derived from Gidoni et al. 2019 (European Nucleotide Archive accession number ERP108501) representing naive B cell IGH repertoires. We used repertoires containing more than 5,000 clonotypes and processed them in the same way as our data. IGHV gene frequencies were calculated as the number of unique clonotypes to which a particular IGHV gene was annotated by MiXCR divided by the total number of clonotypes identified in the sample. To assess IGHV gene segments over- and under-represented in studied subsets, we utilized edgeR package v0.4.4 (Robinson et al., 2010) with the 'trended' dispersion model. To evaluate pairwise similarity between repertoires based on IGHV gene segment frequency distributions, we utilized Jensen-Shannon divergence, calculated using the following formula:

$$JS(P,Q) = \frac{1}{2}\sum_i p_i log_2 p_i + \frac{1}{2}\sum_i q_i log_2 q_i - \sum_i \left(\frac{p_i+q_i}{2} log_2\left(p_i + q_i\right)\right)$$

where P and Q represent distributions of IGHV gene segment in two repertoires, and p_i and q_i represent frequencies of individual member *i* (IGHV gene segment). *In silico* repertoires used for the calculation of background clonal overlap were generated with OLGA software v1.0.2 (Sethna et al. 2019) under standard settings utilizing the built-in model. For clonal overlap calculation, we downsized repertoires to a fixed number of clonotypes. For **Fig. 1B,** the 14,000 most abundant clonotypes were considered in Bmem, 600 in PBL, and 300 in PL. For **Fig. 1C**, we considered 5,000 clonotypes for all cell subsets. Clonotypes with identical CDR3 amino acid sequence and the same IGHV gene segment detected in both analyzed samples were considered shared. Clonotypes shared between repertoires of at least two individuals were termed as public.

Assignment of clonal lineages

Change-O v0.4.4 (Gupta et al. 2015) was utilized to assign clonal groups, defined as groups of clonotypes with the same V segment, CDR3 length, and at least 85% similarity in CDR3 nucleotide sequence. Before clonal group assignment, we excluded all clonotypes with counts equal to 1. Clonal groups represent observed subsets of clonal lineages originating from a single BCR ancestor, so for simplicity, we use the term 'clonal lineages'. To study evolutionary dynamics of clonal lineages, we joined all replicas, three time-points (T1, T2, and T3), and cell subsets for each patient into a single dataset and excluded clonotypes that were presented by a single UMI. Phylogenetic analysis was performed on four patients for whom we had samples at all time-points, and on clonal lineages containing at least 20 unique clonotypes as in (Nourmohammad et al. 2019).

Clusterization of clonal lineages in HBmem and LBmem clusters

We performed principal component analysis on six scaled variables of clonal lineage composition: fractions of Bmem, PBL, and PL, and fractions of IgM, IgG, and IgA. The IgE isotype was not detected in clonal lineages involved in phylogenetic analysis, so we did not include it as a variable. HBmem and LBmem clusters were defined using the K-means clustering algorithm.

Metric of persistence of clonal lineages

We estimated the frequency of a clonal lineage in the repertoire at a given time-point as the ratio of the number of unique clonotypes in the clonal lineage detected at this time-point to the overall number of unique clonotypes detected at this time-point. If the clonal lineage was not detected at some time-point, we assigned its frequency to pseudocount, as it would be a single clonotype detected from this time-point. To estimate persistence of clonal lineage frequency in the repertoire over time we defined the persistence metric:

$$P = \frac{1}{\frac{1}{2}\left(\frac{f_{max}}{f_i} + \frac{f_{max}}{f_j}\right)},$$

where f_{max} is the maximum frequency of the clonal lineage in the three time-points and $f_{i,j}$ are its frequencies in the other two (**Fig. 3D**). Persistence is equal to 1 if the frequency remains consistent at all three time points. If a clonal lineage was detected just once in the experiment and frequencies at other two time points were assigned to pseudocounts, the persistence approaches zero.

Reconstruction of clonal lineage germline sequence

We used MiXCR-derived reference V, D, and J segment sequences to reconstruct IGH germline sequences for each clonal lineage, concatenating only those sequence fragments which were present at CDR3 junctions of original MiXCR-defined clonotypes. Thus, random nucleotide insertions were disregarded, making them appear as gaps in the alignment of lineage clonotypes with the germline sequence. We excluded them from all parts of the phylogenetic analysis where germline sequence was required.

Reconstruction of clonal lineage phylogeny and MRCA

For phylogenetic analysis of clonal lineages, we aligned clonotypes with reconstructed germline sequences using MUSCLE version 3.8.31 with 400 gap open penalty (Edgar 2004). Next, we reconstructed the clonal lineage's phylogeny with RAxML version 8.2.11, using the GTRGAMMA evolutionary model and germline sequence as an outgroup, and computed marginal ancestral states (Stamatakis 2014). The ancestral sequence of the node closest to the root of the tree, represented by the germline sequence, is the MRCA of the sampled clonotypes. It can match the germline sequence or differ by some amount due to SHM, reflecting the starting point of subsequent evolution of observed clonotypes. This allowed us to distinguish between SHMs fixed in the clonal lineage on the way from the germline sequence to the MRCA (G-MRCA SHMs) versus polymorphisms within the observed part of lineage. The G-MRCA p-distance in **Fig. 4B** was measured as a fraction of diverged positions between germline and MRCA sequences.

McDonald-Kreitman (MK) test

The (MK) test is designed to detect the effects of positive or negative selection on population divergence from another species or its ancestral state (McDonald and Kreitman 1991). It is based on the comparison of ratios of nonsynonymous to synonymous substitutions observed in

674 diverged and polymorphic sites, and estimates the fraction of diverged amino acid substitutions
675 fixed by positive selection:

676 $$\alpha = 1 - \frac{P_n}{P_s} \cdot \frac{D_s}{D_n},$$

677 where P_n and P_s respectively represent nonsynonymous and synonymous polymorphisms, and
678 D_n and D_s respectively represent nonsynonymous and synonymous divergences fixed in the
679 population. Under neutral evolution, nonsynonymous and synonymous changes are equally
680 likely to be fixed or appear in the population as polymorphisms, so $\frac{D_n}{D_s} = \frac{P_n}{P_s}$ and α = 0. Positive
681 selection favors adaptive nonsynonymous changes to be fixed, and increases $\frac{D_n}{D_s}$ relative to $\frac{P_n}{P_s}$
682 , resulting in α > 0. Negative selection has the opposite effect and produces α < 0.

683 To detect selection in the origin of clonal lineages, we considered G-MRCA SHM as divergent
684 changes, and the remaining SHM in a clonal lineage after the MRCA as polymorphic ones (**Fig.**
685 **5A**). If we observed different nucleotides in the germline sequence and MRCA at a site that was
686 also polymorphic, we considered it as divergent only if the germline variant was not among the
687 polymorphisms (**Supplementary Table S2**, examples of codons q and r). Codons with unknown
688 germline state were excluded from the MK test (**Supplementary Table S2**, example of codon j).
689 To perform the MK test on joined HBmem or LBmem cluster variation, we summed variation of
690 all clonal lineages of the same cluster in each category (D_n,D_s, P_n, P_s). Calculations of α of
691 distinct clonal lineages for comparison of its distributions between two clusters were
692 complicated by zero G-MRCA distance in some clonal lineages, mostly belonging to the
693 HBmem cluster. We dealt with this using three approaches, presented in **Supplementary Table**
694 **S3**. In the first, we added pseudocounts to D_n and D_s in each clonal lineage, so that for clonal
695 lineages with zero G-MRCA distance, $\frac{D_n}{D_s} = 1$. In the second, we excluded clonal lineages with
696 zero G-MRCA distance from the analysis, still adding pseudocounts to D_n and D_s in each clonal
697 lineage in cases where the G-MRCA distance consists of just one nonsynonymous or
698 synonymous substitution. In the third, we compared only those clonal lineages that had at least
699 one nonsynonymous and at least one synonymous substitution on the G-MRCA branch. We
700 also calculated the MK test on joined variation for all types of exclusion criteria to check its
701 robustness; however, there is no need to exclude clonal lineages in the case of the joined test
702 (**Supplementary Table S3**). In the first approach clonal lineages with zero G-MRCA distance
703 always produced negative α and biased median α to negative values as well. Medians of α in
704 the second and third approaches were more consistent with results of the test on joined
705 variation. However, in the third approach, the filter excluded most of the HBmem cluster, and so
706 in the main test we presented results of the second approach (**Fig. 5B**). To check the
707 significance of deviation of α from neutral expectations, we used an exact Fisher test as in the
708 original MK pipeline (McDonald and Kreitman 1991).

πNπS

To calculate πNπS we identified SHMs in each clonal lineage relative to the reconstructed MRCA sequence. In multiallelic sites (with multiple SHMs observed, see codon *i* in **Supplementary Table S2** as an example) we considered each variant as an independent SHM event. πN and πS were calculated as the number of nonsynonymous and synonymous SHMs in a clonal lineage, normalized to the number of nonsynonymous and synonymous sites in the MRCA sequence respectively. The resulting πNπS value is the ratio between πN and πS:

$$\pi N \pi S = \frac{N}{N_S} : \frac{S}{S_S}$$

where N and S are the number of nonsynonymous or synonymous SHMs, respectively, observed in the clonal lineage and N_S and $S_S$are the number of nonsynonymous or synonymous sites, respectively, in the MRCA sequence of the clonal lineage, calculated as in (Gojobori 1986).

Site frequency spectrum

Site frequency spectrum (SFS) reflects the distribution of SHM frequencies in the clonal lineage. We calculated the frequency of each SHM as a number of unique clonotypes carrying the SHM relative to the overall number of unique clonotypes in the lineage. To visualize SFS, we binned SHM frequencies into 20 equal intervals from 0 to 1 with a step size of 0.05, and counted SHM density in each bin as the number of SHMs in a given frequency bin normalized to the overall number of SHMs detected in the lineage. To obtain the cluster average SFS, we took the mean of clonal lineages of the same cluster in each frequency bin.

Normalized πNπS in bins of SHM frequencies

To compare ratios of nonsynonymous and synonymous SHMs of different frequencies between two clusters, we calculated normalized πNπS in bins of SHM frequency. For this purpose we used a smaller number of frequency bins (0; 0.2; 0.4; 0.6; 0.8; 1) to reduce the probability of bins without observed SHMs. To deal with the remaining empty bins, we added pseudocounts to nonsynonymous and synonymous SHMs in each frequency bin. Thus, normalized πNπS in the *i*-th SHM frequency bin was calculated as following:

$$\text{normalized } \pi N \pi S = \frac{(N_i+1)/(\sum_{i=1}^{5} N_i+5)/N_S}{(S_i+1)/(\sum_{i=1}^{5} S_i+5)/S_S}$$

where N_i and S_i are the number of nonsynonymous and synonymous SHMs, respectively, in the *i*-th frequency bin, $\sum_{i=1}^{5} N_i$ and $\sum_{i=1}^{5} S_i$ are respectively the overall number of nonsynonymous and synonymous SHMs observed in the clonal lineage (the sum of SHMs in all frequency bins), and N_S and S_S are the number of nonsynonymous and synonymous sites respectively in the MRCA sequence of the clonal lineage, calculated as in (Gojobori 1986). To compare distributions of

normalized πNπS between two clusters of clonal lineages in the five frequency bins, we used the Mann-Whitney test with Bonferroni-Holm multiple testing correction.

Data analysis and visualization

All analysis was performed using R language (R Core Team 2018) and visualized with the ggplot2 package (Ginestet 2011). Ggtree package was used to visualize phylogenetic trees of clonal lineages (Yu et al. 2017). The code for repertoire analysis is available at https://github.com/amikelov/igh_subsets; the code for clonal lineage analysis is available at https://github.com/EvgeniiaAlekseeva/Clonal_group_analysis.

Data availability

Sequencing data have been deposited in the ArrayExpress database (www.ebi.ac.uk/arrayexpress, acc. num. E-MTAB-11193).

Acknowledgements

We are grateful to our donors. We are grateful to Alexey Neverov for helpful discussion of inference of selection. The study was supported by a grant from the Ministry of Science and Higher Education of the Russian Federation, grant #075-15-2020-807 (to D.M.C), and by a grant from the Russian Foundation for Basic Research (RBFR), project number 20-34-90153 (to E.I.A).

Author contribution

A.I.M. and I.V.Z. designed the study, collected samples and prepared cDNA libraries; M.A.T. participated in optimization of cDNA library preparation; E.A.K. and D.B.S. performed FACS sorting; A.I.M. performed repertoire data processing and analysis; E.I.A. and G.A.B. designed evolutionary analysis of clonal lineages; E.I.A. performed evolutionary analysis of clonal lineages; M.A.S. and D.M.C. participated in study design and discussions; A.I.M., E.I.A., I.V.Z. and G.A.B. wrote the manuscript with the contribution of all authors.

Competing Interests Statement

Authors declare no competing interests.

References

Bashford-Rogers, R. J. M., L. Bergamaschi, E. F. McKinney, D. C. Pombal, F. Mescia, J. C. Lee, D. C. Thomas, et al. 2019. “Analysis of the B Cell Receptor Repertoire in Six Immune-Mediated Diseases.” *Nature* 574 (7776): 122–26. https://doi.org/10.1038/s41586-019-1595-3.

771 Bolotin, Dmitriy A, Stanislav Poslavsky, Igor Mitrophanov, Mikhail Shugay, Ilgar Z Mamedov,
772 Ekaterina V Putintseva, and Dmitriy M Chudakov. 2015. "MiXCR: Software for
773 Comprehensive Adaptive Immunity Profiling." *Nature Methods* 12 (5): 380–81.
774 https://doi.org/10.1038/nmeth.3364.
775 Bonsignori, Mattia, Hua-Xin Liao, Feng Gao, Wilton B. Williams, S. Munir Alam, David C.
776 Montefiori, and Barton F. Haynes. 2017. "Antibody-Virus Co-Evolution in HIV Infection:
777 Paths for HIV Vaccine Development." *Immunological Reviews* 275 (1): 145–60.
778 https://doi.org/10.1111/imr.12509.
779 Bourcy, Charles F. A. de, Cesar J. Lopez Angel, Christopher Vollmers, Cornelia L. Dekker, Mark
780 M. Davis, and Stephen R. Quake. 2017. "Phylogenetic Analysis of the Human Antibody
781 Repertoire Reveals Quantitative Signatures of Immune Senescence and Aging."
782 *Proceedings of the National Academy of Sciences* 114 (5): 1105–10.
783 https://doi.org/10.1073/pnas.1617959114.
784 Briney, Bryan, Anne Inderbitzin, Collin Joyce, and Dennis R. Burton. 2019. "Commonality
785 despite Exceptional Diversity in the Baseline Human Antibody Repertoire." *Nature* 566
786 (7744): 393–97. https://doi.org/10.1038/s41586-019-0879-y.
787 Chaudhary, Neha, and Duane R. Wesemann. 2018. "Analyzing Immunoglobulin Repertoires."
788 *Frontiers in Immunology* 9 (March): 462. https://doi.org/10.3389/fimmu.2018.00462.
789 Davydov, Alexey N., Anna S. Obraztsova, Mikhail Y. Lebedin, Maria A. Turchaninova, Dmitriy B.
790 Staroverov, Ekaterina M. Merzlyak, George V. Sharonov, et al. 2018. "Comparative
791 Analysis of B-Cell Receptor Repertoires Induced by Live Yellow Fever Vaccine in Young
792 and Middle-Age Donors." *Frontiers in Immunology* 9 (October): 2309.
793 https://doi.org/10.3389/fimmu.2018.02309.
794 De Silva, Nilushi S., and Ulf Klein. 2015. "Dynamics of B Cells in Germinal Centres." *Nature*
795 *Reviews Immunology* 15 (3): 137–48. https://doi.org/10.1038/nri3804.
796 Edgar, R. C. 2004. "MUSCLE: Multiple Sequence Alignment with High Accuracy and High
797 Throughput." *Nucleic Acids Research* 32 (5): 1792–97.
798 https://doi.org/10.1093/nar/gkh340.
799 Gadala-Maria, Daniel, Gur Yaari, Mohamed Uduman, and Steven H. Kleinstein. 2015.
800 "Automated Analysis of High-Throughput B-Cell Sequencing Data Reveals a High
801 Frequency of Novel Immunoglobulin V Gene Segment Alleles." *Proceedings of the*
802 *National Academy of Sciences* 112 (8): E862–70.
803 https://doi.org/10.1073/pnas.1417683112.
804 Gaebler, Christian, Zijun Wang, Julio C. C. Lorenzi, Frauke Muecksch, Shlomo Finkin, Minami
805 Tokuyama, Alice Cho, et al. 2021. "Evolution of Antibody Immunity to SARS-CoV-2."
806 *Nature* 591 (7851): 639–44. https://doi.org/10.1038/s41586-021-03207-w.
807 Galson, Jacob D, Elizabeth A Clutterbuck, Johannes Trück, Maheshi N Ramasamy, Márton
808 Münz, Anna Fowler, Vincenzo Cerundolo, Andrew J Pollard, Gerton Lunter, and Dominic
809 F Kelly. 2015. "BCR Repertoire Sequencing: Different Patterns of B−cell Activation after
810 Two Meningococcal Vaccines." *Immunology & Cell Biology* 93 (10): 885–95.
811 https://doi.org/10.1038/icb.2015.57.
812 Garimilla, Swetha, Doan C. Nguyen, Jessica L. Halliley, Christopher Tipton, Alexander F.
813 Rosenberg, Christopher F. Fucile, Celia L. Saney, et al. 2019. "Differential
814 Transcriptome and Development of Human Peripheral Plasma Cell Subsets." *JCI Insight*
815 4 (9): e126732. https://doi.org/10.1172/jci.insight.126732.
816 Ghraichy, Marie, Valentin von Niederhäusern, Aleksandr Kovaltsuk, Jacob D Galson, Charlotte
817 M Deane, and Johannes Trück. 2021. "Different B Cell Subpopulations Show Distinct
818 Patterns in Their IgH Repertoire Metrics." *ELife* 10 (October): e73111.
819 https://doi.org/10.7554/eLife.73111.
820 Gidoni, Moriah, Omri Snir, Ayelet Peres, Pazit Polak, Ida Lindeman, Ivana Mikocziova, Vikas
821 Kumar Sarna, et al. 2019. "Mosaic Deletion Patterns of the Human Antibody Heavy

822 Chain Gene Locus Shown by Bayesian Haplotyping." *Nature Communications* 10 (1):
823 628. https://doi.org/10.1038/s41467-019-08489-3.
824 Ginestet, Cedric. 2011. "Ggplot2: Elegant Graphics for Data Analysis: Book Reviews." *Journal*
825 *of the Royal Statistical Society: Series A (Statistics in Society)* 174 (1): 245–46.
826 https://doi.org/10.1111/j.1467-985X.2010.00676_9.x.
827 Gojobori, Nei. 1986. "Simple Methods for Estimating the Numbers of Synonymous and
828 Nonsynonymous Nucleotide Substitutions." *Molecular Biology and Evolution*,
829 September. https://doi.org/10.1093/oxfordjournals.molbev.a040410.
830 Grimsholm, Ola, Eva Piano Mortari, Alexey N. Davydov, Mikhail Shugay, Anna S. Obraztsova,
831 Chiara Bocci, Emiliano Marasco, et al. 2020. "The Interplay between CD27dull and
832 CD27bright B Cells Ensures the Flexibility, Stability, and Resilience of Human B Cell
833 Memory." *Cell Reports* 30 (9): 2963-2977.e6.
834 https://doi.org/10.1016/j.celrep.2020.02.022.
835 Gupta, Namita T., Jason A. Vander Heiden, Mohamed Uduman, Daniel Gadala-Maria, Gur
836 Yaari, and Steven H. Kleinstein. 2015. "Change-O: A Toolkit for Analyzing Large-Scale B
837 Cell Immunoglobulin Repertoire Sequencing Data: Table 1." *Bioinformatics* 31 (20):
838 3356–58. https://doi.org/10.1093/bioinformatics/btv359.
839 Hoehn, Kenneth B., Jackson S. Turner, Frederick I. Miller, Ruoyi Jiang, Oliver G. Pybus, Ali H.
840 Ellebedy, and Steven H. Kleinstein. 2021. "Human B Cell Lineages Engaged by
841 Germinal Centers Following Influenza Vaccination Are Measurably Evolving." Preprint.
842 Immunology. https://doi.org/10.1101/2021.01.06.425648.
843 Horns, Felix, Christopher Vollmers, Cornelia L. Dekker, and Stephen R. Quake. 2019.
844 "Signatures of Selection in the Human Antibody Repertoire: Selective Sweeps,
845 Competing Subclones, and Neutral Drift." *Proceedings of the National Academy of*
846 *Sciences* 116 (4): 1261–66. https://doi.org/10.1073/pnas.1814213116.
847 Laserson, Uri, Francois Vigneault, Daniel Gadala-Maria, Gur Yaari, Mohamed Uduman, Jason
848 A. Vander Heiden, William Kelton, et al. 2014. "High-Resolution Antibody Dynamics of
849 Vaccine-Induced Immune Responses." *Proceedings of the National Academy of*
850 *Sciences* 111 (13): 4928–33. https://doi.org/10.1073/pnas.1323862111.
851 Mandric, Igor, Jeremy Rotman, Harry Taegyun Yang, Nicolas Strauli, Dennis J. Montoya,
852 William Van Der Wey, Jiem R. Ronas, et al. 2020. "Profiling Immunoglobulin Repertoires
853 across Multiple Human Tissues Using RNA Sequencing." *Nature Communications* 11
854 (1): 3126. https://doi.org/10.1038/s41467-020-16857-7.
855 McDonald, John H., and Martin Kreitman. 1991. "Adaptive Protein Evolution at the Adh Locus in
856 Drosophila." *Nature* 351 (6328): 652–54. https://doi.org/10.1038/351652a0.
857 Mitsunaga, Erin M., and Michael P. Snyder. 2020. "Deep Characterization of the Human
858 Antibody Response to Natural Infection Using Longitudinal Immune Repertoire
859 Sequencing." *Molecular & Cellular Proteomics* 19 (2): 278–93.
860 https://doi.org/10.1074/mcp.RA119.001633.
861 Neher, R. A., and O. Hallatschek. 2013. "Genealogies of Rapidly Adapting Populations."
862 *Proceedings of the National Academy of Sciences* 110 (2): 437–42.
863 https://doi.org/10.1073/pnas.1213113110.
864 Nei, Masatoshi, and Sudhir Kumar. 2000. *Molecular Evolution and Phylogenetics*. Oxford ; New
865 York: Oxford University Press.
866 Nielsen, Rasmus. 2005. "Molecular Signatures of Natural Selection." *Annual Review of*
867 *Genetics* 39 (1): 197–218. https://doi.org/10.1146/annurev.genet.39.073003.112420.
868 Nielsen, Sandra C.A., Fan Yang, Katherine J.L. Jackson, Ramona A. Hoh, Katharina Röltgen,
869 Grace H. Jean, Bryan A. Stevens, et al. 2020. "Human B Cell Clonal Expansion and
870 Convergent Antibody Responses to SARS-CoV-2." *Cell Host & Microbe* 28 (4): 516-
871 525.e5. https://doi.org/10.1016/j.chom.2020.09.002.
872 Nourmohammad, Armita, Jakub Otwinowski, Marta Łuksza, Thierry Mora, and Aleksandra M

873 Walczak. 2019. “Fierce Selection and Interference in B-Cell Repertoire Response to
874 Chronic HIV-1.” Edited by Thomas Leitner. *Molecular Biology and Evolution* 36 (10):
875 2184–94. https://doi.org/10.1093/molbev/msz143.
876 Peng, Wujian, Song Liu, Jingye Meng, Jiali Huang, Jianrong Huang, Donge Tang, and Yong
877 Dai. 2019. “Profiling the TRB and IGH Repertoire of Patients with H5N6 Avian Influenza
878 Virus Infection by High-Throughput Sequencing.” *Scientific Reports* 9 (1): 7429.
879 https://doi.org/10.1038/s41598-019-43648-y.
880 Perez-Andres, M., B. Paiva, W. G. Nieto, A. Caraux, A. Schmitz, J. Almeida, R. F. Vogt, et al.
881 2010. “Human Peripheral Blood B-Cell Compartments: A Crossroad in B-Cell Traffic.”
882 *Cytometry Part B: Clinical Cytometry* 78B (S1): S47–60.
883 https://doi.org/10.1002/cyto.b.20547.
884 R Core Team. 2018. “R: A Language and Environment for Statistical Computing. R Foundation
885 for Statistical Computing,.” https://www.R-project.org/.
886 Sakharkar, Mrunal, C. Garrett Rappazzo, Wendy F. Wieland-Alter, Ching-Lin Hsieh, Daniel
887 Wrapp, Emma S. Esterman, Chengzi I. Kaku, et al. 2021. “Prolonged Evolution of the
888 Human B Cell Response to SARS-CoV-2 Infection.” *Science Immunology* 6 (56):
889 eabg6916. https://doi.org/10.1126/sciimmunol.abg6916.
890 Sethna, Zachary, Yuval Elhanati, Curtis G Callan, Aleksandra M Walczak, and Thierry Mora.
891 2019. “OLGA: Fast Computation of Generation Probabilities of B- and T-Cell Receptor
892 Amino Acid Sequences and Motifs.” Edited by Bonnie Berger. *Bioinformatics* 35 (17):
893 2974–81. https://doi.org/10.1093/bioinformatics/btz035.
894 Shah, Hemangi B., Kenneth Smith, Jonathan D. Wren, Carol F. Webb, Jimmy D. Ballard,
895 Rebecka L. Bourn, Judith A. James, and Mark L. Lang. 2019. “Insights From Analysis of
896 Human Antigen-Specific Memory B Cell Repertoires.” *Frontiers in Immunology* 9
897 (January): 3064. https://doi.org/10.3389/fimmu.2018.03064.
898 Shugay, Mikhail, Olga V Britanova, Ekaterina M Merzlyak, Maria A Turchaninova, Ilgar Z
899 Mamedov, Timur R Tuganbaev, Dmitriy A Bolotin, et al. 2014. “Towards Error-Free
900 Profiling of Immune Repertoires.” *Nature Methods* 11 (6): 653–55.
901 https://doi.org/10.1038/nmeth.2960.
902 Soto, Cinque, Robin G. Bombardi, Andre Branchizio, Nurgun Kose, Pranathi Matta, Alexander
903 M. Sevy, Robert S. Sinkovits, Pavlo Gilchuk, Jessica A. Finn, and James E. Crowe.
904 2019. “High Frequency of Shared Clonotypes in Human B Cell Receptor Repertoires.”
905 *Nature* 566 (7744): 398–402. https://doi.org/10.1038/s41586-019-0934-8.
906 Stamatakis, Alexandros. 2014. “RAxML Version 8: A Tool for Phylogenetic Analysis and Post-
907 Analysis of Large Phylogenies.” *Bioinformatics* 30 (9): 1312–13.
908 https://doi.org/10.1093/bioinformatics/btu033.
909 Stavnezer, Janet, Jeroen E.J. Guikema, and Carol E. Schrader. 2008. “Mechanism and
910 Regulation of Class Switch Recombination.” *Annual Review of Immunology* 26 (1): 261–
911 92. https://doi.org/10.1146/annurev.immunol.26.021607.090248.
912 Turchaninova, M A, A Davydov, O V Britanova, M Shugay, V Bikos, E S Egorov, V I Kirgizova,
913 et al. 2016. “High-Quality Full-Length Immunoglobulin Profiling with Unique Molecular
914 Barcoding.” *Nature Protocols* 11 (9): 1599–1616. https://doi.org/10.1038/nprot.2016.093.
915 Vidarsson, Gestur, Gillian Dekkers, and Theo Rispens. 2014. “IgG Subclasses and Allotypes:
916 From Structure to Effector Functions.” *Frontiers in Immunology* 5 (October).
917 https://doi.org/10.3389/fimmu.2014.00520.
918 Wu, Yu-Chang, David Kipling, Hui Sun Leong, Victoria Martin, Alexander A. Ademokun, and
919 Deborah K. Dunn-Walters. 2010. “High-Throughput Immunoglobulin Repertoire Analysis
920 Distinguishes between Human IgM Memory and Switched Memory B-Cell Populations.”
921 *Blood* 116 (7): 1070–78. https://doi.org/10.1182/blood-2010-03-275859.
922 Yang, Xiujia, Minhui Wang, Jiaqi Wu, Dianchun Shi, Yanfang Zhang, Huikun Zeng, Yan Zhu, et
923 al. 2021. “Large-Scale Analysis of 2,152 Ig-Seq Datasets Reveals Key Features of B

Cell Biology and the Antibody Repertoire." *Cell Reports* 35 (6): 109110. https://doi.org/10.1016/j.celrep.2021.109110.

Yu, Guangchuang, David K. Smith, Huachen Zhu, Yi Guan, and Tommy Tsan−Yuk Lam. 2017. "Ggtree: Package for Visualization and Annotation of Phylogenetic Trees with Their Covariates and Other Associated Data." Edited by Greg McInerny. *Methods in Ecology and Evolution* 8 (1): 28–36. https://doi.org/10.1111/2041-210X.12628.

Zvyagin, I V, I Z Mamedov, O V Tatarinova, E A Komech, E E Kurnikova, E V Boyakova, V Brilliantova, et al. 2017. "Tracking T-Cell Immune Reconstitution after TCRαβ/CD19-Depleted Hematopoietic Cells Transplantation in Children." *Leukemia* 31 (5): 1145–53. https://doi.org/10.1038/leu.2016.321.