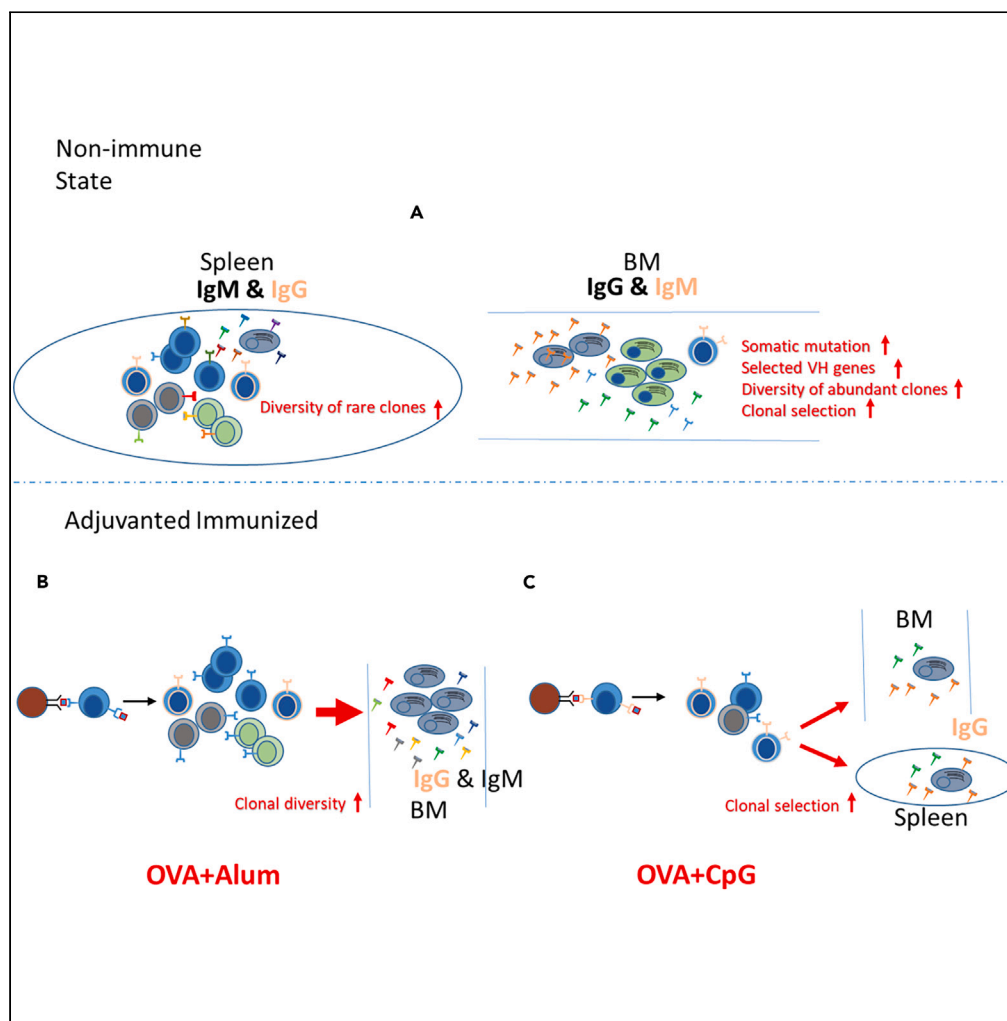


Article

Characterizing adjuvants' effects at murine immunoglobulin repertoire level



Feng Feng, Rachel Yuen, Yumei Wang, Axin Hua, Thomas B. Kepler, Lee M. Wetzler

ffeng@bu.edu (F.F.)
lwetzler@bu.edu (L.M.W.)

Highlights

An IgSeq pipeline was introduced for generating high-quality Ig repertoire data

Murine IGHV repertoires of different tissue and isotype compartments were profiled

At a nonimmune state, repertoires of distinct compartments showed different profiles

Upon immunization, Alum increased clone diversity and CpG promoted clone selection



Article

Characterizing adjuvants' effects at murine immunoglobulin repertoire level

Feng Feng,^{1,*} Rachel Yuen,¹ Yumei Wang,¹ Axin Hua,¹ Thomas B. Kepler,^{1,2} and Lee M. Wetzler^{1,3,4,*}

SUMMARY

Generating large-scale, high-fidelity sequencing data is challenging and, furthermore, not much has been done to characterize adjuvants' effects at the repertoire level. Thus, we introduced an IgSeq pipeline that standardized library prep protocols and data analysis functions for accurate repertoire profiling. We then studied systemically effects of CpG and Alum on the Ig heavy chain repertoire using the ovalbumin (OVA) murine model. Ig repertoires of different tissues (spleen and bone marrow) and isotypes (IgG and IgM) were examined and compared in IGHV mutation, gene usage, CDR3 length, clonal diversity, and clonal selection. We found Ig repertoires of different compartments exhibited distinguishable profiles at the non-immunized steady state, and distinctions became more pronounced upon adjuvanted immunizations. Notably, Alum and CpG effects exhibited different tissue- and isotype-preferences. The former led to increased diversity of abundant clones in bone marrow, and the latter promoted the selection of IgG clones in both tissues.

INTRODUCTION

With the advent of technological advancements in the last two decades, immunoglobulin sequencing (IgSeq), allowing for sampling immunoglobulin variable region genes (IgVRG) at an unprecedented depth, has proved invaluable to study adaptive immune responses in health¹ and disease,^{2,3} and for diagnostic⁴ and therapeutic⁵ purposes. However, generating large-scale, high-fidelity, and high-quality sequencing data is challenging^{6,7}; there is room for further improvement in methods for both data generation and data analysis.^{8,9}

Adjuvants are critical in almost all vaccines to increase antigen immunogenicity and enhance the magnitude and durability of the response. Despite their extensive use in vaccines for more than a century, the understanding of adjuvants' action mechanisms remains partial.¹⁰ As a result, only very few adjuvants have been licensed for human use. Among them are aluminum (Alum) salts, the first and most commonly used adjuvant. Alum adjuvants are particulates and induce antibody and CD4⁺ T helper cell responses.¹⁰ It has long been accepted that Alum exerts its effects through creating an *in situ* depot, which as a delivery system allows for slow release of antigens from the site of immunization and enhanced antigen uptake by DCs.¹¹ However, recent reports suggested that a short-term depot was not necessary for the effect of aluminum adjuvants.¹² Alum also, potentially, enhances the immune response through several other mechanisms. It induces tissue release uric acid, which activates the intracellular NLRP3 inflammasome and increases pro-inflammatory IL-1 β secretion.^{10,11} However, the data are contradictory regarding the importance of the NLRP3 pathway in Alum's adjuvant effect.^{10,11} Moreover, Alum can increase recruitments of various immune cells, including DCs, natural killer cells, and IL-4 secreting eosinophils which in turn promote the Th2 responses commonly seen with the use of Alum.^{10,13} Unmethylated deoxycytidyl-deoxyguanosine (CpG) oligonucleotide ligands are another group of adjuvants used in licensed vaccines. By mimicking bacterial DNA signatures, CpG triggers cells expressing the toll-like receptor 9 (TLR9). There are at least four different types of CpG, distinguishable by their nucleotide sequences and effects on immune responses.¹⁴ Yet, they all influence many immune cells, especially DCs and macrophages. CpG enhances pDC migration and activates them to produce type 1-IFN plus a few other cytokines. It also boosts B cell differentiation, proliferation, and antibody production (such as IgM).^{14,15} Lastly, CpG directly or indirectly stimulates NK cells and Th1 CD4⁺ cells, resulting in NK activation and an improved antigen specific CD4 and CD8⁺ T cell response.^{10,14,15} Again, despite these recent insights, the immunological mechanism and characterization of adjuvant effects is still incomplete. Furthermore, the adjuvant effects on the immune repertoire are not well studied.

B cells in different tissues exhibit great functional and phenotypic heterogeneity. The spleen, the largest secondary lymphoid organ, plays a critical role in bridging the innate and adaptive immunity,¹⁶ and makes up in a large portion the peripheral naive B cell pool, including transition B cells, follicular B cells, marginal-zone B cells.¹⁷ It is also a major hub for long-lived memory B cells¹⁸ and possibly some antibody-secreting plasma cells, among which are long-lived,¹⁷ in the spleen red pulp. On the other hand, bone marrow (BM) is both a primary lymphoid organ for B cell development and maturation and also a vital long-term niche for terminally differentiated plasma cells^{19,20} plus other types of

¹Department of Microbiology, Boston University, Boston, MA 02118, USA

²Department of Mathematics and Statistics, Boston University, Boston, MA 02118, USA

³Department of Medicine, Boston University School of Medicine, Boston University, Boston, MA 02118, USA

⁴Lead contact

*Correspondence: ffeng@bu.edu (F.F.), lwetzler@bu.edu (L.M.W.)

<https://doi.org/10.1016/j.isci.2023.108749>



immune cells. Such a specialized niche is not static but dynamically responds to physiological and pathological challenges and, in turn, regulates B cell lymphopoiesis and plasma cell survival/turnover.^{19–22} In addition, B cells of different isotypes also represent heterogeneous cell subsets with distinct intrinsic properties in terms of their location, origin, developmental stage, differential potential, and effector function. Recent work has reported changes in the IGH repertoire of B cells of different isotypes with metrics such as gene usage, mutation, and clonality.^{2,23,24} Thus, it would be wise to assume B cells of varying tissues and isotypes to form distinct compartments.

The B cell Ig repertoire has, by some estimates, the theoretical potential to comprise up to 10^{14} unique molecules in humans²⁵ and 10^{13} in mice.²⁶ Such a diverse repertoire is essential to achieve immunity against possible encountered antigens during one's lifetime. Traditionally, B cell responses were evaluated by assessing the quantity and function of one or a few specific antibodies in individuals. Unfortunately, those measures reflect little about changes at the B cell repertoire level. Characterization of total repertoire changes has proven insightful for understanding vaccine responses.^{2,3,24} In this work, we immunized mice with ovalbumin (OVA) mixed with either Alum or CpG to systemically study the effects of these adjuvants on the Ig repertoire. We developed a standardized Ig heavy chain library prep protocol and an improved data processing pipeline to generate high-fidelity and high-quality IgSeq data. Heavy-chain repertoires were profiled for four different tissue-isotype compartments from mice under varying immunization conditions (PBS, OVA alone, OVA+CpG, OVA+Alum). We examined sequence mutation frequency, IGHV usage, and CDR3 length of these Ig repertoires. In addition, we developed two metrics to characterize the Ig repertoire at the clone level, rescaled Hill numbers for clone diversity and clone selection strength. We found that Ig repertoires showed statistically significant alterations in these properties among different compartments at the unimmunized steady state and further diverged after immunization. Finally, we characterized the Ig repertoire profiles of different tissues, isotypes, and adjuvant immunizations by unsupervised machine learning approaches. We found that OVA immunization with alum or CpG gave rise to repertoires distinct from each other and from those under immunization with unadjuvanted OVA, and the effects displayed different tissue and isotype preferences. Our data suggested possible action mechanisms of adjuvants: Alum resulted in increased diversity of abundant clones in BM repertoires of both isotypes; CpG led to the increased selection of IgG clones and affected BM and spleen compartments similarly.

RESULTS

IGH sequencing library prep and data preprocessing pipeline

For library prep, we adopted the 5' rapid amplification of cDNA ends (5'RACE) approach (Figure 1B),^{27–33} which is simple, unbiased, and requires only minimal amounts of material.²⁷ We incorporated in the protocol many features that had been proved to improve the quality and accuracy of RNA sequencing. In the cDNA reverse transcription step, we added 5' biotin blockers to the template-switching oligos (TSO)³⁴ and ran reactions with cycles at a higher temperature (50°C).²⁹ The former was to reduce the background concatemer formation ("hedgehog effects"), and the latter was to promote the unfolding of RNA secondary structure for a better yield of cDNA products. Moreover, the TSO contained a unique molecular identifier (UMI) that can be used at the data analysis step to correct PCR bias.³⁵ The UMI contains spacers with either fixed nucleotides or a partial degeneracy (Figure 1D). Such design aimed to reduce artifacts in library prep such as non-specific amplification, strand invasion, and primer-dimer.^{35–38} At the PCR amplification step, isotype-specific primers also contain at the 5-prime end degenerated nucleotides (7nts long), which in design increases the complexity of the read-2 sequences.³⁹ That is necessary since the read-2 sequences start at the junction between the IGHJ segment and the constant region, and both parts consist of low diversity/complexity sequences (Figure S2). Increasing diversity at the beginning of the read-2 resulted in a significantly higher quality of sequencing (Figure S1; Table S1). Lastly, we incorporated the unique dual indexes (distinct pairings) for sample multiplexing to avoid the indexing hopping.⁴⁰ Both amplification and adaptor ligation PCRs were monitored in real-time and terminated when PCR products reached a defined threshold to avoid over-amplification.

The data preprocessing pipeline were developed to best utilize the features incorporated in the library prep method for increasing quantification accuracy (Figure 1C). Typical IgSeq data analysis pipelines require the assembly of paired reads (pair-read joining).^{9,41,42} Sequence lengths, however, vary depending on factors such as primer location, V(D)J segment usage and recombination, the inclusion of leader regions, etc.^{35,43} Also, the assembly of two reads typically requires an overlap between them. Filtering out reads that fail to overlap leads to a bias toward short Ig and CDR3 sequences (Figure S3). Therefore, we developed a pipeline that does not require overlap-based joining. One immediate benefit was more usable reads obtained through the new pipeline (Figure 1E). On the same sequencing dataset, pair-read joining excluded ~85% of raw reads. In addition, pair-read joining relying on a short sequence overlap between paired reads might result in erroneous assembly. These assemblies resulted in a lower yield in the Ig filtering step (60% Ig genes for pair-joining data vs. 90% Ig genes for the new pipeline, Figure 1E). After the UMI deduplication step, pair-end joining resulted in $\sim 5 \times 10^3$ unique Ig sequences per sample, and the processing with no pair-end joining yielded $\sim 5 \times 10^4$ unique Ig sequences each. Lastly, the new pipeline allowed us to identify, on average, about 2.3×10^3 clones per sample, which was 8-fold more than the number of clones obtained with the pair-end joining processing (Figure 1E).

In the second step of the pipeline, we filtered out the non-Ig genes using bowtie2 with an Ig gene segment database.⁴⁴ Libraries prepared with 5'RACE can result in non-specific amplification of non-Ig genes that may skew downstream analyses, this step eliminates these off-target reads. Next, we ran the UMI deduplication to correct sequencing and PCR errors/bias. Our deduplication method is specifically suited to IgSeq, as follows. Each UMI is extended to include the first 25 nucleotides of the sequence and UMIs are merged based on a directional network-based algorithm⁴⁵; multiple sequence alignments are used to identify sequence mismatches within each UMI group; groups with more than 5% mismatches or belonging to different isotypes are split; UMI groups with one single sequence are discarded; the consensus sequence for each UMI group is generated taking quality scores into account. All these steps were designed to correct sequencing and PCR errors (such as index hopping, non-specific amplification, strand invasion, and primer-dimer^{35–38}) and increase quantification accuracy.

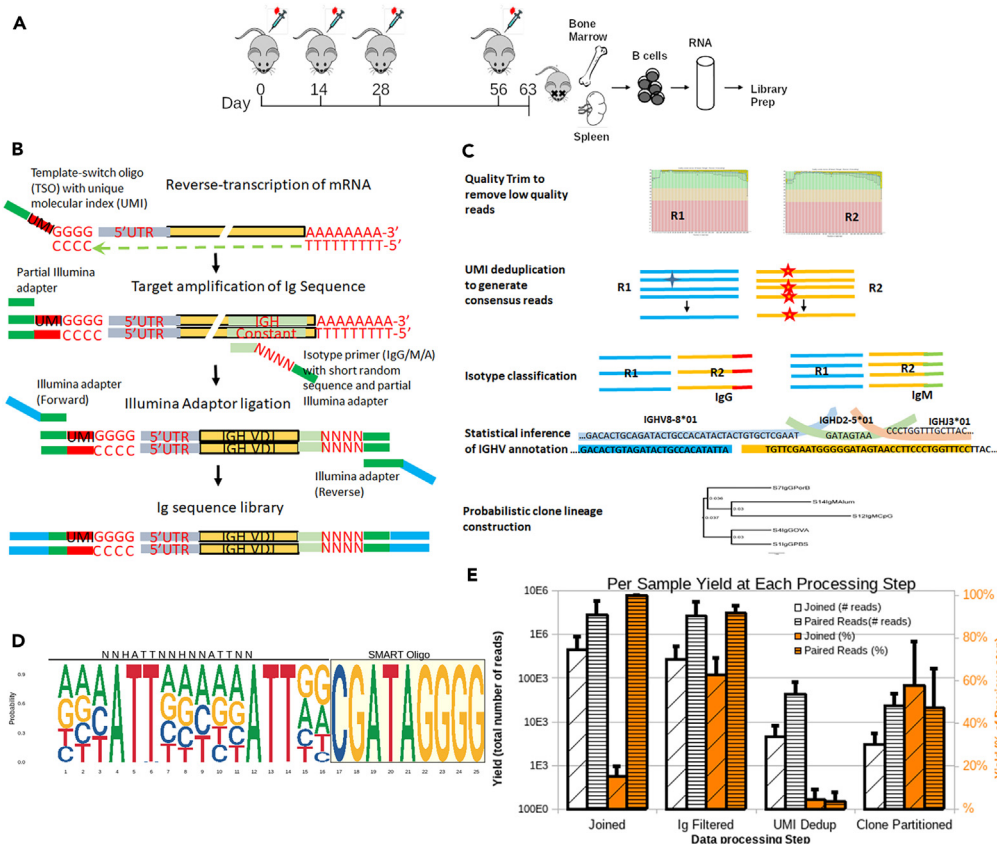


Figure 1. Sequencing library preparation protocol and data preprocessing pipeline

We standardized the library preparation and data preprocessing pipeline by incorporating features and steps that had been proved to improve the quality and accuracy of RNA sequencing.

(A) Each group of three mice received four doses of immunizations (PBS, OVA, OVA+CpG and OVA+Alum). One week after the fourth injection, spleen and bone marrow were harvested. Splenic B cells and bone marrow cells were prepared and total RNA was isolated.

(B) The 5'RACE library prep protocol started with the TSOs containing 5'-biotin blockers and UMIs including spacers and RT-PCR cycles running at a higher temperature. The target Ig amplification PCR was achieved with a reverse isotype-specific primer containing short random nucleotide to increase the sequence diversity in the beginning of Read 2 sequencing. Both PCRs were monitored at real time to avoid over-amplification. Pair end sequencing was done on an Illumina HiSeq2500 system.

(C) The data preprocessing pipeline does not require pair-read joining, and in all steps separated reads were used as input. We also developed and implemented new algorithm to de-duplicate UMI group reads and generate consensus sequences. Ig heavy chain VDJ annotation and clonal analysis were carried out using the Cloanlyst software.

(D) The sequence logo plot for the TSO with UMIs. The UMI is different from a typical one by including spacers with reduced degeneracy.

(E) Sequence yield of each sample at different steps between the pair-end joining protocol and non-joining pipeline in this work. Open bars are for the sequence yields of the typical pair-joining protocol and filled bars are for the non-joining separated reads pipeline. White bars are yields in number of sequences and orange yields in percentages. In the end, with the non-joining pipeline on average we obtained about 10-fold more sequences and 8-fold more clones for each sample. (Data are represented as mean \pm SEM).

Our approach produced isotype-aware sequences. We implemented an in-house C++ program to separate different isotypes. It is alignment-based and allows mismatches. Recombination parameter estimation and clonal partitioning were done using our software, Cloanlyst, which uses explicit likelihood models for V(D)J recombination and somatic mutation.⁴⁶

In summary, we standardized library prep and data preprocessing pipeline for high-quality repertoire sequencing. With that, we prepared IGH repertoire libraries from spleens and bone marrows of 15 mice and then obtained 84 million raw paired reads on sequencing. These reads yielded 1.2 million deduplicated Ig sequences assigned to 667,000 clones, which belong to 60 samples of four different tissue-isotype compartments (spleen IgM, spleen IgG, BM IgM, and BM IgG).

Adjuvants impacted IGHV mutation frequency in a tissue- and isotype-dependent fashion

Somatic hypermutation (SHM), the driver of affinity maturation, has been used as a surrogate for the amount of antigen stimulation that a B cell clone has received during development.^{8,47} Here, we determined the influence that adjuvants have on the level of somatic hypermutation in

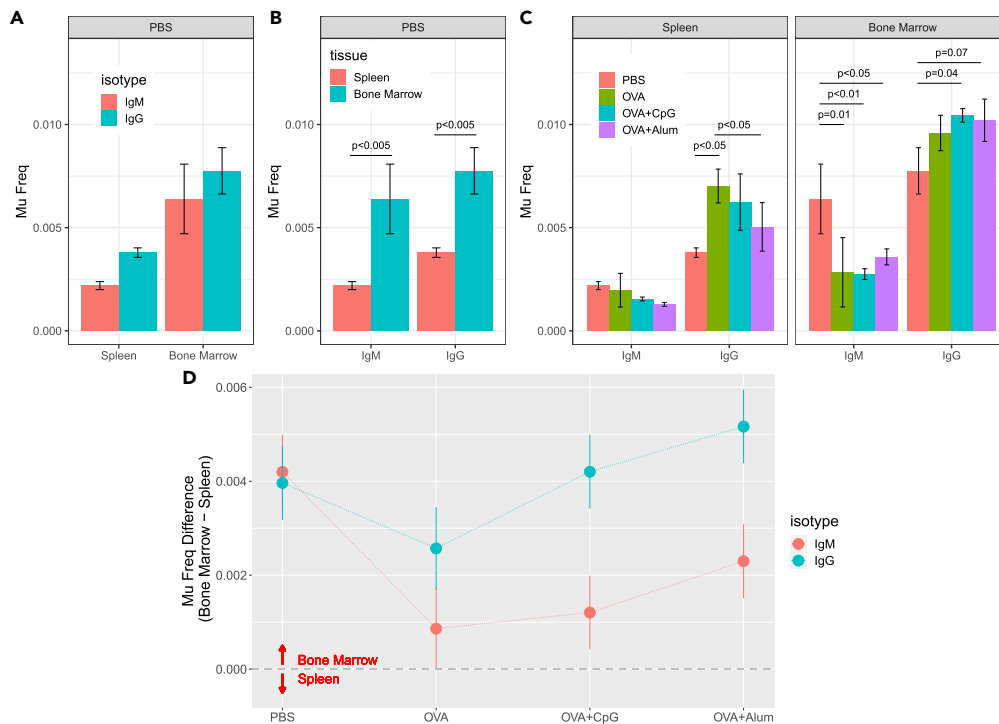


Figure 2. IGHV mean mutation frequency is affected by OVA immunizations and adjuvants in a tissue- and isotype-dependent fashion

IGHV annotation was done by aligning to germline sequences with the software tool, Cloanlyst. Mutations were identified and expressed as per-nucleotide mutation frequency. The distributions of IGHV mutations and means and standard deviations were obtained and used for statistical tests.

(A) IGHV mutation frequencies in different isotype compartments in non-immunized mice (PBS controls).

(B) IGHV mutation frequencies in different tissue in non-immunized mice (PBS controls).

(C) IGHV mutation frequencies were affected by OVA immunization and the effects were isotype- and tissue-dependent.

(D) Mean mutation frequency difference between bone marrow and spleen. The difference was calculated as the BM mutation frequency subtracted by that of spleen for each OVA immunization group of different isotypes. A positive value means the BM compartment has a higher mutation frequency than the respective spleen compartment, a negative value means higher mutation frequency in the spleen, and zero means identical mutation frequency between two tissues. (Data are represented as mean \pm SEM. Statistical p values were corrected by FDR and significant p values were indicated in figure).

different tissue-isotype compartments following immunization with OVA with or without CpG or alum. At the steady state (PBS controls), we observed altered mutation frequencies in different isotype or tissue compartments (IgG vs. IgM or BM vs. spleen). The IgG compartments in both tissues exhibited higher mean mutation frequencies than the IgM compartment, although no statistical significance was achieved ($p > 0.05$, Figure 2A; Figure S4A). Such a trend has been seen before.^{48–50} The linking between isotype and hypermutation was likely a result of a common mechanism involving activation-induced cytidine deaminase (AID).^{51,52} Upon comparing tissue differences, BM sequences exhibited a significantly higher mean mutation frequency than splenic sequences independent of isotypes (Figure 2B, $p < 0.005$). We expected such changes since the spleen, including its marginal zone, is a major reservoir for naive and memory B cells^{18,53,54}; BM provides a preferred niche for plasma cells derived from the late germinal center output and enriched for high-affinity and highly mutated Ig sequences.^{53,55,56}

Next, we observed that an OVA immunization (either alone or adjuvanted) increased the mean mutation frequency of IgG sequences but decreased that of IgM sequences compared with the PBS controls (Figure 2C). This might suggest an effect of OVA immunization on promoting IgMs with high mutations to switch into IgGs. In addition, immunization effects further exhibited a tissue dependency (Figure 2C). OVA alone yielded a higher mean mutation frequency in the spleen IgG compartment ($p < 0.05$ for both OVA vs. PBS and OVA vs. OVA+Alum), while OVA immunizations with adjuvant (either Alum or CpG) led to a higher mean mutation frequency in BM IgG ($p < 0.05$ for OVA+CpG vs. PBS and $p = 0.07$ for OVA+Alum vs. PBS).

Furthermore, Alum and CpG impacted the tissue-related mutation accumulation differently. There was a pre-existing tissue-related mutation frequency discrepancy, with BM consisting of sequences of a higher mean mutation frequency than that of spleen sequences independent of isotype (Figure 2A; Figure S4B). Upon immunizations in the IgG compartment, OVA alone decreased such a discrepancy by preferentially increasing mutations of splenic sequences (Figure S4B), CpG maintained it by increasing BM and spleen sequence mutations similarly ($p < 0.005$, Figure S4B), and Alum enlarged it by accumulating more BM mutations ($p < 0.005$, Figure S4B). On the other hand, in the IgM compartment, immunizations reduced the mutation frequency difference between tissues by significantly decreasing the mutations in the BM IgM compartment (Figure S4B). Such tissue-dependent immunization effects became more evident when we plotted the frequency discrepancy between BM and spleen across varying immunization conditions (Figure 2D).

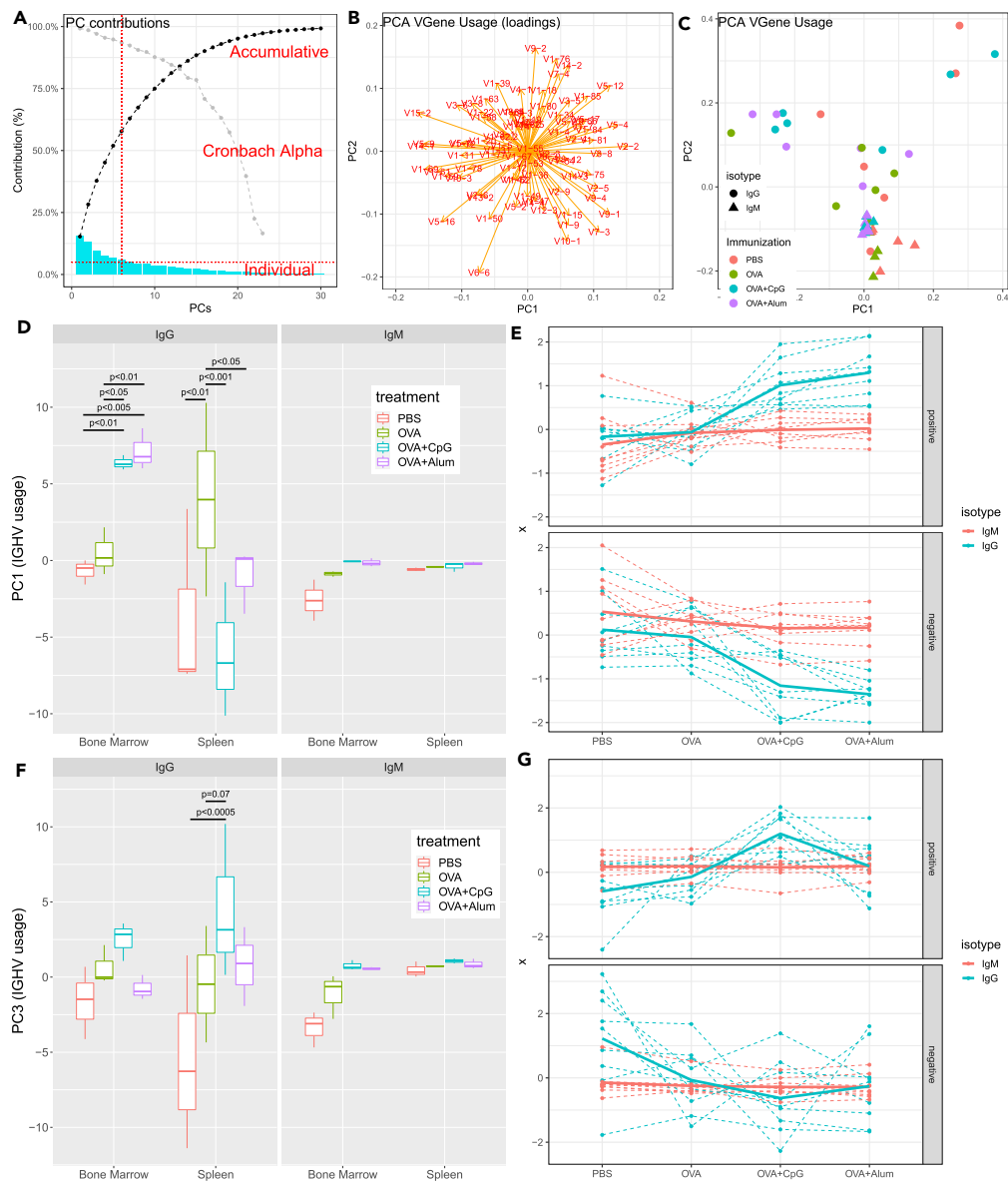


Figure 3. OVA immunization affects IGHV gene usage and adjuvants modify the tissue distribution of the effects

IGHV genes were annotated and gene usage was expressed as the percentage of each IGHV gene segment in each sample. The data were transformed into compositional data format. PCA was applied to reduce data dimensions and group correlating gene usages. Statistical analyses were done on top PCs with large contributions.

(A) The PCA scree plot showed the individual and accumulative PC contributions. The top six PCs with largest contributions (each explaining >5% total variation) have an accumulative contribution of about 58%. We also plotted the Cronbach Alpha value of each PC (gray dashed line), which is a measure of internal PC consistency.

(B) The PC loading plot. It showed correlations of IGHV gene usages projected on the top two PC plane. Genes are highly correlated if they point to the same directly, and contribute more to a PC component if they have larger loadings.

(C) Sample distribution on the first two PC space. The samples were labeled by their isotype (point shape) and immunization group (color).

(D) The PC1 distributions by different isotype, tissue and immunization status. PC1 explained about 15% of total variations and showed a significant immunization effect in a tissue- and isotype-dependent way.

(E) The raw IGHV gene usage patterns to verify the PC1 distribution. We plotted the top 10 raw gene usages (centered by the mean of each group) with largest contributions to PC1. Both positively (top) or negatively (bottom) contributing gene usages were shown.

(F) The PC3 distributions by different isotype, tissue and immunization status. PC3 explained about 9.5% of total variations and showed a significantly different pattern induced by the OVA+CpG immunization in the spleen IgG compartment.

Figure 3. Continued

(G) The raw IGHV gene usage patterns to verify the PC3 distributions. We plotted the top 10 raw gene usages (centered by the mean of each group) with largest contributions to PC3. Both positively (top) or negatively (bottom) contributing gene usages were shown. (Data are represented as mean \pm SEM. Statistical p values were corrected by FDR and significant p values were indicated in figure).

Overall, our data suggested: there were tissue- and isotype-related mutation frequency discrepancies at the steady-state, with higher mutations found in BM and IgG compartments; immunizations increased the mean mutation frequency of IgG sequences but decreased that of IgM sequences; OVA alone affected sequence mutations mainly in the spleen; the CpG immunization impacted mutations of BM and spleen sequences similarly; Alum more efficiently increased mutations of BM IgG sequences.

IGHV gene usage and CDR3 length showed a tissue- and isotype-related difference and affected by adjuvanted immunizations

IGHV gene segment frequencies have been applied to distinguish stages of immune responses, describe age-related alterations, and define general B cell populations.^{57–59} In this work, instead of identifying altered frequencies of individual gene segments, we were interested in discovering correlated gene segment usage changes under different conditions. These changes could reflect a common or convergent development of Igs driven by biological or immunological regulators. The analysis started with identifying germline VDJ genes by a maximum likelihood algorithm developed previously by our group.⁴⁶ It is alignment-based and uses a gene-segment library derived from the IMGT.⁶⁰ Germline VDJ genes were annotated for each consensus sequence, and counts of different IGHV segments in each sample were obtained as the first step of the analysis. Then counts were transformed into compositions, an appropriate and required format for high-throughput sequencing data in microbiome studies.^{61,62} Next, we performed the principal component decomposition of the CLR transformed data. Principal component analysis (PCA) proved effective in capturing genuine biological differences among samples by relating genes with similar expression pattern.⁶³ We hypothesized that PCA of IGHV gene usage data produces a limited number of robust principal components, and many of them reflect the biological/immunological factor effects.

PCA on the IGHV gene usage data identified a group of six robust principal components (PC), each of which explained more than 5% of the total variations and altogether contributed \sim 58% of the total (Figure 3A; Figure S5). These PCs also had large Cronbach Alpha values (>0.75), which indicate a high internal consistency as a group. By relating IGHV genes of similar patterns, these PCs revealed underlying effects or regulations by factors such as immunization status, tissue location, or isotype identity (Figures 3B and 3C; Figure S6). PC1, explaining 15% of the total variations in IGHV gene compositions, reflected mainly an immunization effect (Figure 3D). The immunization effect was also tissue- and isotype-dependent ($p = 0.033$ for the three-way interaction test in ANOVA). OVA alone affected PC1 mainly in the spleen IgG compartment, while OVA plus adjuvants (both CpG and Alum) resulted in significant changes of PC1 in the BM IgG compartment. Such a differential IGHV usage in different compartments upon adjuvanted immunizations indicated a possible mechanism underlying adjuvants' effect, promoting homing of specific B cells to the BM niche. Furthermore, such an effect was most likely secondary during the immune response since the changes were limited to the IgG compartment. The top contributing IGHV segments to PC1 were highly correlated and almost exclusively contributed to PC1 (Figure 3E; Figure S7). They contained both up-regulated (IGHV2-2, IGHV5-4, IGHV5-12, etc.) and down-regulated gene segments (IGHV15-2, IGHV14-1, IGHV5-9, etc.) (Table S2).

Similarly, we found that PC2 (explaining 13% of the total variations) exhibited statistically significant isotype, tissue, and immunization-related changes. First, we found significantly different PC2 values between the IgG and IgM compartments (Figures S8A–S8C). The significant PC2 difference at the steady state indicated a naturally biased usage of this group of IGHV segments between different isotype compartments. Such isotype-related usage bias reached the maximum in the spleen at the steady state and became diminished upon immunizations or BM relocation (Figure S8C). In addition, the immunization effects on PC2 values were statistically significant only in the IgG compartment. The OVA+Alum immunization resulted in significantly increased PC2 values in BM IgG ($p < 0.001$) but decreased PC2 values in spleen IgG ($p < 0.001$, Figures S8C and S8D) compared with PBS controls and OVA alone samples. On the other hand, OVA+CpG only led to increased PC2 values compared with the PBS and OVA alone groups in the BM IgG compartment. Taken together, PC2 included a group of IGHV genes (Figure S9) whose usages are normally biased between different isotypes and were impacted distinctly by the Alum or CpG immunization in different compartments.

Lastly, the PC3 group of IGHV genes (explaining 9.5% of total variations) demonstrated a different treatment effect induced only by the CpG immunization. In this case, OVA+CpG immunization resulted in a significant impact on IGHV usage in the spleen IgG compartment compared with other groups (Figures 3F and 3G). Changes were also seen in the BM IgG compartment but without significance. The top contributing IGHV segments of PC3 consisted of up-regulated and down-regulated gene usages (Figure S10; Table S2).

CDR3, the most variable part of an Ig, is the principal determinant of the specificity and affinity of an antibody.⁶⁴ Its length distributions and amino acid properties have been applied to characterize functional Ig repertoires, e.g., long CDR3s in the case of broadly neutralizing antibodies against HIV.^{65,66} Here, we investigated how OVA immunization and adjuvants affected IGH CDR3 lengths. Interestingly, CDR3 length changes were similar to the IGHV gene segment usage results. First of all, statistically significant differences were limited to the IgG compartments upon immunization. OVA with either Alum or CpG resulted in significantly longer CDR3s than PBS or the OVA alone group in the BM IgG compartment ($p < 0.05$, Figure 4), and OVA+CpG also led to an increased mean CDR3 length in the spleen IgG compartment ($p < 0.005$, Figure 4). As a result, under the OVA+Alum condition, we observed significantly longer IgG CDR3s in BM than those in the spleen. But in the case of OVA+CpG, we found equally long CDR3s in IgGs of both tissues. Lastly, CDR3 lengths in IgM compartments were not affected

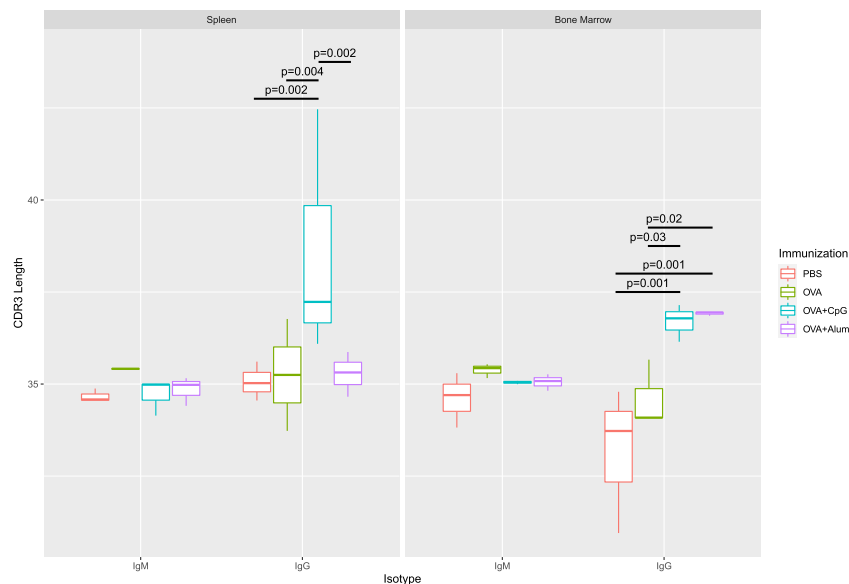


Figure 4. CDR3 length is affected by OVA adjuvant immunizations in a tissue- and isotype-dependent fashion

IGHV genes were annotated and CDR3 regions was identified. The CDR3 length distribution of each sample was determined, and the mean CDR3 lengths were compared and plotted by different compartments and immunization status. (Data are represented as mean \pm SEM. Statistical p values were corrected by FDR and significant p values were indicated).

significantly by OVA immunizations. Again, these changes of mean CDR3 length were similar to the IGHV gene usage alterations (Figure S11) but showed no correlations to the sequence mutation frequency (Figure S12). One thing also worth mentioning was that at the steady state (PBS controls), CDR3s were shorter in IgG than in IgM compartments ($p < 0.05$ for the IgG vs. IgM comparison in BM, Figure S11B), as were reported previously.²

Ig repertoires of different tissue-isotype compartments displayed distinct clone diversity profiles and CpG and Alum affected them differently

Diversity is a frequently used metric in ecology describing both the species variety and the relative abundance in an assemblage. Hill numbers, also known as the effective numbers of species, have been employed increasingly to quantify and assess biological diversity on the basis of sampling data.⁶⁷ They incorporate a mathematically unified group of diversity indices, differing among themselves only by an order q that defines the sensitivity of the number to the abundance or relative abundance of species (or clones in our work). When the order is small, diversity values are computed by weighting rare (small) clones more. As the order increases, diversities are calculated by weighting abundant (big or dominant) clones more.⁶⁸ Diversity of order zero, 0D , is insensitive to the clone size/abundance and referred to as the clone richness (i.e., total number of clones), 1D is the well-known Shannon entropy, and 2D is the Simpson concentration or the inverse Simpson index. In practice, 1D is considered a measure of the effective number of common clones in a compartment, and 2D or higher order diversities are of dominant clones.⁶⁹ In this study, we worked on diversities with orders q between 0 and 4 since the diversity profile changes little beyond $q = 3$ or 4.⁶⁹

Clone abundance data, generated by the maximum likelihood estimation of clone lineages using Cloanlyst,⁴⁶ were used for estimating the asymptotic diversity Hill numbers. One concern about Hill numbers is that sample size affects the estimation: incomplete sampling underestimates the diversity values of low orders.⁶⁹ By resampling the clonal data at varying sample sizes, we showed that the diversity estimates are sensitive to sample size at low orders $q < 1.5$ (Figure S14). To overcome such bias, we determined the lowest sequence depth among all samples and then ran random subsampling 100 times of each sample with that number to estimate the diversity profile. With these profiles, we then started studying factor effects. First, upon comparing them between isotypes, we found that clone diversities of IgM compartments were much higher than those of IgG compartments at all orders, $q = 1$ to 4 (Figure 5A; Figure S15). However, when checking the clone size distributions, we found something contradictory. Most IgM compartments consisted of small clones with 1 or 2 members, but IgG compartments contained a significant proportion of larger clones (e.g., 15% of the clones in the spleen IgG compartment consist of more than five members) (Figure 5B; Figure S13). Intuitively, we expected the IgG compartment to have more abundant clones, i.e., higher effective numbers or greater diversities at higher orders. The contradiction arose because diversity depends on two parameters: clone richness and clone evenness (relative abundance/size of clones); the diversity profile of the IgM compartment was more skewed by the high clone richness than it was in the IgG compartment. To control the effect of clone richness, we rescaled the diversity profile by dividing by its clonal richness (Figures 5C–5E). This way, we control for the effect of clone richness and make all diversity profile curves starting from the same point, ${}^0D = 1$. In addition, the rescaled diversity value at the order of 1 (1D) equals to the exponential of the Pielou evenness index.

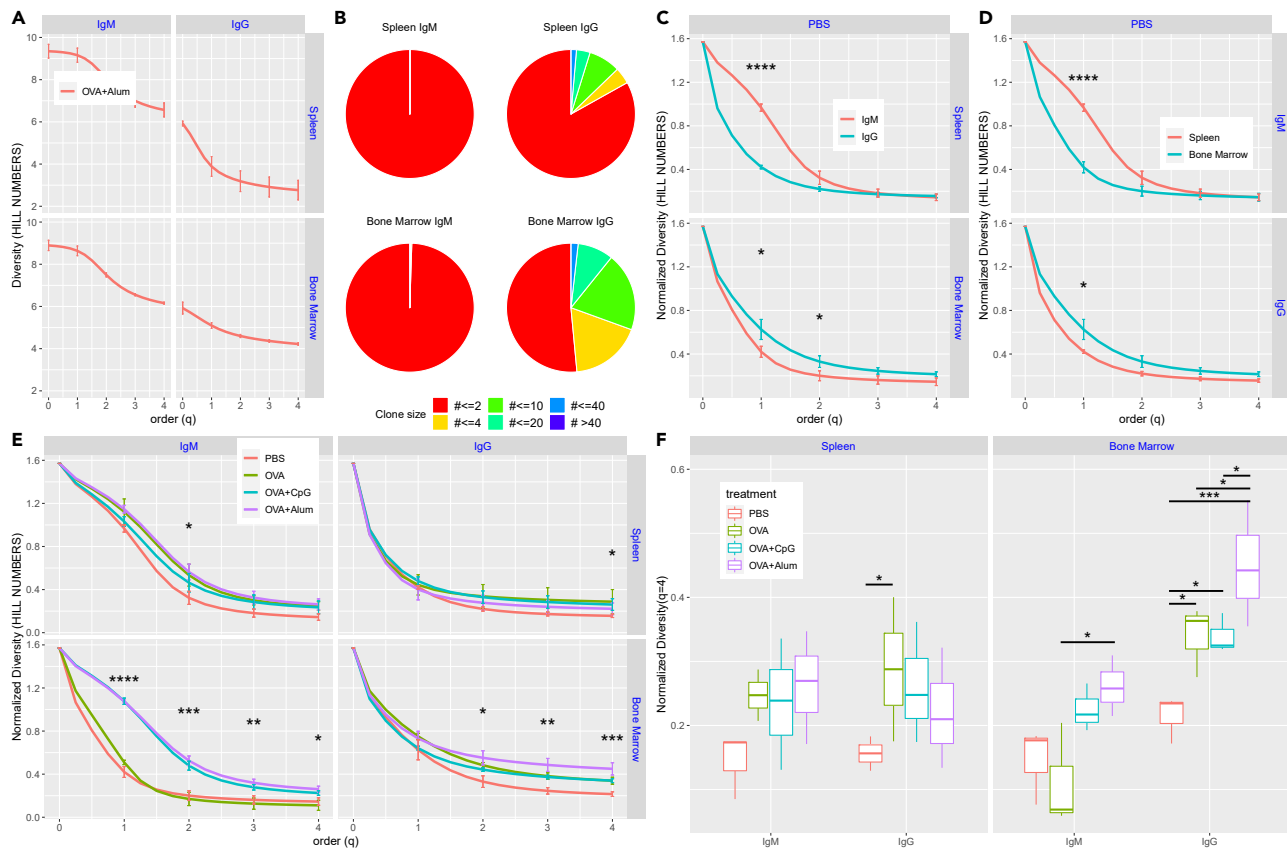


Figure 5. Clonal diversity as Hill numbers

Ig sequence data were annotated and partitioned into clones. Diversity profiles as Hill numbers were estimated based on clone abundance data with orders between 0 and 4. They were also scaled to control the effect of clone richness and focus on clone evenness (relative clone abundance/size).

(A) Diversity profiles of different isotype compartments under the OVA+Alum immunization. Each line is for one isotype compartment in different tissue. Standard errors are plotted for diversities at the order of 0, 1, 2, 3 and 4.

(B) Example clone size distribution pie charts. Here, it shows the data for one mouse immunized with OVA+Alum, and results for four compartments of different isotype and tissue combinations were shown.

(C) Scaled diversity profiles of different isotype compartment at the steady state. Each line is for one isotype compartment in different tissues. Standard errors are plotted for diversities at the order of 0, 1, 2, 3 and 4.

(D) Scaled diversity profiles of different tissue samples at the steady state. Each line is for one tissue sample in different isotype compartments. Standard errors are plotted for diversities at the order of 0, 1, 2, 3 and 4.

(E) Scaled diversity profiles of samples under different immunization status. Each line is for one samples under different immunization in different tissue-isotype compartments. Standard errors are plotted for diversities at the order of 0, 1, 2, 3 and 4.

(F) Scaled diversity distribution and comparison of diversity values at the order $q = 4$ for samples under different immunization conditions in different tissue-isotype compartments. (Data are represented as mean \pm SEM. Statistical p values were corrected by FDR and significant p values were indicated. *, $p < 0.05$; **, $p < 0.01$, ***, $p < 0.001$, ****, $p < 0.0001$).

Clone diversity profiles before or after rescaling exhibited quite distinct patterns when comparing repertoires of different isotypes, tissues, and immunizations (Figures S16 and 5E). We focused on the rescaled diversity profiles to be less confounded with clone richness. We first compared the steady-state (PBS controls) diversity profiles between different isotypes and tissues. In spleen, higher diversity values were observed for IgM clones than those of IgG clones; in BM, it was IgG clones showing higher diversities. When comparing between different tissues, BM clones exhibited higher diversities in the IgG compartment, and spleen clones were of higher diversities in the IgM compartment (Figures 5C and 5D). Furthermore, the observed difference was statistically significant only at the small order $q = 1$ or 2, indicating changes occurred to clones of smaller size.

Next, OVA alone increased clone diversities compared to the non-immune state profiles, with a tissue and isotype dependency. For IgM clones, OVA alone increased diversities only in the spleen but only significantly at smaller orders ($p < 0.05$ at $q = 1$ or 2), while for IgG clones, it increased diversities significantly at larger orders and in both spleen and BM ($p < 0.05$ at $q = 4$ in spleen and at $q = 2, 3, 4$ in BM, Figures 5E and 5F; Figure S17). The OVA alone immunization did not change the pattern of diversity difference between tissues or isotypes observed at the unimmunized state but simply resulted in enlarged gaps (Figure S16A). Under OVA alone immunization and compared between isotypes, the

IgM diversity profile showed significantly higher values ($p < 0.001$) for small orders but slightly lower values than the IgG profile for large orders in the spleen (Figure S16A); the IgG profile was significantly higher at all orders ($p > 0.05$ for all comparisons at $q = 1$ to 4) in BM (Figure S16A). Additionally, when compared between tissues, the spleen clones showed significantly higher diversity values in the IgM compartment than the BM clones for all orders ($p > 0.05$ for all comparisons); it was the BM that contained significantly higher diversity values in the IgG compartment than the spleen but only for smaller orders $q = 1$ ($p < 0.05$, Figure S16B; Figure S17).

Similar to OVA alone, in the spleen, OVA+Alum increased diversities of IgM clones at small orders. However, it affected spleen IgG diversities to a lesser degree (Figure 5E; Figure S16). Significant changes induced by OVA+Alum came from BM clone diversity profile. It resulted in significantly higher BM clone diversities – for both the IgG and IgM isotype – than those of OVA alone or PBS groups (Figures 5E and 5F; Figure S16). As seen above, the increases in IgM clone diversities were mainly at smaller orders ($p < 0.0001$ for all comparisons at $q = 1$ and 2, Figure S17), and the increases in IgG were at higher orders ($p < 0.05$ at $q = 3$ and 4). Consequently, the IgM clone diversity profiles in spleen and BM become almost identical when compared between tissues (Figure S16B), while BM had higher IgM clone diversities at smaller orders but higher IgG clone diversities at higher orders when compared between isotypes (Figure S16A).

Finally, the OVA+CpG immunization gave rise to diversity profiles similar to those yielded either by OVA alone or by OVA+Alum, depending on the tissue and isotype. It raised BM IgM diversities significantly at lower orders resulting in a similar profile to the one caused by OVA+Alum, while for the IgG diversities, it led to profiles more resembling the ones by OVA alone with increasing values at higher orders in both BM and spleen (Figures 5C–5F; Figure S16).

Assessing clone selection strength and adjuvants' effects on it

We developed a method to assess the level of selection a B cell clone experienced throughout its development. It involves comparing two parameter estimates, mean mutation frequency and intra-clonal pairwise sequence dissimilarity. The rationale is that a clone evolving under selection expands some members with a high affinity for the specific driving antigen; thus, it grows more mutations but less dissimilarity among those (Figure S18A). Otherwise, it evolves with both parameters advancing proportionally (Figures S18B and S18C). This reasoning can be verified by visualizing two parameters on an x-y plot (Figure S19). Under a naive condition (the IgM compartment in PBS control), the two parameters of most clones showed values in good proportion, and clones settled on the plot following a straight trend line (Figure S19). Based on such a distribution, we set up a threshold to separate clones. Clones in the first group stayed above the threshold line and had a high mean mutations frequency and high sequence similarity, representing clones likely being selected through its development. In the second group, clones were located below the threshold and consisted of members with balanced mutations and dissimilarity, representing clones evolving under no selection. To quantify and compare the clone selection metric among samples, we selected the top 40 clones (based on the size)⁷⁰ from every sample and calculated the percentage of clones under selection (above the defined threshold on the two-parameter plot) (Figures S19, 6A, and 6B). Then we carried out statistical analyses on these percentage data.

Under all immunization conditions and in both tissues, the IgG compartment included a significantly higher percentage of clones under selection than the IgM compartment (Figure S20A). Such observation matched our expectation since IgG B cells were memory cells educated through the germinal center reaction; IgM B cells were either inexperienced naive cells or exited early from germinal centers during an immune response.^{53,71} Furthermore, the BM IgG compartment contained a significantly higher percentage of selected clones than the spleen IgG compartment either in PBS controls or under immunizations (Figure S20B). These clones in BM most likely consisted of PCs exited from the germinal center at the later stage of an immune response and migrated to the BM niche for long-term survival and antibody production.^{53,72} Lastly, adjuvanted immunization (Alum or CpG) significantly increased the percentage of clones comprised of selected members in BM IgG compartments, with CpG resulting in the highest proportion ($p < 0.01$ and $p < 0.05$ for OVA+CpG compared with PBS and OVA alone, respectively, and $p < 0.05$ for OVA+Alum vs. PBS) (Figure 6B).

Unsupervised machine learning to characterize Ig repertoire profile of different isotypes, tissues, and immunizations

We used unsupervised machine learning to identify associations among the various features of the Ig repertoire and the immunizations. The data are high-dimensional and of varied type: each sample is characterized by isotype, tissue of origin, and immunization status, as well as by mean mutation frequency, usage frequency over more than 130 V-gene segments, mean CDR3 length, Hill numbers of order 0–4, and mean selection strength. To deal effectively with such heterogeneous data, we first carried out a multiple factor analysis (MFA), a PCA-based multivariate data analysis intended to balance the influence of different variable types (quantitative or qualitative) by groups.⁷³ Without this balancing, the analyses would, for example, be dominated by gene segment usage.

The MFA transformed these variables into 15 principal components, the top two of which accounted for about 60% of the total variation (39% and 20% for PC1 and PC2, respectively; Figure S21). We used unsupervised clustering on the data in transformed PCA coordinates. Multiple clustering algorithms on all 15 PCs were carried out, including agglomerative hierarchical clustering (HCL), partition around medoids (PAM), spectral clustering, and affinity propagation clustering (APC). All algorithms yielded similar results (Figures S22 and S23).

Projection onto PC1 and PC2 shows that IgG+ and IgM+ samples are well-separated across PC1, while spleen and bone marrow divide roughly along PC2 (Figures 7A and 7D). In the IgG+ compartment, the samples are separated by tissue specificity, but such separation is not evident among the IgM+ samples (Figure 7A and 7B). The impact of different immunizations depends strongly on tissue and isotype (Figure 7C). In the BM IgG compartment, in particular, samples of different immunization groups aggregated into distinct clusters separated from each other (Figure 7C; Figures S22 and S23). In addition, the samples appear to be roughly ordered across PC1, starting with the PBS samples at the left, then moving right, OVA, and then the two adjuvanted immunization groups together. In the BM IgM compartment,

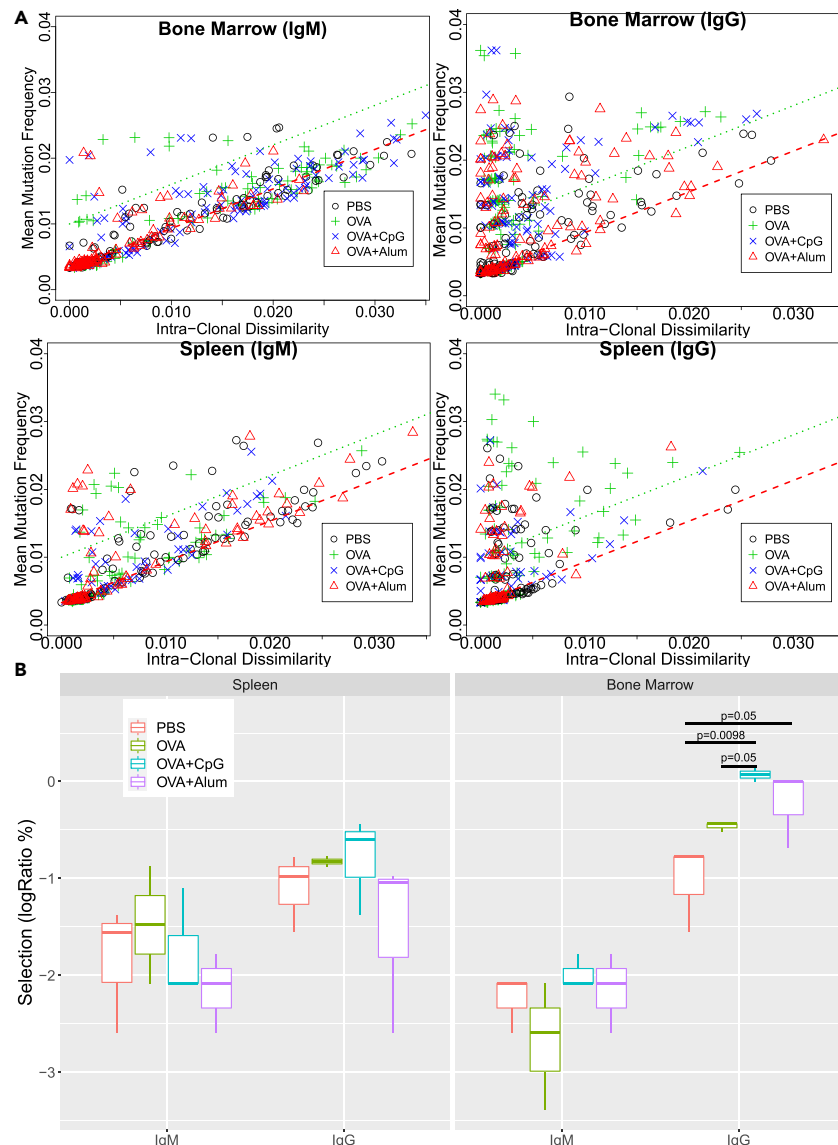


Figure 6. Evaluation of clonal selections and their changes by OVA immunization with adjuvants

We developed a new metric to quantify the changes of clone phylogenetic trees to reflect the effect of selection. It involves two measurements, mean clonal mutation frequency and pairwise intra-clonal dissimilarity. Under no selection, a clone evolves to increase these two parameters in proportion, and two parameters show as a straight line on an x-y 2D plot. Otherwise, some parts of a clone tree are selected, and the linear relationship between the two parameter is distorted. To analyze and compare the selection strength metric, we plotted the two parameters of the largest 40 clones of each sample (A) and defined a threshold (the green dashed line), above which are clones showing some level of selection. For each sample, on the basis of such threshold we can obtain the percentage of selected clones among the largest clones. The percentages were compared statistically and results were shown in (B). (Statistical p values were corrected by FDR. Significant p values are shown for comparing OVA immunization effects).

we mainly saw two clumps, one for samples immunized in the presence of adjuvants (CpG and Alum) and the other for PBS and OVA alone. Patterns in spleen compartments were less clearly marked, with seemingly OVA alone and OVA+CpG groups organized into separated clusters (Figure 7C; Figures S22 and S23).

Next, we projected supplementary categorical variables (i.e., grouping variables such as isotype, tissue, and immunization) on new PC dimensions to compare average changes among different categories. MFA allows for such projections by averaging information across samples in each category Figure 7D shows the projections of these category variables on the top two PC plane. First, same as seen above, the isotype variable showed the largest variation, i.e., IgG and IgM categories stayed far separated with much distance, primarily along the PC1 direction. The tissue variable was distributed more along the PC2 dimension, almost orthogonal to the isotype variable (the PC1 dimension). The immunization variable correlated with tissue along the PC2 direction, except for the OVA+CpG category. The OVA+Alum group stayed very

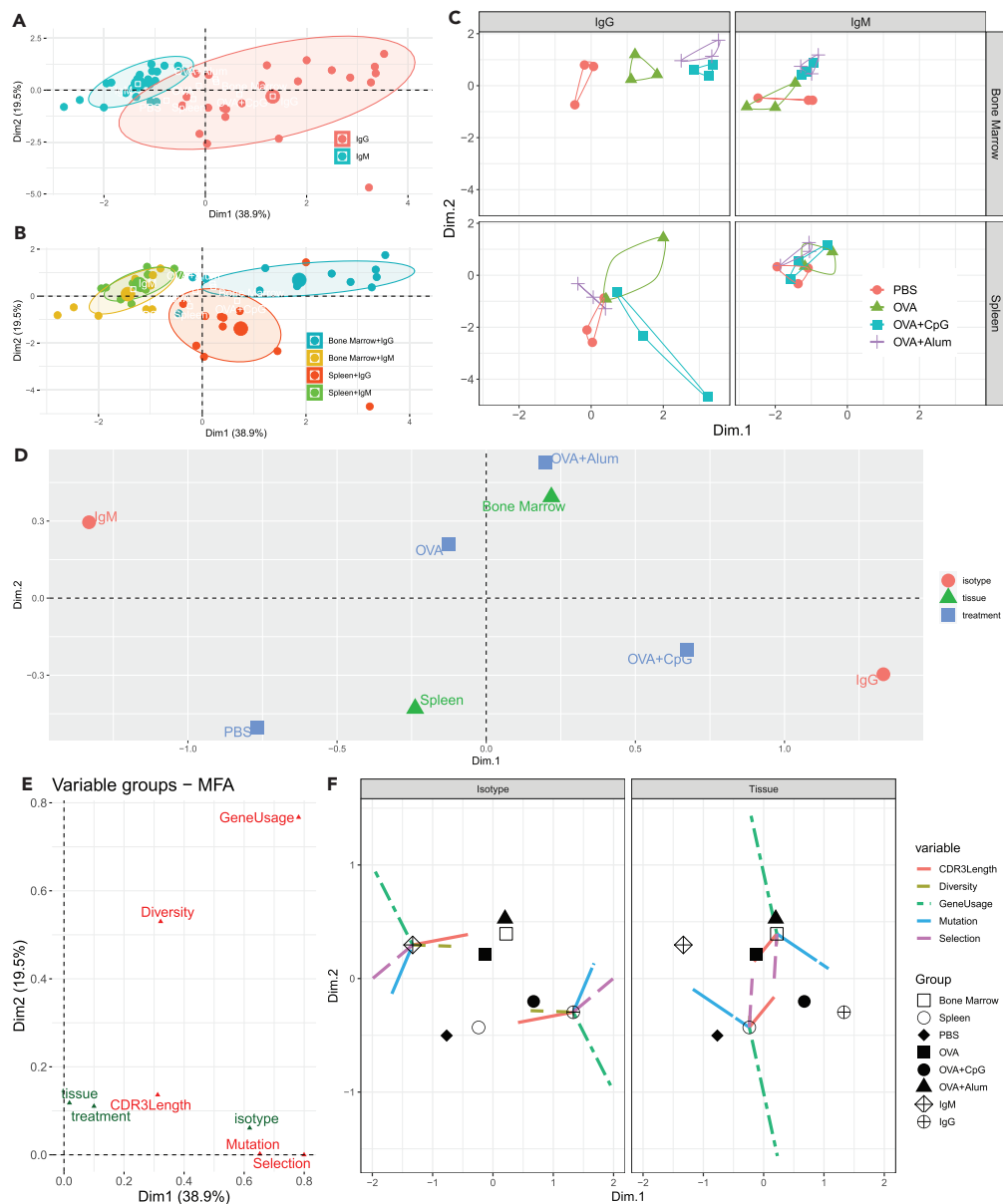


Figure 7. Unsupervised machine learning to characterize Ig repertoire changes

Multiple factor analysis (MFA) was carried out on the groups of variables including IGHV gene usage, mutation frequency, CDR3 length, clone diversity (expansion) and clone selection level. Sample data were then projected on the first two dimensions of MFA arranged by their isotype (A), tissue location (B), and immunization conditions (C). MFA can also summarize and project supplementary category variables (isotype, tissue and immunization) into the new PC dimensions (D) to show their correlations. Furthermore, the correlations between the variable groups and the first PC dimensions were shown in (E). Finally, the partial point graphs of supplementary category variables in MFA were shown in (F), in which supplementary categorical variables (such as isotype, tissue and immunization) were plotted as points with different shapes at the barycenter and variable groups were connected by lines with each center. The figure was arranged into sub-plots by isotype, tissue (F) and immunization (Figure S26) for easy visualization.

close to BM but equally distant to IgM and IgG, meaning the changes induced by OVA+Alum were more in BM and almost equally of both isotypes. On the other hand, the OVA+CpG group resided very close to IgG and slightly more adjacent to BM, meaning the changes caused by OVA+CpG were more of the IgG isotype and equally in both tissues. Such a pattern of changes matched what we saw in the individual sample projection data (Figures 7A–7C). The OVA alone group stayed more adjacent to BM and IgM, meaning the changes of this category occurred more in BM and IgM compartments. However, it was also close to the origin of the coordinates, indicating its effects were lower in magnitude than other groups. Lastly, the PBS category remained more adjacent to IgM and spleen, meaning without vaccination the Ig

repertoire contained alterations mainly in the spleen IgM compartment. Furthermore, we extended our analyses and observations to the top 3-PC space (altogether explaining about 78% of the total variations, [Figure S21](#)) and found a similar pattern of changes ([Figures S24 and S25](#) and [Videos S1 and S2](#)).

Isotype, tissue, and treatment associations with the natural variables were investigated by projecting each variable group onto specified planes in the PC space ([Figure 7E](#)). This procedure revealed that PC1 is dependent largely on gene usage, mutation, and selection and PC2 depends on gene usage and diversity. In this way, we found that isotype, separated in PC1, was correlated with gene usage, mutation, and selection. Tissue of origin, separated in PC2, was correlated with gene usage and diversity ([Figure 7E](#); [Figure S26](#)). Finer detail can be found by Partial Point Analysis⁷⁴ ([Figure 7F](#); [Figure S27](#)). For example, although the clone selection levels of the BM IgG and BM IgM compartments were quite different from those of the respective spleen IgG and spleen IgM compartments ([Figure 6B](#)), the overall clonal selection levels of BM and spleen were similar after averaging those of different isotypes in each tissue. That emphasized the necessity of studying Ig repertoires in separated tissue-isotype compartments.

Similarly, when analyzing treatment groups, we find that the PBS control and OVA immunized groups (with or without adjuvants) varied in diversity, CDR3 length, and gene usage, but not in mutation frequency. The apparent similarity of the clone selection levels between the PBS group and the other categories resulted from averaging over isotype. Differences between OVA alone and adjuvanted OVA immunizations also were attributable mainly to diversity and gene usage changes. When comparing OVA+CpG with OVA+Alum, all variable groups contributed similarly with almost equally distant partial points, except for the mutation frequency variable ([Figure S27](#)), indicating they differed in IGHV usage, CDR3 length, diversity, and clone selection levels but exhibited similar mutation frequencies.

DISCUSSION

We developed an in-house Ig sequencing library preparation and data analysis pipeline for accurately profiling immunoglobulin heavy chain repertoires. The library preparation protocol is simple and generates UMI-labeled isotype-aware reads of high quality and, along with the subsequent data processing, increases the yield of unbiased Ig sequences. One important way in which this yield is enhanced is, in contrast with the common practice,^{9,41,42} to omit paired-end joining. This increased the yield of usable reads 10-fold and eliminated errors caused by spurious short overlaps during paired-end joining. In addition, we developed new methods and metrics and adapted others to characterize the Ig repertoire: simplifying IGHV gene usage information using PCA, introducing diversity profiles using rescaled Hill numbers, and developing a two-parameter clonal selection metric based on the intra-clonal distribution of mutations. We combined these analyses with unsupervised machine learning to characterize the effects of immunization in different tissue-isotype compartments and the contribution of repertoire properties to these effects. We also compared our pipeline with other widely used and general purpose IgSeq pipeline to show these features in the supplementary document ([Table S3](#)).

We applied these techniques to characterize the impact of two commonly used adjuvants, CpG and Alum, on the murine heavy chain Ig repertoire in an OVA immunization model. The dosing and timing were determined based on the levels of serum anti-OVA IgG, which were significantly higher in the two adjuvanted group than those in PBS and OVA alone controls ([Figure S28](#)). We collected data from four compartments distinguished by tissue (BM or spleen) and isotype (IgM or IgG). The hypothesis was that these compartments contain functionally and phenotypically distinct B cells, and that the different immunization formulations affect them differently. Indeed, we found that the Ig repertoires from different tissue-isotype compartments differed in many respects and that immunization had a further tissue- and isotype-dependent impact. Isotype accounted for the highest variation in the repertoire, with 40% of the overall variance. IgM and IgG repertoires differed in IGHV mutation frequency, selection level, and IGHV gene usage. At the steady state in PBS controls, the IgG repertoire consisted of sequences using different IGHV segments and showing higher mutation frequency and contained more clones experiencing selection than the IgM repertoire. Tissue and immunization jointly resulted in changes explaining about 20% of the overall variations and orthogonal to the isotype effect. Repertoires from different tissues differed in IGHV gene usage and mutation frequency. At the non-immune steady state (PBS controls), the repertoire from BM consisted of sequences utilizing different sets of IGHV genes and accumulating more mutations than the spleen repertoire. These repertoire changes related to isotypes or tissues matched what we expected based on functional and phenotypic properties of B cells in these compartments.^{18,53–56}

Interestingly, distinct immunization conditions altered the Ig repertoire differently, and their effects exhibited different tissue and isotype dependencies. The non-immune naive repertoire (PBS controls) showed pre-existing tissue and isotype bias in gene usage, mutation frequency, and diversity, as mentioned above. Moreover, compared to OVA immunized repertoires (either adjuvanted or OVA alone), the naive repertoire utilized a distinct group of IGHV gene segments and comprised more rare clones (higher diversity values of small orders). The OVA alone immunization resulted in a repertoire different from the naive one, and its effect was more pronounced in BM or the IgM compartment. Repertoires under two adjuvanted immunizations differed from those under PBS and OVA alone immunization and distinguished from each other as well. OVA+Alum led to a repertoire profile staying near BM and in roughly an equal distance to both isotypes on the MFA factor map, indicating it causes changes more in BM of both IgG and IgM compartments. However, OVA+CpG resulted in a profile close to the IgG isotype and equally distant to both tissues on the factor map, meaning it invoked alterations in the IgG compartment of both tissues. Overall, OVA alone could accumulate significantly more mutations than the PBS control, to a level slightly lower than adjuvanted immunization did. Compared with OVA alone, adjuvanted immunizations did not impact the gene mutation but changed gene usage, clone diversity, and clone selection of the Ig repertoire significantly. Alum resulted in more abundant clones in BM for both isotypes, presumably by promoting the BM migration of these clones or survival of them in BM. On the other hand, CpG more efficiently increased the clone selection but affected to a lesser degree the diversity of abundant clones and exerted its effect on the IgG compartments in both spleen and BM. Based on these

observations, we proposed distinct adjuvant effect models (the quantity vs. quality model, [Figure S29](#)). These two models are not mutually exclusive and need more evident to confirm.

In summary, we characterized the Ig repertoire profiles in different compartments and under varying immunization conditions. These observations were not surprising but confirmed previous functional and phenotypic studies. In addition, we found that CpG and Alum resulted in distinct Ig repertoire profiles, probably due to differences on the cell types they affect and their mechanisms of action (via TLR9 for CpG vs. more of a non-specific effect for Alum). Looking ahead, we will want to expand the list of properties of Ig repertoire to be studied, such as CDR3 amino acid composition, IGHJ gene usage, etc., along with comparison to both TLR dependent and independent adjuvants. In addition, there are more B cell compartments contributing to protective immunity depending on the route and location of vaccine or infection, such as draining lymph nodes, mucosal B cell, and gut and other tissue-resident B cells. Future work should characterize those repertoires, ideally at different time points during an immune response and including other isotypes. The pipeline we have developed can be easily applied to study the Ig repertoire of other species (human, rat, and so on).

Limitations of the study

We want to bring up the fact that plasma cells (PCs) contain much more Ig mRNA transcripts than other B cells, such as naive or transitional B cells; thus, when mixed together, PCs contribute more to shape the overall BCR repertoire. Such skew is inherent to the bulk mRNA sequencing and can only be overcome by single-cell or DNA sequencing. On the other hand, we treated this as a feature to reflect the nature of distinct distributions of B cell subsets in different compartments, e.g., more PCs in bone marrow and more naive B cells in spleen, and ascribed the repertoire changes to these distributions in combine with other factors. Another limitation is that the current work is cross-sectional and that the timing of sample harvest was determined by our previous work. Ideally, a longitudinal study with multiple time points will surely provide more complete insights about the effects of Ig repertoire on the vaccine responses and will be our next step.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- [KEY RESOURCES TABLE](#)
- [RESOURCE AVAILABILITY](#)
 - Lead contact
 - Materials availability
 - Data and code availability
- [EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS](#)
- [METHOD DETAILS](#)
 - Mouse immunization
 - Cell collection and IGH sequencing library prep
 - Data preprocessing
 - Heavy chain variable region gene usage
 - Mutation frequency and CDR3 length
 - Rescaled clone diversity with hill numbers
 - Strength of selection
 - Unsupervised machine learning
- [QUANTIFICATION AND STATISTICAL ANALYSIS](#)

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.isci.2023.108749>.

ACKNOWLEDGMENTS

This work was supported by NIH/NIAID research contract 1U19AI117892-01 (TBK, PI) and NIH/NIAID research grant 2R01AI040944-19A1 (Wetzler, PI).

AUTHOR CONTRIBUTIONS

L.M.W., T.B.K., and F.F. conceived of the idea. F.F., R.Y., and Y.W. carried out the experiments. F.F., T.B.K., and A.H. ran the data analyses and wrote the manuscript. All authors have given approval to the final version of the manuscript.

DECLARATION OF INTERESTS

The authors declare that there is no conflict of interest.

Received: March 20, 2023
Revised: September 29, 2023
Accepted: December 12, 2023
Published: December 14, 2023

REFERENCES

- Soto, C., Bombardi, R.G., Branchizio, A., Kose, N., Matta, P., Sevy, A.M., Sinkovits, R.S., Gilchuk, P., Finn, J.A., and Crowe, J.E., Jr. (2019). High frequency of shared clonotypes in human B cell receptor repertoires. *Nature* 566, 398–402.
- Galson, J.D., Trück, J., Fowler, A., Clutterbuck, E.A., Münz, M., Cerundolo, V., Reinhard, C., van der Most, R., Pollard, A.J., Lunter, G., and Kelly, D.F. (2015). Analysis of B Cell Repertoire Dynamics Following Hepatitis B Vaccination in Humans, and Enrichment of Vaccine-specific Antibody Sequences. *EBioMedicine* 2, 2070–2079.
- Galson, J.D., Schaetzle, S., Bashford-Rogers, R.J.M., Raybould, M.I.J., Kovaltsuk, A., Kilpatrick, G.J., Minter, R., Finch, D.K., Dias, J., James, L.K., et al. (2020). Deep sequencing of b cell receptor repertoires from covid-19 patients reveals strong convergent immune signatures. *Front. Immunol.* 11.
- Tan, Y.-C., Kongpachith, S., Blum, L.K., Ju, C.-H., Lahey, L.J., Lu, D.R., Cai, X., Wagner, C.A., Lindstrom, T.M., Sokolove, J., and Robinson, W.H. (2014). Barcode-Enabled Sequencing of Plasmablast Antibody Repertoires in Rheumatoid Arthritis. *Arthritis Rheumatol.* 66, 2706–2715.
- Cha, E., Klinger, M., Hou, Y., Cummings, C., Ribas, A., Faham, M., and Fong, L. (2014). Improved survival with T cell clonotype stability after anti-CTLA-4 treatment in cancer patients. *Sci. Transl. Med.* 6, 238ra70.
- Georgiou, G., Ippolito, G.C., Beausang, J., Busse, C.E., Wardemann, H., and Quake, S.R. (2014). The promise and challenge of high-throughput sequencing of the antibody repertoire. *Nat. Biotechnol.* 32, 158–168.
- Mathonet, P., and Ullman, C.G. (2013). The application of next generation sequencing to the understanding of antibody repertoires. *Front. Immunol.* 4, 265.
- Boyd, S.D., and Crowe, J.E., Jr. (2016). Deep sequencing and human antibody repertoire analysis. *Curr. Opin. Immunol.* 40, 103–109.
- Yaari, G., and Kleinstein, S.H. (2015). Practical guidelines for B-cell receptor repertoire sequencing analysis. *Genome Med.* 7, 121.
- Pulendran, B., S Arunachalam, P., and O'Hagan, D.T. (2021). Emerging concepts in the science of vaccine adjuvants. *Nat. Rev. Drug Discov.* 20, 454–475.
- HogenEsch, H., O'Hagan, D.T., and Fox, C.B. (2018). Optimizing the utilization of aluminum adjuvants in vaccines: you might just get what you want. *NPJ Vaccines* 3, 51.
- De Gregorio, E., Caproni, E., and Ulmer, J.B. (2013). Vaccine adjuvants: mode of action. *Front. Immunol.* 4, 214.
- Reed, S.G., Orr, M.T., and Fox, C.B. (2013). Key roles of adjuvants in modern vaccines. *Nat. Med.* 19, 1597–1608.
- Montamat, G., Leonard, C., Poli, A., Klimek, L., and Ollert, M. (2021). CpG adjuvant in allergen-specific immunotherapy: finding the sweet spot for the induction of immune tolerance. *Front. Immunol.* 12.
- Bode, C., Zhao, G., Steinhagen, F., Kinjo, T., and Klinman, D.M. (2011). CpG DNA as a vaccine adjuvant. *Expert Rev. Vaccines* 10, 499–511.
- Mebius, R.E., and Kraal, G. (2005). Structure and function of the spleen. *Nat. Rev. Immunol.* 5, 606–616.
- Lewis, S.M., Williams, A., and Eisenbarth, S.C. (2019). Structure and function of the immune system in the spleen. *Sci. Immunol.* 4, eaau6085.
- Mamani-Matsuda, M., Cosma, A., Weller, S., Faili, A., Staib, C., Garçon, L., Hermine, O., Beyne-Rauzy, O., Fieschi, C., Pers, J.O., et al. (2008). The human spleen is a major reservoir for long-lived vaccinia virus-specific memory B cells. *Blood* 111, 4653–4659.
- Mercier, F.E., Ragu, C., and Scadden, D.T. (2011). The bone marrow at the crossroads of blood and immunity. *Nat. Rev. Immunol.* 12, 49–60.
- Zhao, E., Xu, H., Wang, L., Kryczek, I., Wu, K., Hu, Y., Wang, G., and Zou, W. (2012). Bone marrow and the control of immunity. *Cell. Mol. Immunol.* 9, 11–19.
- Moreau, J.M., Berger, A., Nelles, M.E., Mielnik, M., Furlonger, C., Cen, S.Y., Besla, R., Robbins, C.S., and Paige, C.J. (2015). Inflammation rapidly reorganizes mouse bone marrow B cells and their environment in conjunction with early IgM responses. *Blood* 126, 1184–1192.
- Slamanig, S.A., and Nolte, M.A. (2021). The Bone Marrow as Sanctuary for Plasma Cells and Memory T-Cells: Implications for Adaptive Immunity and Vaccinology. *Cells* 10, 1508.
- Ghraichy, M., von Niederhäusern, V., Kovaltsuk, A., Galson, J.D., Deane, C.M., and Trück, J. (2021). Different B cell subpopulations show distinct patterns in their IgH repertoire metrics. *Elife* 10, e73111.
- Galson, J.D., Clutterbuck, E.A., Trück, J., Ramasamy, M.N., Münz, M., Fowler, A., Cerundolo, V., Pollard, A.J., Lunter, G., and Kelly, D.F. (2015). BCR repertoire sequencing: different patterns of B-cell activation after two Meningococcal vaccines. *Immunol. Cell Biol.* 93, 885–895.
- Saada, R., Weinberger, M., Shahaf, G., and Mehr, R. (2007). Models for antigen receptor gene rearrangement: CDR3 length. *Immunol. Cell Biol.* 85, 323–332.
- Greiff, V., Menzel, U., Miho, E., Weber, C., Riedel, R., Cook, S., Valai, A., Lopes, T., Radbruch, A., Winkler, T.H., and Reddy, S.T. (2017). Systems Analysis Reveals High Genetic and Antigen-Driven Predetermination of Antibody Repertoires throughout B Cell Development. *Cell Rep.* 19, 1467–1478.
- Choi, N.M., Loguercio, S., Verma-Gaur, J., Degner, S.C., Torkamani, A., Su, A.I., Oltz, E.M., Artyomov, M., and Feeney, A.J. (2013). Deep sequencing of the murine IgH repertoire reveals complex regulation of nonrandom V gene rearrangement frequencies. *J. Immunol.* 191, 2393–2402.
- He, L., Sok, D., Azadnia, P., Hsueh, J., Landais, E., Simek, M., Koff, W.C., Poignard, P., Burton, D.R., and Zhu, J. (2014). Toward a more accurate view of human B-cell repertoire by next-generation sequencing, unbiased repertoire capture and single-molecule barcoding. *Sci. Rep.* 4, 6778.
- Picelli, S., Faridani, O.R., Björklund, A.K., Winberg, G., Sagasser, S., and Sandberg, R. (2014). Full-length RNA-seq from single cells using Smart-seq2. *Nat. Protoc.* 9, 171–181.
- Rubelt, F., Bolen, C.R., McGuire, H.M., Vander Heiden, J.A., Gadala-Maria, D., Levin, M., Euskirchen, G.M., Mamedov, M.R., Swan, G.E., Dekker, C.L., et al. (2016). Individual heritable differences result in unique cell lymphocyte receptor repertoires of naive and antigen-experienced cells. *Nat. Commun.* 7, 11112.
- Sun, L., Kono, N., Toh, H., Xue, H., Sano, K., Suzuki, T., Aina, A., Orba, Y., Yamagishi, J., Hasegawa, H., et al. (2019). Identification of Mouse and Human Antibody Repertoires by Next-Generation Sequencing. *J. Vis. Exp.* 1.
- Vázquez Bernat, N., Corcoran, M., Hardt, U., Kaduk, M., Phad, G.E., Martin, M., and Karlsson Hedestam, G.B. (2019). High-Quality Library Preparation for NGS-Based Immunoglobulin Germline Gene Inference and Repertoire Expression Analysis. *Front. Immunol.* 10, 660.
- Picelli, S., Björklund, Å.K., Faridani, O.R., Sagasser, S., Winberg, G., and Sandberg, R. (2013). Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat. Methods* 10, 1096–1098.
- Kapteyn, J., He, R., McDowell, E.T., and Gang, D.R. (2010). Incorporation of non-natural nucleotides into template-switching oligonucleotides reduces background and improves cDNA synthesis from very small RNA samples. *BMC Genom.* 11, 413.
- Khan, T.A., Friedensohn, S., Gorter de Vries, A.R., Straszewski, J., Ruscheweyh, H.J., and Reddy, S.T. (2016). Accurate and predictive antibody repertoire profiling by molecular amplification fingerprinting. *Sci. Adv.* 2, e1501371.
- Cocquet, J., Chong, A., Zhang, G., and Veitia, R.A. (2006). Reverse transcriptase template switching and false alternative transcripts. *Genomics* 88, 127–131.
- Li, B., Li, B., Luo, J., Tang, P., Mao, J., and Wu, X. (2013). Simultaneous reduction in noise and cross-contamination artifacts for dual-energy X-ray CT. *BioMed Res. Int.* 2013, 417278.
- Tang, D.T.P., Plessey, C., Salimullah, M., Suzuki, A.M., Calligaris, R., Gustinich, S., and Carninci, P. (2013). Suppression of artifacts and barcode bias in high-throughput transcriptome analyses utilizing template switching. *Nucleic Acids Res.* 41, e44.
- Wu, L., Wen, C., Qin, Y., Yin, H., Tu, Q., Van Nostrand, J.D., Yuan, T., Yuan, M., Deng, Y., and Zhou, J. (2015). Phasing amplicon sequencing on Illumina Miseq for robust environmental microbial community analysis. *BMC Microbiol.* 15, 125.
- MacConaill, L.E., Burns, R.T., Nag, A., Coleman, H.A., Slevin, M.K., Giorda, K., Light, M., Lai, K., Jarosz, M., McNeill, M.S., et al.

- (2018). Unique, dual-indexed sequencing adapters with UMIs effectively eliminate index cross-talk and significantly improve sensitivity of massively parallel sequencing. *BMC Genom.* 19, 30.
41. Vander Heiden, J.A., Yaari, G., Uduman, M., Stern, J.N.H., O'Connor, K.C., Hafler, D.A., Vigneault, F., and Kleinstein, S.H. (2014). pRESTO: a toolkit for processing high-throughput sequencing raw reads of lymphocyte receptor repertoires. *Bioinformatics* 30, 1930–1932.
 42. Zhang, J., Kobert, K., Flouri, T., and Stamatakis, A. (2014). PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics* 30, 614–620.
 43. Larimore, K., McCormick, M.W., Robins, H.S., and Greenberg, P.D. (2012). Shaping of human germline IgH repertoires revealed by deep sequencing. *J. Immunol.* 189, 3221–3230.
 44. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359.
 45. Smith, T., Heger, A., and Sudbery, I. (2017). UMI-tools: modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. *Genome Res.* 27, 491–499.
 46. Kepler, T.B. (2013). Reconstructing a B-cell clonal lineage. *F1000Res.* 2, 103.
 47. Rechavi, E., Lev, A., Lee, Y.N., Simon, A.J., Yinon, Y., Lipitz, S., Amariglio, N., Weisz, B., Notarangelo, L.D., and Somech, R. (2015). Timely and spatially regulated maturation of B and T cell repertoire during human fetal development. *Sci. Transl. Med.* 7, 276ra25.
 48. Dogan, I., Bertocci, B., Vilmont, V., Delbos, F., Mègret, J., Storck, S., Reynaud, C.A., and Weill, J.C. (2009). Multiple layers of B cell memory with different effector functions. *Nat. Immunol.* 10, 1292–1299.
 49. Pape, K.A., Taylor, J.J., Maul, R.W., Gearhart, P.J., and Jenkins, M.K. (2011). Different B cell populations mediate early and late memory during an endogenous immune response. *Science* 331, 1203–1207.
 50. Petrova, V.N., Muir, L., McKay, P.F., Vassiliou, G.S., Smith, K.G.C., Lyons, P.A., Russell, C.A., Anderson, C.A., Kellam, P., and Bashford-Rogers, R.J.M. (2018). Combined Influence of B-Cell Receptor Rearrangement and Somatic Hypermutation on B-Cell Class-Switch Fate in Health and in Chronic Lymphocytic Leukemia. *Front. Immunol.* 9, 1784.
 51. Kinoshita, K., and Honjo, T. (2001). Linking class-switch recombination with somatic hypermutation. *Nat. Rev. Mol. Cell Biol.* 2, 493–503.
 52. Muramatsu, M., Kinoshita, K., Fagarasan, S., Yamada, S., Shinkai, Y., and Honjo, T. (2000). Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. *Cell* 102, 553–563.
 53. Palm, A.K.E., and Henry, C. (2019). Remembrance of Things Past: Long-Term B Cell Memory After Infection and Vaccination. *Front. Immunol.* 10, 1787.
 54. Tangye, S.G., and Tarlinton, D.M. (2009). Memory B cells: effectors of long-lived immune responses. *Eur. J. Immunol.* 39, 2065–2075.
 55. Roth, K., Oehme, L., Zehentmeier, S., Zhang, Y., Niesner, R., and Hauser, A.E. (2014). Tracking plasma cell differentiation and survival. *Cytometry A.* 85, 15–24.
 56. Zehentmeier, S., Roth, K., Cseresnyes, Z., Sercan, Ö., Horn, K., Niesner, R.A., Chang, H.D., Radbruch, A., and Hauser, A.E. (2014). Static and dynamic components synergize to form a stable survival niche for bone marrow plasma cells. *Eur. J. Immunol.* 44, 2306–2317.
 57. Hou, D., Ying, T., Wang, L., Chen, C., Lu, S., Wang, Q., Seeley, E., Xu, J., Xi, X., Li, T., et al. (2016). Immune Repertoire Diversity Correlated with Mortality in Avian Influenza A (H7N9) Virus Infected Patients. *Sci. Rep.* 6, 33843.
 58. Martin, V., Bryan Wu, Y.C., Kipling, D., and Dunn-Walters, D. (2015). Ageing of the B-cell repertoire. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 370, 20140237.
 59. Boyd, S.D., Gaëta, B.A., Jackson, K.J., Fire, A.Z., Marshall, E.L., Merker, J.D., Maniar, J.M., Zhang, L.N., Sahaf, B., Jones, C.D., et al. (2010). Individual variation in the germline Ig gene repertoire inferred from variable region gene rearrangements. *J. Immunol.* 184, 6986–6992.
 60. Lefranc, M.P., Giudicelli, V., Ginestoux, C., Jabado-Michaloud, J., Folch, G., Bellahcene, F., Wu, Y., Gemrot, E., Brochet, X., Lane, J., et al. (2009). IMGT, the international ImMunoGeneTics information system. *Nucleic Acids Res.* 37, D1006–D1012.
 61. Gloor, G.B., Macklaim, J.M., Pawlowsky-Glahn, V., and Egozcue, J.J. (2017). Microbiome Datasets Are Compositional: And This Is Not Optional. *Front. Microbiol.* 8, 2224.
 62. Morton, J.T., Sanders, J., Quinn, R.A., McDonald, D., Gonzalez, A., Vázquez-Baeza, Y., Navas-Molina, J.A., Song, S.J., Metcalf, J.L., Hyde, E.R., et al. (2017). Balance Trees Reveal Microbial Niche Differentiation. *mSystems* 2, e00162-16.
 63. Fehrmann, R.S.N., Karjalainen, J.M., Krajewska, M., Westra, H.J., Maloney, D., Simeonov, A., Pers, T.H., Hirschhorn, J.N., Jansen, R.C., Schultes, E.A., et al. (2015). Gene expression analysis identifies global gene dosage sensitivity in cancer. *Nat. Genet.* 47, 115–125.
 64. Xu, J.L., and Davis, M.M. (2000). Diversity in the CDR3 region of V(H) is sufficient for most antibody specificities. *Immunity* 13, 37–45.
 65. Kepler, T.B., Liao, H.X., Alam, S.M., Bhaskarabhatla, R., Zhang, R., Yandava, C., Stewart, S., Anasti, K., Kelsoe, G., Parks, R., et al. (2014). Immunoglobulin gene insertions and deletions in the affinity maturation of HIV-1 broadly reactive neutralizing antibodies. *Cell Host Microbe* 16, 304–313.
 66. Yu, L., and Guan, Y. (2014). Immunologic Basis for Long HCDR3s in Broadly Neutralizing Antibodies Against HIV-1. *Front. Immunol.* 5, 250.
 67. Chao, A., Gotelli, N.J., Hsieh, T.C., Sander, E.L., Ma, K.H., Colwell, R.K., and Ellison, A.M. (2014). Rarefaction and extrapolation with Hill numbers: a framework for sampling and estimation in species diversity studies. *Ecol. Monogr.* 84, 45–67.
 68. Jost, L. (2006). Entropy and diversity. *Oikos* 113, 363–375.
 69. Chao, A., and Jost, L. (2015). Estimating diversity and entropy profiles via discovery rates of new species. *Methods Ecol. Evol.* 6, 873–882.
 70. Rosenfeld, A.M., Meng, W., Chen, D.Y., Zhang, B., Granot, T., Farber, D.L., Hershberg, U., and Luning Prak, E.T. (2018). Computational Evaluation of B-Cell Clone Sizes in Bulk Populations. *Front. Immunol.* 9, 1472.
 71. Elsner, R.A., and Shlomchik, M.J. (2020). Germinal Center and Extrafollicular B Cell Responses in Vaccination, Immunity, and Autoimmunity. *Immunity* 53, 1136–1150.
 72. Allen, C.D.C., Okada, T., Tang, H.L., and Cyster, J.G. (2007). Imaging of germinal center selection events during affinity maturation. *Science* 315, 528–531.
 73. Escofier, B., and Pagès, J. (1994). Multiple factor analysis (AFMULT package). *Comput. Stat. Data Anal.* 18, 121–140.
 74. Abdi, H., Williams, L.J., and Valentin, D. (2013). Multiple factor analysis: principal component analysis for multitable and multiblock data sets. *WIREs Computational Stats.* 5, 149–179.
 75. Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120.
 76. Aitchison, J. (1982). The Statistical Analysis of Compositional Data. *J. Roy. Stat. Soc. B* 44, 139–160.
 77. Chao, A., and Chiu, C.-H. (2016). Species Richness: Estimation and Comparison. In *Wiley StatsRef: Statistics Reference Online*, N. Balakrishnan, T. Colton, B. Everitt, W. Piegorsch, F. Ruggeri, and J.L. Teugels, eds., pp. 1–26.
 78. Lê, S., Josse, J., and Husson, F. (2008). FactoMineR: An R Package for Multivariate Analysis. *J. Stat. Soft.* 25, 18.
 79. Galili, T. (2015). dendextend: an R package for visualizing, adjusting and comparing trees of hierarchical clustering. *Bioinformatics* 31, 3718–3720.
 80. Karatzoglou, A., Smola, A., Hornik, K., and Zeileis, A. (2004). kernlab - An S4 Package for Kernel Methods in R. *J. Stat. Softw.* 11, 1–20.
 81. Bodenhofer, U., Kothmeier, A., and Hochreiter, S. (2011). APCluster: an R package for affinity propagation clustering. *Bioinformatics* 27, 2463–2464.
 82. Konopka, T. (2020). Umap: Uniform Manifold Approximation and Projection.
 83. Fox, J., and Weisberg, S. (2019). *An R Companion to Applied Regression*, 33rd Edition (Sage).
 84. Lenth, R.V. (2021). *Emmeans: Estimated Marginal Means, Aka Least-Squares Means*.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
Class B CpG oligonucleotide - Murine TLR9 ligand	InvivoGen	ODN 1826
Aluminum hydroxide Gel	Sigma	A8222
RNAzol RT	Molecular Research Center	RN 190
EasySep mouse B cell isolation kit	Stemcell Technologies	19854
Maxima H Minus Reverse Transcriptase	ThermoFisher Scientific	EP0752
AMPure XP Reagent	Beckman Coulter Life Sciences	A63881
KAPA HiFi HotStart Real-Time Library Amp Kit	Roche	KK2702
Experimental models: Organisms/strains		
Mouse:C57BL/6J:C57BL/6J	The Jackson Laboratory	Strain #:000664 RRID:IMSR_JAX:000664
Oligonucleotides		
Forward template-switching oligos	This paper	/5Biosg/-GTCAGATGTGTATAAGAGACAG NNHATTNNHNNATTNNCGATAGrGrG-3'
Reverse template-switching oligos	This paper	5'-GTGTCACGTACAGAGTCATC(T) ₃₀ VN-3'
WTA forward primer	This paper	5'-TCGTCGGCAGCGTCAGATGTGTATA AGAGACAG-3'
Reverse IgG1/2-specific primer	This paper	5'-GTCTCGTGGGCTCGGAGATGTGTATA AGAGACAGNNNNCAGATCCAGGG CCAGTGGATAGA*C-3'
Reverse IgG3-specific primer	This paper	5'-GTCTCGTGGGCTCGGAGATGTGTATA AGAGACAGNNNTGCAGCCAGGG ACCAAGGGATAGA*C
Reverse IgM-specific primer	This paper	5'-GTCTCGTGGGCTCGGAGATGTGTAT AAGAGACAGNNNNGGAAGAC ATTGGGAAGGACTGA*C-3'
Illumina adapters and dual multiplexing indexes forward primer	Illumina	AATGATACGGCGACCACCGAGATCTAC AC-[i5 index]-TCGTCGGCAGCGT*C
Illumina adapters and dual multiplexing indexes reverse primer	Illumina	CAAGCAGAAGACGGCATAACGAG AT-[i7 index]-GTCTCGTGGGCTCG*G
Software and algorithms		
bcl2fastq v2.20	Illumina	N/A
Trimmomatic v0.32	Bolger et al. ⁷⁵	http://www.usadellab.org/cms/?page=trimmomatic
bowtie2 v2.4.5	Langmead et al. ⁴⁴	http://bowtie-bio.sourceforge.net/bowtie2/index.shtml
R v3.6.3	The R Project for Statistical Computing	https://cran.r-project.org/
Cloanalyst	Kepler, TB ⁴⁶	N/A
Isotype classification	This paper	https://github.com/BULQI/IgSeq
UMI dedup	This paper	https://github.com/BULQI/IgSeq
Deposited data		
Ig RNA-Seq Raw data	This paper	Gene Expression Omnibus (GEO) with the accession number GEO: GSE228856 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE228856)

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Lee M. Wetzler (lwetzler@bu.edu).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- Immunoglobulin RNA-seq data have been deposited at GEO and are publicly available as of the date of publication. Accession numbers are listed in the [key resources table](#).
- All original code has been deposited at github and is publicly available as of the date of publication.
- Any additional information required to reanalyze the data reported in this paper is available from the [lead contact](#) upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Male C57BL/6J mice were purchased from The Jackson Laboratory (Bar Harbor, ME, USA). Groups of three mice started to receive subcutaneous immunizations at the age of 6–8 weeks. All mice were maintained in the Association for Assessment and Accreditation of Laboratory Animal Care International accredited facility at Boston University School of Medicine Laboratory Animal Science Center. All animal experiments were conducted under the approved Institutional Animal Care and Use Committee (IACUC) protocol at Boston University.

METHOD DETAILS

Mouse immunization

C57BL/6J mice of 6–8 weeks old started to receive one of the four subcutaneous immunizations in a volume of 100 μ L with 10 μ g ovalbumin (OVA) plus adjuvant or vehicle control (PBS): PBS, OVA, OVA + CpG (10 μ g, Invivogen, Cat# ODN1826) and OVA + Alum (200 μ g Sigma, Cat# A8222). A total of four immunizations were applied at two-week intervals between the first three injections and a four-week interval between the third and the fourth dose ([Figure 1A](#)). All mice were maintained in the Association for Assessment and Accreditation of Laboratory Animal Care International accredited facility at Boston University School of Medicine Laboratory Animal Science Center.

Cell collection and IGH sequencing library prep

A week after the fourth immunization, when the anti-OVA response were significantly higher in the two adjuvanted groups than controls ([Figure S28](#)), BMs and spleens were harvested. The spleen was disrupted, and cell suspension was obtained by passing a 70 μ m mesh nylon strainer. Splenic B cells were enriched with the EasySep mouse B cell isolation kit (STEMCELL Technologies, Cambridge, MA, Cat# 19854). Bone marrow cells were obtained by flushing femora and tibiae with RPMI medium, and whole tissue cells were used without enrichment. Red blood cells were lysed from samples with ACK (150 mM NaH₄CL, 50 mM KHCO₃) lysis buffer. Splenic B cells and bone marrow tissue cells were treated with RNAzol RT (Molecular Research Center, Cat# RN 190), and total RNA was isolated according to the manufacturer's instruction and saved at -80° C.

We adopted a 5' rapid amplification of cDNA ends (5'RACE) approach for library prep ([Figure 1B](#)).^{27–33} cDNA synthesis was done with two primers, forward template switching oligos (TSO) with unique molecular identifiers (UMI)/5Biosg/-GTCAGATGTGTATAAGA GACAGNNHATTNNHNNATTNNCGATAGrGrG-3' and reverse 5'-GTGTCACGTACAGAGTCATC(T)₃₀VN-3'. Two hundred nanograms of total RNA was reverse-transcribed in the presence of 100 μ M primers, 10 mM dNTP mix, and 40U/ μ L Maxima H Minus transcriptase (Thermo Scientific, Cat# EP0752). PCR was run on a thermal cycler with a cycle: 30 min at 50 $^{\circ}$ C for the initial step; 2 min at 50 $^{\circ}$ C and 2 min at 42 $^{\circ}$ C for 10 cycles; 15 min at 80 $^{\circ}$ C to inactivate the transcriptase.

cDNA was cleaned with AmpureXP beads (Beckman Coulter Life Sciences, Indianapolis, IN, Cat# A63881) and amplified in the presence of 0.4 μ M primer mix, WTA forward primer: 5'-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG-3' and reverse isotype-specific (IgG1/2: 5'-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGNNNNCAGATCCAGGGGCCAGTGGATAGA*C-3'; IgG3: 5'-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGNNNNNTGCAGCCAGGGACCAAGGGATAGA*C; IgM: 5'-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGNNNNGGGAAGACATTTGGGAAGGACTGA*C-3'). Then the last PCR was carried out to add Illumina adapters and dual multiplexing indexes (10nts long) with 0.125 μ M primer pairs, forward: AATGATACGGCGACCACCGAGATCTACAC-[i5 index]-TCGTCGGCAGCGT*C and reverse: CAAGCAGAAGACGGCATACGAGAT-[i7 index]-GTCTCGTGGGCTCG*G (Integrated DNA Technologies, Iowa, USA). Both PCR amplifications were done with KAPA Real-time library Amp kit (Roche, Wilmington, MA) with a cycle condition: 45 s at 98 $^{\circ}$ C for initial denaturation; 15 s at 98 $^{\circ}$ C, 30 s at 66 $^{\circ}$ C and 60 s at 72 $^{\circ}$ C for up to 22 cycles; 2 min at 72 $^{\circ}$ C for the last extension. PCRs were monitored in real-time and terminated when reaching a defined threshold to avoid over-amplification. Final PCR products were cleaned up by beads, measured for DNA concentrations by a Qubit 4 fluorometer (ThermoFisher Scientific), and then multiple samples were pooled with equal amounts. Finally, the library was size-selected between 400 and 1000 bp by Blue Pippin (LABGENE Scientific, Switzerland) before being sequenced on Illumina HiSeq2500 with 2 \times 250 bp chemistry. Libraries were loaded at 6.0p.m. with 5% Phix control v3 (Illumina) added.

Data preprocessing

We preprocessed the sequence data using procedures developed in-house, including sample demultiplexing, quality trimming, search filtering, UMI deduplication, isotype determination, recombination parameter identification, and clonal partitioning (Figure 1C). Sample demultiplexing was done using bcl2fastq (Illumina) with no mismatch allowed. Quality trimming was done using Trimmomatic with a sliding window of 4 nucleotides and a minimum quality score of 20⁷⁵. We ran bowtie2 comparing the sequences to an Ig heavy-chain database to filter out non-IG genes^{44,60} (Scheme S1 in the Supplemental Information). Isotype classification was carried out using an in-house developed C++ program with flexibility for mismatches and offset isotype sequences. To meet the specific needs of Ig sequencing data, we implemented an in-house algorithm for deduplicating UMIs in python (Scheme S2 in the Supplemental Information). Sequences within each group have identical UMIs, share the first 25 nucleotides identically, have less than 5% mismatches, and have identical isotype. To increase the quantification accuracy a directional network-based method was adopted to account for sequencing errors in UMIs.⁴⁵ Consensus sequences were generated on the basis of the multiple alignments of sequences of each group and taking into consideration of sequence quality. Singleton sequences were excluded from the analysis. Finally estimation of recombination parameters and clonal partitioning were done using Cloanalyzer⁴⁶ (see details in the next section). All scripts and software tools are available at <https://github.com/BULQI/IgSeq>.

Heavy chain variable region gene usage

The IGHV counts for each sample were converted to compositional data using the R compositions library,⁷⁶ and further transformed to center log-ratios (CLR) to map the Aitchison simplex into real space.⁷⁶ The principal component analysis (PCA) was performed using the R stats package. The coordinates in the first few PCs were used as covariates in multi-factor analysis of variance (ANOVA). *Post hoc* analyses were performed with the false discovery rate (FDR) correction.

Mutation frequency and CDR3 length

Mutations were identified by comparing IgVRG sequences with their inferred IGHV genes through the end of framework region 4 only (excluding CDR3). The mean and standard deviation of mutation frequencies were recorded for each of forty-eight groups. The mean and standard deviation of CDR3 lengths (nucleotides) for each sample were calculated. We then ran the three-way ANOVA on mean mutation frequencies and CDR3 lengths to test the effects of three factors (tissue, isotype, and immunization) (see the [quantification and statistical analysis](#) subsection for details).

Rescaled clone diversity with hill numbers

Clonal diversity was quantified using Hill numbers,⁴¹ estimated using the methods described in Chao et al. work.⁶⁹ To further reduce bias in the estimates from finite sample size, we made comparisons by pseudorandomly subsampling each sample to the size of the smallest sample a total of 100 times each. The 'SpadeR' R package⁷⁷ was used to estimate diversity profiles as Hill numbers of order q between 0 and 4,⁶⁹ which were further rescaled by the clonal richness and transformed by the arcsine square root function. This rescaling was intended for down weighting the effect of clone richness on the diversity measure but emphasized the relative clone abundance information (or the clone evenness). Finally, the three-way ANOVA and *post hoc* analysis were performed to test for factor effects.

Strength of selection

To assess the level of selection that occurred within a B cell clone during its development, we compare two statistical characters of the associated phylogenetic tree: pairwise sequence dissimilarity (PSD) and mean mutation frequency (MMF). Our expectation is that clones experiencing stronger selection will exhibit lower ratio of sequence dissimilarity to mean mutation frequency. We calculated the MMF and PSD of the forty largest clones in each sample and made scatterplots. For comparisons, we established a threshold ratio to define the percentage of selected clones in each sample (Figure S19). Finally, the percentages of clones with selection were transformed to compositional data and used for statistical analysis.

Unsupervised machine learning

To identify and summarize the feature variations characteristic of different treatments, we carried out multiple factor analysis (MFA) and unsupervised clustering on the complete assembled dataset. For each sample, there were five quantitative variables: mutation frequency, IGHV gene usage (multicomponent), CDR3 length, Hill numbers (multicomponent), and clonal selection level; these were used in the MFA. In addition, there were three supplementary categorical variables: tissue, isotype, and treatment (immunization). We conducted MFA using the 'FactorMineR' package⁷⁸ and visualized data using the 'factoextra' R package.⁷⁸ Unsupervised clustering was carried out on the PCs generated by MFA. Three clustering algorithms were performed, agglomerative hierarchical clustering (HC), spectral clustering, and affinity propagation clustering (APC). HC was done using the base R library and visualized using the 'dendroextra' R package,⁷⁹ spectral clustering was done using the 'kernlab' R package,⁸⁰ and APC was done using the 'APCluster' R package.⁸¹ Results were visualized on a 2D figure projected from high dimensional data with the 'umap' R package.⁸²

QUANTIFICATION AND STATISTICAL ANALYSIS

We ran three-factor ANOVA to test for effects of tissue (spleen and bone marrow), isotype (IgM and IgG), and immunization (PBS, OVA, OVA+CpG, and OVA+Alum) on repertoire properties (mutation, gene usage, etc.). The sample size for each group is three mice with an exception of two mice in the OVA group of spleen samples. The data were transformed to obtain normality and homoscedasticity when necessary. For IGHV gene usage data, we transformed gene counts into compositional data,⁷⁶ which were further center-log ratio (CLR) transformed to map the Aitchison simplex into real space. Similarly, normalized clone diversity data were arcsine transformed, and the CLR transformation was done on the clonal selection data, which are percentages of clones under selection in each sample.

Both type II and III ANOVA tables were obtained by the R 'car' package⁸³ because of unequal sample sizes of different groups. For tests yielding significant differences ($p < 0.05$) in ANOVA, we ran post hoc analysis by the simple effect test with contrasts using the false discovery rate (FDR) to control for the multiple comparison error. The follow-up analyses were done using the R 'emmeans' package.⁸⁴ All R code and scripts are available at <https://github.com/BULQI/IgSeq>.