Human antibody immune responses are personalized by selective removal of MHC-II peptide epitopes

Authors

Matias Gutiérrez-González[1], Ahmed S. Fahad[1], Matt Ardito[2], Padma Nanaware[3], Liying Lu[3], Erica Normandin[4], Bharat Madan[1], Jacob Tivin[2], Emily Coates[4], Amy R. Henry[4], Farida Laboune[4], Barney S. Graham[4], Daniel C. Douek[4], Julie E. Ledgerwood[4], John R. Mascola[4], William D. Martin[2], Lawrence J. Stern[3], Annie S. De Groot[2,5], Brandon J. DeKosky[1,6,*]

Affiliations

[1] Department of Pharmaceutical Chemistry, The University of Kansas, Lawrence, KS 66044, USA

[2] EpiVax, Inc., 188 Valley Street, Suite 424, Providence, RI 02909, United States

[3] Department of Pathology, University of Massachusetts Medical School, Worcester, MA, USA

[4] Vaccine Research Center, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland, USA

[5] Center for Vaccines and Immunology, University of Georgia, Athens, Georgia, USA

[6] Department of Chemical Engineering, The University of Kansas, Lawrence, Kansas, USA

[*]Corresponding author. Email: dekosky@ku.edu

Summary

Human antibody responses are established by the generation of combinatorial sequence diversity in antibody variable domains, followed by iterative rounds of mutation and selection via T cell recognition of antigen peptides presented on MHC-II. Here, we report that MHC-II peptide epitope deletion from B cell receptors (BCRs) correlates with antibody development *in vivo*. Large-scale antibody sequence analysis and experimental validation of peptide binding revealed that MHC-II epitope removal from BCRs is linked to genetic signatures of T cell help, and donor-specific antibody repertoire modeling demonstrated that somatic hypermutation selectively targets the personalized MHC-II epitopes in antibody variable regions. Mining of class-switched sequences and serum proteomic data revealed that MHC-II epitope deletion is associated with antibody class switching and long-term secretion into serum. These data suggest that the MHC-II peptide epitope content of a BCR is an important determinant of antibody maturation that shapes the composition and durability of humoral immunity.

Keywords

B cell development, antibodies, B cell sequencing, somatic hypermutation, MHC-II peptide epitopes

Highlights

- Antibody somatic hypermutation selectively removes MHC-II peptide epitopes from B cell receptors.
- Antibodies with lower MHC-II epitope content show evidence of greater T cell help, including class-switching.
- MHC-II peptide epitope removal from a BCR is linked to long-term antibody secretion in serum.
- MHC-II genotype provides a personalized selection pressure on human antibody development.

Introduction

Human antibody adaptive immune responses are somatically generated by a Darwinian selection process via the generation of high genetic diversity in B lineage cells, followed by iterative rounds of selection with continued diversification. As B cells develop, first heavy chain V-(D)-J recombination occurs, followed by the light chain V-J recombination, to achieve tremendous combinatorial antibody diversity. The selection of antibodies with optimal characteristics from this highly diverse pool is achieved by several well-described mechanisms. First, self-reactive antibodies are negatively selected prior to the generation of the fully mature B cells (also called the naïve B cell population) [1]. Next, B cells migrate to germinal centers and capture foreign protein antigens via B cell receptor (BCR)-mediated endocytosis and present antigen-derived peptides on Major Histocompatibility Class II (MHC-II) to CD4+ helper T cells in the course of classical T cell-dependent antibody maturation [2, 3]. In this process, captured antigen and BCR are endocytosed together and shuttled into the MHC-II peptide processing pathway for cell surface presentation as linear peptides in the peptide-binding grooves of MHC-II proteins [4, 5]. T cells recognize the peptides displayed on MHC-II proteins via T cell receptor (TCR) interactions. The display of peptide:MHC-II (pMHC-II) on B cells provides the critical molecular targets for the TCRs of activating CD4+ helper T cells to recognize and provide stimulatory signals that induce somatic hypermutation, antibody class-switching, and eventual transition to plasmablasts/plasma cells for long-lived antibody production [3, 5].

Despite decades of study related to B cell developmental checkpoints, several critical questions remain in B cell development mechanisms. In particular, it is unclear why only some of the antibodies that bind to foreign antigens with high affinity are selected for clonal expansion, class-switching, and maturation to plasma cells. The humoral immune compartment is highly polarized

and has capacity to contain relatively few (<10,000) representatives of unique antibody clones at a concentration above their affinity constant (K_D); the vast majority of the $>10^7$ unique antibody sequences present in our cellular immune repertoires are not present in serum at an adequate concentration for functional activity [6, 7]. These data also suggest that the memory B cell (mBC) population targets a broader range of antigens than are recognized by serum antibodies [6, 8]. Plasma cells constitute the last stage in B cell development, when plasma cells stop dividing, downregulate surface MHC-II expression, and can persist in bone marrow and secrete antibodies continuously for many years. It remains unclear what molecular mechanisms lead to robust selection for long-lived serum antibodies versus memory B cell persistence in the cellular repertoire, although available evidence strongly suggests that some type of B cell imprinting process determines B cell fate [9-12].

Surface display of antigen-derived MHC-II epitopes is one critical determinant of B cell fate due to the need for B cells to obtain help from antigen-specific CD4+ helper T cells. The affinity of antigen peptides for binding to MHC-II plays a major role in regulating immune responses to foreign proteins, including monoclonal antibody drugs [13-15]. MHC-II molecules are encoded by three human leukocyte antigen (HLA) loci: HLA-DR, -DQ, and -DP. Of these, HLA-DR is the most polymorphic [16], and is usually expressed at higher levels [17, 18]. It is unclear why anti-antibody (or anti-idiotype) immune responses are not highly prevalent due to the very high diversity of somatically mutated human antibodies, including the substantial untemplated diversity of CDR3 regions, although highly homologous antibody sequences (including T regulatory cell epitopes, or Tregitopes) have been suggested to play a role in reducing anti-antibody immunity [19-21]. Methods for computational MHC binding prediction have continually improved in recent years, particularly for HLA-DR [22], and recent high-throughput proteomic elution data have

92 provided large experimental datasets as benchmarks to enhance prediction accuracy [23, 24].
93 Moreover, peptides derived from BCR proteins are commonly detected as self-peptides in MHC-
94 II elution experiments [25-27]. Despite these advances, the landscape of potential MHC-II peptide
95 epitope content in healthy antibody repertoires has not yet been evaluated, partially due to the
96 relevantly recent invention of methods for repertoire-scale analysis of complete, natively paired
97 antibody heavy and light chains [28, 29].

98 Given the high importance of MHC-II epitopes in controlling B cell selection via MHC-II
99 interactions, we hypothesized that MHC-II epitopes in BCR-encoded peptides could influence
100 antibody selection and maturation. To explore these features, we analyzed potential MHC-II
101 epitopes in the variable region sequences of human antibody repertoires to understand how
102 antibody repertoire features correlate with MHC-II epitopes and may be influenced by a person's
103 unique HLA gene profile. Our analysis of seven natively paired heavy and light chain antibody
104 repertoires from healthy human donors revealed that antibodies show hallmarks of selective
105 removal of MHC-II peptide epitopes via somatic hypermutation throughout antibody
106 development. By studying the MHC-II epitope content of BCRs along with molecular signatures
107 of CD4+ T cell help (e.g., somatic hypermutation, antibody isotype class-switching, and serum
108 proteomic detection), we found that the preferential deletion of MHC-II epitopes from the antibody
109 variable regions was associated with B cells achieving the critical T cell help needed for robust
110 and long-lived antibody immune memory. These data reveal a new mechanism regulating human
111 antibody immunity and provide insights for the design of new vaccines and therapeutics associated
112 with long-term immune memory.
113

Results

We began by characterizing MHC-II peptide epitope content in healthy human antibody variable region sequences using high-throughput computational MHC-II peptide epitope prediction. We collected seven paired heavy and light chain datasets from antigen-experienced B cells of healthy donors, with a total of 250,645 high-quality consensus sequences of natively paired heavy and light chain antibody lineages. We analyzed these immune repertoires using multiple pMHC-II affinity prediction algorithms to determine how the features of antibody development correlated with changes in potential MHC-II peptide epitope content of BCRs (**Fig. 1A**). First, we used the commercially available EpiMatrix MHC-II epitope prediction platform to characterize aggregate predicted HLA-DR epitope content based on eight human HLA-DR gene supertypes. EpiMatrix reports a T cell epitope score, where a higher score indicates higher content of putative MHC-II peptide epitopes within the analyzed protein sequence [30]. Strikingly, we noted that all donors showed reduced MHC-II peptide epitope content (i.e., reduced EpiMatrix scores) that was correlated with increasing somatic hypermutation (SHM), and the correlation was statistically significant in all donors (Spearman correlation test, adjusted p-value < 0.05). These data demonstrated that SHM reduces pMHC-II affinities in antibody peptides at a repertoire level (**Figs. 1B**, **S1A**). Subsequent analysis of antibody repertoire data fractionated by paired antibody heavy and light chain V-genes showed that changes in MHC-II peptide epitope content were concentrated in certain V-gene combinations (**Figs. 1C**, **1D**, **S1B**, **S1C**), and each V-gene shows a different initial distribution of MHC-II peptide epitope content (**Fig. S2**). While each donor showed a unique pattern of V-genes with the highest reductions in MHC-II peptide epitope content, some V-genes were repeatedly observed as statistically significant across donors. Nearly all statistically

136 significant V-gene changes showed removal of MHC-II peptide epitopes as SHM levels increased
137 (**Fig. 1, Fig. S1**).

138 We next sought to understand the molecular drivers of decreased MHC-II peptide epitope
139 content based on personalized HLA gene profiles. We applied the netMHCIIpan algorithm to
140 model individual MHC-II binding affinities of every peptide in our antibody datasets, according
141 to the known HLA gene profiles that were available for donors 1 to 5 (**Fig. 2**) [31]. We found that
142 several predicted high-affinity HLA-DR-binding peptides were encoded by antibody germline
143 genes, and these MHC-II peptide epitopes were being mutated during antibody somatic
144 hypermutation (**Fig. 2A**, **S3**). Thus, somatic hypermutation caused deletion of MHC-II peptide
145 epitopes from B cell receptors, and the correlations that we observed in **Figure 1** could be traced
146 to specific peptides with a high germline (unmutated) affinity for the donor's MHC-II genes. When
147 comparing V-genes between germline and high SHM antibody sequences, the removal of high-
148 affinity MHC-II peptide epitopes by SHM was readily apparent (**Figs. 2B**, **2C**, **2D**, **2E**, **S4A**, **S4B,**
149 **S4C, S4D**). Thus, the reduction in MHC-II peptide epitope content that we observed with
150 increasing SHM was predominantly driven by the deletion of high-affinity peptides that had been
151 present since the earliest stages of antibody development.

152 We next sought to experimentally confirm the loss of peptide affinities that were observed via
153 *in silico* affinity modeling. We validated peptide affinity changes for key driver epitopes of MHC-
154 II epitope deletion using *in vitro* pMHC-II affinity assays (**Fig. 2G**). These data showed that, as in
155 prior studies, large-scale *in silico* predictions of peptide binding to MHC-II are generally accurate,
156 especially for the DRB1 gene used in the current study [31]. Next, we mined the Immune Epitope
157 Database (IEDB) to identify antibody peptides eluted from human MHC-II in immunopeptidomic
158 assays to see if our detected peptides successfully process inside endosomes and displayed on

MHC-II *in vivo* [32, 33]. We identified a large number of naturally-processed peptides that were experimentally confirmed in IEDB and appeared to be targets of preferential mutations that reduce peptide affinity via SHM, including peptides that were mutated in antibody sequence data such as IGHV3-23$_{73\text{-}93}$ and IGHV1-18$_{73\text{-}92}$ (**Fig. 2E**, **Fig. S5A**, **S5B**). Interestingly, donor antibody repertoires also contained some of the same peptides that were eluted from HLA-DP and HLA-DQ molecules(**Fig. S5B)**; numerous IEDB-validated peptides overlapped between DRB and DP/DQ binding **(Fig. S5C**). Thus we confirmed that some of the key peptides analyzed in our study are presented on human MHC-II in previously reported proteomic datasets.

Once we realized that antibody peptides with high affinity for DRB binding were being targeted for mutations and MHC-II epitope removal, we shifted our focus to patient-specific analyses to explore these high-affinity MHC-II peptide epitopes encoded by germline IGHV and IGKV/IGLV genes (**Fig. S6**). MHC-II peptide epitopes often require multiple amino acid matches with appropriate spacing for binding to the MHC-II cleft, and we reasoned that the reduced T cell content observed with increasing SHM could be introduced as an indirect consequence of SHM mutational pattern preferences, rather than by active selection pressure. To test this alternate hypothesis, we reasoned that if MHC-II peptide epitopes are removed by SHM to a greater degree in experimentally-derived patient repertoires than in carefully matched *in silico* simulations (which account for SHM DNA motif targets, but not for any HLA-dependent MHC-II peptide epitope selection pressure), then we could conclude that MHC-II epitope removal was a result of active selection *in vivo.* We thus began large-scale *in silico* experiments simulating antibody repertoires using established somatic hypermutation models (**Figs. 3, S7**). We compared two different SHM models to the experimentally-derived sequence data: one *in silico* SHM model customized by the 5-mer DNA base targeting patterns in each individual patient's experimentally-derived antibody

repertoire, and a second *in silico* model based on 5-mer DNA bases in universal out-of-frame human B cell receptor data. Our out-of-frame model controls for the nucleotide targeting preferences of human activation-induced cytidine deaminase (AID), the enzyme responsible for SHM, as antibody DNA sequences with out-of-frame V-(D)-J junctions cannot be expressed or functionally selected, and it was constructed from approximately 56,000 genomic out-of-frame antibody sequences compiled from 114 donors [34, 35]. In contrast, the patient-specific in-frame antibody SHM model encompassed local AID 5-mer nucleotide preferences, in addition to biophysical restrictions on permissible DNA/amino acid mutations in functional B cell receptors, as along with any positive selection for 5-mer DNA mutations within a patient's immune system. By comparing MHC-II peptide epitope deletion metrics in experimentally-derived antibody data versus *in silico* simulations, we found that in most cases the replacement-silent (R-S) model and universal out-of-frame (OoF) models showed a lower number of statistically significant IGHV and IGKV/IGLV gene pairs with decreased MHC-II peptide epitope content compared to experimentally-derived donor data (**Figs. 3C, 3D, S8**). Often, one donor HLA-DRB1 allele showed a greater degree of MHC-II epitope loss than the other allele. Comprehensive SHM computational models did not recreate the same degree of personalized MHC-II peptide epitope deletion observed in experimentally-derived donor data (**Fig. 3E**), confirming that the SHMs deleting pMHC-II epitopes *in vivo* were functionally selected and would not arise simply as a consequence of AID targeting preference. These data demonstrate the SHM preferentially deletes pMHC-II epitopes from BCR variable regions.

Next, we tested whether MHC-I peptide epitopes were also being preferentially deleted. We predicted peptide K_D for donor-matched MHC-I molecules to compare relative MHC-I and MHC-II peptide affinity changes as a result of antibody somatic hypermutation. Because some peptides

bind to both MHC-I and MHC-II, we binned peptide epitopes according to binding for MHC-I, MHC-II, or both MHC-I+MHC-II to determine how T cell epitope removal via SHM affected the different MHC classes separately. In contrast to our analyses of MHC-II, the peptides predicted to bind to MHC-I showed very few statistically significantly changes when removing peptides that were shared epitopes with MHC-II ($p<0.001$, Wilcoxon rank sum test, **Fig. 3F**, *upper panel*). Moreover, unique MHC-I peptides showed a weaker correlation between K_D fold-change and SHM compared to shared MHC-I/MHC-II peptides ($p < 0.05$, Wilcoxon rank sum test, **Fig. 3F**, *lower panel*). In contrast, we observed no significant difference between shared MHC-I/MHC-II peptides and MHC-II-restricted peptides. These data demonstrated that peptides binding to MHC-II were targeted for preferential deletion from antibody variable regions via SHM, but peptides that bound to MHC-I did not show similar preferential removal via SHM. Thus, SHM appears to selectively target MHC-II peptide epitopes for deletion.

Next, we analyzed our data by antibody isotype bins to further understand how MHC-II peptide epitope removal correlated with key markers of B cell development and CD4+ T cell help. Like SHM, antibody class switching is induced by AID and is strongly correlated with CD4+ T cell help obtained via pMHC-II:TCR interactions [36]. We found that the greatest correlation of MHC-II peptide epitope deletion with SHM was observed in class-switched IgG and IgA repertoires (**Figs. 4A, 4B**). Analysis of class-switched data provided a clear association between MHC-II peptide epitope removal from antibody gene sequences with antibody class-switching, an important hallmark of effective CD4+ T cell help.

Finally, we sought to understand how MHC-II peptide epitope content in BCRs is associated with elicitation of antibodies into the serum immune compartment. Serum antibodies are secreted by plasmablasts and long-lived plasma cells, and recent advances in antibody sequencing,

228 computational mining of BCR NGS data, and proteomic mass spectrometry have enabled the
229 identification of individual antibody clonal lineages in human serum [7, 37-39]. We performed
230 HLA-DRB1 MHC-II peptide binding affinity predictions using cellular-derived and serum-
231 derived antibody repertoire data from recent studies of influenza vaccination [40, 41]. We found
232 that antibodies identified in serum exhibited lower MHC-II peptide epitope content than the
233 antibodies present in the donor-matched cellular repertoire (**Figs. 4C**). Thus, a lower MHC-II
234 epitope content in the BCR was associated with B cell maturation to plasmablasts and plasma cells
235 for secretion of antibodies at appreciable concentrations into the blood compartment. We also
236 tracked the MHC-II peptide epitope content of anti-influenza antibodies with different temporal
237 persistence in human serum. We found that antibodies detected in serum at multiple time points
238 showed lower MHC-II peptide epitope content relative to antibodies observed only at a single time
239 point (**Fig. 4D**), implying that lower MHC-II peptide epitope content is associated with longer
240 antibody-secreting cell life spans *in vivo*. These analyses of serum antibody data, together with our
241 observations that class-switched IgG and IgA compared with donor-matched IgM repertoires,
242 suggested that human BCRs are functionally selected to remove MHC-II epitopes via somatic
243 hypermutation as a component of natural human antibody development.
244
245 **Discussion**
246 This study reveals that antibody maturation and somatic hypermutation are closely associated
247 with the removal of MHC-II peptide epitope content in antibody and BCR molecules. We observed
248 strong selection for the removal of MHC-II peptide epitopes by SHM in class-switched BCRs, and
249 also in antibodies secreted persistently in human serum. These data reveal a previously unreported

mechanism for the personalization of antibody immune responses via functional selection according to each individual's unique HLA MHC-II gene profile (**Fig. 4E**).

Our study employed *in silico* and statistical techniques using computational HLA-DRB1 MHC-II peptide binding predictions, which have been demonstrated to be generally accurate in several recent studies [42, 43]. To validate *in silico* results, we confirmed our findings with experimental validation of key MHC-II peptide predictions (**Fig. 2G**), by analysis of eluted peptides reported in the IEDB (**Fig. S5**), and by retrospective analysis of serum antibody data reported in prior studies (**Figs. 4C, 4D**) [40]. We focused on HLA-DRB1 genes, which have the highest observed prevalence among MHC-II receptor genes in immunopeptidome assays and IEDB datasets, and are the best-characterized MHC-II receptor genes for computational peptide affinity predictions. We note that not all donors showed the same extent of HLA-DRB1 genetic selection (**Fig. 3**). Variability between individuals could result from the influence of HLA-DP and HLA-DQ genes providing additional MHC-II epitope selection pressures, that were not encompassed by our study of HLA-DRB1 peptide epitopes. Many T-dependent antigens can elicit HLA-DP and HLA-DQ responses, although we also note that some peptide binding overlap exists between different HLA molecules. Improved *in silico* tools for predicting peptide processing, as well as the incorporation of HLA-DP and HLA-DQ modeling, will enhance future large-scale studies of pMHC-II content in antibody repertoires.

Our data suggest that reduced MHC-II epitope content in BCRs could be an important correlate of durable human antibody immunity. These findings are supported by our observations that BCRs in class-switched isotypes (e.g., IgA and IgG that require high levels of T cell help) show stronger rates of MHC-II peptide epitope removal than the IgM compartment (**Figs. 3A**, **3B**, **S9**). We also observed that lower BCR MHC-II peptide epitope content was associated with higher serum

antibody prevalence, suggesting that HLA-DRB1 peptide epitope deletion may support B cell trafficking to a long-lived plasma cell niche by enhancing the acquisition of T cell help (**Fig. 4E**) [44]. Certain heavy and light chain V-genes showed higher rates of HLA-DRB1 peptide epitope removal than other V-gene pairs (**Fig. 2B**), reflecting the different baseline levels of MHC-II peptide epitopes in antibody germline genes (**Figs. S2, S3**). These data suggest that MHC-II epitope deletion is targeted toward those V-genes that contain germline-encoded MHC-II epitopes, as would be expected to occur in a functional selection mechanism. Low MHC-II epitope content in a B cell receptor could help that B cell present more MHC-II epitopes from antigen, thereby enhancing CD4+ T cell help for that B cell (**Fig. 4E**). This selection mechanism offers several important advantages *in vivo*. First, selection of lower MHC-II peptide epitope content reduces the propensity of an individual's secreted antibodies to induce CD4+ T-cell dependent anti-idiotype antibody immune responses in non-templated regions (e.g., from pMHC-II derived from CDR3 loops, or that may arise as a result of SHM), reducing the risk of immune responses to somatically generated antibody proteins. Perhaps more importantly, low MHC-II peptide epitope content in an antibody could help dendritic cells present a greater fraction of MHC-II peptides derived from antigen (and fewer peptides derived from the BCR) after immune complex capture and processing. These findings have important implications for vaccine design and antibody drug therapeutics. As one example in HIV vaccine development, where targeted elicitation of specific lineage mutations are being pursued, these data suggest an important HLA-dependent selection pressure guiding SHM, and that antibody mutations may accumulate differently in patients with different HLA gene profiles due to MHC-II-based selection pressure [45, 46]. In addition, our findings lend further support to ongoing efforts to mitigate anti-drug antibody responses by removal of MHC-II peptide epitopes from the monoclonal antibody drug variable regions [47, 48]

One limitation of our study is that we analyzed only the HLA-DRB1 gene, due to its high representation in quantitative peptide:MHC-II proteomic elution studies and established predictive peptide binding accuracy [31]. Future studies will further analyze human HLA-DP and HLA-DQ genes, which have lower peptide elution prevalence in immunopeptidomic assays but still make important contributions to human immunity. We will also study the influence of SHM on previously reported regulatory MHC-II epitopes [19]. We recognize that T-cell independent B cell activation pathways also exist (especially for antigens with repeated structural motifs and that lack MHC-II epitopes, for example the regularly ordered polysaccharides in bacterial cell walls). However, most foreign antigens generate T-dependent immunity and we anticipate that the majority of human B cells are selected via T-dependent mechanisms. Follow-up studies will investigate dysregulation of MHC-II antibody selection pathways for specific antigens (including T-dependent and T-independent) in mouse models, and similar analyses of clinical samples from patients with autoimmune diseases known to disrupt antibody developmental checkpoints [49-51].

In summary, here we identified a previously unreported correlation between lower MHC-II peptide epitope content in BCRs and the signatures of T cell help throughout antibody development. These data suggest that an MHC-II-based selection pressure influences antibody selection *in vivo*, and may represent an important factor shaping the durability of serological immunity in humans [9, 44, 52].

Acknowledgments

We thank L. Santambrogio for insightful comments and Bill Flegel for help with HLA typing. This work was supported by NIH grants 1R01AI141452, R21AI143407, R21AI144408, and DP5OD023118. We also thank Grace L. Chen, MD, Adam DeZure, MD, Nina Berkowitz, Maria Burgos Florez, Abidemi Ola, Cynthia Hendel, Floreliz Mendoza, Ingelise Gordon, Jamie Saunders, Jennifer Cunningham, Kathy Zephir, Lasonji Holman, Laura Novik, Pamela Costner, Sarah Plummer, Xioalin Wang, William Whalen, and Catina Evan from the Vaccine Research Center (VRC) Clinical Trials program for help coordinating and collecting samples.

Author Contributions

M.G.G., A.S.F, M.A, P.N, W.D, L.J.S, A.S.G and B.J.D., designed the experiments; M.G.G, A.S.F., M.A, P.N., L.L, E.N, B.M, J.T, E.C, A.R.H, and F.L performed the experiments; M.G.G, A.S.F., M.A, P.N., E.N, B.M, J.T, D.C.D, J.E.L., B.S.G, J.R.M., W.D.M., L.J.S., A.S.G, and B.J.D. analyzed the data; and M.G.G. and B.J.D. wrote the manuscript with feedback from all authors.

Disclosures

M.A., J.T., W.M., and A.S.G. are employees of Epivax, Inc., which commercializes the EpiMatrix prediction tool.

Methods

Resource Availability

Lead Contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Dr. Brandon DeKosky (dekosky@ku.edu).

Materials Availability

No new reagents were generated in this study.

Data and Code Availability

Raw NGS antibody sequence data used for the study are deposited in the NCBI Short Read Archive under accession numbers: XXXX, XXXX, XXXX, XXXX.

Experimental Model and Subject Details

Human Subjects

For cellular antibody MHC-II content, a total of seven datasets were analyzed. These include previously published data (Donors 1,2,4,6 and 7) [53][Fahad, DeKosky et al., *Front. Immunol.*, Accepted 2021], and new unpublished datasets (Donors 3 and 5). All human samples were collected under the Vaccine Research Center's (VRC)/National Institutes of Allergy and Infectious Diseases (NIAID)/ National Institutes of Health (NIH) sample collection protocol, VRC 200 (NCT00067054) in compliance with the NIH IRB approved procedures. All subjects met protocol eligibility criteria and agreed to participate in the study by signing the NIH IRB approved informed

355 consent. Research studies with these samples were conducted by protecting the rights and privacy
356 of the study participants.

357 For cellular and serum antibody datasets, data was retrieved from previously published Ig-Seq
358 and BCR-Seq data [40, 41]. The first dataset consists of IgG/A/M from B cell receptors and serum
359 IgG antibody sequences that were obtained from donors after influenza vaccination, and is
360 available in MassIVE (https://massive.ucsd.edu/ProteoSAFe/static/massive.jsp) under accession
361 ID MSV000080184. The published dataset comprise serum antibodies that were purified by
362 affinity chromatography with inactivated components of the 2011–2012 IIV3 vaccine at days 0,
363 28 and 180 post-vaccination and analyzed via proteomic mass spectrometry [40]. The second
364 dataset contains clonotypes that were detected in serum as a response to repeated flu vaccinations
365 during several years (MassIVE ID MSV000083120). In this case, the original study contemplated
366 persistent, intermediate and transient categories; which were changed to single observation
367 (transient in the original study) and multiple observations (persistent and intermediate) [41].

368

369 ***Cell Lines***

370 Drosophila S2 cells were grown at 100 rpm in 27 °C incubator, with SF900 II serum-free medium
371 (Thermo Fisher cat #10902096) and penicillin-streptomycin (100 U/ml Thermo Fisher cat #
372 15140148). HLA-DR1 protein production was induced by addition of 1mM CuSO4 and culture
373 supernatants were collected after 6 days.

Method Details

Emulsion Overlap Extension RT-PCR

Natively paired antibody heavy and light chains sequencing was carried out as previously described [54]. B cell isolation from cryopreserved PBMCs was carried out using Memory B cells Isolation Kit (MACS/Miltenyi Biotec, Bergisch Gladbach, Germany). Next, cells were stimulated *in vitro* using IL-2, IL-21, and co-cultured with 3T3-CD40L fibroblasts for 5 days [55]. Following cell stimulation, single cells were captured in emulsion droplets, lysed, and their mRNA captured with oligo(dT)-coated magnetic beads. Native heavy and light chains were obtained by an overlap-extension RT-PCR and resulting cDNA libraries were sent for Illumina sequencing.

Antibody Sequence Analysis

Illumina 2x300 bp sequencing was analyzed as previously described [55]. Briefly, Illumina reads were quality filtered and aligned into full reads. V(D)J annotation was carried out using IgBlast [56], and productive sequences were paired by CDR-H3 match. Isotype assignment was carried out by matching of constant region sequences to isotype barcodes. Consensus sequences of paired heavy and light chain clusters were generated as previously reported to remove NGS errors prior to MHC-II peptide epitope content predictions [29, 54, 57].

For serum and cellular antibody repertoire data, reported protein sequences were mapped to clonotypes by generating consensus VH sequences using the reported cluster identifier in the data, with a 80% identity threshold using usearch version 6.1.544 [58], and V(D)J annotation was carried out using IgBlast. Serum antibodies were retrieved from BCR-seq data by matching reported CDR-H3 sequences with the available BCR-seq data.

MHC Peptide Epitope Content Prediction

The EpiMatrix tool (EpiVax, Rhode Island, USA) was used for aggregate MHC-II peptide epitope / T cell epitope predictions [30]. EpiMatrix uses main HLA II DRB1 "supertypes" to predict overall protein epitope content [59]. Higher scores in the EpiMatrix output indicate a higher probability of T cell dependent immunogenicity of foreign protein antigens. The alleles analyzed were DRB1*01:01, DRB1*03:01, DRB1*04:01,DRB1*07:01, DRB1*08:01, DRB1*09:01, DRB1*11:01, DRB1*1302 and DRB1*15:01. The output data includes aggregate epitope score by chain, normalized by length, and total antibody epitope content. We used the complete antibody epitope content, not corrected for Treg epitope content as a measure of immunogenicity. Spearman Rho correlations between complete antibody epitope scores and SHM were calculated, and a linear model was fitted to calculate slopes.

For individual MHC-II peptide epitopes, netMHCIIpan 3.1 with default options was used, working with a subset of 38 representative HLA-DRB1 molecules DRB1*01:01, DRB1*01:02, DRB1*01:03, DRB1*03:01, DRB1*03:02, DRB1*04:01, DRB1*04:02, DRB1*04:03, DRB1*04:04, DRB1*04:05, DRB1*04:06, DRB1*04:07, DRB1*04:08, DRB1*07:01, DRB1*08:01, DRB1*08:02, DRB1*08:03, DRB1*08:04, DRB1*09:01, DRB1*10:01, DRB1*11:01, DRB1*11:02, DRB1*11:03, DRB1*11:04, DRB1*12:01, DRB1*12:02, DRB1*13:01, DRB1*13:02, DRB1*13:03, DRB1*13:05, DRB1*14:01, DRB1*14:02, DRB1*14:06, DRB1*15:01, DRB1*15:02, DRB1*15:03, DRB1*16:01, and DRB1*16:02 [31]. netMHCIIpan output was parsed using pandas for further processing. The equilibrium dissociation constant (K_D) or rank of 15-mers was considered for analysis. As a consequence of this, higher netMHCIIpan K_D reflect a lower level of MHC-II peptide epitope content. Peptide K_D's were predicted for donor repertoire MHC-II peptide epitopes, and for a database of germline heavy and

light chain V genes. Germline (unmutated) peptides with K_D <1,000 nM for tested alleles were used as a search database for peptides in antibody repertoires by matching V-gene usage and index position within the protein [60]. Peptides hits were grouped according to parent antibody, considering heavy or light chains, and the geometric mean of K_D fold-change between donor and germline peptides was calculated. The same procedure for calculating the geometric mean of K_D fold-change was applied in a grouping by complete antibody (paired heavy and light chain) sequences. Next, data was grouped by common VH gene and VK/L gene pairs, and Spearman correlation between the geometric mean of K_D fold-change and SHM was calculated for each allele.

For position-based MHC-II peptide epitope content, we selected the top 5% of sequences in terms of SHM burden from the selected V-gene subset, and the geometric mean of the rank of each peptide at the same position was calculated for a subset of the antibody repertoire. The same approach was carried out for germline sequences. This information was also retrieved for selection of candidate peptides, shown as logo plots.

For MHC-I peptide epitope prediction, netMHCpan 4.1 [24] was used, using donor-matched HLA-A, -B and -C genes and predicting binding affinity for 9-mers. Germline (unmutated) peptides with K_D<500 nM for tested alleles were used as a search database for peptides in antibody repertoires by matching V-gene usage and index position within the protein [61].

Determination of MHC-I/MHC-II shared epitopes was done by matching 9-mer peptides (MHC-I peptide epitopes, 6 HLA-I alleles) into 15-mers (MHC-II peptide epitopes, 2 HLA-II DRB1 alleles). Epitopes with a positive match were considered as shared epitopes and the unmatched peptides were considered as exclusively MHC-I or MHC-II peptide epitopes. From MHC-I or MHC-II all (no matching), shared and unique databases, peptides were aggregated into

443 parent antibodies as described previously. The number of significant VH:VL gene pairs was
444 compared for MHC-I and MHC-II peptides separately.
445
446 ***In silico Repertoire Modeling and Analysis***
447 Repertoire modeling was carried out with immuneSIM for V(D)J recombination modeling,
448 and ShaZam for SHM modeling [62, 63]. First, V gene frequency was extracted from donor data
449 and used to build V gene distribution table. This table was used to modify the vdj_list parameter
450 in the immuneSIM function, which controls the frequency of different V-genes in the modeled
451 repertoire. D and J gene distributions were maintained in their default settings. Using the custom
452 V gene distribution frequencies, a naïve repertoire of the same size of the parent repertoire was
453 generated, using the immuneSIM function, with no mutations allowed. Heavy and light chains
454 were generated separately. The naïve dataset was mutated using SHM models from repertoire data
455 generated by ShaZam, using the createTargetingModel function, which allows the determination
456 of a 5-mer targeting model based on sequence data and gene annotation. Two SHM models were
457 generated, one considering donor antibody repertoire data, and other built from out-of-frame (OoF)
458 sequences from genomic antibody sequencing studies, comprising 115 donors, and 56,278
459 sequences [34, 35]. Briefly, data from two genomic antibody sequencing studies were retrieved
460 for the generation of a new OoF targeting mutational model. One dataset comprised large-scale
461 genomic BCR sequencing of healthy donors (NCBI BioProject Accession number PRJNA491287)
462 and another from a genomic B cell sequencing study of CAPRISA cohort donor CAP256
463 (Accession number SRP124539) [35]. After aggregating data form the 115 donors, all NGS reads
464 were quality filtered and aligned using MiXCR [64] to identify a combined total of 56,278 out-of-
465 frame antibody sequences that were used to build a mutational targeting model using ShaZam.

First, nucleotide sequences were analyzed with IgBlast , and the output was parsed into a Shazam-compatible database using Change-O [62]. Compiled out-of-frame aggregate donor data and personalized in-frame donor antibody databases were transformed individually into a 5-mer mutational targeting model using the create TargetingModel function from ShaZam to generate the out-of-frame model (OoF) or the personalized replacement-silent mutational models, respectively. After mutational model generation, each sequence was mutated individually using the shmulateSeq function from ShaZam. The number of mutations per sequence was selected to match the distribution of SHM observed in personalized donor data, on an individual donor dataset basis. After each repertoire generation, sequences were annotated using IgBlast and paired following donor distributions of SHM between heavy and light chains. Each simulation was compared to parent repertoire to verify appropriate V-gene distributions and SHM content to match the experimental data. From the IgBlast output, amino acid sequences were extracted for MHC-II peptide epitope prediction using netMHCIIpan.

We generated 30 complete simulated antibody repertoires for each of donors #1-5, for each of the two mutational models generated as described in the previous paragraphs. Thus we simulated a total of 150 personalized replacement-silent antibody repertoires *in silico*, and 150 out-of-frame mutational model antibody repertoires, for a total of 300 simulated repertoires with an average of 27,000 antibodies each. Thus, we generated approximately 8,100,000 antibodies *in silico* and used the resulting data to explore hypotheses related to mutational targeting in experimental antibody data compared with simulated datasets.

Personalized MHC-II Peptide Epitope Content Analysis

netMHCIIpan MHC-II peptide epitope content predictions were analyzed following the same procedures as donor repertoires. Isotype assignments for modeled repertoire subsets were made by matching the SHM distribution by isotype observed in donor-matched experimental data. Next, heavy and light chain V-gene pairs with n<9 antibodies were removed from analysis. To compare MHC-II peptide epitope removal in experimental and computationally modeled antibody repertoire data, isotype-switched, statistically significant VH:VKL gene pairs were retrieved from donor data and modeled repertoires. These pairs in donor and modeled data were matched in modeled and donor data, respectively. The number of significant VH:VKL pairs was also compared between donor and modeled repertoires. For modeled data, the average number of statistically significant pairs was calculated by dividing the total number of statistically significant pairs across all modeled repertoires by the number of repertoires (n=30 modeled repertoires per donor & model type). Volcano plots of modeled repertoires in Figure 3D and S8 were made by selecting the modeled replicate with an average Spearman Rho value closest to the median Spearman Rho value of the 30 modeled repertoires from that donor and model type. Statistical significance was determined by calculating the Spearman correlation and retrieving p-values, that were adjusted for multiple comparisons using the Benjamini-Holchberg method. For donor matched vs. mismatched HLA comparisons in Figure S8, the mean Spearman Rho for all V-gene pairs was calculated for each of the 38 analyzed HLA alleles. The mean Spearman rho for all 38 alleles was shown, with each allele colored according to its supertype family.

IEDB Data Mining

We searched for experimentally validated antibody-derived peptides in the Immune Epitope Database (www.iedb.org). The search was limited to linear epitopes from human origin, with

511 experimental validation of the binding to MHC II, and from immunoglobulin sequences. After
512 removing peptides derived from constant regions and T cell receptors, and selecting assays for
513 HLA-DRB1 molecules, a final database of 448 peptides was obtained. These peptides were
514 searched for matches in the germline and donor database. As IEDB-validated peptides are of
515 variable length and whereas our germline/donor peptide databases are exclusively 15-mers, the
516 presence of validated peptides as substrings in 15-mers was considered a match that is certain to
517 contain the MHC-II peptide binding core. The same procedure was carried out for HLA-DP and
518 HLA-DQ molecules, and a total of 187 peptides were found in the IEDB-validated peptide
519 database.
520
521 ***Serum Antibody Analysis***
522 From consensus sequences, MHC-II peptide epitope prediction was carried out using
523 netMHCIIpan and processed as previously indicated. For the cellular and serum antibody analysis
524 in Fig. 4C [40], data were mined from a recent study of serum antibody prevalence in healthy
525 donors and antibodies with an total extracted-ion chromatogram (XIC) peak area on the top 50%
526 for any of the time points analyzed.
527 Since the HLA alleles for these donors are unknown, all 38 alleles were considered for analysis.
528 To this end, mean K_D was compared between cellular and serum repertoires, and between multiple
529 observation and single observation antibodies .
530
531 ***HLA-DR1 Binding Assay***
532 HLA-DR1 (DRA*01:01/DRB1*01:01) extracellular domains were expressed in Drosophila
533 S2 cells and purified by immunoaffinity chromatography with LB3.1 antibody followed by

Superdex200 (GE Healthcare) size exclusion chromatography as described [65, 66]. Ig-derived peptides and influenza HA306-318-derived probe peptide Ac-PRYVKQNTLRLAT were synthesized (21st Century Biochemicals, Marlboro, MA). The probe peptide was labeled with Alexa Fluor 488 tetrafluorophenyl ester (Invitrogen, Eugene, OR) through primary amine of K5. Peptide binding was monitored using a fluorescence polarization assay [67]. The DR1 concentration used was selected by titrating DR1 against fixed labeled peptide concentration (25 nM) and choosing the concentration of DR1 that showed ~50% maximum binding. For calculating IC50 values, 100 nM DR1 was incubated with 25 nM Alexa488-labeled HA306–318 probe peptide, in combination with a serial dilution of test peptides, beginning at 100 µM followed by 2-fold dilutions. The reaction mixture was incubated at 37 °C. The capacity of each test peptide to compete for binding of probe peptide was measured by FP after 72 h at 37 °C. FP values were converted to fraction bound by calculating [(FP_sample - FP_free)/(FP_no_comp - FP_free)], where FP_sample represents the FP value in the presence of test peptide; FP_free represents the value for free Alexa488-conjugated HA306–318; and FP_no_comp represents values in the absence of competitor peptide. We plotted fraction bound versus concentration of test peptide and fit the curve to the equation y = bottom + (top – bottom)/(1 +[pep]/IC50), where [pep] is the concentration of test peptide, y is the fraction of probe peptide bound at that concentration of test peptide, IC50 is the 50% inhibitory concentration of the test peptide, top is the maximum fraction of probe peptide bound, and bottom is the minimum fraction of probe peptide bound.

Quantification and Statistical Analysis

Statistical analyses were performed using R. When multiple comparisons were performed, we adjusted the p values using the Benjamini-Horchberg method from the stats package. Sample

557 distributions were compared using the Kolmorogov-Smirnov test from the stats package. All
558 correlations were calculated using the Spearman method from the stats package. Levenshtein
559 distances between donor and germline peptide were calculated using the stringdist package.
560 Peptide binding curve fitting was carried out using the nls() function from the stats package,
561 following the equation y = bottom + (top – bottom)/(1 +[pep]/IC50), as previously described. IC50
562 and standard deviation values were reported. Differences in the mean Spearman correlation for
563 VH:VL matched gene pairs was carried out using a paired t-test from the stats package. Differences
564 in the number of significant VH:VL gene pairs for MHC-I and MHC-II were calculated using the
565 Wilcoxon rank sum test from the stats package.
566
567 **Key Resources Table**
568

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
LB3.1	ATCC	Cat#ATCC® HB-298
Bacterial and Virus Strains		
Biological Samples		
Healthy adult PMBC (Donors 3 and 5)	Vaccine Research Center (VRC)	Sample collection protocol, VRC 200 (NCT00067054)
Chemicals, Peptides, and Recombinant Proteins		
HA306-318-derived probe peptide Ac-PRYVKQNTLRLAT	21st Century Biochemicals, Marlboro, MA	http://www.21stcenturybio.com/custom_peptide_synthesis.htm
Alexa Fluor 488 tetrafluorophenyl ester	Invitrogen, Eugene, OR	Cat#A37570
Ig-derived peptides	21st Century Biochemicals, Marlboro, MA	http://www.21stcenturybio.com/custom_peptide_synthesis.htm
Critical Commercial Assays		
Memory B cells Isolation Kit	MACS/Miltenyi Biotec, Bergisch Gladbach, Germany	Cat#130-093-546

Deposited Data		
Paired VH:VL sequencing from two healthy donors (Donors 6 and 7)	DeKosky, BJ. , Lungu, OI. et al (2015)	BioProject (PRNJA315079)
Paired VH:VL sequencing from three healthy donors (Donors 1, 2 and 4)	Ahmed, F., DeKosky, BJ., et al. (2021, Accepted)	PRJNA682833
Paired VH:VL sequencing from three healthy donors (Donors 3 and 5)	This study	XXXXXXXX
VH sequencing from cellular and serum antibodies (Serum study 1)	Lee, J., Boutz, DR. et al (2016)	MassIVE (MSV000080184)
Longitudinal VH sequencing of serum antibodies (Serum study 2)	Lee, J., Paparoditis, P. et al (2019)	MassIVE (MSV000083120)
Experimental Models: Cell Lines		
Drosophila S2 cells		
Experimental Models: Organisms/Strains		
Oligonucleotides		
See Table S1	McDaniel, JR., DeKosky, BJ,. et al (2016)	McDaniel, JR., DeKosky, BJ,. et al (2016)
Recombinant DNA		
Software and Algorithms		
R version 3.53	R Core Team (2019)	https://www.R-project.org/
Pandas version 0.25.3	The pandas development team (2020)	https://pandas.pydata.org
Tidyverse version 1.3.0	Wickham et al., (2019)	https://www.tidyverse.org
MixCR version 3.012	Bolotin, D.A. (2015)	https://mixcr.readthedocs.io/en/master/index.html
IgBlast version 1.16	Ye, J., et al (2013)	https://www.ncbi.nlm.nih.gov/igblast/
Usearch version 6.1.544	R.C. Edgar (2010)	https://drive5.com/usearch/
netMHCIIpan version 3.1	Andreatta, M., et al (2015)	http://www.cbs.dtu.dk/services/NetMHCIIpan-3.1/
netMHCpan version 4.1	Reynisson, B., et al (2020)	http://www.cbs.dtu.dk/services/NetMHCpan/
EpiMatrix	Schafer, J.R., et al (1998)	https://epivax.com/i
ImmuneSim version 0.8.7	Weber, C.R., et al (2020)	https://immunesim.readthedocs.io/en/latest/
ShaZam version 0.23	Gupta, N.T., et al (2015)	https://shazam.readthedocs.io/en/stable/

Stringdist version 0.9.5.5	MPJ van der Loo (2014)	https://github.com/markvanderloo/stringdist
Other		

569

570

Main Figures

Figure 1. Decreased MHC-II peptide epitope content is correlated with SHM in B cell receptors, with stronger effects in certain V-genes. A. Overview of MHC-II peptide epitope characterization in natively paired heavy and light chain human antibody sequence repertories. Paired heavy and light chain antibody repertoire data were generated by ultra-high throughput single cell sequencing of B cells from healthy donor PBMCs. An overlap-extension RT-PCR pairs antibody heavy and light chain variable region (VH and VL) transcripts for NGS analysis. V(D)J annotation and somatic hypermutation (SHM) assignment was carried out using IgBlast. MHC-II peptide epitope content of BCR variable regions was analyzed for antibody sequence repertoires using the EpiMatrix and netMHCIIpan algorithms. MHC-II peptide epitope content metrics were cross-referenced with SHM and antibody isotype to characterize relationships between MHC-II peptide epitope content and sequence-based markers of B cell development. **B**. Scatter plots of EpiMatrix MHC-II binding prediction scores vs. SHM, based on aggregate data for human supertype alleles DRB1*01:01, DRB1*03:01, DRB1*04:01, DRB1*07:01, DRB1*08:01, DRB1*11:01, DRB1*13:02 and DRB1*15:01. Each point represents an antibody sequence; points are colored according to data density (yellow: high, purple: low). Linear regressions are shown in red. *p*-value of the Spearman correlation is indicated. **C**. Volcano plots of spearman ρ vs. Benjamini-Hochberg adjusted *p*-values for MHC-II peptide epitope content vs. SHM, for antibody repertoires binned by IGHV and IGKV/IGLV gene pairs. Statistically significant pairs are shown in blue, and other gene pairs are shown in gray. **D**. Scatter plots of selected IGHV gene and IGKV/IGLV gene pairs for SHM vs. predicted binding scores. Linear regression lines are shown in blue.

Figure 2. V-gene dependence is driven by the deletion of high affinity peptides present in germline sequences. A. Repertoire-scale data analysis schematic using netMHCIIpan to identify patient-specific MHC-II peptide epitopes according to known donor HLA genes. **B.** Volcano plots of Spearman ρ vs. Benjamini-Hochberg adjusted p-values for antibody SHM vs. geometric mean K_D fold-change from germline K_D, as predicted by netMHCIIpan. Data were grouped by IGHV gene and IGKV/IGLV gene pairings and analyzed for peptides derived from germline-encoded MHC-II binding peptides (predicted germline K_D <1,000 nM). Statistically significant IGHV gene and IGKV/IGLV combinations are shown in blue, other gene pairs are shown in gray. **C.** Scatter plots of antibody data for selected IGHV and IGKV/IGLV gene pairs displaying antibody SHM vs. predicted peptide geomean K_D fold-change from germline K_D. Linear regressions are shown in blue. **D.** Geometric mean of the rank percentage, as defined by netMHCIIpan of each putative peptide across the IGHV sequence, comparing germline IGHV gene (black) and high SHM (top 5%, blue) from the IGHV gene-controlled repertoire. **E.** Logograms of high affinity germline-encoded peptide residues comparing germline and high SHM antibodies at those residues (top 5%). *n* represents the number of unique peptides displayed in the high SHM subset. **F**. netMHCIIpan K_D prediction for peptides shown in the logograms, using one of the donor-specific HLA-DRB1 alleles. Peptides from Donors 1-3 are shown. **G**. Experimental validation of peptide binding affinity to HLA II DRB1 molecules, using a competition assay with peptides derived from Donor 1. IC_{50} was calculated using a log-logistic equation. Somatic hypermutations are highlighted in bold script.

Figure 3. Sequence data comparisons with *in silico* SHM models, and a separate analysis of MHC-I vs. MHC-II epitope content, both demonstrate the preferential deletion of human MHC-II peptide epitopes by SHM. A. VH gene usage between experimentally-derived Donor 1 data and Donor 1 modeled antibody repertoires, incorporating both the donor-specific Replacement-Silent (R-S) SHM model based on Donor 1's repertoire data, and the universal Out-of-Frame (OoF) SHM model. Gene usage is shown as frequency of the total antibody repertoire. The same data for additional donors is provided in **Figure S7A**. **B.** Distribution of SHM between Donor 1 experimentally-derived data and *in silico* modeled repertoires. Black dots represent outliers. The same data for additional donors is provided in **Figure S7B. C.** Number of statistically significant (adjusted $p<0.05$) IGHV and IGKV/IGLV gene pairs in experimentally-derived donor data, divided by the average number of significant gene pairs in donor-matched modeled R-S repertoires (n=30 modeled RS repertoires for each donor). Values >1 indicate that experimentally-derived donor data has more statistically significant heavy:light gene pairs with deleted MHC-II peptide epitopes from the antibody variable region via SHM. **D.** Volcano plots of Spearman ρ vs. Benjamini-Hochberg adjusted p-values for SHM vs. geometric mean K_D fold-change from germline K_D in IGHV and IGKV/IGLV gene pairs, as predicted by netMHCIIpan, for isotype-switched antibody sequences. Data were calculated for peptides derived from germline-encoded high-affinity binders (<1,000 nM). Statistically significant IGHV and IGKV/IGLV gene pairs are shown in blue, other gene pairs are shown in gray. Experimental donor data and R-S models are shown. For R-S simulations, 30 repertoires were modeled for each donor for each simulation type, and the model closest to the median Spearman Rho of all 30 simulations is shown. **E.** Isotype-switched VH:VKL gene pairs with a significant correlation between K_D change and SHM in donor data and modeled repertoires were retrieved. For donor data, the gene pair list was matched in the modeled repertoires, and vice versa. The Spearman rho correlation was compared between donor and modeled repertoires using a paired t-test. **F**. *Upper:* The number of significant VH:VL gene pairs for MHC-I vs. MHC-II peptide epitopes; each point is a different MHC gene:donor combination. Peptide epitopes were binned as being both an MHC-I+MHC-II (shared) epitope, a unique MHC-I, or a unique MHC-II epitope, based on donor genotype. *Lower:* Comparison of Spearman correlations (K_D fold-change vs SHM) between MHC peptide epitope bins for significant VH:VL gene pairs. *:$p<0.05$, ***:$p < 0.001$, N.S: Not significant, Wilcoxon rank sum test.

Figure 4. Isotype class switching and antibody secretion as long-lived serum IgG are correlated with lower MHC-II peptide epitope content in BCRs. A. Antibody repertoires were fractionated by isotype, and Spearman correlations were calculated for each repertoire subset. EpiMatrix binding scores are shown as aggregate binding score for supertype alleles DRB1*01:01, DRB1*03:01, DRB1*04:01, DRB1*07:01, DRB1*08:02, DRB1*11:01, DRB1*13:02 and DRB1*15:01. Each point represents a BCR sequence, and points are colored by data density (yellow: high, purple: low). Linear regressions are shown in red; *p*-value of the Spearman correlation is indicated. **B**. Volcano plot of Spearman ρ vs. Benjamini-Hochberg adjusted p-values for SHM vs. MHC-II binding score for repertoires grouped by isotype. Data are shown for all seven donors. **C.** Geometric mean of the K_D comparison for antibody variable region peptides encoded by cellular vs. serum antibody repertoires, determined using netMHCIIpan. K_D for complete antibodies was obtained from peptides derived from germline peptides with K_D <1,000 nM. 'Serum' antibody clones were detected in human blood via serum proteomics in a previously reported study; 'Cellular' antibody sequences were restricted to the cellular compartment [40].

Differences between groups were analyzed using a t-test. Each point represents the BCR repertoire MHC-II peptide geomean K_D for a human HLA allele (modeled for 38 human alleles, because donor HLAs are unknown); and alleles with adjusted $p<0.05$ are shown in red. **D.** *Left* Single vs. Multiple observation antibodies from longitudinal serum repertoire data, plotted as described in Panel C. Multiple observation antibody clones were detected at multiple time points via serum proteomics, whereas single observation antibodies were detected only at a single time point [41]. *Right* Geomean K_D fold-change comparison between Multiple vs. Single observation serum antibodies **E.** Proposed mechanism of *in vivo* selection for BCRs with lower MHC-II peptide epitopes. Unmutated B cells in germinal centers often express unmutated BCRs that encode high-affinity MHC-II peptides. These high-affinity MHC-II peptides from the BCR can display on surface MHC-II after endocytosis of the BCR-antigen complex and compete with antigen-derived peptides for MHC-II surface presentation. Competition between BCR MHC-II peptides and antigen MHC-II peptides provides a selective pressure for B cells to mutate high-affinity MHC-II peptide epitopes in the BCR variable region to enhance CD4+ T cell help. Efficient T cell help leads to further SHM, isotype switching, and the generation of long-lived plasma cells that secrete an antibody repertoire with decreased MHC-II peptide epitope content.

Supplementary Figures

Figure S1. SHM correlates with decreased MHC-II peptide epitope content in B cell receptors, with stronger effects in certain V-genes. A. Scatter plots of somatic hypermutation levels (SHM) and EpiMatrix prediction of MHC II binding, as aggregate binding score for supertype alleles DRB1*01:01, DRB1*03:01, DRB1*04:01, DRB1*07:01, DRB1*08:02, DRB1*11:01, DRB1*13:02 and DRB1*15:01. Each point represents an antibody sequence; points are colored according to data density (yellow: high, purple: low). Linear regressions are shown in red. *p*-value of the Spearman correlation is indicated. **B**. Volcano plots of spearman ρ vs. Benjamini-Hochberg adjusted *p*-values for SHM vs. MHC-II peptide epitope content, for antibodies repertoires grouped by IGHV and IGKV/IGLV gene pairs. Statistically significant pairs are shown in blue, and other gene pairs are shown in gray. **C**. Scatter plots of selected IGHV gene and IGKV/IGLV gene pairs for SHM vs. predicted binding scores. Linear regression lines are shown in blue.

Figure S2. Germline MHC-II peptide epitope content varies by IGHV and IGLV/IGKV genes. Predicted MHC-II binding score was calculated using EpiMatrix for complete donor repertoires, and divided into V-gene subsets. Higher scores indicate higher content of MHC-II DRB1 peptide epitopes in the germline V-gene. V-genes were plotted in alphanumerical order, and the mean of scores (black points) and range (gray lines) are displayed together.

Figure S3. Germline MHC-II peptide epitope content varies according to HLA-DRB1 gene profile. MHC-II peptide epitope content was predicted for a database of germline-encoded VH, VK and VL genes for each HLA-DRB1 allele encoded by donors in this study netMHCIIpan. The geometric means of the rank percentage for all IGHV- and IGKV/IGLV genes were calculated (black line) and the range of ranks (0.01%-100%) for peptides centered in each residue is shown in shaded gray. A lower rank indicates higher peptide:MHC-II binding affinity.

Figure S4. V-gene dependence is driven by deletion of high affinity peptides present in germline sequences. A. Volcano plots of Spearman ρ vs. Benjamini-Hochberg adjusted p-values for antibody SHM vs. geometric mean K_D fold-change from germline K_D, as predicted by netMHCIIpan. Data were grouped by IGHV gene and IGKV/IGLV gene pairings and analyzed for peptides derived from germline-encoded MHCII binding peptides (predicted germline K_D <1,000 nM). Statistically significant IGHV gene and IGKV/IGLV combinations are shown in blue, other gene pairs are shown in gray. **B.** Scatter plots of antibody data for selected IGHV and IGKV/IGLV gene pairs displaying antibody SHM vs. predicted peptide geomean K_D fold-change from germline K_D. Linear regressions are shown in blue. **C.** Geometric mean of the rank percentage, as defined by netMHCIIpan of each putative peptide across the IGHV sequence, comparing germline IGHV gene (black) and high SHM (top 5%, blue) from the IGHV gene-controlled repertoire. **D.** Logograms of high affinity germline-encoded peptide residues comparing germline and high SHM antibodies at those residues (top 5%). n represents the number of unique peptides displayed in the high SHM subset. **E**. netMHCIIpan K_D prediction for peptides shown in the logograms, using one of the donor-specific HLA-DRB1 alleles. Donors 4 and 5 are shown. **Figure S5. Experimental observation of key antibody peptides in immunopeptidomic assay data in IEDB. A**. Observations of IGHV-derived peptides experimentally confirmed to be immune epitopes and displayed by residue position. Data was retrieved from the Immune Epitope Database and analysis

resource (IEDB, www.iedb.org). **B**. Presence of confirmed MHC-II peptide epitopes in antibody repertoires. Peptides eluted from MHC-II molecules were retrieved from IEDB and used as a search database to mine donor repertoire data. IEDB peptides present both as substrings entirely contained within antibody 15-mers, and complete 15-mer matches, were accepted. **C**. Overlap between confirmed HLA-DRB1 peptides and HLA-DP/DQ peptides from antibody V-genes found in IEDB. Antibody peptides detected in the IEDB HLA-DRB1 database were searched in the HLA-DP/DQ database, accepting only complete matches.

Figure S6. netMHCIIpan data analysis, computational repertoire modeling, and personalized repertoire analytics. A. Data processing using netMHCIIpan. *Upper panel* The presence of MHC-II peptide epitopes was determined in donor data for the complete set of 38 HLA alleles. HLA typing was also carried out. *Middle panel* Somatic hypermutation models ShaZam and immuneSIM were used to simulate 30 repertoires, with the same number of BCR sequences as experimentally-derived donor data. SHM distribution and V-gene frequencies were calculated. *Lower panel* The subset of peptides with MHC-II K_D<1,000 nM to any of the 38 alleles were selected to generate a database of potential predicted binders. **B**. Using the germline peptide database, peptides at the same position within the V-region were extracted from experimentally-derived donor data or simulated repertoires and grouped according to parent antibody V-gene. The fold-change between repertoire-scale BCR geomean[peptide:MHC-II K_D] and germline geomean[peptide:MHC-II K_D] was calculated and aggregated by V-gene. The Spearman correlation between K_D fold-change and SHM was calculated for each V-gene. These data was used for the plots shown in **Figure 2A**. **C**. Using correlation data from **B**, significant (adjusted p <0.05) and strong ($\rho > 0.5$) correlations were extracted and averaged by allele. Alleles were plotted according to their individual geomean Spearman ρ scores, with the larger circles corresponding to each of the donor's two HLA-DRB1 alleles.

Figure S7. IGHV gene usage and SHM distribution for each experimentally-derived BCR repertoire data, universal Out-of-frame (OoF) modeled repertoire data, and donor-specific Replacement-Silent (RS) modeled repertoire data. A. IGHV gene usage between Donors 2-5 experimentally-derived repertoires and OoF and RS modeled repertoires. Gene usage is shown as frequency of total repertoire. **B.** SHM distribution for Donors 2-5 experimentally-derived repertoires and OoF and RS modeled repertoires. Black dots represent outliers.

Figure S8. Somatic hypermutations selectively delete MHC-II peptide epitopes. A. Levenshtein distance between donor and germline peptide was calculated as a measure of mutational load. The number of mutations was plotted against the K_D fold-change between donor and germline peptides for donor-matched alleles. Outliers were removed for visualization but not for calculation of quartiles for boxplot generation **B.** Volcano plots of Spearman ρ vs. Benjamini-Hochberg adjusted p-values for SHM vs. geometric mean K_D fold-change from germline K_D, as predicted by netMHCIIpan. Data were calculated for peptides derived from germline-encoded high-affinity binders (< 1,000 nM). Statistically significant IGHV and IGKV/IGLV gene pairs are shown in blue, other gene pairs are shown in gray. Donor, OoF and R-S models are shown. For OoF and R-S simulations, 30 repertoires were modeled for each donor, and the model closest to the median Spearman Rho of all 30 simulations is shown. **C.** Number of statistically significant (adjusted $p < 0.05$) IGHV and IGKV/IGLV gene pairs in experimentally-derived donor data, divided by the average number of significant gene pairs in donor-matched modeled OoF

repertoires (n=30 modeled OoF repertoires for each donor). Values >1 indicate that experimentally-derived donor data has more statistically significant gene pairs that show decreased MHC-II peptide epitope content by SHM. "All alleles" reports the average of all 10 HLA-DRB1 alleles from the 5 donors, "Top alleles" reports the average of the top HLA-DRB1 allele collected from each donor. **D.** Spearman Rho comparison of aggregated HLA molecules. Alleles were clustered according to supertypes as defined in [60]. The Spearman ρ geometric mean was calculated for every allele, and then for all supertypes. Each color represents a different supertype. Supertypes with donor-matched HLA molecules are shown as bigger circles. **E.** Isotype-switched VH:VKL gene pairs with a significant correlation between K_D change and SHM in donor data and modeled repertoires were retrieved. For donor data, the gene pair list was matched in the modeled repertoires, and vice versa. Spearman rho correlations were compared between donor and modeled repertoires using a paired t-test.

Figure S9. Isotype class switching is correlated with preferential removal of MHC-II peptide epitopes from BCRs. A. Antibody repertoires were fractionated by isotype, and Spearman correlations were calculated for each repertoire subset. EpiMatrix binding scores are shown as aggregate binding score for supertype alleles DRB1*01:01, DRB1*03:01, DRB1*04:01, DRB1*07:01, DRB1*08:01, DRB1*11:01, DRB1*13:02 and DRB1*15:01. Each point represents a BCR sequence, and points are colored by data density (yellow: high, purple: low). Linear regressions are shown in red; *p*-value of the Spearman correlation is indicated.

Supplementary Table 1 Primers used for paired heavy and light chain overlap extension RT-PCR

Oligonucleotide name	Oligonucleotide sequence	Source
IgM Constant Region	CGCAGTAGCGGTAAACGGCCACAGGAGACGAGGGGGAAA	McDaniel, JR., DeKosky, BJ,. et al (2016)
IgG Constant Region	CGCAGTAGCGGTAAACGGCAGGGYGCCAGGGGGAAGAC	McDaniel, JR., DeKosky, BJ,. et al (2016)
IgA Constant Region	CGCAGTAGCGGTAAACGGCCGGGAAGACCTTGGGGCTGG	McDaniel, JR., DeKosky, BJ,. et al (2016)
IgLC Constant Region	GCGGATAACAATTTCACACAGGTCCTCAGAGGAGGGYGGGAA	McDaniel, JR., DeKosky, BJ,. et al (2016)
IgKC Constant Region	GCGGATAACAATTTCACACAGGGATGAAGACAGATGGTGCAG	McDaniel, JR., DeKosky, BJ,. et al (2016)
VH1 FR1 Region	TATTCCCATCGCGGCGCCAGGTCCAGCTKGTRCAGTCTGG	McDaniel, JR., DeKosky, BJ,. et al (2016)
VH157 FR1 Region	TATTCCCATCGCGGCGCCAGGTGCAGCTGGTGSARTCTGG	McDaniel, JR., DeKosky, BJ,. et al (2016)
VH2 FR1 Region	TATTCCCATCGCGGCGCCAGRTCACCTTGAAGGAGTCTG	McDaniel, JR., DeKosky, BJ,. et al (2016)
VH3 FR1 Region	TATTCCCATCGCGGCGCGAGGTGCAGCTGKTGGAGWCY	McDaniel, JR., DeKosky, BJ,. et al (2016)
VH4 FR1 Region	TATTCCCATCGCGGCGCCAGGTGCAGCTGCAGGAGTCSG	McDaniel, JR., DeKosky, BJ,. et al (2016)
VH4-DP63 FR1 Region	TATTCCCATCGCGGCGCCAGGTGCAGCTACAGCAGTGGG	McDaniel, JR., DeKosky, BJ,. et al (2016)
VH6 FR1 Region	TATTCCCATCGCGGCGCCAGGTACAGCTGCAGCAGTCA	McDaniel, JR., DeKosky, BJ,. et al (2016)
VH3N FR1 Region	TATTCCCATCGCGGCGCTCAACACAACGGTTCCCAGTTA	McDaniel, JR., DeKosky, BJ,. et al (2016)
VK1 FR1 Region	GCGCCGCGATGGGAATAGCTAGCCGACATCCRGDTGACCCAGTCTCC	McDaniel, JR., DeKosky, BJ,. et al (2016)
VK2 FR1 Region	GCGCCGCGATGGGAATAGCTAGCCGATATTGTGMTGACBCAGWCTCC	McDaniel, JR., DeKosky, BJ,. et al (2016)

VK3 FR1 Region	GCGCCGCGATGGGAATAGCTAGCCGAAATTGTRWTGACRCA GTCTCC	McDaniel, JR., DeKosky, BJ,. et al (2016)
VK5 FR1 Region	GCGCCGCGATGGGAATAGCTAGCCGAAACGACACTCACGCA GTCTC	McDaniel, JR., DeKosky, BJ,. et al (2016)
VL1 FR1 Region	GCGCCGCGATGGGAATAGCTAGCCCAGTCTGTSBTGACGCAG CCGCC	McDaniel, JR., DeKosky, BJ,. et al (2016)
VL1459 FR1 Region	GCGCCGCGATGGGAATAGCTAGCCCAGCCTGTGCTGACTCAR YC	McDaniel, JR., DeKosky, BJ,. et al (2016)
VL15910 FR1 Region	GCGCCGCGATGGGAATAGCTAGCCCAGCCWGKGCTGACTCA GCCMCC	McDaniel, JR., DeKosky, BJ,. et al (2016)
VL2 FR1 Region	GCGCCGCGATGGGAATAGCTAGCCCAGTCTGYYCTGAYTCA GCCT	McDaniel, JR., DeKosky, BJ,. et al (2016)
VL3 FR1 Region	GCGCCGCGATGGGAATAGCTAGCCTCCTATGWGCTGACWCA GCCAA	McDaniel, JR., DeKosky, BJ,. et al (2016)
VL-DPL16 FR1 Region	GCGCCGCGATGGGAATAGCTAGCCTCCTCTGAGCTGASTCAG GASCC	McDaniel, JR., DeKosky, BJ,. et al (2016)
VL3-38 FR1 Region	GCGCCGCGATGGGAATAGCTAGCCTCCTATGAGCTGAYRCAG CYACC	McDaniel, JR., DeKosky, BJ,. et al (2016)
VL6 FR1 Region	GCGCCGCGATGGGAATAGCTAGCCAATTTTATGCTGACTCAG CCCC	McDaniel, JR., DeKosky, BJ,. et al (2016)
VL78 FR1 Region	GCGCCGCGATGGGAATAGCTAGCCCAGDCTGTGGTGACYCA GGAGCC	McDaniel, JR., DeKosky, BJ,. et al (2016)
IgG Constant Region nested	NNNNATGGGCCCTGSGATGGGCCCTTGGTGGARGC	McDaniel, JR., DeKosky, BJ,. et al (2016)
IgM Constant Region nested	NNNNATGGGCCCTGGGTTGGGGCGGATGCACTCC	McDaniel, JR., DeKosky, BJ,. et al (2016)
IgA Constant Region nested	NNNNATGGGCCCTGCTTGGGGCTGGTCGGGGATG	McDaniel, JR., DeKosky, BJ,. et al (2016)
IgKC Constant Region nested	NNNNGTGCGGCCGCAGATGGTGCAGCCACAGTTC	McDaniel, JR., DeKosky, BJ,. et al (2016)
IgLC Constant Region nested	NNNNGTGCGGCCGCGAGGGYGGGAACAGAGTGAC	McDaniel, JR., DeKosky, BJ,. et al (2016)

795

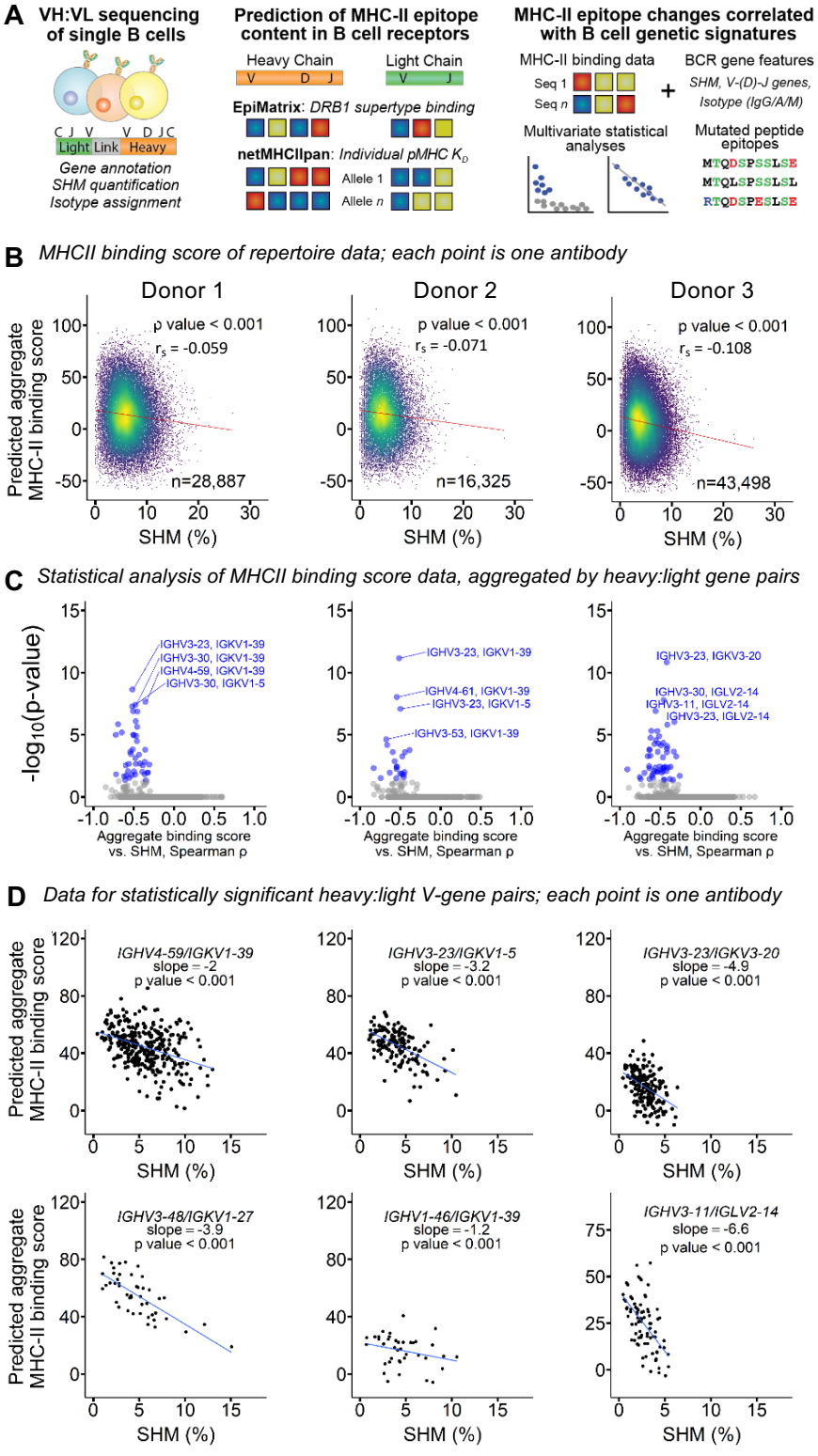
References

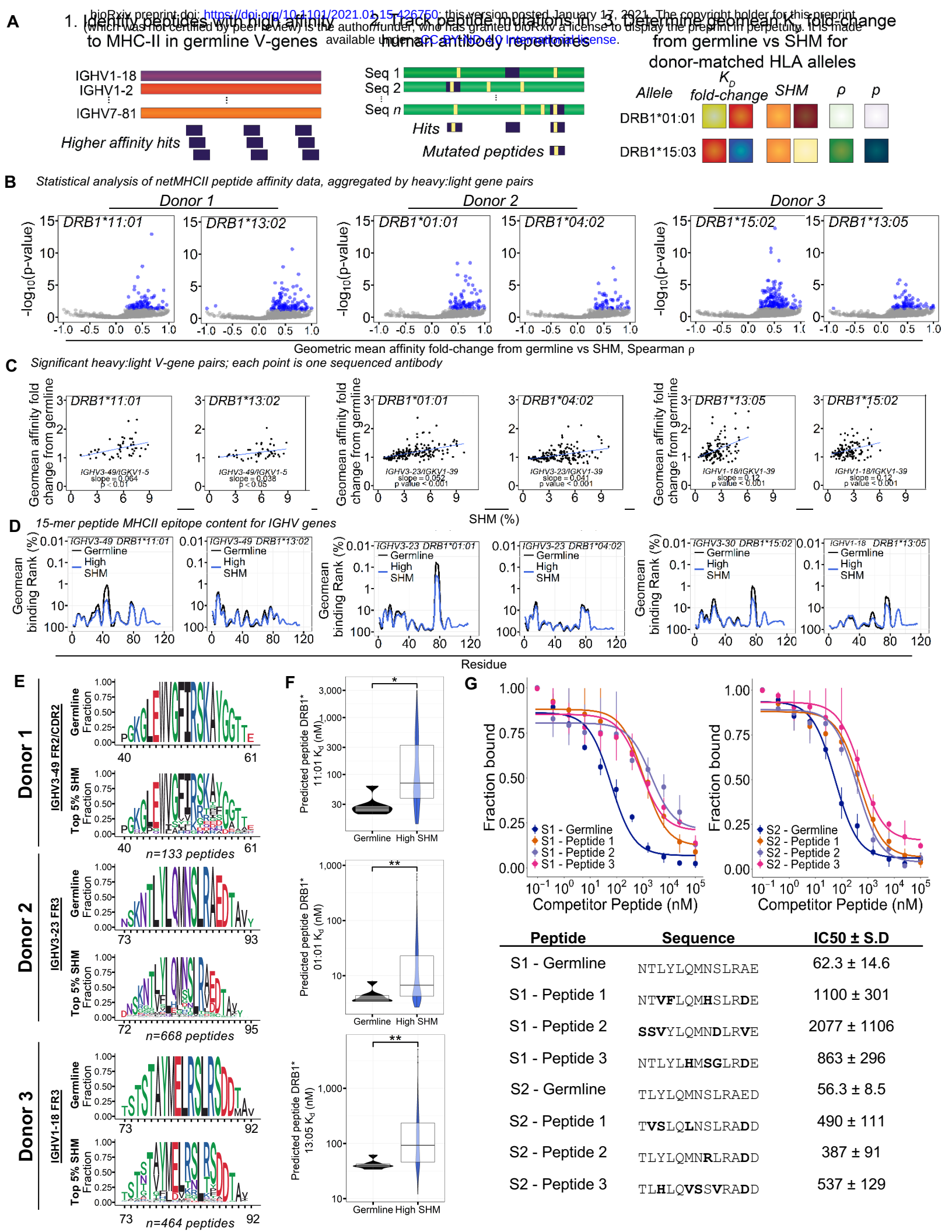
1. Nemazee, D., *Mechanisms of central tolerance for B cells.* Nat Rev Immunol, 2017. **17**(5): p. 281-294.
2. Gitlin, A.D., Z. Shulman, and M.C. Nussenzweig, *Clonal selection in the germinal centre by regulated proliferation and hypermutation.* Nature, 2014. **509**(7502): p. 637-40.
3. Mesin, L., J. Ersching, and G.D. Victora, *Germinal Center B Cell Dynamics.* Immunity, 2016. **45**(3): p. 471-482.
4. Batista, F.D. and M.S. Neuberger, *B cells extract and present immobilized antigen: implications for affinity discrimination.* EMBO J, 2000. **19**(4): p. 513-20.
5. Cyster, J.G. and C.D.C. Allen, *B Cell Responses: Cell Interaction Dynamics and Decisions.* Cell, 2019. **177**(3): p. 524-540.
6. Wine, Y., et al., *Serology in the 21st century: the molecular-level analysis of the serum antibody repertoire.* Curr Opin Immunol, 2015. **35**: p. 89-97.
7. Lavinder, J.J., et al., *Identification and characterization of the constituent human serum antibodies elicited by vaccination.* Proc Natl Acad Sci U S A, 2014. **111**(6): p. 2259-64.
8. Purtha, W.E., et al., *Memory B cells, but not long-lived plasma cells, possess antigen specificities for viral escape mutants.* J Exp Med, 2011. **208**(13): p. 2599-606.
9. Amanna, I.J., N.E. Carlson, and M.K. Slifka, *Duration of humoral immunity to common viral and vaccine antigens.* N Engl J Med, 2007. **357**(19): p. 1903-15.
10. Amanna, I.J. and M.K. Slifka, *Mechanisms that determine plasma cell lifespan and the duration of humoral immunity.* Immunol Rev, 2010. **236**: p. 125-38.
11. Melchers, F., *Checkpoints that control B cell development.* J Clin Invest, 2015. **125**(6): p. 2203-10.
12. Slifka, M.K. and I.J. Amanna, *Role of Multivalency and Antigenic Threshold in Generating Protective Antibody Responses.* Front Immunol, 2019. **10**: p. 956.
13. Cassotta, A., et al., *A single T cell epitope drives the neutralizing anti-drug antibody response to natalizumab in multiple sclerosis patients.* Nat Med, 2019. **25**(9): p. 1402-1407.
14. Jurewicz, M.M. and L.J. Stern, *Class II MHC antigen processing in immune tolerance and inflammation.* Immunogenetics, 2019. **71**(3): p. 171-187.
15. Lamberth, K., et al., *Post hoc assessment of the immunogenicity of bioengineered factor VIIa demonstrates the use of preclinical tools.* Sci Transl Med, 2017. **9**(372).
16. Robinson, J., et al., *IPD-IMGT/HLA Database.* Nucleic Acids Res, 2020. **48**(D1): p. D948-D955.
17. Alcaïde-Loridan, C., et al., *Differential expression of MHC class II isotype chains.* Microbes Infect, 1999. **1**(11): p. 929-34.
18. Peretti, M., et al., *Expression of the three human major histocompatibility complex class II isotypes exhibits a differential dependence on the transcription factor RFXAP.* Mol Cell Biol, 2001. **21**(17): p. 5699-709.
19. De Groot, A.S., et al., *Activation of natural regulatory T cells by IgG Fc-derived peptide "Tregitopes".* Blood, 2008. **112**(8): p. 3303-11.
20. Jawa, V., et al., *T-Cell Dependent Immunogenicity of Protein Therapeutics Pre-clinical Assessment and Mitigation-Updated Consensus and Review 2020.* Front Immunol, 2020. **11**: p. 1301.

21. Cousens, L.P., et al., *In vitro and in vivo studies of IgG-derived Treg epitopes (Tregitopes): a promising new tool for tolerance induction and treatment of autoimmunity.* J Clin Immunol, 2013. **33 Suppl 1**: p. S43-9.
22. Peters, B., M. Nielsen, and A. Sette, *T Cell Epitope Predictions.* Annu Rev Immunol, 2020. **38**: p. 123-145.
23. Barra, C., et al., *Footprints of antigen processing boost MHC class II natural ligand predictions.* Genome Med, 2018. **10**(1): p. 84.
24. Reynisson, B., et al., *NetMHCpan-4.1 and NetMHCIIpan-4.0: improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data.* Nucleic Acids Res, 2020. **48**(W1): p. W449-W454.
25. Wang, Q., et al., *Immunogenic HLA-DR-Presented Self-Peptides Identified Directly from Clinical Samples of Synovial Tissue, Synovial Fluid, or Peripheral Blood in Patients with Rheumatoid Arthritis or Lyme Arthritis.* J Proteome Res, 2017. **16**(1): p. 122-136.
26. Sordé, L., et al., *Tregitopes and impaired antigen presentation: Drivers of the immunomodulatory effects of IVIg?* Immun Inflamm Dis, 2017. **5**(4): p. 400-415.
27. Huszthy, P.C., et al., *B cell receptor ligation induces display of V-region peptides on MHC class II molecules to T cells.* Proc Natl Acad Sci U S A, 2019. **116**(51): p. 25850-25859.
28. DeKosky, B.J., et al., *In-depth determination and analysis of the human paired heavy- and light-chain antibody repertoire.* Nat Med, 2015. **21**(1): p. 86-91.
29. Wang, B., et al., *Functional interrogation and mining of natively paired human V.* Nat Biotechnol, 2018. **36**(2): p. 152-155.
30. Schafer, J.R., et al., *Prediction of well-conserved HIV-1 ligands using a matrix-based algorithm, EpiMatrix.* Vaccine, 1998. **16**(19): p. 1880-4.
31. Andreatta, M., et al., *Accurate pan-specific prediction of peptide-MHC class II binding affinity with improved binding core identification.* Immunogenetics, 2015. **67**(11-12): p. 641-50.
32. Dhanda, S.K., et al., *IEDB-AR: immune epitope database-analysis resource in 2019.* Nucleic Acids Res, 2019. **47**(W1): p. W502-W506.
33. Roche, P.A. and K. Furuta, *The ins and outs of MHC class II-mediated antigen processing and presentation.* Nat Rev Immunol, 2015. **15**(4): p. 203-16.
34. Nielsen, S.C.A., et al., *Shaping of infant B cell receptor repertoires by environmental factors and infectious disease.* Sci Transl Med, 2019. **11**(481).
35. Johnson, E.L., et al., *Sequencing HIV-neutralizing antibody exons and introns reveals detailed aspects of lineage maturation.* Nat Commun, 2018. **9**(1): p. 4136.
36. Honjo, T., K. Kinoshita, and M. Muramatsu, *Molecular mechanism of class switch recombination: linkage with somatic hypermutation.* Annu Rev Immunol, 2002. **20**: p. 165-96.
37. Wine, Y., et al., *Molecular deconvolution of the monoclonal antibodies that comprise the polyclonal serum response.* Proc Natl Acad Sci U S A, 2013. **110**(8): p. 2993-8.
38. Lavinder, J.J., et al., *Next-generation sequencing and protein mass spectrometry for the comprehensive analysis of human cellular and serum antibody repertoires.* Curr Opin Chem Biol, 2015. **24**: p. 112-20.
39. Chen, J., et al., *Proteomic Analysis of Pemphigus Autoantibodies Indicates a Larger, More Diverse, and More Dynamic Repertoire than Determined by B Cell Genetics.* Cell Rep, 2017. **18**(1): p. 237-247.

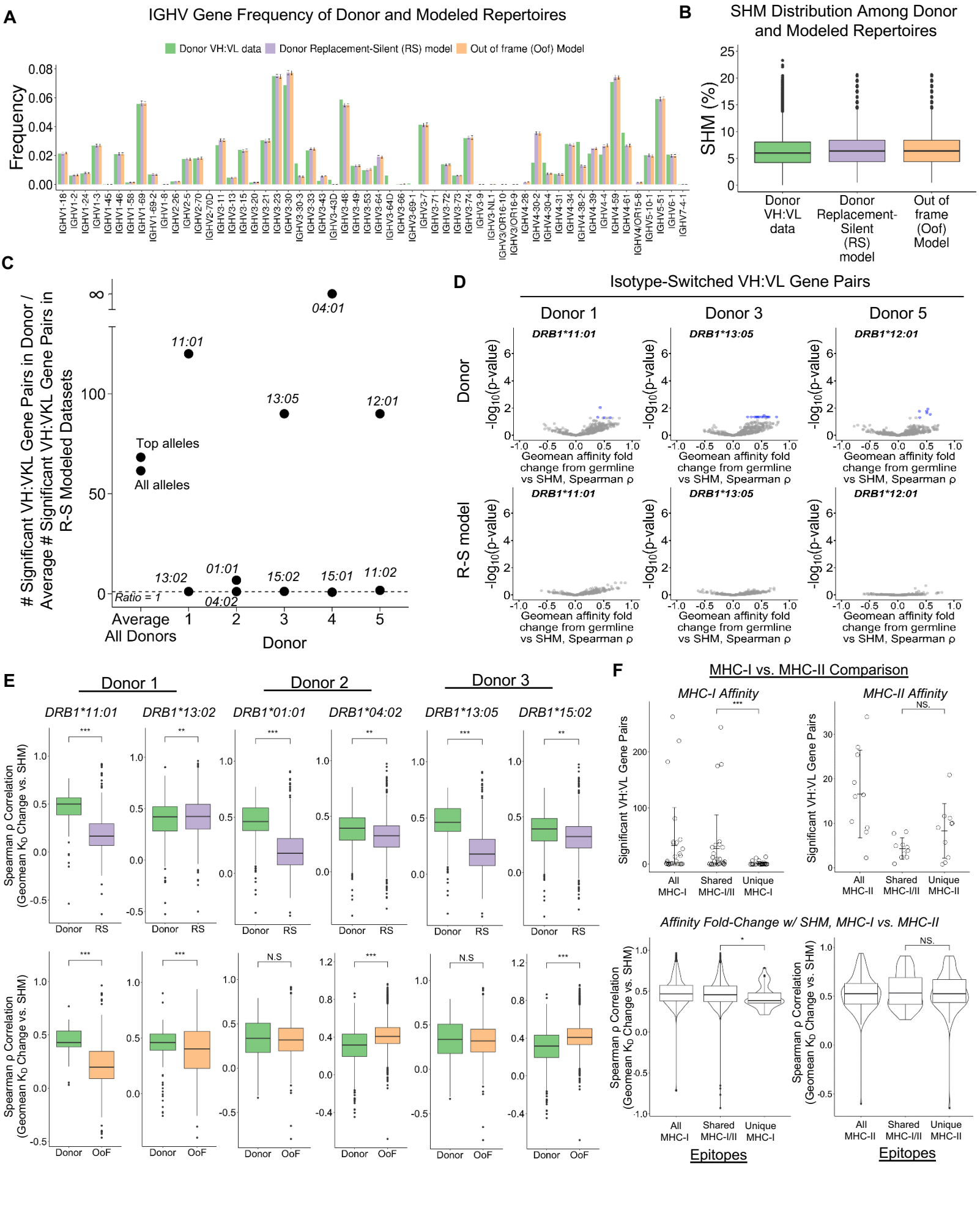
40. Lee, J., et al., *Molecular-level analysis of the serum antibody repertoire in young adults before and after seasonal influenza vaccination.* Nat Med, 2016. **22**(12): p. 1456-1464.
41. Lee, J., et al., *Persistent Antibody Clonotypes Dominate the Serum Response to Influenza over Multiple Years and Repeated Vaccinations.* Cell Host Microbe, 2019. **25**(3): p. 367-376.e5.
42. Andreatta, M., et al., *An automated benchmarking platform for MHC class II binding prediction methods.* Bioinformatics, 2018. **34**(9): p. 1522-1528.
43. Zhao, W. and X. Sher, *Systematically benchmarking peptide-MHC binding predictors: From synthetic to naturally processed epitopes.* PLoS Comput Biol, 2018. **14**(11): p. e1006457.
44. Halliley, J.L., et al., *Long-Lived Plasma Cells Are Contained within the CD19(-)CD38(hi)CD138(+) Subset in Human Bone Marrow.* Immunity, 2015. **43**(1): p. 132-45.
45. Kwong, P.D. and J.R. Mascola, *HIV-1 Vaccines Based on Antibody Identification, B Cell Ontogeny, and Epitope Structure.* Immunity, 2018. **48**(5): p. 855-871.
46. Bonsignori, M., et al., *Antibody-virus co-evolution in HIV infection: paths for HIV vaccine development.* Immunol Rev, 2017. **275**(1): p. 145-160.
47. De Groot, A.S. and W. Martin, *Reducing risk, improving outcomes: bioengineering less immunogenic protein therapeutics.* Clin Immunol, 2009. **131**(2): p. 189-201.
48. Vaisman-Mentesh, A., et al., *The Molecular Mechanisms That Underlie the Immune Biology of Anti-drug Antibody Formation Following Treatment With Monoclonal Antibodies.* Front Immunol, 2020. **11**: p. 1951.
49. Suurmond, J., et al., *Loss of an IgG plasma cell checkpoint in patients with lupus.* J Allergy Clin Immunol, 2019. **143**(4): p. 1586-1597.
50. Yurasov, S., et al., *Defective B cell tolerance checkpoints in systemic lupus erythematosus.* J Exp Med, 2005. **201**(5): p. 703-11.
51. Meffre, E. and K.C. O'Connor, *Impaired B-cell tolerance checkpoints promote the development of autoimmune diseases and pathogenic autoantibodies.* Immunol Rev, 2019. **292**(1): p. 90-101.
52. Antia, A., et al., *Heterogeneity and longevity of antibody memory to viruses and vaccines.* PLoS Biol, 2018. **16**(8): p. e2006601.
53. DeKosky, B.J., et al., *Large-scale sequence and structural comparisons of human naive and antigen-experienced antibody repertoires.* Proc Natl Acad Sci U S A, 2016. **113**(19): p. E2636-45.
54. McDaniel, J.R., et al., *Ultra-high-throughput sequencing of the immune receptor repertoire from millions of lymphocytes.* Nat Protoc, 2016. **11**(3): p. 429-42.
55. Lagerman, C.E., et al., *Ultrasonically-guided flow focusing generates precise emulsion droplets for high-throughput single cell analyses.* J Biosci Bioeng, 2019. **128**(2): p. 226-233.
56. Ye, J., et al., *IgBLAST: an immunoglobulin variable domain sequence analysis tool.* Nucleic Acids Res, 2013. **41**(Web Server issue): p. W34-40.
57. DeKosky, B.J., et al., *High-throughput sequencing of the paired human immunoglobulin heavy and light chain repertoire.* Nat Biotechnol, 2013. **31**(2): p. 166-9.
58. Edgar, R.C., *Search and clustering orders of magnitude faster than BLAST.* Bioinformatics, 2010. **26**(19): p. 2460-1.
59. Lund, O., et al., *Definition of supertypes for HLA molecules using clustering of specificity matrices.* Immunogenetics, 2004. **55**(12): p. 797-810.

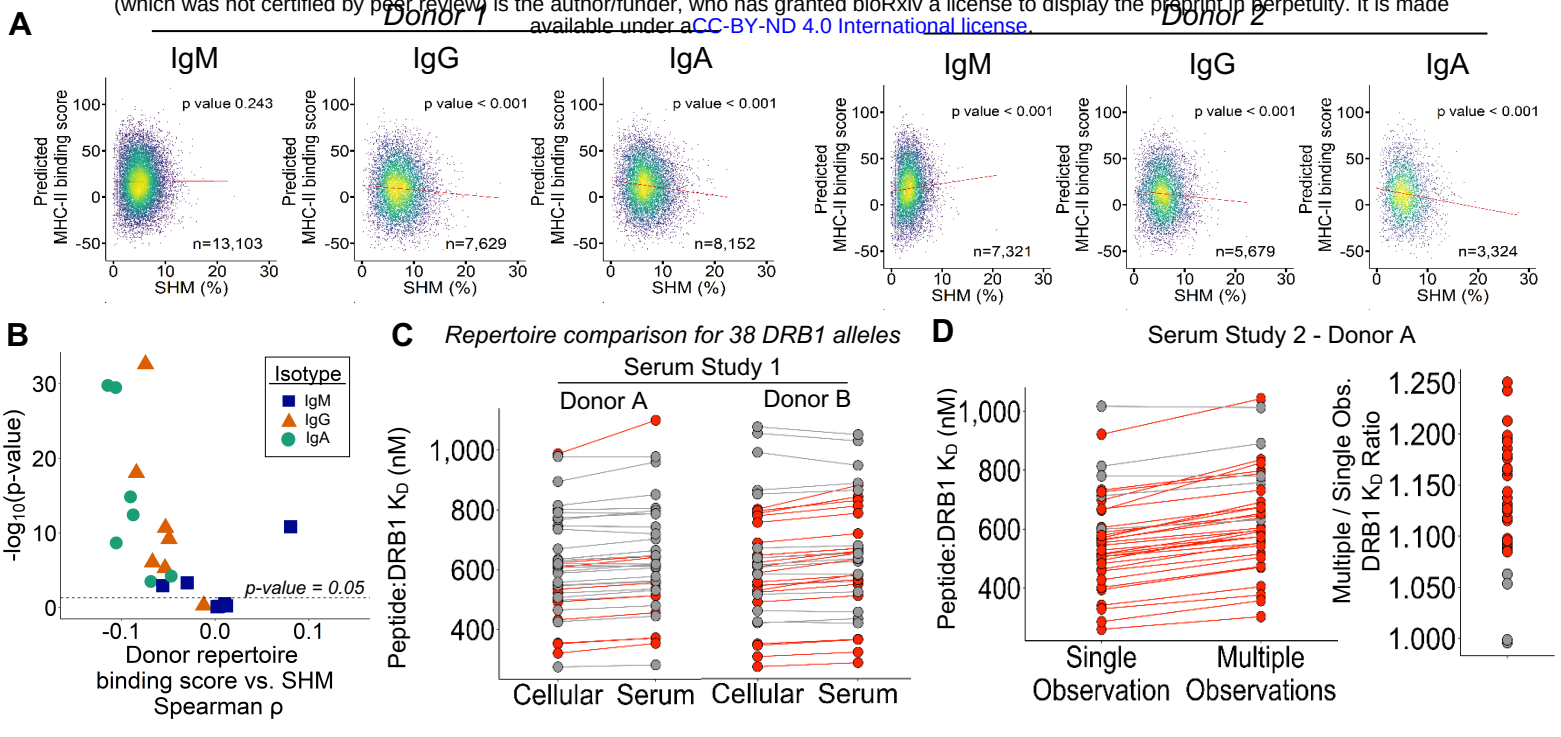
60. Southwood, S., et al., *Several common HLA-DR types share largely overlapping peptide binding repertoires.* J Immunol, 1998. **160**(7): p. 3363-73.
61. Sette, A., et al., *The relationship between class I binding affinity and immunogenicity of potential cytotoxic T cell epitopes.* J Immunol, 1994. **153**(12): p. 5586-92.
62. Gupta, N.T., et al., *Change-O: a toolkit for analyzing large-scale B cell immunoglobulin repertoire sequencing data.* Bioinformatics, 2015. **31**(20): p. 3356-8.
63. Weber, C.R., et al., *immuneSIM: tunable multi-feature simulation of B- and T-cell receptor repertoires for immunoinformatics benchmarking.* Bioinformatics, 2020. **36**(11): p. 3594-3596.
64. Bolotin, D.A., et al., *MiXCR: software for comprehensive adaptive immunity profiling.* Nat Methods, 2015. **12**(5): p. 380-1.
65. Stern, L.J. and D.C. Wiley, *The human class II MHC protein HLA-DR1 assembles as empty alpha beta heterodimers in the absence of antigenic peptide.* Cell, 1992. **68**(3): p. 465-77.
66. Sloan, V.S., et al., *Mediation by HLA-DM of dissociation of peptides from HLA-DR.* Nature, 1995. **375**(6534): p. 802-6.
67. Yin, L. and L.J. Stern, *Measurement of Peptide Binding to MHC Class II Molecules by Fluorescence Polarization.* Curr Protoc Immunol, 2014. **106**: p. 5.10.1-5.10.12.

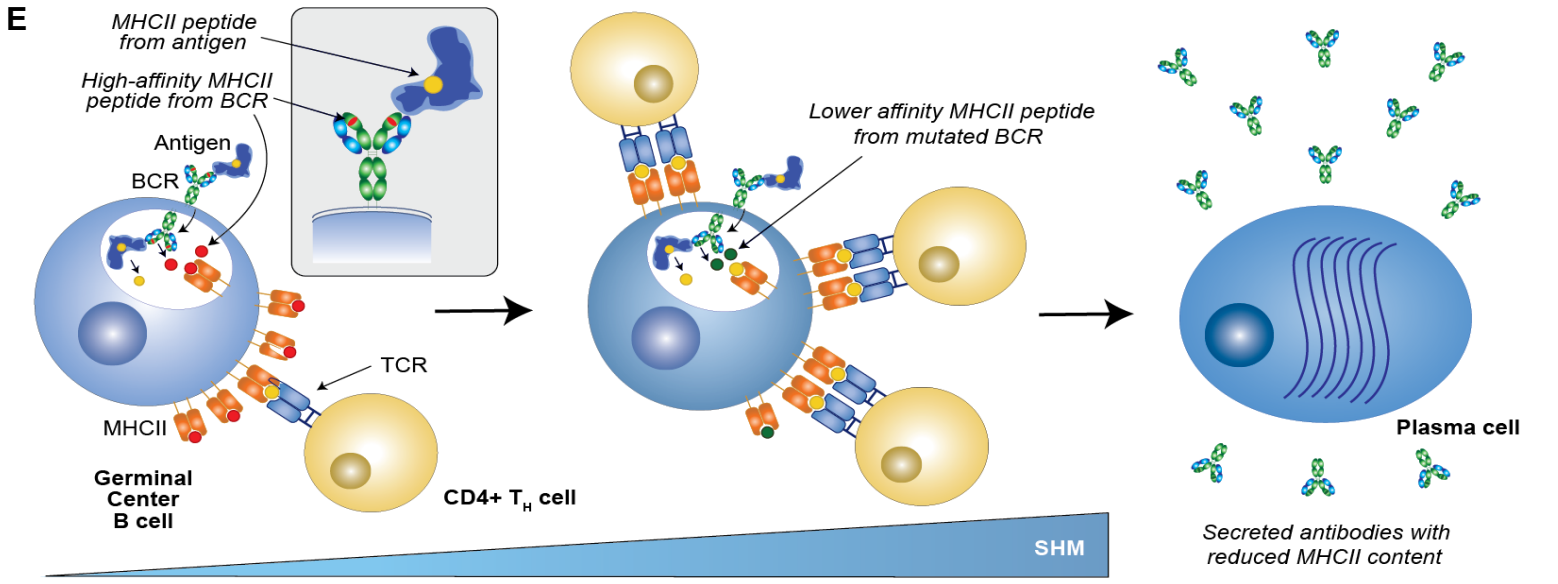




Peptide	Sequence	IC50 ± S.D
S1 - Germline	NTLYLQMNSLRAE	62.3 ± 14.6
S1 - Peptide 1	NT**VF**LQM**H**SLR**D**E	1100 ± 301
S1 - Peptide 2	**SSV**YLQMN**D**LR**V**E	2077 ± 1106
S1 - Peptide 3	NTLYL**H**M**SG**LR**D**E	863 ± 296
S2 - Germline	TLYLQMNSLRAED	56.3 ± 8.5
S2 - Peptide 1	T**VS**LQ**L**NSLRA**D**D	490 ± 111
S2 - Peptide 2	TLYLQMN**R**LRA**D**D	387 ± 91
S2 - Peptide 3	TL**H**LQ**VS**S**V**RA**D**D	537 ± 129







1. Unmutated BCRs contain high-affinity MHCII peptides that compete with antigen-derived peptides for MHCII display on the cell surface.

2. GC B cells are selected for their ability to secure T cell help. Mutations that reduce MHCII affinity of BCR peptides allow greater Ag presentation on MHCII, for enhanced CD4+ T cell help and clonal expansion.

3. T cell help also leads to plasmabast plasma cell differentiation. The sreum antibody repertoire has reduced CD4 T cell epitope affinity, resultin gin lower patient-specific immunogenicity of secreted mAbs.