Supplemental Material

Mitigation of T-cell dependent immunogenicity by reengineering Factor VIIa analogue

Wojciech Jankowski[1], Joseph McGill[1], H.A. Daniel Lagassé[1], Stepan Surov[1,2], Gary Bembridge[3], Campbell Bunce[3], Edward Cloake[3], Mark H. Fogg[3], Katarzyna Jankowska[4], Abdul Khan[5], Joseph Marcotrigiano[5], Mikhail V. Ovanesov[1], Zuben E. Sauna[1]

Supplemental Methods……………………………………………………………………………1-3
Supplemental Figures S1-S6 ……………………………………………………………………4-8
Supplemental Tables S1-S3 ……………………………………………………………………9-11
References…………………………………………………………………………………………12

Supplemental Methods

Supplemental methods to Figure S1

Creation of a single score for each positional change:

The scoring function used to determine the viability of all amino acid substitutions at each position incorporated three factors: (i) A population-based score judging the in-silico probability of an immune response. (ii) A score judging the evolutionary pressure for conservation of an amino acid at each position in the sequence. (iii) A score assessing functional consequence of each amino acid substitution based on a third-party algorithm, EVmutation[1] .

Immunogenicity Score:

The immunogenicity score is equal to for any amino acid change defined as:

$$\sum_{p\varepsilon\, P}\sum_{a\in A} f_a \cdot T_t$$

Where:

P is the set of all possible peptides containing the positional change

A is the set of all HLA-DRB1 alleles considered

f_a *is the frequency of allele a in the population of interest*

T_t *is a binary operator testing whether the predicted binding affinity of peptide p and allele a is less than threshold t*

The Immunogenicity score is then compared to the Immunogenicity score for the original protein over the same positions but with no amino acid substitutions. If the amino acid substitution has a lower score than the original protein, it is considered a candidate for substitution.

Conservation score:

For each column in the sequence of interest, a score was calculated by examining the divergence between a baseline probability distribution and the probability distribution of amino acids in that column. The baseline amino acid distribution was derived from the Uniprot-Knowledge-base- SwissProt database. The distribution is the frequency of amino acids within the database (553,474 proteins). This background distribution represents an expected distribution of amino acids with no evolutionary pressure.

The set of sequences examined included a wide distribution of species. The sequences for FVII in these species were aligned using MUSCLE and the 31 amino acid sequences of interest were extracted from these sequences.

Jensen Shannon Divergence:

$$\lambda = \frac{1}{2}$$

$$RE_{p_c,q} = \sum_{\alpha \varepsilon A.A.} p_C(\alpha) log \frac{p_c(\alpha)}{q(\alpha)}$$

$$r = \lambda p_c + (1 - \lambda) q$$

$$D_C^{JS} = \lambda RE_{p_c,r} + (1 - \lambda) RE_{(q,r)}$$

Where:

$p_c(\alpha) = probability\ distribution\ of amino\ acids\ of\ the\ samples$

$q(\alpha) = probability\ distribution\ of amino\ acids\ UniProtKB - Swiss - Prot$

$c\ is\ the\ column\ being\ observed$

$\alpha\ is\ one\ particular\ amino\ acid\ being\ observed$

The divergence between the background distribution and the observed distribution was evaluated using the Jensen-Shannon-Divergence score. This score has similar properties to the Kullback-Leibler divergence that is often used to evaluate the divergence of probability distance[2] ; however, it is symmetric and bounded by 0 and 1. It was found to be a favorable tool for identifying functionally important residues in proteins.

A position in the peptide is considered to be not conserved if the amino acids at that position could be seen as more likely to occur as random chance rather than as a result of evolutionary pressure. A position with a random assortment of amino acids would have a lower Jensen-Shannon-Divergence score than a position that was more homogeneous as it is closer to the baseline random distribution of peptides in the proteome. These low scoring positions can be considered to have less evolutionary pressure for amino acid conservation.

A Conservation score was calculated for each position on the peptide. During analysis, only a subset of positions in the region of interest was considered. Positions which are greater than one mean absolute deviation below the median are considered to be the less conserved positions within the region of interest and are retained for further consideration.

EVmutation Score

The EVmutation score was obtained by using the precomputed single mutation scores available at the EVmutation website [https://marks.hms.harvard.edu/EVmutation/index.html]. Scores of roughly 0 or higher are considered stable mutations. For our analysis, we kept scores that were greater than -0.2 in order to have a small amount of leeway around 0.

Overall Consideration:

Using this method, the number of possible changes to be made was reduced from 285 to 123 after the immunogenicity step, 42 after the conservation step and 5 after the EVmutation step. The python and R code used for the selection of the optimal positions is available upon request

Supplemental Figures

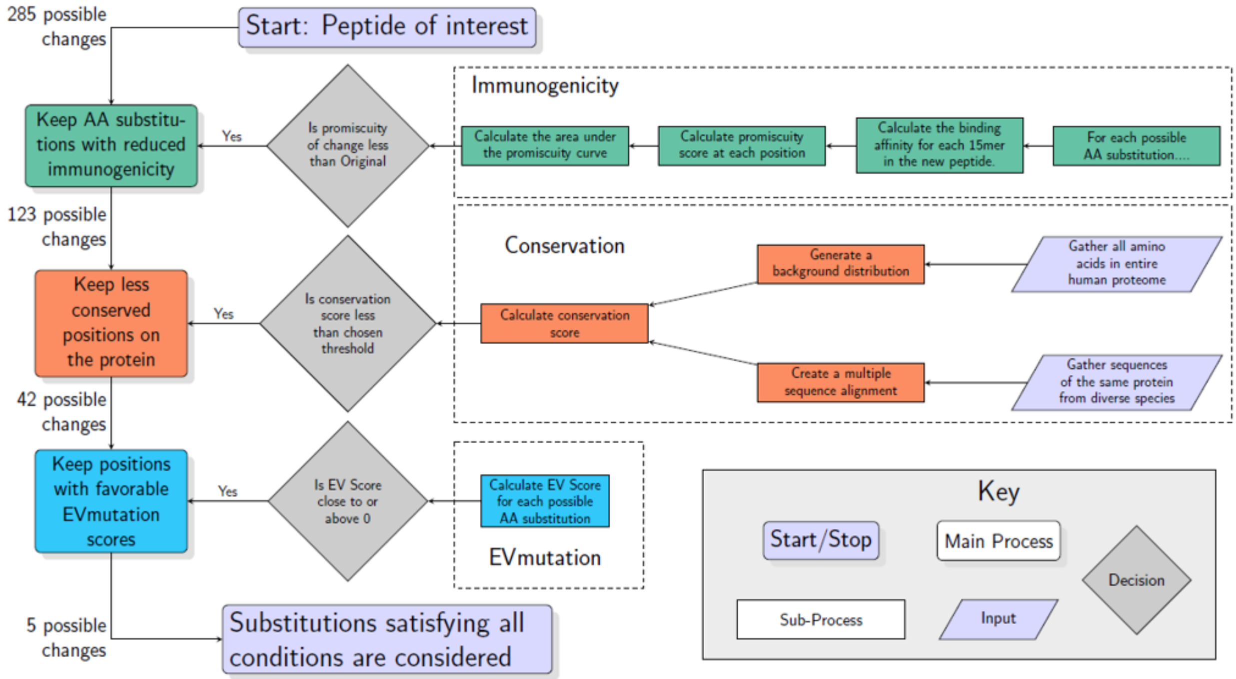

Figure S1. The Deimmunization algorithm (*RID*). *RID* is a set of steps intended to filter the possible search space for amino acid changes. There are three filters: 1. The use of the promiscuity scores to determine whether an amino acid change will result in reduced engagement with HLA variants at the population level (depicted in green on the workflow). 2. The use of a conservation score applied to amino acid substitutions that result in lower promiscuity scores, to identify primary sequence locations that should not be altered as they are likely to be functionally important (depicted in orange on the workflow). 3. The determination of an EVmutation score[25] for amino acid substitutions that result in lower promiscuity scores and are ***not*** conserved to identify amino acid changes that could be functionally deleterious (depicted in blue on the workflow).

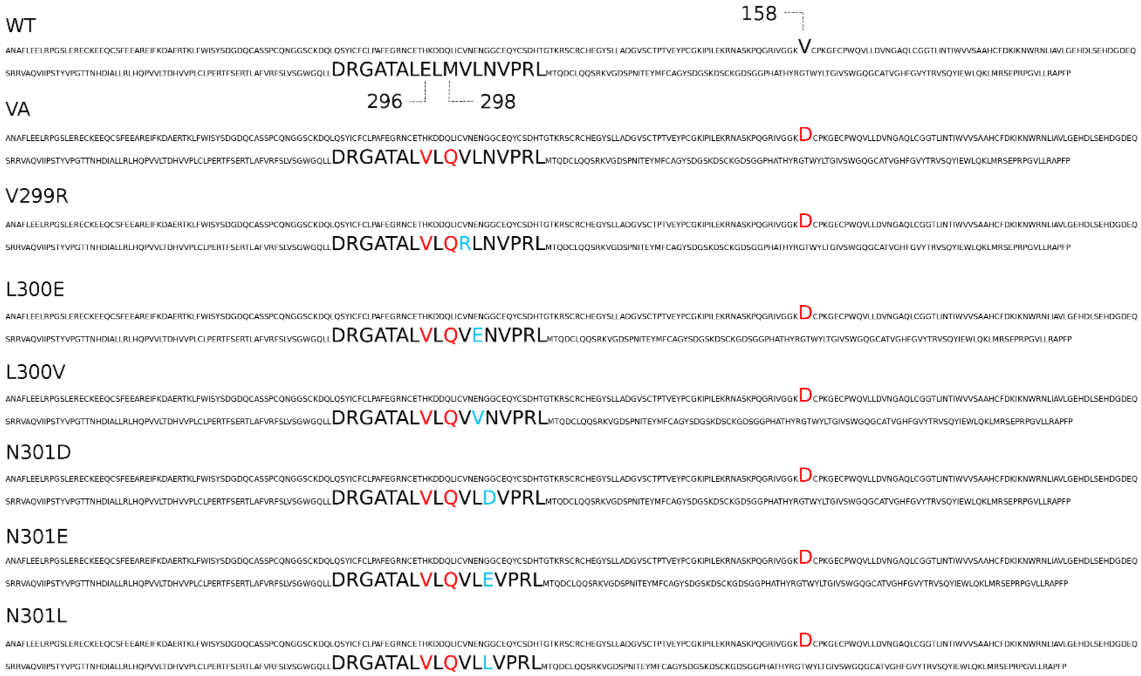

Figure S2. Sequence alignment. Amino acid sequence alignment of wild-type FVIIa, vatreptacog alfa (VA), and the 6 variants analyzed. Substitutions inherent to VA are shown in red. Additional substitutions are shown in blue. Larger fonts are used to highlight the substitution in VA at position 158 as well as the substitutions made in the immunogenic region of interest surrounding positions 296 and 298.

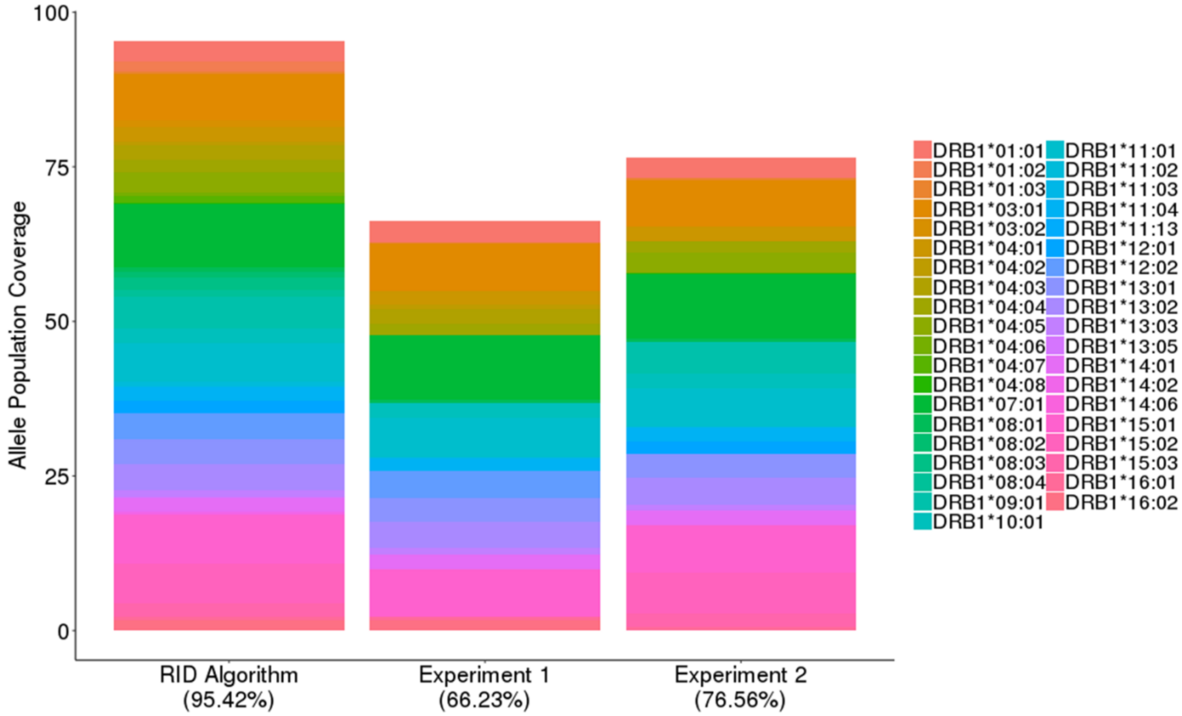

Figure S3. Frequencies of HLA-DRB1 variants used in this study. The relative frequency of individual HLA-DRB1 variants used in predicting immunogenicity using RID (left) and the relative frequencies of HLA-DRB1 variants identified in the 50 subjects used in each of the two assays (center and right). The total coverage of alleles in the world population is also indicated.

Figure S4. T-cell Responses. Cells from 50 subjects were subjected to a 3H-thymidine incorporation T-cell proliferation assay and IL-2-ELISpot assay. A stimulation index > 1.9 is considered a positive responder for either assay. Positive donors are shown in red for the T-cell Proliferation assay and in blue for the ELISpot assay.

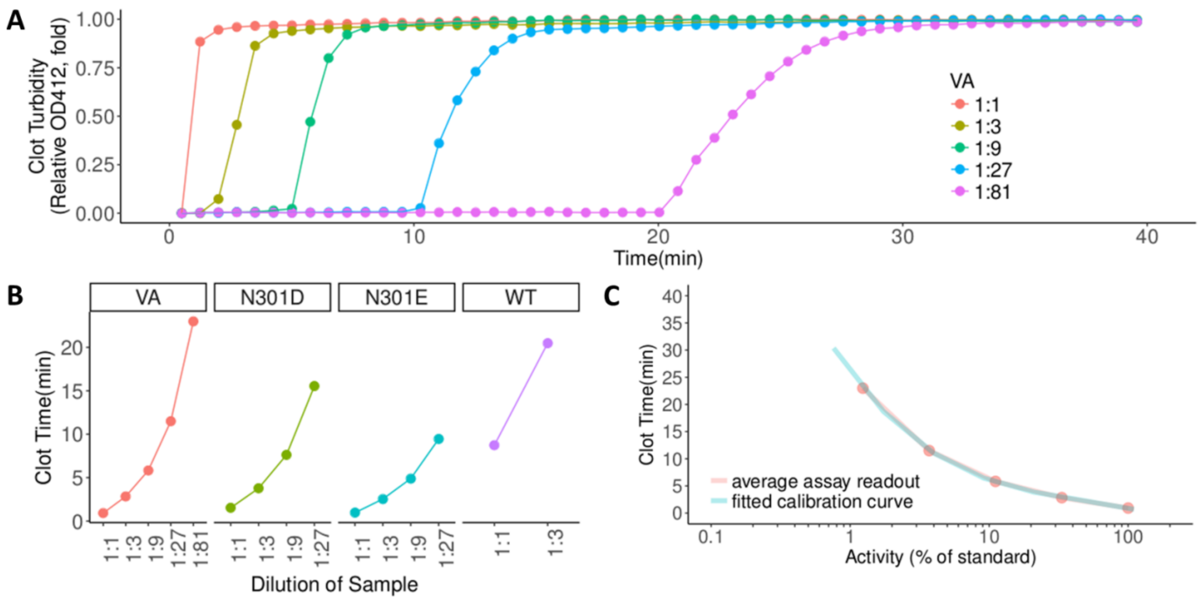

Figure S5. Quantification of FVIIa-like activity by clotting assay. (A) FVIIa samples were serially diluted in Tris-BSA buffer (pH 7.4) and added to a commercially available plasma from congenital hemophilia A patients pre-mixed with 4 µM of phospholipid (PS:PC) vesicles. Clot formation was monitored through optical density (clot turbidity) at 412nm. (B) Clot time values were calculated for each dilution of the FVIIa samples from the kinetic data as the point when optical density reached half of its maximal value. (C) Calibration curve used to calculate relative activities of FVIIa variants by comparing the clot times for several vatreptacog alfa (VA) serial dilutions.

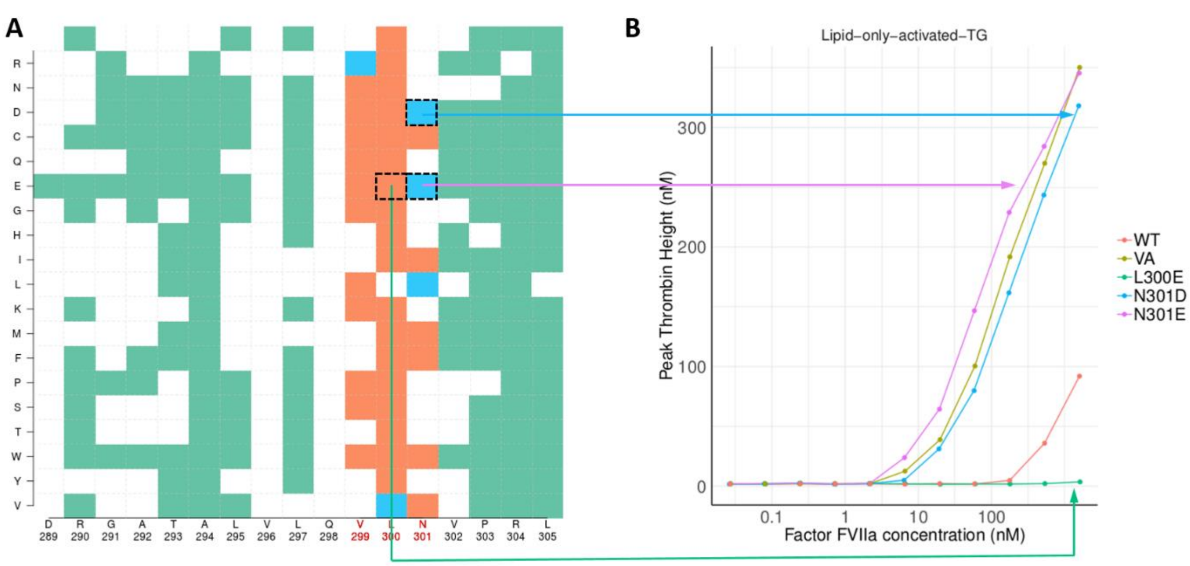

Figure S6. Validation of the selected screening candidates. (A) Selected mutations N301D and N301E (blue squares), predicted to retain function comparable to VA and the mutation L300D (orange square) predicted to be non-functional. (B) Activities of wild-type FVIIa, the vatreptacog alfa (VA) discontinues variant and the mutants L300E, N301D and N301E determined using the Thrombin Generation assay in the presence of phospholipid vesicles (4 µM).

Supplemental Tables

	African	European	Native American	North American	NorthEast Asian	SouthEast Asian	World
DRB1 0101	2.586	7.737	4.644	7.352	3.391	0.932	3.393
DRB1 0102	4.01	1.297	3.304	1.663	0.049	0.158	1.579
DRB1 0103	0.25	1.041	0.712	1.032	0.015	0.041	0.336
DRB1 0301	6.969	11.43	7.12	11.152	4.291	5.109	7.574
DRB1 0302	6.315	0.036	1.19	0.913	0.004	0.015	1.093
DRB1 0401	1.932	7.51	2.114	7.389	0.691	0.419	2.303
DRB1 0402	0.083	1.067	1.234	0.87	0.034	0.188	0.609
DRB1 0403	0.194	1.459	1.604	0.783	2.424	3.109	2.415
DRB1 0404	0.8	3.563	4.592	3.29	1.006	0.603	1.914
DRB1 0405	1.565	0.729	2.456	1.084	8.254	5.913	3.296
DRB1 0406	0.023	0.061	0.106	0.171	3.13	0.634	0.682
DRB1 0407	0.373	0.977	6.26	0.96	0.287	0.1	1.032
DRB1 0408	0.058	0.356	0.13	0.322	0.033	0.078	0.144
DRB1 0701	10.06	13.764	10.827	12.523	5.483	7.317	10.299
DRB1 0801	0.356	2.014	1.13	1.925	0.036	0.122	0.664
DRB1 0802	0.087	0.136	6.392	0.107	1.761	0.213	0.936
DRB1 0803	0.035	0.296	0.201	0.56	7.19	3.111	2.06
DRB1 0804	5.464	0.213	1.088	0.92	0.009	0.047	1.131
DRB1 0901	2.968	0.999	1.276	2.081	13.079	8.062	5.089
DRB1 1001	1.942	1.579	1.586	1.024	1.364	3.995	2.34
DRB1 1101	8.479	5.763	4.138	6.019	5.223	4.494	6.278
DRB1 1102	3.855	0.246	1.251	0.77	0.005	0.019	0.773
DRB1 1103	0.068	0.548	0.365	0.5	0.013	0.024	0.204
DRB1 1104	0.623	3.255	2.897	2.469	0.163	0.707	2.105
DRB1 1201	3.753	1.5	1.181	2.041	4.004	0.845	1.981
DRB1 1202	0.281	0.386	0.111	0.668	7.212	17.077	4.278
DRB1 1301	5.404	5.702	4.68	5.338	1.141	1.656	3.998
DRB1 1302	7.35	4.714	4.11	5.092	5.208	1.757	4.228
DRB1 1303	3.317	1.017	1.327	1.337	0.018	0.08	1.076
DRB1 1305	0.035	0.23	0.328	0.209	0	0.02	0.112
DRB1 1401	1.868	2.499	1.862	2.544	3.045	2.256	2.275
DRB1 1402	0.079	0.035	2.304	0.039	0.027	0.044	0.274
DRB1 1406	0.009	0.026	2.57	0.019	0.441	0.038	0.262
DRB1 1501	2.725	12.598	6.347	11.787	9.103	4.656	7.779
DRB1 1502	0.26	2.064	1.09	0.758	3.715	19.553	6.564
DRB1 1503	11.85	0.05	1.182	1.674	0.003	0.031	2.173
DRB1 1601	0.192	1.326	0.959	1.184	0.027	0.072	0.546
DRB1 1602	1.385	0.238	2.127	0.547	2.608	1.628	1.619

Table S1. Percentage of each HLA-DRB1 allele in different human populations. These frequency distributions were used in silico calculations of weighted promiscuity scores.

Allele	Assay 1 VA/WT	Assay 2 VA/WT/N301D/N301E	Population
DRB1_01_01	0.1300	0.0918	0.0339
DRB1_01_03	0.0100	0.0204	0.0034
DRB1_03_01	0.1200	0.0714	0.0757
DRB1_04_01	0.0600	0.0714	0.0230
DRB1_04_02	0.0200	0.0000	0.0061
DRB1_04_03	0.0400	0.0000	0.0242
DRB1_04_04	0.0500	0.0510	0.0191
DRB1_04_05	0.0000	0.0102	0.0330
DRB1_04_08	0.0000	0.0204	0.0014
DRB1_07_01	0.1000	0.1224	0.1030
DRB1_08_01	0.0300	0.0306	0.0066
DRB1_09_01	0.0000	0.0102	0.0509
DRB1_10_01	0.0200	0.0102	0.0234
DRB1_11_01	0.0300	0.0204	0.0628
DRB1_11_03	0.0200	0.0102	0.0020
DRB1_11_04	0.0400	0.0306	0.0211
DRB1_12_01	0.0000	0.0102	0.0198
DRB1_12_02	0.0100	0.0000	0.0428
DRB1_13_01	0.0800	0.0918	0.0400
DRB1_13_02	0.0500	0.0918	0.0423
DRB1_13_03	0.0200	0.0102	0.0108
DRB1_14_01	0.0067	0.0102	0.0228
DRB1_14_04	0.0100	0.0000	0.0100
DRB1_15_01	0.1300	0.1633	0.0778
DRB1_15_02	0.0000	0.0204	0.0656
DRB1_15_03	0.0000	0.0102	0.0217
DRB1_16_01	0.0100	0.0204	0.0055
DRB1_16_02	0.0100	0.0000	0.0162

Table S2. Percentage of each HLA-DRB1 allele in each of the two T-cell proliferation experiments as well as the respective frequency of that allele in the World population.

Assay 1

VA/WT

	African	European	Native_American	North_American	NorthEast_Asian	SouthEast_Asian	World
DRB1_01_01	2.586	7.737	4.644	7.352	3.391	0.932	3.393
DRB1_01_03	0.25	1.041	0.712	1.032	0.015	0.041	0.336
DRB1_03_01	6.969	11.43	7.12	11.152	4.291	5.109	7.574
DRB1_04_01	1.932	7.51	2.114	7.389	0.691	0.419	2.303
DRB1_04_02	0.083	1.067	1.234	0.87	0.034	0.188	0.609
DRB1_04_03	0.194	1.459	1.604	0.783	2.424	3.109	2.415
DRB1_04_04	0.8	3.563	4.592	3.29	1.006	0.603	1.914
DRB1_04_05							
DRB1_04_08							
DRB1_07_01	10.06	13.764	10.827	12.523	5.483	7.317	10.299
DRB1_08_01	0.356	2.014	1.13	1.925	0.036	0.122	0.664
DRB1_09_01							
DRB1_10_01	1.942	1.579	1.586	1.024	1.364	3.995	2.34
DRB1_11_01	8.479	5.763	4.138	6.019	5.223	4.494	6.278
DRB1_11_03	0.068	0.548	0.365	0.5	0.013	0.024	0.204
DRB1_11_04	0.623	3.255	2.897	2.469	0.163	0.707	2.105
DRB1_12_01							
DRB1_12_02	0.281	0.386	0.111	0.668	7.212	17.077	4.278
DRB1_13_01	5.404	5.702	4.68	5.338	1.141	1.656	3.998
DRB1_13_02	7.35	4.714	4.11	5.092	5.208	1.757	4.228
DRB1_13_03	3.317	1.017	1.327	1.337	0.018	0.08	1.076
DRB1_14_01	1.868	2.499	1.862	2.544	3.045	2.256	2.275
DRB1_15_01	2.725	12.598	6.347	11.787	9.103	4.656	7.779
DRB1_15_02							
DRB1_15_03							
DRB1_16_01	0.192	1.326	0.959	1.184	0.027	0.072	0.546
DRB1_16_02	1.385	0.238	2.127	0.547	2.608	1.628	1.619
Totals	**56.864**	**89.21**	**64.486**	**84.825**	**52.496**	**56.242**	**66.233**

Assay 2

VA/WT/N301D/N301E

	African	European	Native_American	North_American	NorthEast_Asian	SouthEast_Asian	World
DRB1_01_01	2.586	7.737	4.644	7.352	3.391	0.932	3.393
DRB1_01_03	0.25	1.041	0.712	1.032	0.015	0.041	0.336
DRB1_03_01	6.969	11.43	7.12	11.152	4.291	5.109	7.574
DRB1_04_01	1.932	7.51	2.114	7.389	0.691	0.419	2.303
DRB1_04_02							
DRB1_04_03							
DRB1_04_04	0.8	3.563	4.592	3.29	1.006	0.603	1.914
DRB1_04_05	1.565	0.729	2.456	1.084	8.254	5.913	3.296
DRB1_04_08	0.058	0.356	0.13	0.322	0.033	0.078	0.144
DRB1_07_01	10.06	13.764	10.827	12.523	5.483	7.317	10.299
DRB1_08_01	0.356	2.014	1.13	1.925	0.036	0.122	0.664
DRB1_09_01	2.968	0.999	1.276	2.081	13.079	8.062	5.089
DRB1_10_01	1.942	1.579	1.586	1.024	1.364	3.995	2.34
DRB1_11_01	8.479	5.763	4.138	6.019	5.223	4.494	6.278
DRB1_11_03	0.068	0.548	0.365	0.5	0.013	0.024	0.204
DRB1_11_04	0.623	3.255	2.897	2.469	0.163	0.707	2.105
DRB1_12_01	3.753	1.5	1.181	2.041	4.004	0.845	1.981
DRB1_12_02							
DRB1_13_01	5.404	5.702	4.68	5.338	1.141	1.656	3.998
DRB1_13_02	7.35	4.714	4.11	5.092	5.208	1.757	4.228
DRB1_13_03	3.317	1.017	1.327	1.337	0.018	0.08	1.076
DRB1_14_01	1.868	2.499	1.862	2.544	3.045	2.256	2.275
DRB1_15_01	2.725	12.598	6.347	11.787	9.103	4.656	7.779
DRB1_15_02	0.26	2.064	1.09	0.758	3.715	19.553	6.564
DRB1_15_03	11.85	0.05	1.182	1.674	0.003	0.031	2.173
DRB1_16_01	0.192	1.326	0.959	1.184	0.027	0.072	0.546
DRB1_16_02							
Totals	**75.375**	**91.758**	**66.725**	**89.917**	**69.306**	**68.722**	**76.559**

Table S3. Percentage of each HLA-DRB1 allele in different human populations for each of the two T-cell proliferation experiments.

References

1. Hopf TA, Ingraham JB, Poelwijk FJ, et al. Mutation effects predicted from sequence co-variation. *Nat Biotechnol.* 2017;35(2):128-135.
2. Capra JA, Singh M. Predicting functionally important residues from sequence conservation. *Bioinformatics.* 2007;23(15):1875-1882.