

Millions of dots,
each with a unique story



A Reassessment of IgM Memory Subsets in Humans

This information is current as of September 22, 2015.

Davide Bagnara, Margherita Squillario, David Kipling, Thierry Mora, Aleksandra M. Walczak, Lucie Da Silva, Sandra Weller, Deborah K. Dunn-Walters, Jean-Claude Weill and Claude-Agnès Reynaud

J Immunol published online 9 September 2015
<http://www.jimmunol.org/content/early/2015/09/09/jimmunol.1500753>

-
- Supplementary Material** <http://www.jimmunol.org/content/suppl/2015/09/09/jimmunol.1500753.DCSupplemental.html>
- Subscriptions** Information about subscribing to *The Journal of Immunology* is online at: <http://jimmunol.org/subscriptions>
- Permissions** Submit copyright permission requests at: <http://www.aai.org/ji/copyright.html>
- Email Alerts** Receive free email-alerts when new articles cite this article. Sign up at: <http://jimmunol.org/cgi/alerts/etoc>

The Journal of Immunology is published twice each month by The American Association of Immunologists, Inc., 9650 Rockville Pike, Bethesda, MD 20814-3994. Copyright © 2015 by The American Association of Immunologists, Inc. All rights reserved. Print ISSN: 0022-1767 Online ISSN: 1550-6606.



A Reassessment of IgM Memory Subsets in Humans

Davide Bagnara,^{*1} Margherita Squillario,[†] David Kipling,[‡] Thierry Mora,[§]
Aleksandra M. Walczak,[¶] Lucie Da Silva,^{*} Sandra Weller,^{*} Deborah K. Dunn-Walters,^{||}
Jean-Claude Weill,^{*} and Claude-Agnès Reynaud^{*}

From paired blood and spleen samples from three adult donors, we performed high-throughput V_H sequencing of human B cell subsets defined by IgD and CD27 expression: IgD^+CD27^+ (“marginal zone [MZ]”), IgD^-CD27^+ (“memory,” including IgM [“IgM-only”], IgG and IgA) and IgD^-CD27^- cells (“double-negative,” including IgM, IgG, and IgA). A total of 91,294 unique sequences clustered in 42,670 clones, revealing major clonal expansions in each of these subsets. Among these clones, we further analyzed those shared sequences from different subsets or tissues for V_H gene mutation, H-CDR3-length, and V_H/J_H usage, comparing these different characteristics with all sequences from their subset of origin for which these parameters constitute a distinct signature. The IgM-only repertoire profile differed notably from that of MZ B cells by a higher mutation frequency and lower V_H4 and higher J_H6 gene usage. Strikingly, IgM sequences from clones shared between the MZ and the memory IgG/IgA compartments showed a mutation and repertoire profile of IgM-only and not of MZ B cells. Similarly, all IgM clonal relationships (among MZ, IgM-only, and double-negative compartments) involved sequences with the characteristics of IgM-only B cells. Finally, clonal relationships between tissues suggested distinct recirculation characteristics between MZ and switched B cells. The “IgM-only” subset (including cells with its repertoire signature but higher IgD or lower CD27 expression levels) thus appear as the only subset showing precursor–product relationships with $CD27^+$ switched memory B cells, indicating that they represent germinal center–derived IgM memory B cells and that IgM memory and MZ B cells constitute two distinct entities. *The Journal of Immunology*, 2015, 195: 000–000.

Different B cell lineages are mobilized in mice during T-dependent and T-independent immune responses. T-dependent responses against protein Ags drive recruitment of naive follicular B cells into germinal centers (GCs), where somatic hypermutation and isotype switching allow emergence of expanded IgG^+ and IgA^+ memory B cell clones with

higher affinity for immunizing Ags. Recent evidence in mice demonstrates the existence of a mutated IgM memory compartment generated in parallel during the GC response (1, 2), even though the precise conditions of its activation in subsequent Ag challenges are still a matter of debate. In contrast, mouse T-independent responses (e.g., those against polysaccharides from encapsulated bacteria) involve marginal zone (MZ) B cells, a distinct splenic subset with limited recirculation capacities that rapidly differentiate into effector plasma cells secreting IgM as well as IgG and IgA Abs after isotype switching (3).

The murine and human peripheral and splenic B cell compartments differ in many ways. In humans, subsets with mutated Ig genes are markedly expanded, accounting for ~40% of B cells. Mutated B cells are identified by the CD27 marker, even though a small $CD27^-$ compartment that contains IgM^+ , IgG^+ , and IgA^+ has been identified in blood (4–6). Switched $CD27^+$ B cells are GC derived and are capable of accumulating a considerable number of mutations, while it was recently proposed that blood IgA^+CD27^- cells may have a T-independent origin, possibly originating in GALTs (7).

The existence of a human MZ cell lineage is still a topic of debate. Splenic $IgM^+IgD^+CD27^+$ B cells share with mouse MZ B cells a localization at the periphery of B cell follicles and a distinct cell surface phenotype ($CD21^{high}CD23^-IgM^{high}IgD^{low}CD1c^{high}$). However, in contrast to mouse MZ B cells previously described as a splenic resident subset with unmutated B cell receptors, human $IgM^+IgD^+CD27^+$ B cells harbor mutated IgV genes and recirculate (8). On the basis of the presence of such cells in CD40L-deficient patients (albeit in reduced numbers) that cannot initiate GC reactions, we proposed that MZ B cells prediversify their IgR outside immune responses (9). This proposition was further supported by finding splenic MZ B cell Ig gene diversification in infants before the development of functional T-independent responses (10).

^{*}Institut Necker-Enfants Malades, INSERM U1151–Centre National de la Recherche Scientifique Unité Mixte de Recherche 8253, Université Paris Descartes, Faculté de Médecine-Site Broussais, 75993 Paris Cedex 14, France; [†]Dipartimento di Informatica, Bioingegneria, Robotica e Ingegneria dei Sistemi, Università degli Studi di Genova, 16146 Genoa, Italy; [‡]Institute of Cancer and Genetics, School of Medicine, Cardiff University, Cardiff CF14 4XN, United Kingdom; [§]Laboratoire de Physique Statistique, Centre National de la Recherche Scientifique Unité Mixte de Recherche 8550, Université Pierre et Marie Curie and Ecole Normale Supérieure, 75005 Paris, France; [¶]Laboratoire de Physique Théorique, Unité Mixte de Recherche 8549, Centre National de la Recherche Scientifique and Ecole Normale Supérieure, 75005 Paris, France; and ^{||}Department of Immunobiology, Faculty of Life Sciences and Medicine, King’s College London, London SE1 9RT, United Kingdom

¹Current address: Karches Center for Chronic Lymphocytic Leukemia Research, Feinstein Institute for Medical Research, Manhasset, NY.

Received for publication March 31, 2015. Accepted for publication August 6, 2015.

This work was supported by the Ligue Nationale contre le Cancer (“Equipe labélisée”), a European Research Council Advanced Grant (Memo-B), and the Fondation Princesse Grace (to the team “Développement du Système Immunitaire”, INSERM U1151). This work was also supported by the Medical Research Council, the Biotechnology and Biological Sciences Research Council, the Human Frontiers Science program, and the Dunhill Medical Trust (to the King’s College London team).

Address correspondence and reprint requests to Prof. Jean-Claude Weill and Dr. Claude-Agnès Reynaud, Institut Necker-Enfants Malades, INSERM U1151–Centre National de la Recherche Scientifique Unité Mixte de Recherche 8253, Faculté de Médecine Paris Descartes - Site Broussais, Bâtiment Leriche - Porte 9, CS61431, 14 Rue Maria Helena Vieira Da Silva, 75993 Paris Cedex 14, France. E-mail addresses: jean-claude.weill@inserm.fr (J.-C.W.) and claude-agnes.reynaud@inserm.fr (C.-A.R.)

The online version of this article contains supplemental material.

Abbreviations used in this article: DN, double-negative; GC, germinal center; MZ, marginal zone; N, naive.

Copyright © 2015 by The American Association of Immunologists, Inc. 0022-1767/15/\$25.00

The presence of Ig V gene mutations in the IgM⁺IgD⁺CD27⁺ subset, however, is consistent with another interpretation (i.e., they represent IgM memory B cells that exited GCs prior to isotype switching) (11). “IgM-only” CD27⁺ cells, a minor B cell subset with very low IgD expression, are similarly categorized (12). The most striking evidence favoring the memory lineage argument is the observation of a large number of clonal relationships between blood IgG⁺CD27⁺ and IgM⁺IgD⁺CD27⁺ B cells through CDR3-specific PCR amplification (13).

The relative relationship with other B cell subsets was investigated further by extensive phenotypic and molecular profiling and by high-throughput sequencing of Ig genes (7, 14). In the latter study, clonal relationships were found between blood IgM⁺IgD⁺CD27⁺ and other IgM⁺ B cell subsets (IgM-only and IgM⁺IgD⁻CD27⁻) but not with switched CD27⁺ cells, although large-scale sequencing is less powerful than a CDR3-targeted PCR approach.

We addressed the origin of IgM memory subsets in humans using high-throughput sequencing of paired blood and spleen samples. For each subset, we identified a specific repertoire profile, based on V_H gene mutation distribution, H-CDR3 size, and V_H and J_H usage. Surprisingly, we found that all clonal relationships between IgM⁺IgD⁺CD27⁺ and switched CD27⁺ cells, as well as with other IgM⁺ subsets, involved cells displaying the IgM-only repertoire profile, suggesting that IgM memory cells have a more heterogeneous IgD and CD27 expression level than previously thought. This repertoire profile is distinct from MZ B cells. We also observed that all CD27⁻ subsets (IgM⁺, IgG⁺, and IgA⁺) include different subpopulations having multiple origins.

Materials and Methods

Samples

Paired blood and spleen samples were obtained from three organ donors (45-, 48-, and 54-y-old). All died of cerebral vascular accidents, and blood samples were taken before organ removal. The criteria for inclusion as organ donor include a noninfectious serology as well as normal blood parameters for standard metabolic markers and cell composition. These individuals can thus reasonably be considered as “healthy donors” at the time of their blood and spleen sampling. This study was authorized by the French Agence Nationale de la Biomédecine.

Cell sorting

PBMCs and splenocytes obtained by Ficoll–Hypaque density centrifugation were stained with the following anti-human Abs: V500 anti-CD19 (BD Biosciences), PerCPcy5.5 anti-CD38 (BioLegend), PE-cy7 anti-CD24 (BioLegend), FITC anti-IgD (Invitrogen), and allophycocyanin anti-CD27 (BD Biosciences). Dead cells were excluded by Sytox Blue staining (Invitrogen). B cells (1,000–20,000) were sorted using a FACSAria (BD Biosciences) and collected into 30 μl reverse transcription lysis buffer (14).

V_H gene amplification and sequencing

cDNA synthesis and isotype-specific seminested PCR amplification was performed as described previously (14). Purified PCR products were used for high-throughput sequencing on Roche 434 Titanium platform (LGC Genomics). Sequences were processed as reported previously (14).

Selection of unique sequences and H-CDR3 clustering

Identical sequences are predominantly generated by PCR amplification, and not through contribution of identical mRNA species because they present an inverse correlation with the initial cell input (data not shown). Only unique sequences were used for these analyses. Identical sequences were removed, including those differing only by insertions/deletions, which most probably were generated by the sequencing procedure, and by selecting the in-frame one. Among 152,981 sequences determined, 103,066 unique sequences were obtained.

Sequences from each donor were clustered based on CDR3 nucleotide sequences to identify clonally related sequences. CDR3 is defined as the nucleotide region comprised between the invariant Cys residue at the end of

all V_H genes and the invariant Trp from all J_H segments. We empirically selected a 16.5% divergence cutoff for the CDR3 clustering algorithm (i.e., allowing for an 8-nt difference over an average CDR3 length of 16 aa). We arrived at this value by starting from a higher cutoff (20%), and performing a manual evaluation (described below) of a representative selection of the resulting clones. After subsequent clustering at lower cutoffs (0.5% interval), we selected the lower threshold that maintained the assessed clonal associations. It needs to be emphasized that the stringency of this cutoff depends upon the average mutation frequency of the subset analyzed, and accordingly, a higher stringency should be used for the analysis of naive B cells (not included in this study). All clusters with sequences belonging to two or more B cell subsets, isotypes, and/or tissues were manually aligned and, in the case of false positive associations, were assigned to a new cluster. The criteria used to evaluate the relevance of the clustering included the distribution of divergent positions along the CDR3 sequence (in the V, D, and J coding segments versus at the V-D and D-J junctions) as well as the number of shared mutations in the V gene sequence.

Estimate of clone size and B cell sampling

The size of a clone is the number of cells belonging to that clone, and this can be estimated from the number of sequences attributed to that clone in the sequence dataset, using Bayes’ inversion rule. N represents the total number of cells under consideration (e.g., of a given cell type), and M the subset of these cells that belong to a given clone. Likewise, n represents the total number of sequences within that type, and m the subset of these that belong to the clone of interest. The naive estimate $M = N \times m/n$ could be quite inaccurate for very small clones (e.g., clones that have been seen once in the sequence dataset [$m = 1$]). Such clones could be of size of order $\sim N/n$, but they could also be much smaller. These, out of many similarly small clones, could randomly be assigned to the sequence dataset by chance. To better control what M could be based on our knowledge of m, we use Bayes’ rule: $P(M|m) = P(m|M) \times P(M)/P(m)$.

If we assume that the sequenced reads are a random subset of the cell receptors, $P(m|M)$ is given by a Poisson distribution: $P(m|M) = (1/m!) \times \exp(-M \times n/N) (M \times n/N)^m$. $P(M)$, in contrast, represents our prior information about the distribution of clone sizes. These distributions are typically observed to obey power laws, $P(M) \sim M^{-\nu}$ (in this paper we use $\nu = 1$). The result of Bayes’ procedure is a distribution over the possible values of M, $P(M|m)$, rather than a single number. In the plots depicted in Supplemental Fig. 1B, we represent the median value of that distribution, with error bars showing 95% confidence intervals. The higher m is, the more reliable the estimate of M is, as evidenced by the shrinking error bars as m increases.

We used the lower estimate of the real size of each individual clone in the different subsets to estimate the minimal fraction of the total B cell subset that the sampled clones represent (described in Supplemental Fig. 1 and listed in Supplemental Table I).

Statistics

For continuous variables, the unpaired t test was used and for proportions χ^2 test was used; for both, Bonferroni corrections were included. Supplemental Table III lists the results of statistical tests performed for the repertoire analysis reported in Figs. 5 and 6. All repertoire characteristics include error bars representing the 95% confidence interval of the distribution.

Results

We used paired blood and spleen samples from three adult organ donors, who, to the best of our knowledge, did not present any overt pathology. We isolated three different CD19⁺ B cell populations according to surface CD27 and IgD expression: MZ (CD27⁺IgD⁺), classical memory (CD27⁺IgD⁻), and double-negative (DN; CD27⁻IgD⁻) B cells. Surface expression of CD38 and CD24 was used to exclude transitional, GC, and plasma cells from the sorted populations (Fig. 1). Using isotype-specific H chain primers, IgM sequences were amplified from MZ B cells, whereas IgM, IgG, and IgA isotypes were amplified from both memory and DN cells (Mem_M, Mem_G, Mem_A, and DN_M, DN_G, and DN_A, respectively). After eliminating 49,975 duplicates, we obtained 91,294 unique IgH sequences from the three donors. These fell into 42,670 clonotypes after H-CDR3 clustering (Supplemental Table II).

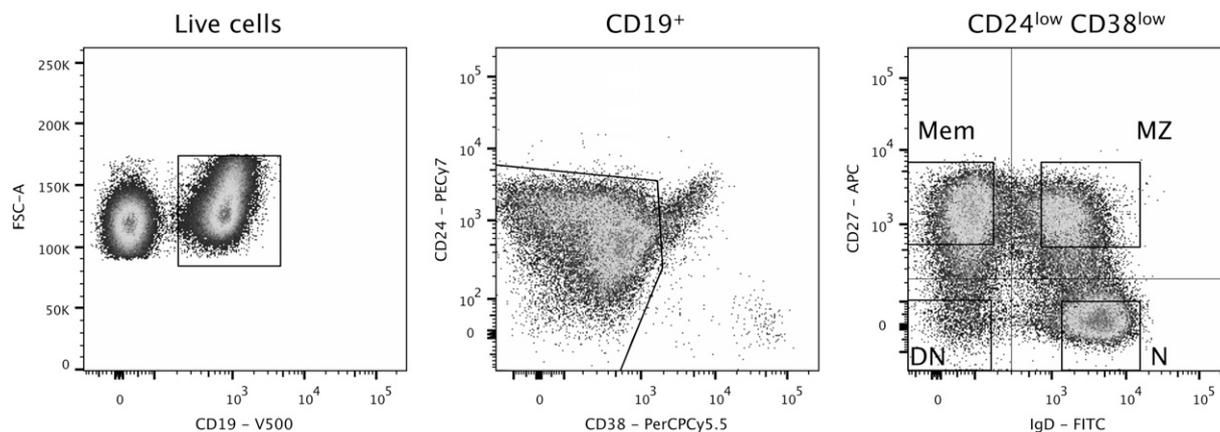


FIGURE 1. Isolation of N, MZ, Mem, and DN B cell subsets. Human peripheral blood and splenic CD19⁺ B cells were stained with CD38 and CD24 to exclude transitional, GC, and plasma cells. The four gated populations were sorted according to CD27 and IgD expression: CD27⁺IgD⁻ Mem cells, CD27⁺IgD⁺ MZ B cells, CD27⁻IgD⁺ N, and CD27⁻IgD⁻ DN B cells. The profile shown corresponds to a spleen sample.

We also isolated IgM sequences from naive B cells (N; CD27⁻IgD⁺) to estimate background mutations and for H-CDR3 length comparisons.

H-CDR3 clustering

The H-CDR3 clustering algorithm has to be sufficiently stringent to minimize the number of false clusters but flexible enough to not exclude sequences that differ because of somatic mutations or sequencing errors. Therefore, we empirically determined a threshold for inclusion of 16.5% sequence divergence and then systematically performed a manual screening of all clusters containing sequences from different populations, isotypes, and/or tissues (see *Materials and Methods*).

To estimate the errors (false clusters) generated by this algorithm, we ran the clustering algorithm on spleen sequences pooled from two different donors and performed a manual screening of the clusters containing sequences from both donors with the same selection criteria. False clustering can be influenced by the average mutation frequency and the number of sequences analyzed. The occurrence of false clusters varied for the different subsets, representing 2.4% for MZ and 0.1% for memory B cells; no shared clones were detected for DN B cells between donors. The fraction of clusters judged as real after manual screening among total clones was 0.6%, and this therefore represents the probability of having unrelated sequences clustering by chance (i.e., similar H chain rearrangements occurring in nonclonally related B cells).

Sampling and estimate of clone sizes and clonal diversity

The adult lymphoid compartment comprises $\sim 10 \times 10^9$ cells ($5\text{--}25 \times 10^9$) in blood and 75×10^9 cells ($70\text{--}80 \times 10^9$) in spleen (15). Taking into account the B cell percentage observed in the lymphoid gate by flow cytometry, the B cell pool was estimated to 10 , 23 , and 24×10^8 B cells in blood and 25 , 55 , and 43×10^9 B cells in spleen for the three donors, respectively. We similarly estimated B cell numbers for each subset from flow cytometry data for each donor, which, in combination with the number of unique sequences obtained for each B cell population, allowed us to evaluate the extent of our sampling as a fraction of total B cells per subset (Supplemental Table I). We collected roughly the same number of sequences for blood and spleen. Because blood is a 20- to 30-fold smaller reservoir for B cells, we achieved better sampling for the blood ($10^{-4}\text{--}10^{-5}$) than for the spleen ($10^{-5}\text{--}10^{-6}$), even though these values are low in both cases (Supplemental Table I).

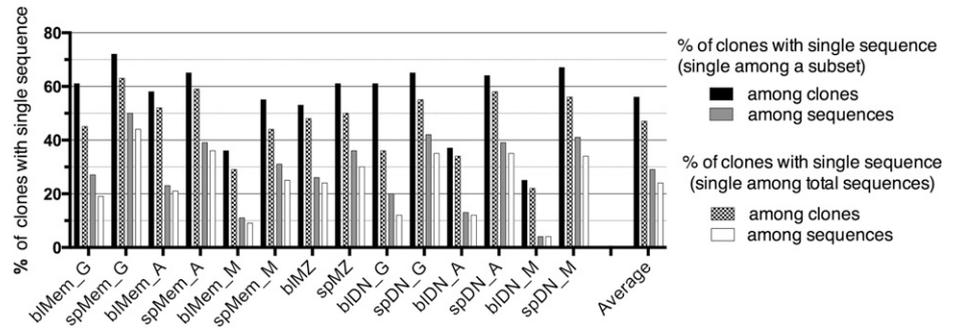
Observing clonally related sequences clearly indicates the presence of expanded clones. Clones defined by only one sequence represented 56% on average of the clones of a given subset (with large variation; see Fig. 2) but, in terms of sequences, only 29% of the sequences of this subset. This proportion is further reduced if the total pool of sequences obtained for one individual is taken into account for defining clones with one sequence and represents then 47% of the clones and 24% of the sequences on average for each subset (Fig. 2).

Using Bayes' rule and Poisson statistics, we extrapolated an estimated size of each sampled clone in each subset-tissue combination (see *Materials and Methods*), an extrapolation whose precision increases with the number of clonally related sequences. From this, we estimated the minimum fraction of the entire B cell subset that all sampled clones represent in vivo (Supplemental Table I and Supplemental Fig. 1 for the calculation procedure). Thus, despite small B cell sampling, sampled clones represent an average of at least 15% of the B cells in vivo in a given subset (Supplemental Table I). For all subsets, B cell clones are larger in spleen compared with blood, consistent with the relative size of their B lymphocyte compartment (Supplemental Fig. 1 and data not shown).

Clonal relationships between subsets and tissues

Through H-CDR3 clustering, we found clones that were present in different B cell populations, isotypes, and tissues. Clustering of all sequences was first performed for each donor separately, and then clones falling within the same category (e.g., clones present in blood and spleen MZ subsets) were pooled for the three donors to obtain weighted averages in frequency estimates. The total numbers of shared clones is listed in Supplemental Table II, and the fraction they represent, related to the total number of clones in the subset/tissue, is depicted in Fig. 3. The fraction of shared clones observed between two given subsets depends on the sampling size, the in vivo clone size distribution, and the total B cell numbers comprising each subset, making the observed fraction of shared clones not readily comparable. Nevertheless, this fraction represents a minimal estimate of the clonal relationships existing in vivo. A high frequency of clonally related sequences is observed between the Mem_A and the DN_A subset, and amounts to 9.5% of Mem_A and 12.5% of DN_A clones (Fig. 3A). Clones shared with MZ B cells also represent a large part of the Mem_M and DN_M subsets. In contrast, even though MZ B cells represent the largest sequence set, they display the lowest frequency of clonal relationships with other subsets. All subsets studied

FIGURE 2. Fraction of clones with single sequences among each subset. The fraction of clones with single sequences among each subset was reported either to the total number of clones or the total number of sequences of this subset. Two different calculations are performed, depending whether clones have single sequence at the level of a given subset (*first two bars*) or at the level of the total sequence pool of the donor (*last two bars*).



presented clones shared between blood and spleen, the highest frequency being observed for MZ B cells (Fig. 3B).

The different B cell subsets can be characterized by mutation frequency and distribution, H-CDR3-length, and usage of V_H family and J_H genes. We found that blood subsets tended to carry more mutations than splenic ones (apart from DN_M; discussed below) and mutation frequencies varied based on individual subsets and isotypes (Fig. 4A). With respect to V_H gene family usage, the V_{H1} family was rarely used by the various IgM-positive subsets, and the frequency of V_{H4} family gene usage appeared to be the most consistent discriminating feature (Fig. 4B). It should be noted that V_H family distribution might be biased because of the amplification reaction (i.e., the frequency of V_{H1} usage is less than in previous reports) (14). Thus, we do not consider our result as an absolute estimate of the V_H family repertoire, although it does serve as a valid tool to compare within B cell subsets as each was subjected to the same V_H amplification protocol. For J_H usage, J_{H6} (the J_H segment with the longest sequence contribution to H-CDR3) was the most variable (Fig. 4C).

Of 42,670 clones, 1,629 clones contained sequences from blood and spleen for the same B cell subset and 1,681 contained sequences from different subsets, allowing us to trace their relationships. We therefore analyzed these different repertoire parameters in clones sharing sequences from different B cell populations, comparing their profiles with those of their subset of origin. For mutation frequency, the only parameter that can differ between sequences within a single clone, we analyzed sequences from each subset within a clone separately (see legend to Fig. 5). In Figs. 5 and 6, V gene mutations are represented both by mutation distribution (with six different 2% mutation intervals from 0–2% to >10%) and by the average mutation frequency. The statistical relevance of the comparisons depicted in these figures is listed in Supplemental Table III.

Clonal relationships between different B-cell subsets

IgM subsets: MZ, IgM-only (Mem_M), and DN (DN_M). The three IgM-positive subsets (Mem_M, MZ, and DN_M) differed from each other for all four parameters. Mutation frequency decreased from Mem_M to MZ and then to DN_M, a difference most noticeable in the mutation distribution plot by the marked different proportion of sequences in the 0–2% mutation range (Fig. 5A, *first row*). MZ and DN_M clones differed from Mem_M based on greater V_{H4} and J_{H6} usage (Fig. 5A, *fourth and fifth rows*). DN_M differed from the other two due to its longer average H-CDR3 length (Fig. 5A, *third row*).

Analyses of clonal relationships between IgM subsets showed that clones shared between each two-population combination (MZ/Mem_M, MZ/DN_M, and Mem_M/DN_M) exhibited mutation distribution, V_H and J_H profiles similar to the Mem_M, even for MZ/DN_M where Mem_M was not part of the comparison

(Fig. 5A). Accordingly, clones shared between MZ and Mem_M diverged from the MZ subset with respect to these three parameters; shared Mem_M/DN_M clones differed from DN_M with respect to mutation frequency and J_H usage; and shared MZ/DN_M clones deviated from MZ and DN_M based on mutation frequency and J_H usage.

A three-dimensional analysis of these different subsets based on mutation frequency and V_{H4} and J_{H6} usage is presented in Fig. 5D. This illustrates the closer proximity of all clones shared between two IgM-positive subsets with Mem_M B cells than with MZ or DN_M B cells. All IgM shared clones thus appeared to display a homogenous profile, similar to the Mem_M subset, regardless of the isolated B cell population from which these sequences originated.

MZ/IgM-only and IgG/IgA memory subsets. Examination of clonal relationships between MZ and switched memory subsets (MZ/Mem_G and MZ/Mem_A) indicated that shared clones displayed a Mem_M profile for V_H and J_H usage as well as for mutation (for the IgM sequences) (Fig. 5B). Accordingly, shared clones diverged from MZ for mutations and V_H repertoire. The switched sequences of the shared clones had a higher mutation frequency than the IgM ones, each matching the corresponding switched subset.

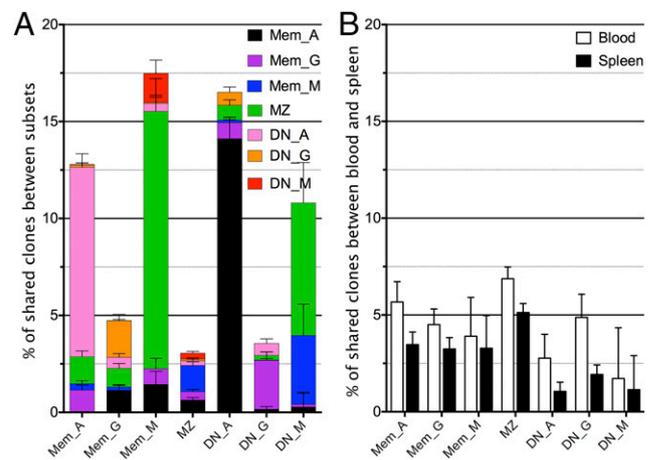


FIGURE 3. Clonal relationships between subsets and tissues. **(A)** Percentage of clones shared between B cell subsets, analyzed for each individual subset and color-coded for the subset sharing these clones. **(B)** Percentage of clones showing clonal relationships between blood and spleen for the same subset among blood clones (□) and spleen clones (■). The error bars represent the binomial proportion confidence interval (95%). Clonal relationships were analyzed for each donor separately, and the total number of clones belonging to the same category (e.g., clones shared between blood and spleen Mem_A subsets) was pooled from the three donors and reported to the total number of clones obtained in the corresponding subset for the three samples [e.g., total clones from blood Mem_A or spleen Mem_A subsets, *first two columns* (B)].

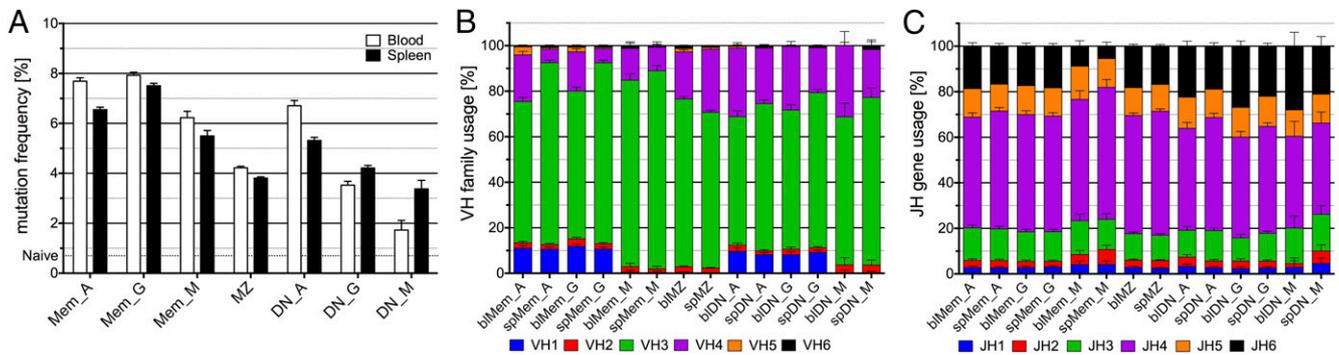


FIGURE 4. V_H mutation frequencies, V_H family, and J_H gene usage of the different B cell subsets. **(A)** V_H mutation frequencies were estimated by comparison with the ImMunoGeneTics database. Mutation frequencies are calculated based on the average mutation value obtained for each clone. The mutation frequency observed for the naive B cell pool (0.7%) is represented by a dotted line and is taken as an estimate of the contribution of each sequencing errors. V_H family **(B)** and J_H gene **(C)** usage.

The clonal relationship between IgM and switched memory subsets was analyzed by combining IgG and IgA clones (38 clones in total). Shared clones (Mem_M/Mem_A-G) presented an IgM-only profile for V_H and J_H usage, with mutation frequency corresponding to the parent isotype-matching subset.

These data suggest that clones shared between the MZ and switched memory subsets present a repertoire profile similar to Mem_M/Mem_A-G shared clones and represent IgM-only B cells that underwent class switching with accumulation of additional mutations. Accordingly, these clones segregated close to the Mem_M subset in the three-dimensional plot (Fig. 5D).

MZ and DN_G/DN_A subsets. Finally, defining clonal relationships between MZ and switched DN subsets revealed that shared clones (MZ/DN_G and MZ/DN_A) exhibit a MZ profile for V_H and J_H usage and, for IgM sequences within these clones, for mutation frequency (Fig. 5C), although the data are limited in amount (MZ/DN_G: 16 clones, MZ/DN_A: 41 clones). Accordingly, both MZ/DN_G and MZ/DN_A shared clones differed from Mem_M for mutation, and MZ/DN_A also differed from Mem_M for V_H repertoire (other parameters were not statistically significant). Switched sequences within shared clones had a mutation frequency corresponding to the isotype-matching DN cells. MZ/DN_G shared clones had H-CDR3 lengths similar to the DN parental subset, whereas differing from both MZ and Mem_M. MZ clones shared with DN_A and DN_G sequences were the only ones segregating close to the MZ subset in the three-dimensional plot (Fig. 5D).

These data suggest that a small fraction of MZ B cells undergoes isotype switching and persists in the DN population. MZ B cell switching to IgG DN cells appears biased for clones bearing longer H-CDR3 than the bulk MZ B cell population.

Clonal relationships between blood and spleen subsets

Memory B cells. Switched memory B cells differed between blood and spleen mainly by V_H4 usage, which was lower in spleen (Fig. 6A, fourth row) and, for Mem_A, for mutation frequency, which was higher in blood (Fig. 6A, first and second rows).

Clones shared between blood and spleen in the switched Mem IgG and IgA subpopulations displayed low V_H4 usage and mutation frequency for IgA, typical of the spleen (Fig. 6A, 6B). Thus, it appeared that switched memory cells in the spleen recirculate in an unbiased manner, at least based on V_H repertoire, whereas the blood counterparts comprise a mixture of splenic and other lymphoid tissue-derived memory B cells with different V_H repertoires, contributing to its higher V_H4 usage. Moreover, for IgA but not for IgG, the total blood-contained

clones had significantly more mutations than the ones shared with spleen, indicating that clones from other lymphoid tissues contributing to the blood repertoire carried Igs with higher mutation loads. The IgM-only memory subset displayed minor differences between tissues; thus analysis of the 43 shared clones was not informative (Fig. 6B).

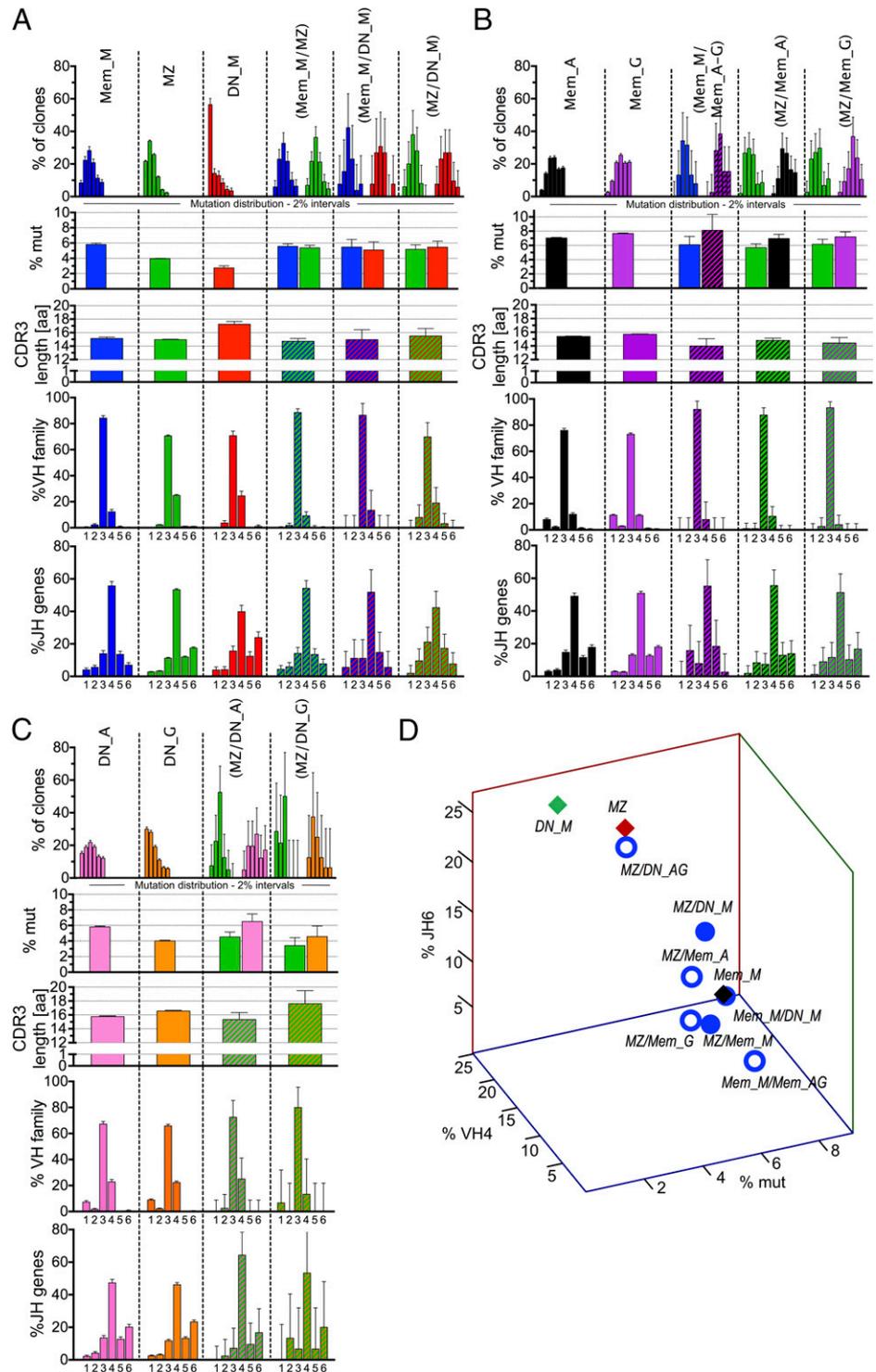
The closer proximity of clones shared between blood and spleen to the spleen compartment for switched memory B cells is illustrated on a two-dimensional plot of mutation frequency and V_H4 usage (Fig. 6C).

MZ B cells. MZ B cells in blood and spleen exhibited a V_H profile opposite to that of the other subsets, with a lower V_H4 usage in the blood. MZ B cells clones shared between blood and spleen displayed a V_H profile closer to the blood (low V_H4) and significantly different from spleen (Fig. 6B), suggesting that splenic MZ B cells have a biased recirculation, consisting of those with a lower V_H4 representation. Accordingly, clones shared between blood and spleen segregated closer to the blood subset in the two-dimensional representation (Fig. 6C). Thus, the blood MZ B cell population appears to be composed of a subset of splenic MZ B cells, suggesting that not all splenic MZ B cells recirculate equally.

DN subsets. The switched DN subsets presented the same differences between blood and spleen as their $CD27^+$ counterpart: higher V_H4 usage (and mutation frequency for DN_A) in blood. Similarly, shared clones for the DN IgG subset showed the same V_H repertoire bias toward the splenic repertoire observed for memory IgG; the shared DN_A clones had in addition the same mutation bias observed for memory IgA, the only two parameters for which differences in the comparison with the two parent tissues reached statistical significance (data not shown). This suggests that switched cells show similar recirculation dynamics for both $CD27^+$ and $CD27^-$, with blood subsets containing both recirculating splenic cells and cells from other tissues.

DN IgM cells did not present a sufficient number of shared clones to perform such analysis. Nevertheless, blood and spleen subsets present consistent differences, notably in mutation distribution and H-CDR3 length (Fig. 7A). In spleen, almost half of the clones displayed unmutated V_H genes with H-CDR3s longer than naive B cells (18.6 versus 16.8 aa) (Fig. 7B). In contrast, the mutated fraction carries H-CDR3s with a length comparable to the other mutated subsets. Segregation of these two parameters suggested that the DN_M subset included B cells having at least two different origins and that part of the mutated DN_M sequences corresponds to IgM-only B cells downmodulating surface $CD27$.

FIGURE 5. Repertoire profile of clones shared between different subsets. Analysis of clones shared between the different IgM B cell subsets (**A**), between IgM memory or MZ and IgG/IgA memory B cells (**B**), and between MZ and DN IgG/IgA B cells (**C**). Each column represents a specific B cell population, with its name indicated on top of each panel: B cell subset, CD27⁺IgD⁻ Mem, CD27⁺IgD⁺ MZ, and CD27⁻IgD⁻ DN B cells; Ig isotype: IgG (G), IgA (A), and IgM (M). MZ are IgM⁺ and only analyzed in (C). Populations indicated in brackets represent the tissue/subset of origin between which the clones analyzed are shared. The repertoire analysis includes five rows, with, from the top: 1) V_H mutation frequency depicted as fraction of clones within six different 2% mutation intervals from 0–2% (*left*) to >10% mutation frequency (*far right*); 2) average V_H mutation frequency; 3) average H-CDR3 aa length; 4) V_H family 1–6 usage v) J_H 1–6 usage. The mutation data for shared clones (*first and second row*) is plotted separately for the sequences belonging to each of the two populations involved because, contrary to other parameters, V_H mutations can vary between B cells belonging to the same clone. All analyses are based on clones, not sequences, and take into account, for mutation frequencies, the average mutation frequency of the clone. (**D**) Three-dimensional plot representation according to mutation frequency, V_H4 and J_H6 usage of clones shared between two IgM-positive subsets (filled blue dots) or between MZ or Mem_M and Mem_A/Mem-G or DN_A/DN_G (open blue dots). In the latter case, data concerning MZ clones shared with DN_A and DN_G subsets were pooled. Parental IgM subsets are represented by diamonds, red for MZ, green for DN_M, and black for Mem_M. Statistics are depicted in Supplemental Table III to facilitate reading, with the different pairwise comparisons indicated. Error bars represent the 95% interval of confidence of data distribution and therefore directly reflect the size of the sample analyzed.



The blood DN IgM subset was composed mainly of clones with low mutation (only twice the background of naive B cells) and long H-CDR3 (Figs. 4A, 5C), suggesting a different recirculation of mutated and unmutated subsets or a downmodulation of the CD27 marker taking place upon splenic location or trafficking of IgM-only B cells.

A schematic representation of the main clonal relationships observed between subsets or tissues is provided in Fig. 8.

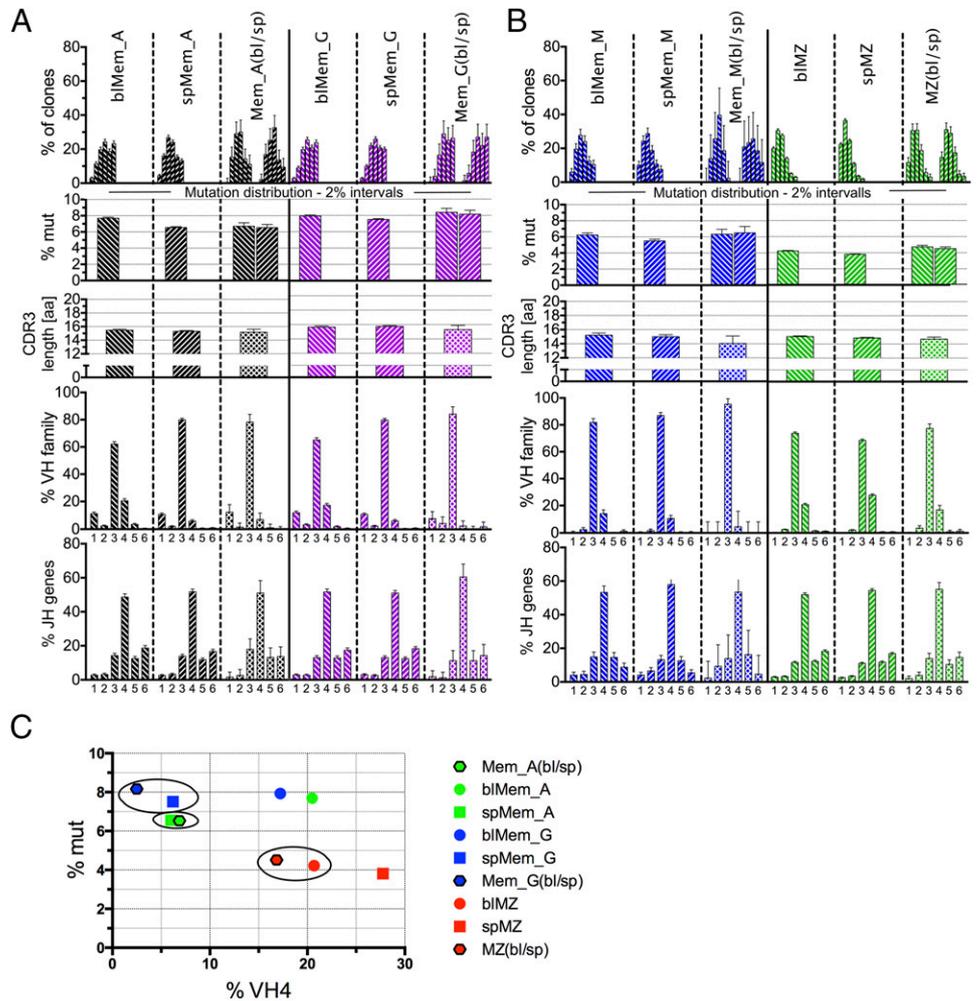
Discussion

We performed high-throughput Ig V_HDJ_H sequencing of human B cell subsets from paired blood and spleen samples, to investigate

clonal relationships between MZ B cells and other memory subsets as well as the recirculation dynamics of different B cell subsets between blood and spleen.

Sequences from seven different B-cell populations were collected from these two compartments based on cell sorting according to expression of surface IgD and CD27 (after excluding GCs or plasma cells) and on isotype-specific PCR amplification: MZ B cells (IgD⁺CD27⁺, IgM⁺), memory B cells (IgD⁻CD27⁺, IgG⁺, IgA⁺ and IgM⁺, i.e., “IgM-only”), and “DN” cells (IgD⁻CD27⁻, IgM, IgG, and IgA). This latter subset, albeit CD27⁻, was previously shown by several groups to include B cells with mutated Ig genes (4, 5).

FIGURE 6. Repertoire profile of clones shared between blood and spleen. Analysis of clones shared between blood and spleen for IgG and IgA (**A**) and IgM subsets (**B**). Abbreviations and criteria of repertoire analysis are depicted in legend to Fig. 5 with the addition of tissue: blood (bl) and spleen (sp). (**C**) Two-dimensional plot representation according to mutation frequency and V_H4 usage of the clones shared between blood and spleen for Mem_G (blue), Mem_A (green), and MZ (red) B cells together with the profile of the parental blood and spleen subsets. Spleen sequences are represented by squares, blood sequences by dots, and sequences shared between blood and spleen by a hexagon framed in black. Circles mark the closest parental subset of the clones shared between both tissues.



Even though we collected and analyzed several thousand unique V_H sequences, they still only represent a small sample of the total B cell pool, which is estimated to be $\sim 10^{-5}$ for blood B cell subsets and $\sim 10^{-6}$ for splenic ones. Surprisingly, despite this low sampling, a large number of nonidentical sequences were clonally related (91,294 sequences segregating into 42,670 clones). As mentioned in similar studies (6), isolation of clonally related V_H sequences indicates that B cell clones have undergone major clonal expansion, with estimated average clone sizes of $\sim 10^5$ cells for blood and 10^6 for spleen. Clones having more than one sequence within a subset represented 40–50% of the total number of clones obtained (i.e., $\sim 70\%$ of the total sequences). Moreover, an average of 16% of clones composed by a single sequence within a subset appeared to have a clonal relative in

another subset or tissue of the same individual, suggesting that expanded clones represent a major fraction of each B cell subset and not a minor one.

Interestingly, after extrapolating to the in vivo size of the B cell clones, we calculated that the sampled clones represented at least 15% of the total B cells in the compartment, in striking contrast with the 10^{-5} – 10^{-6} sampling performed. Estimating the global repertoire diversity of these different subsets was not within the scope of this study because it would require sequencing of multiple samples of the same B cell population to estimate whether expanded clones are distributed along a normal distribution of sizes, a likely possibility for classical memory subsets, or distributed bi- or multimodally, thus corresponding to a much larger repertoire diversity.

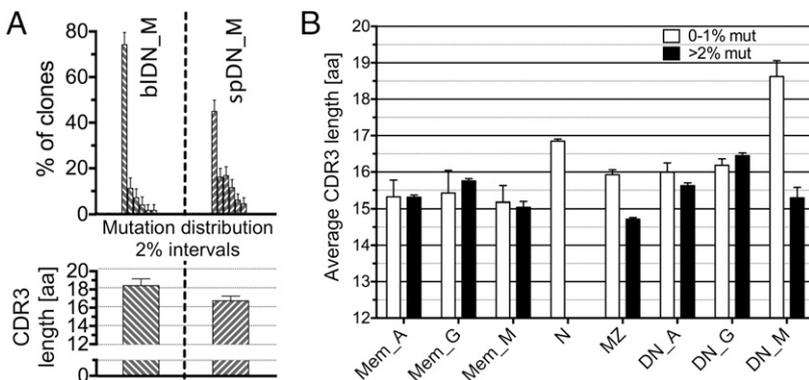


FIGURE 7. Mutation frequency of DN_M subsets and average H-CDR3 aa length in mutated and unmutated splenic clones. (**A**) V_H mutation frequency depicted as fraction of clones within six different 2% mutation intervals, from 0–2% to $>10\%$ mutation frequency (first row); average CDR3 aa (a.a.) length (second row) of blood (bl) and spleen (sp) DN_M subsets. (**B**) Average H-CDR3 length of each splenic B cell subset represented according to their mutation frequency (\square , $<1\%$ mutation; \blacksquare , $>2\%$ mutation). Error bars represent the 95% interval of confidence.

The focus of this study was the investigation of clonal relationships between subsets and tissues. We used an “intermediate” high-throughput approach that generated tens of thousands but not hundreds of thousands of sequences, allowing a relatively easy manual assessment of clones obtained through the clustering program, a critical point for their validity and for the relevance of the conclusions obtained. The number of sequences collected was nevertheless sufficiently large to define a robust repertoire signature for each individual subset and for most combinations of shared clones between two given populations. This subset-specific signature was based on mutation frequency and distribution, V_H repertoire (V_H4 gene family usage that appears to be the most discriminating parameter), J_H repertoire (notably J_H6 usage), and H-CDR3 length.

One-hundred seventy-seven MZ clones among 16,824 analyzed had clonal relationships with switched IgG or IgA memory subsets. Strikingly, the overall mutation frequency and V_H gene profile of these IgM sequences differed from MZ B cells while precisely matching characteristics of IgM-only B cells, clearly showing a mutation level intermediate between MZ and switched memory B cells and a lower V_H4 usage. Clones shared between switched memory and IgM-only B cells also harbored this IgM-only repertoire profile. Similarly, all IgM sequences presenting clonal relationships between subsets (MZ with IgM-only, IgM-only with DN IgM, or even DN IgM with MZ) displayed the unique repertoire profile of IgM-only B cells. These results suggest that IgM memory B cells with IgM-only repertoire characteristics are in fact more heterogeneous than previously estimated for IgD and CD27 expression and therefore fall in the MZ and DN sorting gates (schematized in Fig. 8B).

Clonal relationships between tissues further distinguished MZ from switched B cells. In effect, switched memory B cell clones shared between blood and spleen had the spleen repertoire profile of the corresponding memory subset, suggesting, possibly not unexpectedly, that B cells from multiple lymphoid compartments and harboring a different repertoire profile contribute to the memory recirculating pool. For IgA memory B cells, clones contributed by other lymphoid tissues carry Ig with a higher mutation load, which

would be consistent with a possible mucosal origin, a tissue in which IgA^+ B cells display the highest mutation frequency (16). Recirculating MZ clones had a signature significantly closer to the blood subset, suggesting, in contrast, that recirculation concerns a subset of splenic MZ B cells with a distinct repertoire (Fig. 8A). Our data support the notion that the spleen is the major reservoir for blood $IgM^+IgD^+CD27^+$ B cells in agreement with the observed impact of splenectomy on this recirculating subset (8, 17). However, a minor contribution of other lymphoid tissues cannot be excluded (18), but if existing, it would concern cells with a repertoire similar to recirculating splenic MZ cells, contrary to memory B cells.

Although MZ B cells do not appear affiliated with the switched $CD27^+$ subset, a few clonal relationships involving MZ B cells with an MZ-like mutation profile were observed with $CD27^-$ switched cells, a minor subset showing age-related accumulation (19). Interestingly, the clones shared with DN IgG^+ cells have longer H-CDR3s compared with the MZ subset, indicating that switching and maintenance within the DN pool might involve MZ B cells with distinct specificities. This was not the case for IgA, for which H-CDR3 sizes were in the normal range.

DN memory subsets thus appear to have multiple origins, comprising at least MZ and $CD27^+$ memory related B cells. $CD27^-IgA^+$ cells indeed have a large fraction of clonally related sequences, which are for their major part shared with $CD27^+IgA^+$ cells. Whether such relations represent precursor-product relationships or a reversible modulation of surface CD27 remains to be established. $CD27^-IgG^+$ cells have similar, albeit less numerous clonal relationships with their $CD27^+$ counterparts. The most drastic heterogeneity concerns DN IgM cells that can be split into cells with an IgM-only profile or cells with low mutation frequency and H-CDR3 sizes larger than naive B cells. Their presence in the DN pool likely corresponds to a repertoire selection bias, possibly redirecting cells with autoimmune potential into this $CD27^-$ fraction. Whether specific activation conditions could salvage DN IgM^+ B cells with such unusual H-CDR3 lengths and mobilize them into the memory effector pool is obviously an important issue because such long H-CDR3s are notably the

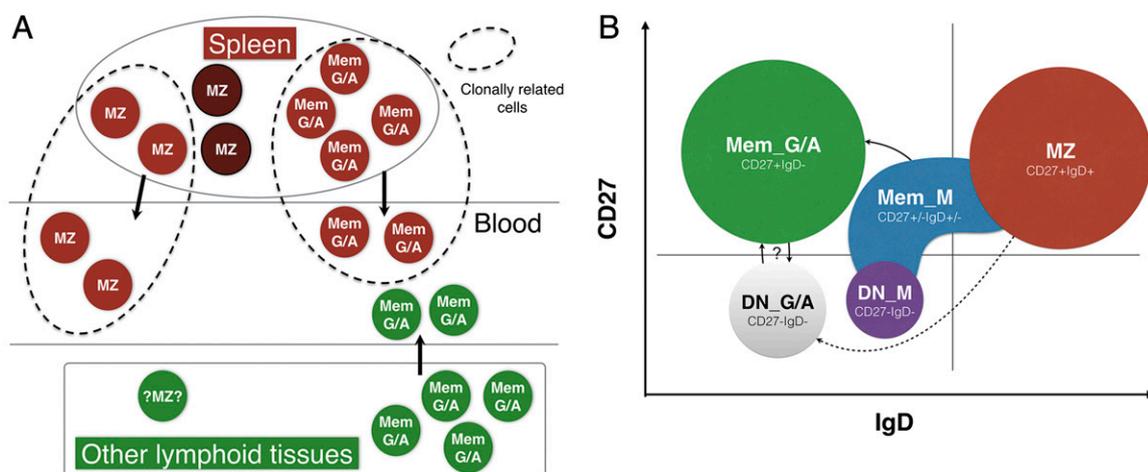


FIGURE 8. Recirculation of spleen MZ and switched memory B cell subsets and clonal relationships between B cell subsets. **(A)** Recirculation between blood and spleen is schematized for MZ and switched memory B cell subsets, as deduced from the repertoire characteristics of clones shared between both tissues (within dotted lines). Brown circles represent splenic and spleen-derived cells and green circles cells from other lymphoid tissues. **(B)** Clonal relationships between subsets. Each subset is represented with a different color, distributed over a theoretical FACS profile for IgD and CD27 expression. Clonal relationships with switched memory B cells are observed for IgM memory B cells, which present the repertoire profile of IgM-only B cells (IgD^-CD27^+) while extending within the MZ and DN gate. MZ B cells only present a few clonal relationships with DN IgG^+ and IgA^+ cells. The switched memory and DN pool have many clones in common, which could correspond either to precursor-product relationships or reversible up- and downmodulation of CD27 expression. Subsets are not drawn on scale because IgM memory B cells represent on average 2–4% of the total B cell pool, switched memory 20% and DN $IgM < 1\%$.

hallmark of autoreactive Igs (20) and broadly neutralizing anti-HIV Abs (21).

Post-GC IgM memory B cells have been documented in mouse and appear to represent a population numerically equivalent to switched ones (1, 22). The issue of their existence, number, and function in humans is, however, more confused. Several reports documented the presence of IgM memory B cells in specific responses, notably in alloimmunized individuals following blood transfusions, but their surface phenotype, IgD⁺ or not, has not been addressed (23). IgM-only B cells are found in AID-deficient patients, who lack switched B cells but have active GC reactions (8, 24). This observation suggested that IgD downmodulation taking place during the GC reaction might be conserved in GC-derived cells that did not isotype switch. However, this observation did not preclude a heterogeneous IgD expression profile. IgM memory B cells, assessed as IgM-only cells, appear in most cases as a minor subset in humans (2–3%), whereas switched memory averages 20% (8). We estimated, based on their distinct mutation profiles, that they may represent up to 15% of MZ clones: these figures would indicate that IgM memory B cells may in fact be broadly distributed over the IgD⁺ and IgD⁻CD27⁺ B cell gates.

In conclusion, the “IgM-only” subset (including cells with its repertoire signature but higher IgD or lower CD27 expression levels) appears as the only subset showing precursor–product relationships with CD27⁺ switched memory B cells, indicating that they represent GC-derived IgM memory B cells. MZ B cells, in contrast, only present a few clones in common with CD27⁻ IgG and IgA cells, involving, in the former case, MZ B cells with a distinct repertoire. Uncovering new surface markers that may allow the unambiguous identification of the MZ and IgM memory compartments will be key toward studying the conditions of their mobilization and better delineating their respective contributions to T-dependent and T-independent immune responses.

Acknowledgments

We thank Nick Chiorazzi for a thorough editing of the manuscript.

Disclosures

The authors have no financial conflicts of interest.

References

- Dogan, I., B. Bertocci, V. Vilmont, F. Delbos, J. Mégret, S. Storck, C. A. Reynaud, and J. C. Weill. 2009. Multiple layers of B cell memory with different effector functions. *Nat. Immunol.* 10: 1292–1299.
- Pape, K. A., J. J. Taylor, R. W. Maul, P. J. Gearhart, and M. K. Jenkins. 2011. Different B cell populations mediate early and late memory during an endogenous immune response. *Science* 331: 1203–1207.
- Balázs, M., F. Martin, T. Zhou, and J. Kearney. 2002. Blood dendritic cells interact with splenic marginal zone B cells to initiate T-independent immune responses. *Immunity* 17: 341–352.
- Wirths, S., and A. Lanzavecchia. 2005. ABCB1 transporter discriminates human resting naive B cells from cycling transitional and memory B cells. *Eur. J. Immunol.* 35: 3433–3441.
- Fecteau, J. F., G. Côté, and S. Néron. 2006. A new memory CD27⁺ B cell population in peripheral blood expressing VH genes with low frequency of somatic mutation. *J. Immunol.* 177: 3728–3736.
- Wu, Y. C., D. Kipling, and D. K. Dunn-Walters. 2011. The relationship between CD27 negative and positive B cell populations in human peripheral blood. *Front. Immunol.* 2: 81.
- Berkowska, M. A., G. J. Driessen, V. Bikos, C. Grosserichter-Wagener, K. Stamatopoulos, A. Cerutti, B. He, K. Biermann, J. F. Lange, M. van der Burg, et al. 2011. Human memory B cells originate from three distinct germinal center-dependent and -independent maturation pathways. *Blood* 118: 2150–2158.
- Weill, J. C., S. Weller, and C. A. Reynaud. 2009. Human marginal zone B cells. *Annu. Rev. Immunol.* 27: 267–285.
- Weller, S., A. Faihi, C. Garcia, M. C. Braun, F. Le Deist, F. G. de Saint Basile, O. Hermine, A. Fischer, C. A. Reynaud, and J. C. Weill. 2001. CD40-CD40L independent Ig gene hypermutation suggests a second B cell diversification pathway in humans. *Proc. Natl. Acad. Sci. USA* 98: 1166–1170.
- Weller, S., M. Mamani-Matsuda, C. Picard, C. Cordier, D. Lecoeuche, F. Gauthier, J. C. Weill, and C. A. Reynaud. 2008. Somatic diversification in the absence of antigen-driven responses is the hallmark of the IgM⁺IgD⁺CD27⁺ B cell repertoire in infants. *J. Exp. Med.* 205: 1331–1342.
- Tangye, S. G., and K. L. Good. 2007. Human IgM⁺CD27⁺ B cells: memory B cells or “memory” B cells? *J. Immunol.* 179: 13–19.
- Seifert, M., M. Przekopowicz, S. Taudien, A. Lollies, V. Ronge, B. Drees, M. Lindemann, U. Hillen, H. Engler, B. B. Singer, and R. Küppers. 2015. Functional capacities of human IgM memory B cells in early inflammatory responses and secondary germinal center reactions. *Proc. Natl. Acad. Sci. USA* 112: E546–E555.
- Seifert, M., and R. Küppers. 2009. Molecular footprints of a germinal center derivation of human IgM⁺(IgD⁺)CD27⁺ B cells and the dynamics of memory B cell generation. *J. Exp. Med.* 206: 2659–2669.
- Wu, Y. C., D. Kipling, H. S. Leong, V. Martin, A. A. Ademokun, and D. K. Dunn-Walters. 2010. High-throughput immunoglobulin repertoire analysis distinguishes between human IgM memory and switched memory B-cell populations. *Blood* 116: 1070–1078.
- Westermann, J., and R. Pabst. 1992. Distribution of lymphocyte subsets and natural killer cells in the human body. *Clin. Investig.* 70: 539–544.
- Dunn-Walters, D. K., M. Hackett, L. Boursier, P. J. Ciclitira, P. Morgan, S. J. Challacombe, and J. Spencer. 2000. Characteristics of human IgA and IgM genes used by plasma cells in the salivary gland resemble those used in duodenum but not those used in the spleen. *J. Immunol.* 164: 1595–1601.
- Cameron, P. U., P. Jones, M. Gorniak, K. Dunster, E. Paul, S. Lewin, I. Woolley, and D. Spelman. 2011. Splenectomy associated changes in IgM memory B cells in an adult spleen registry cohort. *PLoS One* 6: e23164.
- Dono, M., S. Zupo, N. Leanza, G. Melioli, M. Fogli, A. Melagrana, N. Chiorazzi, and M. Ferrarini. 2000. Heterogeneity of tonsillar subepithelial B lymphocytes, the splenic marginal zone equivalents. *J. Immunol.* 164: 5596–5604.
- Bulati, M., S. Buffa, G. Candore, C. Caruso, D. K. Dunn-Walters, M. Pellicanò, Y. C. Wu, and G. Colonna Romano. 2011. B cells and immunosenescence: a focus on IgG⁺IgD⁺CD27⁻ (DN) B cells in aged humans. *Ageing Res. Rev.* 10: 274–284.
- Wardemann, H., S. Yurasov, A. Schaefer, J. W. Young, E. Meffre, and M. C. Nussenzweig. 2003. Predominant autoantibody production by early human B cell precursors. *Science* 301: 1374–1377.
- Kwong, P. D., and J. R. Mascola. 2012. Human antibodies that neutralize HIV-1: identification, structures, and B cell ontogenies. *Immunity* 37: 412–425.
- Taylor, J. J., K. A. Pape, and M. K. Jenkins. 2012. A germinal center-independent pathway generates unswitched memory B cells early in the primary response. *J. Exp. Med.* 209: 597–606.
- Della Valle, L., S. E. Dohmen, O. J. Verhagen, M. A. Berkowska, G. Vidarsson, and C. Ellen van der Schoot. 2014. The majority of human memory B cells recognizing RhD and tetanus resides in IgM⁺ B cells. *J. Immunol.* 193: 1071–1079.
- Weller, S., M. C. Braun, B. K. Tan, A. Rosenwald, C. Cordier, M. E. Conley, A. Plebani, D. S. Kumararatne, D. Bonnet, O. Tournilhac, et al. 2004. Human blood IgM “memory” B cells are circulating splenic marginal zone B cells harboring a prediversified immunoglobulin repertoire. *Blood* 104: 3647–3654.

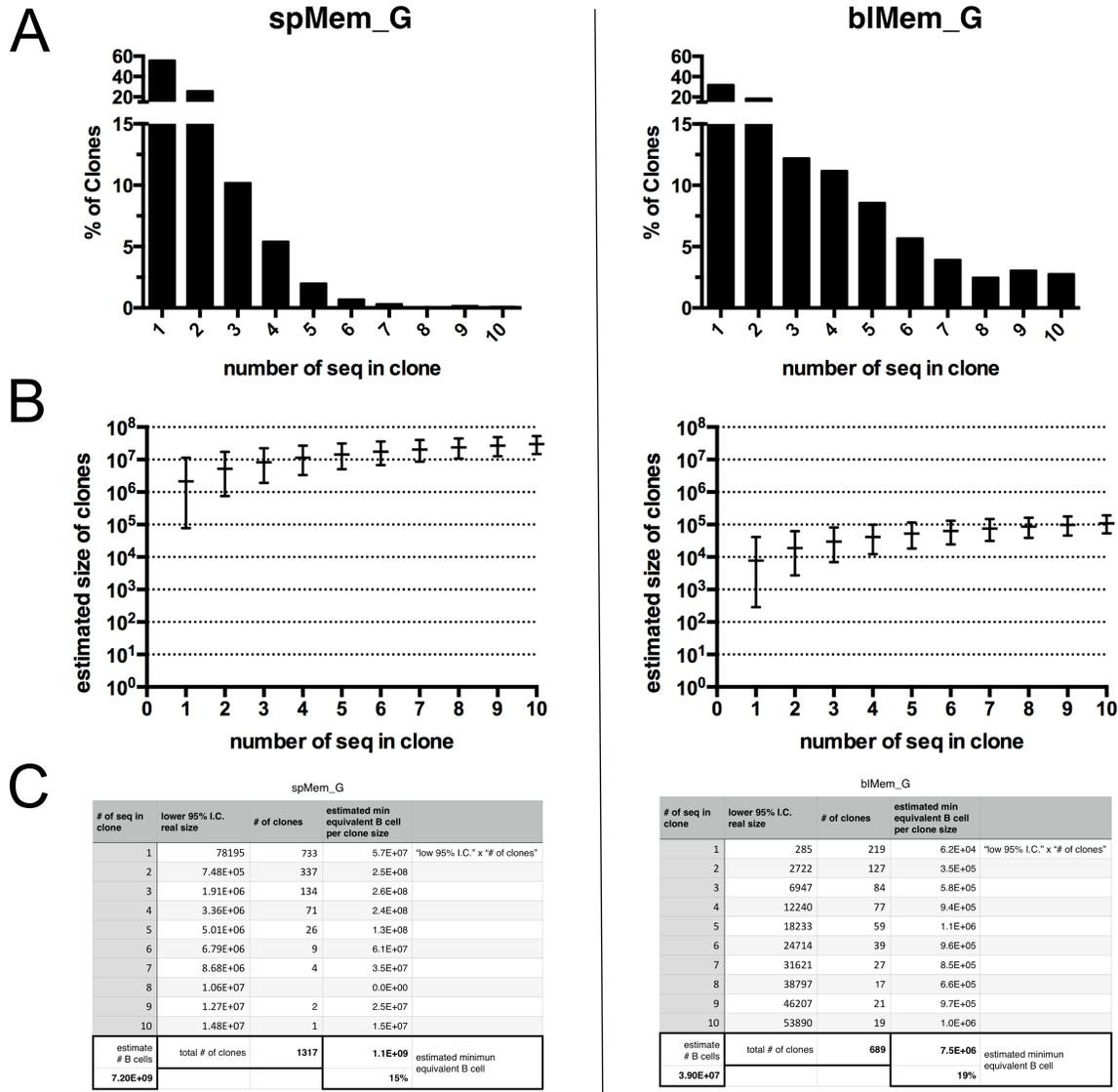


Figure S1. Estimate of clone sizes and fraction of B cell repertoire analyzed .

This figure describes in three steps the calculation performed for one spleen (left part) and one blood (right part) subset chosen as an example:

A. Distribution of clone sizes obtained by CDR3 clustering within a given subset.

B. Using the procedure described in Material and Methods ("Estimate of clone sizes and B cell sampling" section), the values obtained in panel A are extrapolated for the total size of the B cell subset, with error bars representing the 95% interval of confidence of the estimate.

C. The lower 5% values of the estimate of B cell numbers obtained in panel B are tabulated and multiplied by the number of clones in each category. The total number of cells obtained in this calculation, reported to the total size of the B cell subset considered, represents the minimal fraction of B cells within this subset that is clonally related to the sequenced sample and is listed in Table S1 ("minimum equivalent B cell fraction").

Table S1. Estimate size of B cell subsets, number of unique sequence obtained, number of sequenced clones and minimal estimate of the corresponding fraction of the B cell repertoire analyzed for three adult donors.

	estimated population size	number of unique sequences	number of observed clones	minimum equivalent B cell fraction (%)
spMem_G	7.6×10^9	2416	1662	10.7
	2.7×10^9	2319	1309	21.0
	7.2×10^9	2333	1317	15.0
blMem_G	1.2×10^8	1519	615	23.3
	8.0×10^7	2832	1783	18.6
	3.9×10^7	3473	543	39.3
spMem_A	6.8×10^9	2094	1498	10.2
	2.3×10^9	1973	1263	13.3
	6.1×10^9	1608	1184	9.9
blMem_A	1.9×10^8	1636	835	17.6
	1.7×10^8	2741	1488	18.7
	1.2×10^8	2037	484	44.7
spMem_M	1.2×10^9	167	108	11.3
	1.3×10^9	1184	626	16.2
	6.8×10^8	93	72	8.0
blMem_M	7.2×10^7	336	136	23.5
	2.4×10^7	1762	485	36.9
	2.3×10^7	119	51	24.2
spMZ	4.5×10^9	4580	2802	12.5
	1.5×10^{10}	1812	1379	8.0
	1.4×10^{10}	10077	5586	13.0
blMZ	5.4×10^8	2122	1166	14.4
	3.9×10^8	11715	5889	18.7
	5.5×10^7	1547	534	28.0
spDN_G	5.6×10^8	2297	1576	11.6
	4.8×10^8	1477	1008	10.8
	1.5×10^9	2755	1393	16.0
blDN_G	3.6×10^7	453	130	22.3
	2.6×10^7	3553	1257	22.0
	1.4×10^7	686	236	20.7
spDN_A	3.6×10^8	2106	1527	9.6
	1.6×10^8	479	325	11.4
	4.3×10^8	1919	1017	16.2
blDN_A	1.8×10^7	606	134	46.3
	2.6×10^7	2416	1186	28.3
	3.9×10^7	1636	202	60.4
spDN_M	3.3×10^8	460	222	19.2
	1.7×10^8	182	156	5.7
	1.4×10^8	41	39	3.5
blDN_M	4.7×10^7	21	17	6.2
	1.3×10^7	1291	76	13.6
	1.3×10^7	59	24	12.7

Table S2. Number of clones obtained that are either unique to each subset or shared by two different ones*

	biMem_A	spMem_A	biMem_G	spMem_G	biMem_M	spMem_M	biMZ	spMZ	biDN_A	spDN_A	biDN_G	spDN_G	biDN_M	spDN_M
biMem_A	2806	190	34	12	7	8	29	24	472	79	2	2		
spMem_A		3945	16	31	8	8	36	37	67	189	3	3		2
biMem_G			3161	161	5	5	25	22	27	7	36	42		1
spMem_G				4287	3	4	18	19	7	6	22	78	1	
biMem_M					672	43	85	37	3	2			4	13
spMem_M						806	97	56	3		1	1	1	13
biMZ							7589	552	14	6	4	4	2	30
spMZ								9767	8	18	7	5	3	36
biDN_A									1522	44	12	1		
spDN_A										2869	8	17		
biDN_G											1624	91		
spDN_G												3977		
biDN_M													243	8
spDN_M														417

Analysis is performed for each donor separately, and the number of clones obtained in each category is pooled for the three donors

*Numbers in the diagonal represent clones unique to this subset. Numbers outside of the diagonal are clones shared by the two corresponding subsets: since clones can be shared between more than two subsets, the sum of these numbers is bigger than the total number of shared clones.

Table S3. Analysis of statistical significance of repertoire comparisons reported in Figures 5 and 6

			mut		CDR3	VH	JH	
			MZ	Mem_M				
Fig 5A	(Mem_M/MZ)	Mem_M	ns	ns	ns	ns	ns	
		MZ	***	***	ns	***	***	
		Mem_M	Mem_M					
	(Mem_M/DN_M)	Mem_M	ns	ns	ns	ns	ns	
		DN_M	***	***	ns	ns	*	
			MZ	DN_M				
	(MZ/DN_M)	MZ	**	***	ns	ns	***	
		DN_M	***	***	ns	ns	**	
		Mem_M	ns	ns	ns	ns	ns	
	Fig 5B	(MemM/Mem_A-G) ^a	Mem_M	Mem_M	Mem_A-G			
Mem_M			ns	**	ns	ns	ns	
(MZ/Mem_A)		MZ	Mem_M	Mem_A				
		Mem_A	***	***	ns	***	ns	
		Mem_M	***	ns	ns	*	ns	
(MZ/Mem_G)		MZ	MZ	Mem_G				
		Mem_G	***	***	ns	***	ns	
		Mem_M	***	ns	*	**	ns	
Fig 5C		(MZ/DN_A)	MZ	MZ	DN_A			
			DN_A	ns	***	ns	ns	ns
	Mem_M		***	ns	ns	ns	ns	
	(MZ/DN_G)	MZ	MZ	DN_G				
		DN_G	ns ^b	ns ^b	**	ns	ns	
		Mem_M	ns ^b	ns ^b	ns	ns	ns	
		Mem_M	* ^b	ns ^b	*	ns	ns	
	Fig 6A	Mem_A(bl/sp)	bl	***	***	ns	***	ns
			sp	ns	ns	ns	ns	ns
			bl	bl	sp			
Mem_G(bl/sp)		bl	ns	ns	ns	***	ns	
		sp	*	ns	ns	ns	ns	
Fig 6B		Mem_M(bl/sp)	bl	bl	sp			
			sp	ns	ns	ns	ns	ns
			sp	**	ns	ns	ns	ns
	MZ(bl/sp)	bl	bl	sp				
		bl	***	ns	ns	ns	ns	
		sp	***	***	ns	***	ns	

^a statistics are shown for pooled Mem_A and Mem_G samples

^b statistics calculated on the mean mutation frequency and not on the mutation distribution because of the small sample size