

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

3

4 **Authors**

5 Matias Gutiérrez-González¹, Ahmed S. Fahad¹, Matt Ardito², Padma Nanaware³, Liying Lu³,
6 Erica Normandin⁴, Bharat Madan¹, Jacob Tivin², Emily Coates⁴, Amy R. Henry⁴, Farida Laboune⁴,
7 Barney S. Graham⁴, Daniel C. Douek⁴, Julie E. Ledgerwood⁴, John R. Mascola⁴, William D.
8 Martin², Lawrence J. Stern³, Annie S. De Groot^{2,5}, Brandon J. DeKosky^{1,6,*}

9

10 **Affiliations**

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

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

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

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

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

17 ⁶ Department of Chemical Engineering, The University of Kansas, Lawrence, Kansas, USA

18

19 *Corresponding author. Email: dekosky@ku.edu

20 **Summary**

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

33

34 **Keywords**

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

37 **Highlights**

38 • Antibody somatic hypermutation selectively removes MHC-II peptide epitopes from B
39 cell receptors.

40 • Antibodies with lower MHC-II epitope content show evidence of greater T cell help,
41 including class-switching.

42 • MHC-II peptide epitope removal from a BCR is linked to long-term antibody secretion in
43 serum.

44 • MHC-II genotype provides a personalized selection pressure on human antibody
45 development.

46 **Introduction**

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

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

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

80 Surface display of antigen-derived MHC-II epitopes is one critical determinant of B cell fate
81 due to the need for B cells to obtain help from antigen-specific CD4+ helper T cells. The affinity
82 of antigen peptides for binding to MHC-II plays a major role in regulating immune responses to
83 foreign proteins, including monoclonal antibody drugs [13-15]. MHC-II molecules are encoded
84 by three human leukocyte antigen (HLA) loci: HLA-DR, -DQ, and -DP. Of these, HLA-DR is the
85 most polymorphic [16], and is usually expressed at higher levels [17, 18]. It is unclear why anti-
86 antibody (or anti-idiotype) immune responses are not highly prevalent due to the very high
87 diversity of somatically mutated human antibodies, including the substantial untemplated diversity
88 of CDR3 regions, although highly homologous antibody sequences (including T regulatory cell
89 epitopes, or Tregitopes) have been suggested to play a role in reducing anti-antibody immunity
90 [19-21]. Methods for computational MHC binding prediction have continually improved in recent
91 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

114 **Results**

115 We began by characterizing MHC-II peptide epitope content in healthy human antibody
116 variable region sequences using high-throughput computational MHC-II peptide epitope
117 prediction. We collected seven paired heavy and light chain datasets from antigen-experienced B
118 cells of healthy donors, with a total of 250,645 high-quality consensus sequences of natively paired
119 heavy and light chain antibody lineages. We analyzed these immune repertoires using multiple
120 pMHC-II affinity prediction algorithms to determine how the features of antibody development
121 correlated with changes in potential MHC-II peptide epitope content of BCRs (**Fig. 1A**). First, we
122 used the commercially available EpiMatrix MHC-II epitope prediction platform to characterize
123 aggregate predicted HLA-DR epitope content based on eight human HLA-DR gene supertypes.
124 EpiMatrix reports a T cell epitope score, where a higher score indicates higher content of putative
125 MHC-II peptide epitopes within the analyzed protein sequence [30]. Strikingly, we noted that all
126 donors showed reduced MHC-II peptide epitope content (i.e., reduced EpiMatrix scores) that was
127 correlated with increasing somatic hypermutation (SHM), and the correlation was statistically
128 significant in all donors (Spearman correlation test, adjusted p-value < 0.05). These data
129 demonstrated that SHM reduces pMHC-II affinities in antibody peptides at a repertoire level (**Figs.**
130 **1B, S1A**). Subsequent analysis of antibody repertoire data fractionated by paired antibody heavy
131 and light chain V-genes showed that changes in MHC-II peptide epitope content were concentrated
132 in certain V-gene combinations (**Figs. 1C, 1D, S1B, S1C**), and each V-gene shows a different
133 initial distribution of MHC-II peptide epitope content (**Fig. S2**). While each donor showed a unique
134 pattern of V-genes with the highest reductions in MHC-II peptide epitope content, some V-genes
135 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

159 MHC-II *in vivo* [32, 33]. We identified a large number of naturally-processed peptides that were
160 experimentally confirmed in IEDB and appeared to be targets of preferential mutations that reduce
161 peptide affinity via SHM, including peptides that were mutated in antibody sequence data such as
162 IGHV3-23₇₃₋₉₃ and IGHV1-18₇₃₋₉₂ (**Fig. 2E**, **Fig. S5A**, **S5B**). Interestingly, donor antibody
163 repertoires also contained some of the same peptides that were eluted from HLA-DP and HLA-
164 DQ molecules(**Fig. S5B**); numerous IEDB-validated peptides overlapped between DRB and
165 DP/DQ binding (**Fig. S5C**). Thus we confirmed that some of the key peptides analyzed in our
166 study are presented on human MHC-II in previously reported proteomic datasets.

167 Once we realized that antibody peptides with high affinity for DRB binding were being
168 targeted for mutations and MHC-II epitope removal, we shifted our focus to patient-specific
169 analyses to explore these high-affinity MHC-II peptide epitopes encoded by germline IGHV and
170 IGKV/IGLV genes (**Fig. S6**). MHC-II peptide epitopes often require multiple amino acid matches
171 with appropriate spacing for binding to the MHC-II cleft, and we reasoned that the reduced T cell
172 content observed with increasing SHM could be introduced as an indirect consequence of SHM
173 mutational pattern preferences, rather than by active selection pressure. To test this alternate
174 hypothesis, we reasoned that if MHC-II peptide epitopes are removed by SHM to a greater degree
175 in experimentally-derived patient repertoires than in carefully matched *in silico* simulations (which
176 account for SHM DNA motif targets, but not for any HLA-dependent MHC-II peptide epitope
177 selection pressure), then we could conclude that MHC-II epitope removal was a result of active
178 selection *in vivo*. We thus began large-scale *in silico* experiments simulating antibody repertoires
179 using established somatic hypermutation models (**Figs. 3, S7**). We compared two different SHM
180 models to the experimentally-derived sequence data: one *in silico* SHM model customized by the
181 5-mer DNA base targeting patterns in each individual patient's experimentally-derived antibody

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

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

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

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

225 Finally, we sought to understand how MHC-II peptide epitope content in BCRs is associated
226 with elicitation of antibodies into the serum immune compartment. Serum antibodies are secreted
227 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

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

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

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

273 antibody prevalence, suggesting that HLA-DRB1 peptide epitope deletion may support B cell
274 trafficking to a long-lived plasma cell niche by enhancing the acquisition of T cell help (**Fig. 4E**)
275 [44]. Certain heavy and light chain V-genes showed higher rates of HLA-DRB1 peptide epitope
276 removal than other V-gene pairs (**Fig. 2B**), reflecting the different baseline levels of MHC-II
277 peptide epitopes in antibody germline genes (**Figs. S2, S3**). These data suggest that MHC-II
278 epitope deletion is targeted toward those V-genes that contain germline-encoded MHC-II epitopes,
279 as would be expected to occur in a functional selection mechanism. Low MHC-II epitope content
280 in a B cell receptor could help that B cell present more MHC-II epitopes from antigen, thereby
281 enhancing CD4+ T cell help for that B cell (**Fig. 4E**). This selection mechanism offers several
282 important advantages *in vivo*. First, selection of lower MHC-II peptide epitope content reduces the
283 propensity of an individual's secreted antibodies to induce CD4+ T-cell dependent anti-idiotype
284 antibody immune responses in non-templated regions (e.g., from pMHC-II derived from CDR3
285 loops, or that may arise as a result of SHM), reducing the risk of immune responses to somatically
286 generated antibody proteins. Perhaps more importantly, low MHC-II peptide epitope content in an
287 antibody could help dendritic cells present a greater fraction of MHC-II peptides derived from
288 antigen (and fewer peptides derived from the BCR) after immune complex capture and processing.
289 These findings have important implications for vaccine design and antibody drug therapeutics. As
290 one example in HIV vaccine development, where targeted elicitation of specific lineage mutations
291 are being pursued, these data suggest an important HLA-dependent selection pressure guiding
292 SHM, and that antibody mutations may accumulate differently in patients with different HLA gene
293 profiles due to MHC-II-based selection pressure [45, 46]. In addition, our findings lend further
294 support to ongoing efforts to mitigate anti-drug antibody responses by removal of MHC-II peptide
295 epitopes from the monoclonal antibody drug variable regions [47, 48]

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

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

314 **Acknowledgments**

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

322

323 **Author Contributions**

324 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,
325 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,
326 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.
327 analyzed the data; and M.G.G. and B.J.D. wrote the manuscript with feedback from all authors.

328

329 **Disclosures**

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

332 **Methods**

333

334 **Resource Availability**

335 ***Lead Contact***

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

338

339 ***Materials Availability***

340 No new reagents were generated in this study.

341

342 ***Data and Code Availability***

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

345

346 **Experimental Model and Subject Details**

347 ***Human Subjects***

348 For cellular antibody MHC-II content, a total of seven datasets were analyzed. These include
349 previously published data (Donors 1,2,4,6 and 7) [53][Fahad, DeKosky et al., *Front. Immunol.*,
350 Accepted 2021], and new unpublished datasets (Donors 3 and 5). All human samples were
351 collected under the Vaccine Research Center's (VRC)/National Institutes of Allergy and Infectious
352 Diseases (NIAID)/ National Institutes of Health (NIH) sample collection protocol, VRC 200
353 (NCT00067054) in compliance with the NIH IRB approved procedures. All subjects met protocol
354 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.

374 **Method Details**

375 ***Emulsion Overlap Extension RT-PCR***

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

383

384 ***Antibody Sequence Analysis***

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

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

396

397 ***MHC Peptide Epitope Content Prediction***

398 The EpiMatrix tool (EpiVax, Rhode Island, USA) was used for aggregate MHC-II peptide
399 epitope / T cell epitope predictions [30]. EpiMatrix uses main HLA II DRB1 “supertypes” to
400 predict overall protein epitope content [59]. Higher scores in the EpiMatrix output indicate a higher
401 probability of T cell dependent immunogenicity of foreign protein antigens. The alleles analyzed
402 were DRB1*01:01, DRB1*03:01, DRB1*04:01, DRB1*07:01, DRB1*08:01, DRB1*09:01,
403 DRB1*11:01, DRB1*13:02 and DRB1*15:01. The output data includes aggregate epitope score
404 by chain, normalized by length, and total antibody epitope content. We used the complete antibody
405 epitope content, not corrected for Treg epitope content as a measure of immunogenicity. Spearman
406 Rho correlations between complete antibody epitope scores and SHM were calculated, and a linear
407 model was fitted to calculate slopes.

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

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

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

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

438 Determination of MHC-I/MHC-II shared epitopes was done by matching 9-mer peptides
439 (MHC-I peptide epitopes, 6 HLA-I alleles) into 15-mers (MHC-II peptide epitopes, 2 HLA-II
440 DRB1 alleles). Epitopes with a positive match were considered as shared epitopes and the
441 unmatched peptides were considered as exclusively MHC-I or MHC-II peptide epitopes. From
442 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 from 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.

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

479 We generated 30 complete simulated antibody repertoires for each of donors #1-5, for each of
480 the two mutational models generated as described in the previous paragraphs. Thus we simulated
481 a total of 150 personalized replacement-silent antibody repertoires *in silico*, and 150 out-of-frame
482 mutational model antibody repertoires, for a total of 300 simulated repertoires with an average of
483 27,000 antibodies each. Thus, we generated approximately 8,100,000 antibodies *in silico* and used
484 the resulting data to explore hypotheses related to mutational targeting in experimental antibody
485 data compared with simulated datasets.

486

487 ***Personalized MHC-II Peptide Epitope Content Analysis***

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

507

508 ***IEDB Data Mining***

509 We searched for experimentally validated antibody-derived peptides in the Immune Epitope
510 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 a 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

534 Superdex200 (GE Healthcare) size exclusion chromatography as described [65, 66]. Ig-derived
535 peptides and influenza HA306-318-derived probe peptide Ac-PRYVKQNTLRLAT were
536 synthesized (21st Century Biochemicals, Marlboro, MA). The probe peptide was labeled with
537 Alexa Fluor 488 tetrafluorophenyl ester (Invitrogen, Eugene, OR) through primary amine of K5.
538 Peptide binding was monitored using a fluorescence polarization assay [67]. The DR1
539 concentration used was selected by titrating DR1 against fixed labeled peptide concentration (25
540 nM) and choosing the concentration of DR1 that showed ~50% maximum binding. For calculating
541 IC₅₀ values, 100 nM DR1 was incubated with 25 nM Alexa488-labeled HA306–318 probe
542 peptide, in combination with a serial dilution of test peptides, beginning at 100 μM followed by 2-
543 fold dilutions. The reaction mixture was incubated at 37 °C. The capacity of each test peptide to
544 compete for binding of probe peptide was measured by FP after 72 h at 37 °C. FP values were
545 converted to fraction bound by calculating [(FP_{sample} - FP_{free})/(FP_{no_comp} - FP_{free})],
546 where FP_{sample} represents the FP value in the presence of test peptide; FP_{free} represents the
547 value for free Alexa488-conjugated HA306–318; and FP_{no_comp} represents values in the
548 absence of competitor peptide. We plotted fraction bound versus concentration of test peptide and
549 fit the curve to the equation $y = \text{bottom} + (\text{top} - \text{bottom})/(1 + [\text{pep}]/\text{IC50})$, where [pep] is the
550 concentration of test peptide, y is the fraction of probe peptide bound at that concentration of test
551 peptide, IC₅₀ is the 50% inhibitory concentration of the test peptide, top is the maximum fraction
552 of probe peptide bound, and bottom is the minimum fraction of probe peptide bound.
553

554 **Quantification and Statistical Analysis**

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

557 distributions were compared using the Kolmogorov-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 $y = \text{bottom} + (\text{top} - \text{bottom})/(1 + [\text{pep}]/\text{IC50})$, 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	XXXXXXXXXX
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)	<a href="https://github.com/mar
kvanderloo/stringdist">https://github.com/mar kvanderloo/stringdist
Other		

569

570

571 **Main Figures**
572

573 **Figure 1. Decreased MHC-II peptide epitope content is correlated with SHM in B cell**

574 receptors, with stronger effects in certain V-genes. **A.** Overview of MHC-II peptide epitope

575 characterization in natively paired heavy and light chain human antibody sequence repertoires.

576 Paired heavy and light chain antibody repertoire data were generated by ultra-high throughput

577 single cell sequencing of B cells from healthy donor PBMCs. An overlap-extension RT-PCR pairs

578 antibody heavy and light chain variable region (VH and VL) transcripts for NGS analysis. V(D)J

579 annotation and somatic hypermutation (SHM) assignment was carried out using IgBlast. MHC-II

580 peptide epitope content of BCR variable regions was analyzed for antibody sequence repertoires

581 using the EpiMatrix and netMHCIIpan algorithms. MHC-II peptide epitope content metrics were

582 cross-referenced with SHM and antibody isotype to characterize relationships between MHC-II

583 peptide epitope content and sequence-based markers of B cell development. **B.** Scatter plots of

584 EpiMatrix MHC-II binding prediction scores vs. SHM, based on aggregate data for human

585 supertype alleles DRB1*01:01, DRB1*03:01, DRB1*04:01, DRB1*07:01, DRB1*08:01,

586 DRB1*11:01, DRB1*13:02 and DRB1*15:01. Each point represents an antibody sequence; points

587 are colored according to data density (yellow: high, purple: low). Linear regressions are shown in

588 red. *p*-value of the Spearman correlation is indicated. **C.** Volcano plots of spearman ρ vs.

589 Benjamini-Hochberg adjusted *p*-values for MHC-II peptide epitope content vs. SHM, for antibody

590 repertoires binned by IGHV and IGKV/IGLV gene pairs. Statistically significant pairs are shown

591 in blue, and other gene pairs are shown in gray. **D.** Scatter plots of selected IGHV gene and

592 IGKV/IGLV gene pairs for SHM vs. predicted binding scores. Linear regression lines are shown

593 in blue.

594

595 **Figure 2. V-gene dependence is driven by the deletion of high affinity peptides present in**

596 **germline sequences.** **A.** Repertoire-scale data analysis schematic using netMHCIIpan to identify

597 patient-specific MHC-II peptide epitopes according to known donor HLA genes. **B.** Volcano plots

598 of Spearman ρ vs. Benjamini-Hochberg adjusted *p*-values for antibody SHM vs. geometric mean

599 K_D fold-change from germline K_D , as predicted by netMHCIIpan. Data were grouped by IGHV

600 gene and IGKV/IGLV gene pairings and analyzed for peptides derived from germline-encoded

601 MHC-II binding peptides (predicted germline $K_D < 1,000$ nM). Statistically significant IGHV gene

602 and IGKV/IGLV combinations are shown in blue, other gene pairs are shown in gray. **C.** Scatter

603 plots of antibody data for selected IGHV and IGKV/IGLV gene pairs displaying antibody SHM

604 vs. predicted peptide geometric mean K_D fold-change from germline K_D . Linear regressions are shown in

605 blue. **D.** Geometric mean of the rank percentage, as defined by netMHCIIpan of each putative

606 peptide across the IGHV sequence, comparing germline IGHV gene (black) and high SHM (top

607 5%, blue) from the IGHV gene-controlled repertoire. **E.** Logograms of high affinity germline-

608 encoded peptide residues comparing germline and high SHM antibodies at those residues (top

609 5%). *n* represents the number of unique peptides displayed in the high SHM subset. **F.**

610 netMHCIIpan K_D prediction for peptides shown in the logograms, using one of the donor-specific

611 HLA-DRB1 alleles. Peptides from Donors 1-3 are shown. **G.** Experimental validation of peptide

612 binding affinity to HLA II DRB1 molecules, using a competition assay with peptides derived from

613 Donor 1. IC₅₀ was calculated using a log-logistic equation. Somatic hypermutations are highlighted

614 in bold script.

615

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

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

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

678 **Supplementary Figures**

679

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

692

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

698

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

705

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

724 resource (IEDB, www.iedb.org). **B.** Presence of confirmed MHC-II peptide epitopes in antibody
725 repertoires. Peptides eluted from MHC-II molecules were retrieved from IEDB and used as a
726 search database to mine donor repertoire data. IEDB peptides present both as substrings entirely
727 contained within antibody 15-mers, and complete 15-mer matches, were accepted. **C.** Overlap
728 between confirmed HLA-DRB1 peptides and HLA-DP/DQ peptides from antibody V-genes found
729 in IEDB. Antibody peptides detected in the IEDB HLA-DRB1 database were searched in the HLA-
730 DP/DQ database, accepting only complete matches.
731

732 **Figure S6. netMHCIIpan data analysis, computational repertoire modeling, and**
733 **personalized repertoire analytics.** **A.** Data processing using netMHCIIpan. *Upper panel* The
734 presence of MHC-II peptide epitopes was determined in donor data for the complete set of 38
735 HLA alleles. HLA typing was also carried out. *Middle panel* Somatic hypermutation models
736 ShaZam and immuneSIM were used to simulate 30 repertoires, with the same number of BCR
737 sequences as experimentally-derived donor data. SHM distribution and V-gene frequencies were
738 calculated. *Lower panel* The subset of peptides with MHC-II $K_D < 1,000$ nM to any of the 38
739 alleles were selected to generate a database of potential predicted binders. **B.** Using the germline
740 peptide database, peptides at the same position within the V-region were extracted from
741 experimentally-derived donor data or simulated repertoires and grouped according to parent
742 antibody V-gene. The fold-change between repertoire-scale BCR geomean[peptide:MHC-II K_D]
743 and germline geomean[peptide:MHC-II K_D] was calculated and aggregated by V-gene. The
744 Spearman correlation between K_D fold-change and SHM was calculated for each V-gene. These
745 data was used for the plots shown in **Figure 2A**. **C.** Using correlation data from **B**, significant
746 (adjusted $p < 0.05$) and strong ($\rho > 0.5$) correlations were extracted and averaged by allele.
747 Alleles were plotted according to their individual geomean Spearman ρ scores, with the larger
748 circles corresponding to each of the donor's two HLA-DRB1 alleles.
749

750 **Figure S7. IGHV gene usage and SHM distribution for each experimentally-derived BCR**
751 **repertoire data, universal Out-of-frame (OoF) modeled repertoire data, and donor-specific**
752 **Replacement-Silent (RS) modeled repertoire data.** **A.** IGHV gene usage between Donors 2-5
753 experimentally-derived repertoires and OoF and RS modeled repertoires. Gene usage is shown as
754 frequency of total repertoire. **B.** SHM distribution for Donors 2-5 experimentally-derived
755 repertoires and OoF and RS modeled repertoires. Black dots represent outliers.
756

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

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

782
783 **Figure S9. Isotype class switching is correlated with preferential removal of MHC-II peptide**
784 **epitopes from BCRs. A.** Antibody repertoires were fractionated by isotype, and Spearman
785 correlations were calculated for each repertoire subset. EpiMatrix binding scores are shown as
786 aggregate binding score for supertype alleles DRB1*01:01, DRB1*03:01, DRB1*04:01,
787 DRB1*07:01, DRB1*08:01, DRB1*11:01, DRB1*13:02 and DRB1*15:01. Each point represents
788 a BCR sequence, and points are colored by data density (yellow: high, purple: low). Linear
789 regressions are shown in red; *p*-value of the Spearman correlation is indicated.

790
791 **Supplementary Table 1** Primers used for paired heavy and light chain overlap extension RT-
792 PCR

793
794

Oligonucleotide name	Oligonucleotide sequence	Source
IgM Constant Region	CGCAGTAGCGGTAAACGGCCACAGGAGACGAGGGGGAAA	McDaniel, JR., DeKosky, BJ., et al (2016)
IgG Constant Region	CGCAGTAGCGGTAAACGGCAGGGYGCCAGGGGAAGAC	McDaniel, JR., DeKosky, BJ., et al (2016)
IgA Constant Region	CGCAGTAGCGGTAAACGGCCGGGAAGACCTTGGGCTGG	McDaniel, JR., DeKosky, BJ., et al (2016)
IgLC Constant Region	GC GGATAACAATTACACAGGTCCAGAGGGYGGGA	McDaniel, JR., DeKosky, BJ., et al (2016)
IgKC Constant Region	GC GGATAACAATTACACAGGGATGAAGACAGATGGTCA	McDaniel, JR., DeKosky, BJ., et al (2016)
VH1 FR1 Region	TATTCCCATCGCGGCCAGGTCCAGCTKGTRCAGTCTGG	McDaniel, JR., DeKosky, BJ., et al (2016)
VH157 FR1 Region	TATTCCCATCGCGGCCAGGTGCAGCTGGTSARTCTGG	McDaniel, JR., DeKosky, BJ., et al (2016)
VH2 FR1 Region	TATTCCCATCGCGGCCAGRTCACCTGAAGGAGTCTG	McDaniel, JR., DeKosky, BJ., et al (2016)
VH3 FR1 Region	TATTCCCATCGCGGCCAGGTGCAGCTGKTGGAGWCY	McDaniel, JR., DeKosky, BJ., et al (2016)
VH4 FR1 Region	TATTCCCATCGCGGCCAGGTGCAGCTGCAGGAGTC	McDaniel, JR., DeKosky, BJ., et al (2016)
VH4-DP63 FR1 Region	TATTCCCATCGCGGCCAGGTGCAGCTACAGCAGTGGG	McDaniel, JR., DeKosky, BJ., et al (2016)
VH6 FR1 Region	TATTCCCATCGCGGCCAGGTACAGCTGCAGCAGTCA	McDaniel, JR., DeKosky, BJ., et al (2016)
VH3N FR1 Region	TATTCCCATCGCGGCCAGGTCAACACAACGGTCCCAGTTA	McDaniel, JR., DeKosky, BJ., et al (2016)
VK1 FR1 Region	GCGCCGCGATGGGAATAGCTAGCCGACATCCRGDTGACCCA	McDaniel, JR., DeKosky, BJ., et al (2016)
VK2 FR1 Region	GCGCCGCGATGGGAATAGCTAGCCGATATTGTGMTGACBCA	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	GCGCCGCGATGGGAATAGCTAGCCCAGTCTGYYCTGAYTCAG GCCT	McDaniel, JR., DeKosky, BJ., et al (2016)
VL3 FR1 Region	GCGCCGCGATGGGAATAGCTAGCCTCCTATGWGCTGACWCA GCCAA	McDaniel, JR., DeKosky, BJ., et al (2016)
VL-DPL16 FR1 Region	GCGCCGCGATGGGAATAGCTAGCCTCCTTGAGCTGASTCAG GASCC	McDaniel, JR., DeKosky, BJ., et al (2016)
VL3-38 FR1 Region	GCGCCGCGATGGGAATAGCTAGCCTCCTATGAGCTGAYRCAG CYACC	McDaniel, JR., DeKosky, BJ., et al (2016)
VL6 FR1 Region	GCGCCGCGATGGGAATAGCTAGCCAATTATGCTGACTCAG 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	NNNNATGGGCCCTGGTTGGGGCGGATGCACTCC	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	NNNNGTGCGGCCGCAGGGYGGGAACAGAGTGAC	McDaniel, JR., DeKosky, BJ., et al (2016)

796

797

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.

841 21. Cousens, L.P., et al., *In vitro and in vivo studies of IgG-derived Treg epitopes*
842 (*Tregitopes*): a promising new tool for tolerance induction and treatment of
843 autoimmunity. *J Clin Immunol*, 2013. **33 Suppl 1**: p. S43-9.

844 22. Peters, B., M. Nielsen, and A. Sette, *T Cell Epitope Predictions*. *Annu Rev Immunol*,
845 2020. **38**: p. 123-145.

846 23. Barra, C., et al., *Footprints of antigen processing boost MHC class II natural ligand*
847 *predictions*. *Genome Med*, 2018. **10**(1): p. 84.

848 24. Reynisson, B., et al., *NetMHCpan-4.1 and NetMHCIIpan-4.0: improved predictions of*
849 *MHC antigen presentation by concurrent motif deconvolution and integration of MS*
850 *MHC eluted ligand data*. *Nucleic Acids Res*, 2020. **48**(W1): p. W449-W454.

851 25. Wang, Q., et al., *Immunogenic HLA-DR-Presented Self-Peptides Identified Directly from*
852 *Clinical Samples of Synovial Tissue, Synovial Fluid, or Peripheral Blood in Patients with*
853 *Rheumatoid Arthritis or Lyme Arthritis*. *J Proteome Res*, 2017. **16**(1): p. 122-136.

854 26. Sordé, L., et al., *Tregitopes and impaired antigen presentation: Drivers of the*
855 *immunomodulatory effects of IVIg*? *Immun Inflamm Dis*, 2017. **5**(4): p. 400-415.

856 27. Huszthy, P.C., et al., *B cell receptor ligation induces display of V-region peptides on*
857 *MHC class II molecules to T cells*. *Proc Natl Acad Sci U S A*, 2019. **116**(51): p. 25850-
858 25859.

859 28. DeKosky, B.J., et al., *In-depth determination and analysis of the human paired heavy-*
860 *and light-chain antibody repertoire*. *Nat Med*, 2015. **21**(1): p. 86-91.

861 29. Wang, B., et al., *Functional interrogation and mining of natively paired human V*. *Nat*
862 *Biotechnol*, 2018. **36**(2): p. 152-155.

863 30. Schafer, J.R., et al., *Prediction of well-conserved HIV-1 ligands using a matrix-based*
864 *algorithm, EpiMatrix*. *Vaccine*, 1998. **16**(19): p. 1880-4.

865 31. Andreatta, M., et al., *Accurate pan-specific prediction of peptide-MHC class II binding*
866 *affinity with improved binding core identification*. *Immunogenetics*, 2015. **67**(11-12): p.
867 641-50.

868 32. Dhand, S.K., et al., *IEDB-AR: immune epitope database-analysis resource in 2019*.
869 *Nucleic Acids Res*, 2019. **47**(W1): p. W502-W506.

870 33. Roche, P.A. and K. Furuta, *The ins and outs of MHC class II-mediated antigen*
871 *processing and presentation*. *Nat Rev Immunol*, 2015. **15**(4): p. 203-16.

872 34. Nielsen, S.C.A., et al., *Shaping of infant B cell receptor repertoires by environmental*
873 *factors and infectious disease*. *Sci Transl Med*, 2019. **11**(481).

874 35. Johnson, E.L., et al., *Sequencing HIV-neutralizing antibody exons and introns reveals*
875 *detailed aspects of lineage maturation*. *Nat Commun*, 2018. **9**(1): p. 4136.

876 36. Honjo, T., K. Kinoshita, and M. Muramatsu, *Molecular mechanism of class switch*
877 *recombination: linkage with somatic hypermutation*. *Annu Rev Immunol*, 2002. **20**: p.
878 165-96.

879 37. Wine, Y., et al., *Molecular deconvolution of the monoclonal antibodies that comprise the*
880 *polyclonal serum response*. *Proc Natl Acad Sci U S A*, 2013. **110**(8): p. 2993-8.

881 38. Lavinder, J.J., et al., *Next-generation sequencing and protein mass spectrometry for the*
882 *comprehensive analysis of human cellular and serum antibody repertoires*. *Curr Opin*
883 *Chem Biol*, 2015. **24**: p. 112-20.

884 39. Chen, J., et al., *Proteomic Analysis of Pemphigus Autoantibodies Indicates a Larger,*
885 *More Diverse, and More Dynamic Repertoire than Determined by B Cell Genetics*. *Cell*
886 *Rep*, 2017. **18**(1): p. 237-247.

887 40. Lee, J., et al., *Molecular-level analysis of the serum antibody repertoire in young adults*
888 *before and after seasonal influenza vaccination*. Nat Med, 2016. **22**(12): p. 1456-1464.

889 41. Lee, J., et al., *Persistent Antibody Clonotypes Dominate the Serum Response to Influenza*
890 *over Multiple Years and Repeated Vaccinations*. Cell Host Microbe, 2019. **25**(3): p. 367-
891 376.e5.

892 42. Andreatta, M., et al., *An automated benchmarking platform for MHC class II binding*
893 *prediction methods*. Bioinformatics, 2018. **34**(9): p. 1522-1528.

894 43. Zhao, W. and X. Sher, *Systematically benchmarking peptide-MHC binding predictors:*
895 *From synthetic to naturally processed epitopes*. PLoS Comput Biol, 2018. **14**(11): p.
896 e1006457.

897 44. Halliley, J.L., et al., *Long-Lived Plasma Cells Are Contained within the CD19(-*
898 *)CD38(hi)CD138(+) Subset in Human Bone Marrow*. Immunity, 2015. **43**(1): p. 132-45.

899 45. Kwong, P.D. and J.R. Mascola, *HIV-1 Vaccines Based on Antibody Identification, B Cell*
900 *Ontogeny, and Epitope Structure*. Immunity, 2018. **48**(5): p. 855-871.

901 46. Bonsignori, M., et al., *Antibody-virus co-evolution in HIV infection: paths for HIV*
902 *vaccine development*. Immunol Rev, 2017. **275**(1): p. 145-160.

903 47. De Groot, A.S. and W. Martin, *Reducing risk, improving outcomes: bioengineering less*
904 *immunogenic protein therapeutics*. Clin Immunol, 2009. **131**(2): p. 189-201.

905 48. Vaisman-Mentesh, A., et al., *The Molecular Mechanisms That Underlie the Immune*
906 *Biology of Anti-drug Antibody Formation Following Treatment With Monoclonal*
907 *Antibodies*. Front Immunol, 2020. **11**: p. 1951.

908 49. Suurmond, J., et al., *Loss of an IgG plasma cell checkpoint in patients with lupus*. J
909 Allergy Clin Immunol, 2019. **143**(4): p. 1586-1597.

910 50. Yurasov, S., et al., *Defective B cell tolerance checkpoints in systemic lupus*
911 *erythematosus*. J Exp Med, 2005. **201**(5): p. 703-11.

912 51. Meffre, E. and K.C. O'Connor, *Impaired B-cell tolerance checkpoints promote the*
913 *development of autoimmune diseases and pathogenic autoantibodies*. Immunol Rev,
914 2019. **292**(1): p. 90-101.

915 52. Antia, A., et al., *Heterogeneity and longevity of antibody memory to viruses and*
916 *vaccines*. PLoS Biol, 2018. **16**(8): p. e2006601.

917 53. DeKosky, B.J., et al., *Large-scale sequence and structural comparisons of human naive*
918 *and antigen-experienced antibody repertoires*. Proc Natl Acad Sci U S A, 2016. **113**(19):
919 p. E2636-45.

920 54. McDaniel, J.R., et al., *Ultra-high-throughput sequencing of the immune receptor*
921 *repertoire from millions of lymphocytes*. Nat Protoc, 2016. **11**(3): p. 429-42.

922 55. Lagerman, C.E., et al., *Ultrasonically-guided flow focusing generates precise emulsion*
923 *droplets for high-throughput single cell analyses*. J Biosci Bioeng, 2019. **128**(2): p. 226-
924 233.

925 56. Ye, J., et al., *IgBLAST: an immunoglobulin variable domain sequence analysis tool*.
926 Nucleic Acids Res, 2013. **41**(Web Server issue): p. W34-40.

927 57. DeKosky, B.J., et al., *High-throughput sequencing of the paired human immunoglobulin*
928 *heavy and light chain repertoire*. Nat Biotechnol, 2013. **31**(2): p. 166-9.

929 58. Edgar, R.C., *Search and clustering orders of magnitude faster than BLAST*.
930 Bioinformatics, 2010. **26**(19): p. 2460-1.

931 59. Lund, O., et al., *Definition of supertypes for HLA molecules using clustering of specificity*
932 *matrices*. Immunogenetics, 2004. **55**(12): p. 797-810.

933 60. Southwood, S., et al., *Several common HLA-DR types share largely overlapping peptide*
934 *binding repertoires*. *J Immunol*, 1998. **160**(7): p. 3363-73.

935 61. Sette, A., et al., *The relationship between class I binding affinity and immunogenicity of*
936 *potential cytotoxic T cell epitopes*. *J Immunol*, 1994. **153**(12): p. 5586-92.

937 62. Gupta, N.T., et al., *Change-O: a toolkit for analyzing large-scale B cell immunoglobulin*
938 *repertoire sequencing data*. *Bioinformatics*, 2015. **31**(20): p. 3356-8.

939 63. Weber, C.R., et al., *immuneSIM: tunable multi-feature simulation of B- and T-cell*
940 *receptor repertoires for immunoinformatics benchmarking*. *Bioinformatics*, 2020. **36**(11):
941 p. 3594-3596.

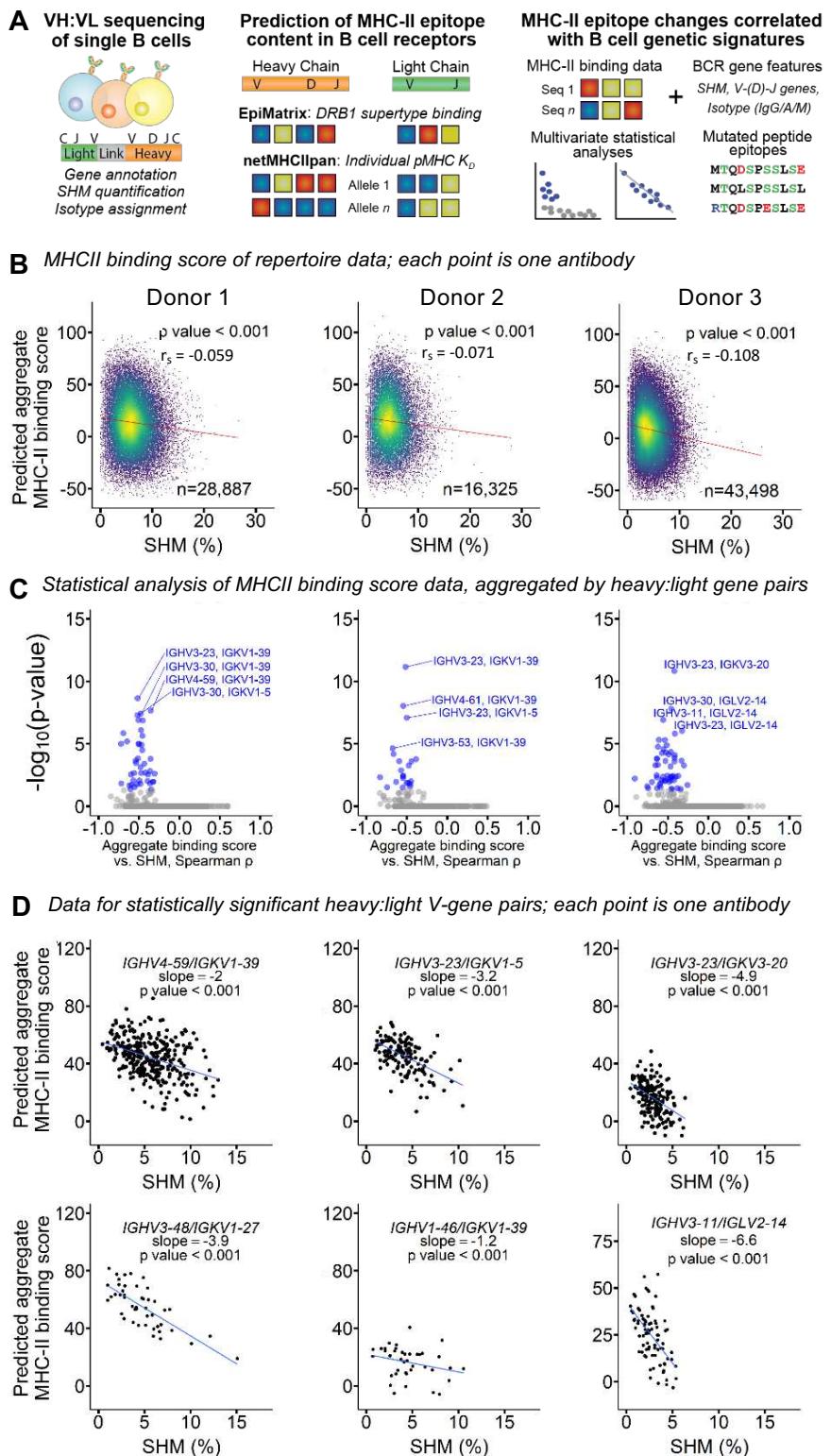
942 64. Bolotin, D.A., et al., *MiXCR: software for comprehensive adaptive immunity profiling*.
943 *Nat Methods*, 2015. **12**(5): p. 380-1.

944 65. Stern, L.J. and D.C. Wiley, *The human class II MHC protein HLA-DR1 assembles as*
945 *empty alpha beta heterodimers in the absence of antigenic peptide*. *Cell*, 1992. **68**(3): p.
946 465-77.

947 66. Sloan, V.S., et al., *Mediation by HLA-DM of dissociation of peptides from HLA-DR*.
948 *Nature*, 1995. **375**(6534): p. 802-6.

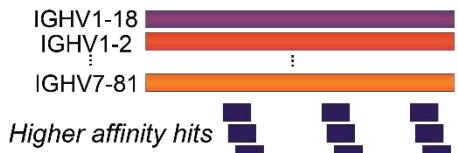
949 67. Yin, L. and L.J. Stern, *Measurement of Peptide Binding to MHC Class II Molecules by*
950 *Fluorescence Polarization*. *Curr Protoc Immunol*, 2014. **106**: p. 5.10.1-5.10.12.

951

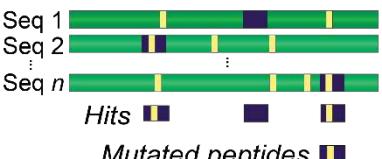


A

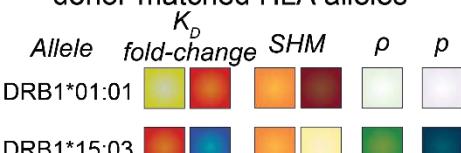
1. Identify peptides with high affinity to MHC-II in germline V-genes



2. Track peptide mutations in human antibody repertoires



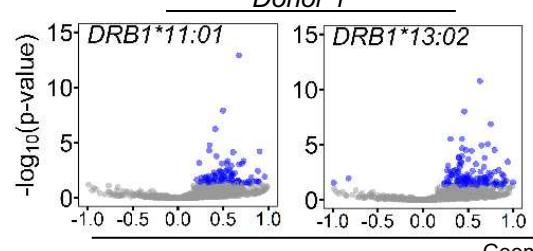
3. Determine peptide K_d fold-change from germline vs SHM for donor-matched HLA alleles



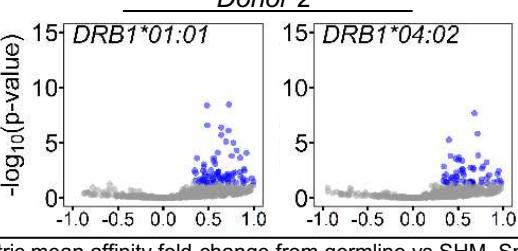
B

Statistical analysis of netMHCII peptide affinity data, aggregated by heavy:light gene pairs

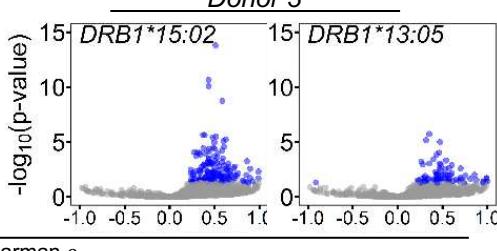
Donor 1



Donor 2

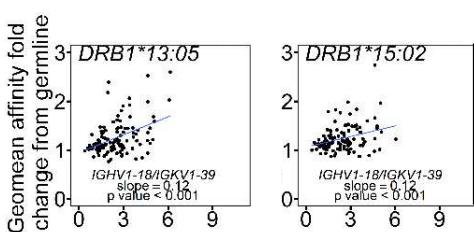
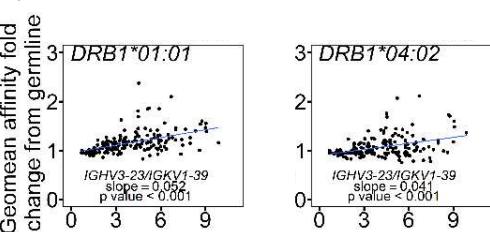
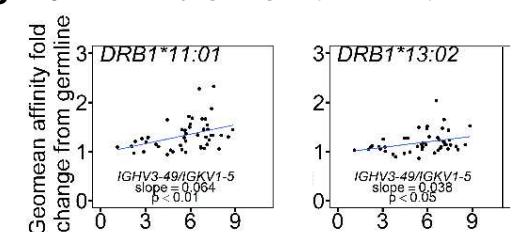


Donor 3



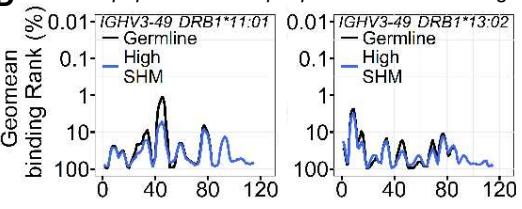
C

Significant heavy:light V-gene pairs; each point is one sequenced antibody

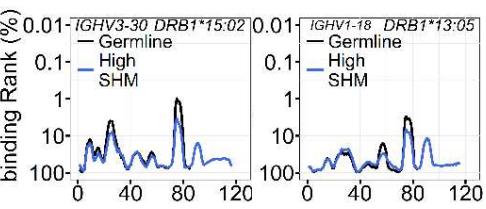
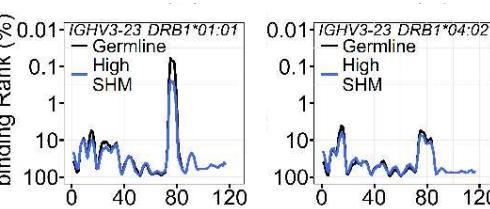


D

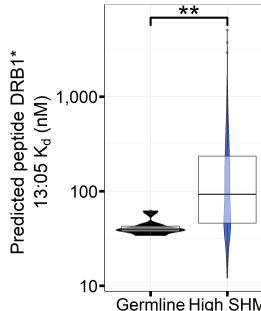
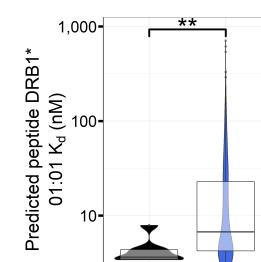
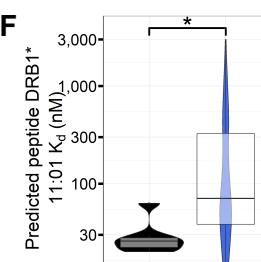
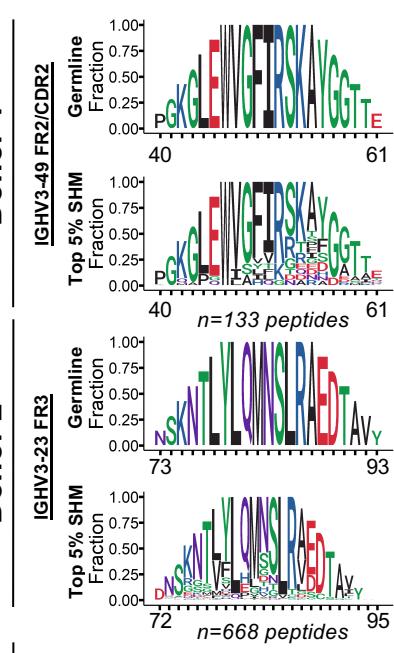
15-mer peptide MHCII epitope content for IGHV genes



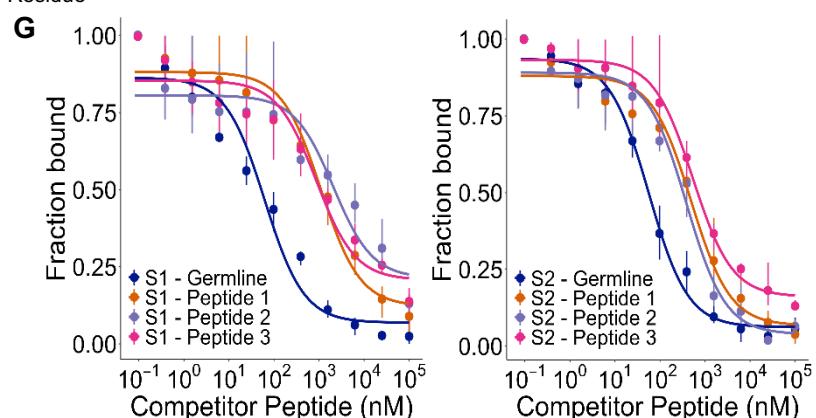
SHM (%)



E



Residue



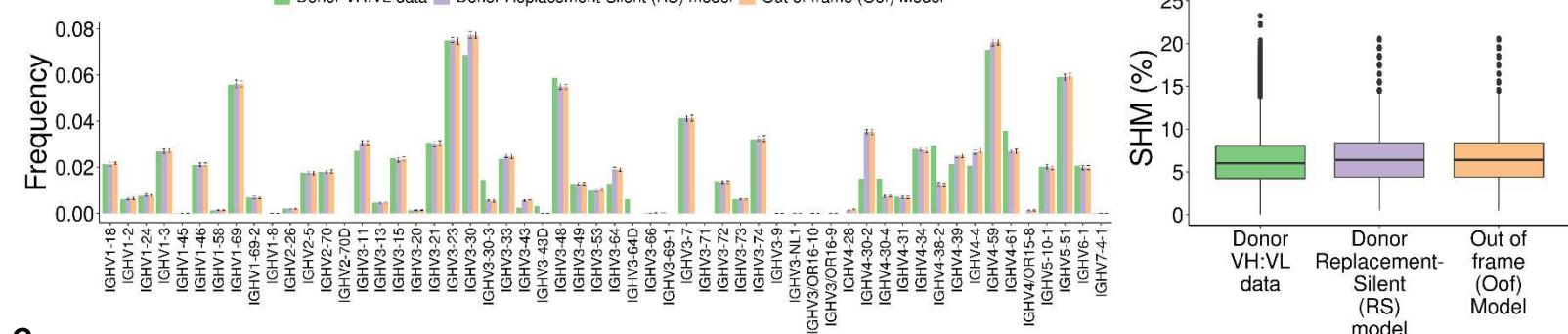
Peptide Sequence IC50 \pm S.D

Peptide	Sequence	IC50 \pm S.D
S1 - Germline	NTLYLQMNSLRAE	62.3 \pm 14.6
S1 - Peptide 1	NTVFLQMHSLRDE	1100 \pm 301
S1 - Peptide 2	SSVYLMQMDLRVE	2077 \pm 1106
S1 - Peptide 3	NTLYLHMSGLRDE	863 \pm 296
S2 - Germline	TLYLQMNSLRAED	56.3 \pm 8.5
S2 - Peptide 1	TVSLOQLNSLRAADD	490 \pm 111
S2 - Peptide 2	TLYLQMNRLLRAADD	387 \pm 91
S2 - Peptide 3	TLHLQVSSVRADD	537 \pm 129

A

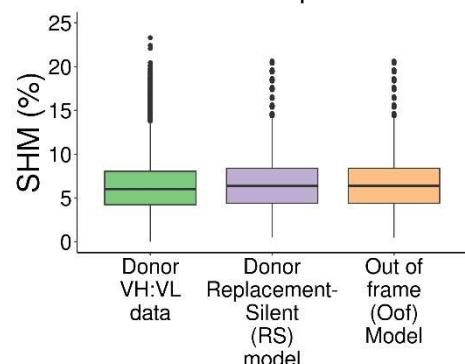
IGHV Gene Frequency of Donor and Modeled Repertoires

Donor VH:VL data Donor Replacement-Silent (RS) model Out of frame (Oof) Model

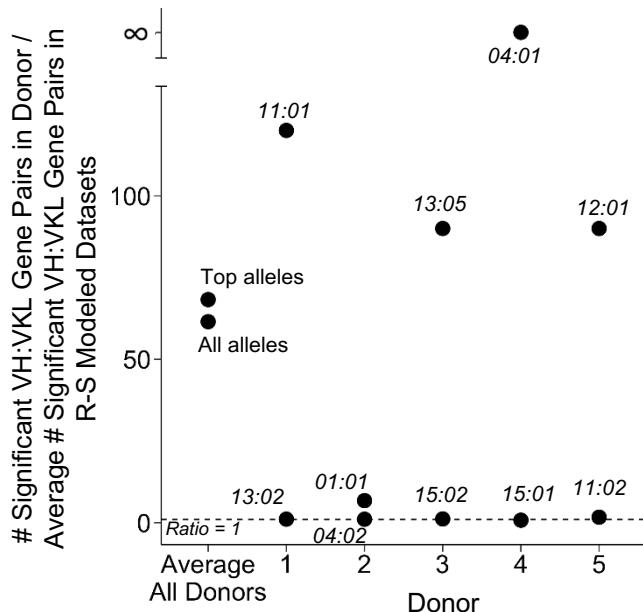


B

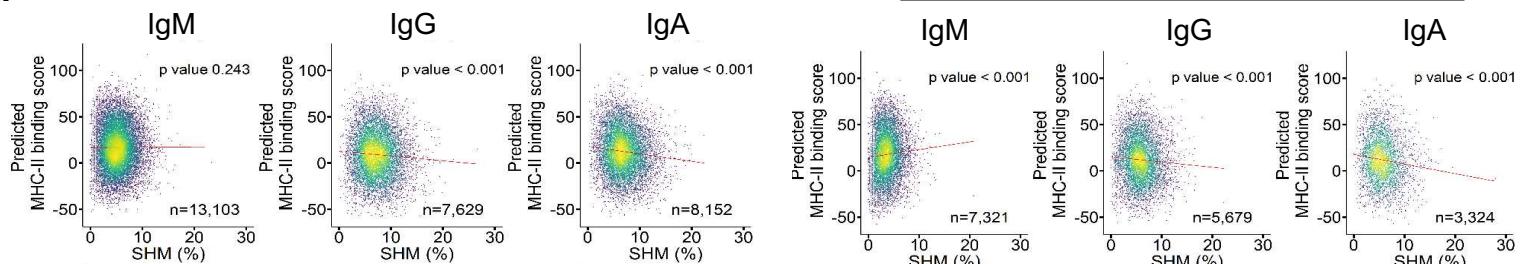
SHM Distribution Among Donor and Modeled Repertoires



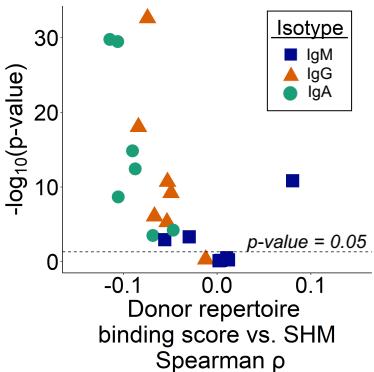
C



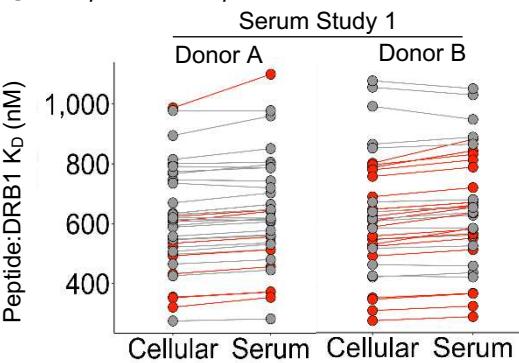
A



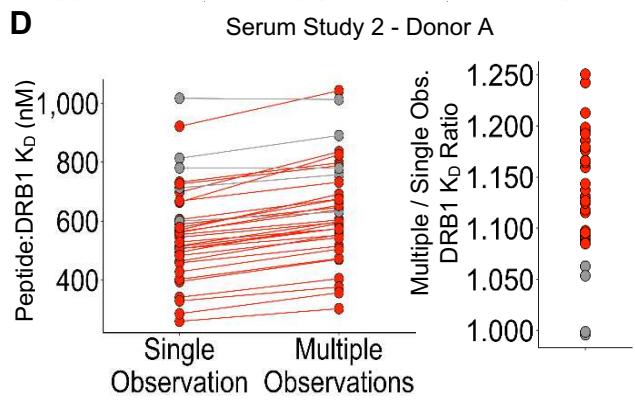
B



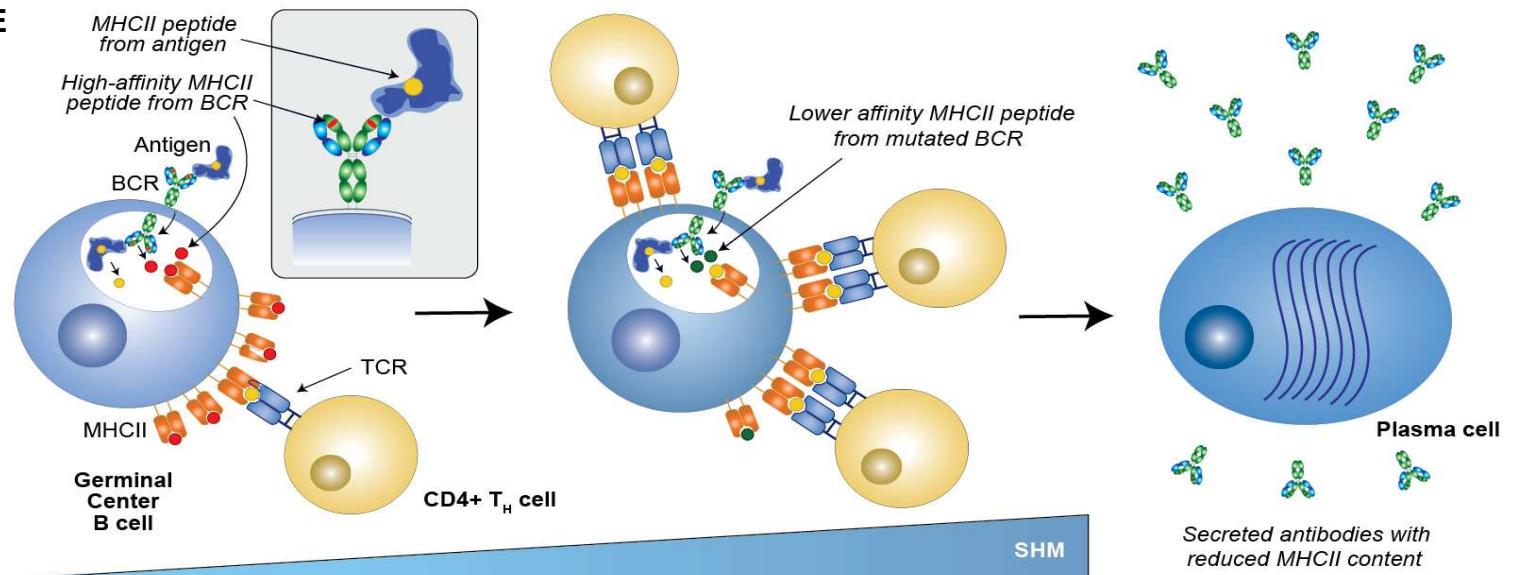
C Repertoire comparison for 38 DRB1 alleles



D



E



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 plasmablast plasma cell differentiation. The serum antibody repertoire has reduced CD4 T cell epitope affinity, resulting in lower patient-specific immunogenicity of secreted mAbs.