Supplementary information for:

Genetic variation in MHC proteins is associated with T cell receptor expression biases

Eilon Sharon[1,2,8], Leah V. Sibener[3,4,5,8], Alexis Battle[6], Hunter B. Fraser[2], K. Christopher Garcia[3,4,7,9] & Jonathan K. Pritchard[1,2,7,9]

1 Department of Genetics, Stanford University, Stanford, CA 94305, USA.
2 Department of Biology, Stanford University, Stanford, CA 94305, USA.
3 Department of Molecular Physiology, Stanford University School of Medicine, Stanford, CA 94305, USA
4 Department of Structural Biology, Stanford University School of Medicine, Stanford, CA 94305, USA
5 Immunology Program, Stanford University, Stanford, CA 94305, USA.
6 Department of Computer Science, Johns Hopkins University, Baltimore, MD 21218, USA
7 Howard Hughes Medical Institute, Stanford University, Stanford, CA 94305, USA.
8 These authors contributed equally to this work.
9 Corresponding authors

List of Tables

List of Figures

closest TCR residues in solved TCR-peptide-MHC complexes.
Supplementary Fig. 29. Distance between MHC class II β chain amino acid residues and the closest peptide residue in solved TCR-peptide-MHC complexes.
Supplementary Fig. 30. MHC residues that are associated with TCR Vα genes expression tend to be closer to the TCR.
Supplementary Fig. 31. Associations of TCR Vα and Vβ genes with variation in MHC proteins are largely independent of SNPs shown previously to associate with the ratio of CD4:CD8 T cells.
Supplementary Fig. 32. MHC amino acid positions posterior probability of influencing any TCR Vα expression is not correlated with its imputation quality.

Supplementary Tables

1	Sequencing Depth	23	Time of Day Blood Drawn
2	Number of Coding Bases	24	Percent Hemoglobin
3	Number of UTR Bases	25	Individual-Specific GC
4	Number of PF Aligned Bases	26	Percent Duplicated Reads
5	Number of BF Bases	27	Median 3Prime Bias
6	Percent Coding Bases	28	Median CV Coverage
7	Percent MRNA Bases	29	Cell Frequency: Tc Cells
8	Percent Usable Bases	30	Globin Flag (Technician)
9	Percent UTR Bases	31	Number of Intergenic Bases
10	Cell Frequency: Mono	32	Number of Intronic Bases
11	Cell Frequency: DC	33	Percent Intergenic Bases
12	RNA Yield	34	Percent Intronic Bases
13	Cell Frequency: Neutro	35	Cell Frequency: Tc_act Cells
14	Individual-Specific Exon Length	36	Sex
15	Median 5PRime Biase	37	Genotyping PC1
16	Median 5Prime to 3Prime Bias	38	Genotyping PC2
17	Cell Frequency: NK Cells	39	Genotyping PC3
18	Cell Frequency: Th Cells	40	# TCR mapped reads
19	Cell Frequency: Platelet Cells	41	# α chain mapped reads
20	Cell Frequency: NK_act Cells	42	# β chain mapped reads
21	Cell Frequency: DC_act Cells	43	# γ chain mapped reads
22	Cell Frequency: B Cells	44	# δ chain mapped reads

Supplementary Table 1. Covariates used in the linear regression. The table lists all the covariates (potential technical and biological confounding factors and the total number of reads that map to each TCR chain genes) that the log reads counts were regressed on. Covariates 1–39 are as described in Battle *et al.*[1]. Covariates 40–44 are the total number of reads that are mapped to TCR and to each of the chains (For Ig genes these covariates were replaced by the equivalent Ig-suitable covariates). The reads that mapped to each chain include reads that are mapped to J and C TCR genes. All sequencing-read count covariates were log-transformed.

Legends of tables that are attached as excel files:

Supplementary Table 2. TCR V-genes expression. Normalized expression of Ig V-genes (rows) in 895 individuals (columns).

Supplementary Table 3. Ig V-genes expression. Normalized expression of Ig V-genes (rows) in 895 individuals (columns).

Supplementary Table 4. TCR and Ig V-genes association with short-range genetic variation and genetic variation in the MHC locus. First tab contains the results of association test between expression of TCR and Ig V-genes and short range (<1Mb) imputed genetic variation. The empirical significance of the strongest association for each gene was computed using 10,000 permutations. 5% FDR was used to control for multiple hypothesis testing (calculated separately for each chain). Second tab contain the results of a similar analysis for long range association with genotyped genetic variability in the extended MHC locus.

Supplementary Table 5. Most significance association between expression of each TCR and Ig V-gene and genotyped SNPs in the MHC locus. The table contains for each V-gene the most significance association between its expression and a measured SNP in the extended MHC locus.

Supplementary Table 6. Nucleotide and amino acid variants in the MHC locus that are independently associated with single TCR Vα or Vβ-gene expression. Variants were selected using a forward stepwise regression (results are shown in Fig. 3A).

Supplementary Table 7. Expression variation of TCR Vα and Vβ-genes explained by indepedent associations with SNP and amino acid variation in the MHC locus. Expression variation explained by the models presented in Figure 3A.

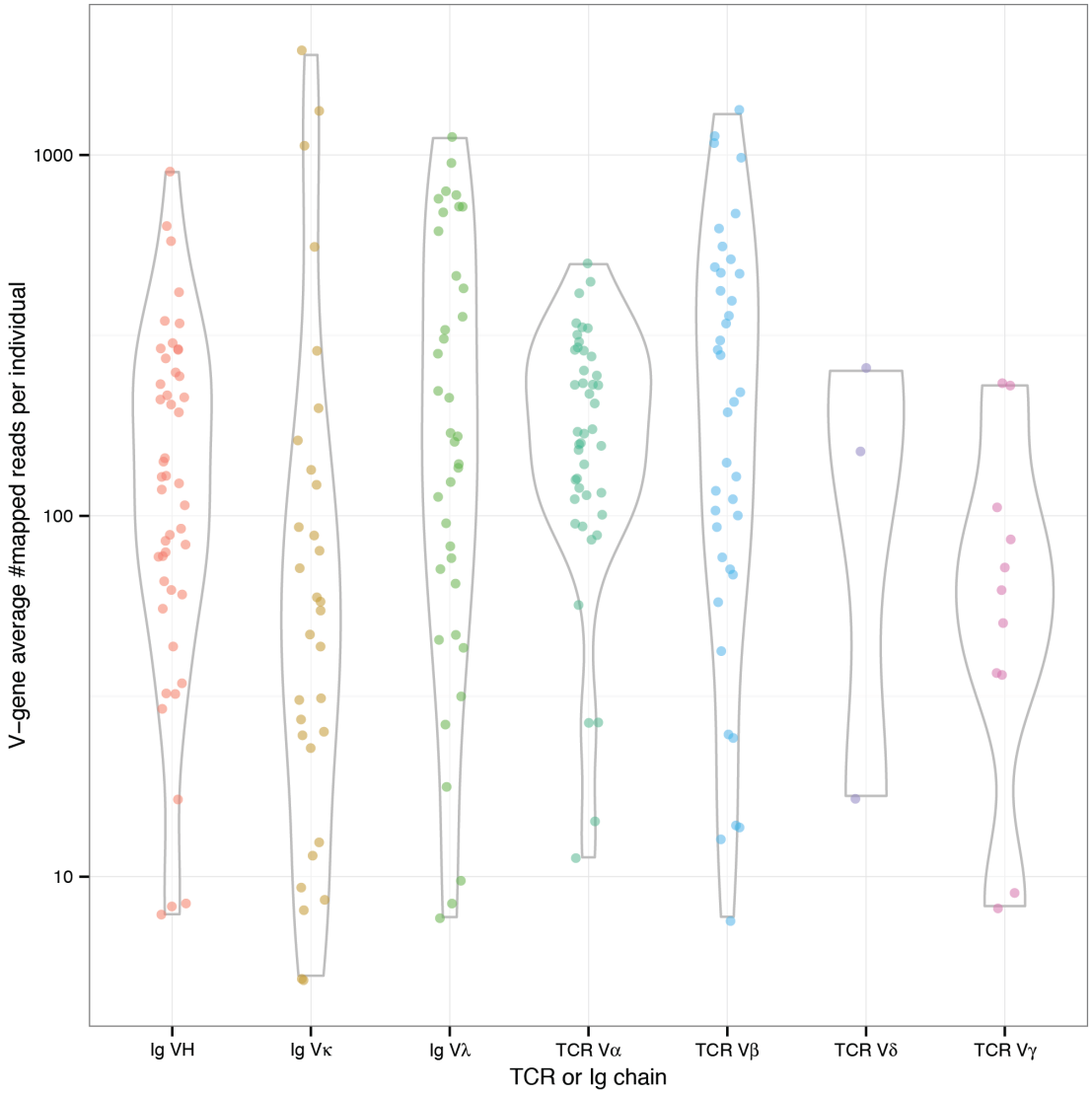
Supplementary Table 8. Expression variation of TCR Vα and Vβ-genes explained by indepedent associations with classical MHC genes 4-digit haplotypes. Expression variation of TCR Vα and Vβ-genes explained by a linear model of classical MHC 4-digit haplotypes. Haplotypes were selected using a conditional analysis (i.e. stepwise forward regression)

Supplementary Table 9. Probabilities that classical MHC gene amino acid positions influences expression of any TCR Vα-gene. Posterior probabilities that variable MHC amino acid positions influence the expression of any TCR Vα-gene, as estimated by the Bayesian model.

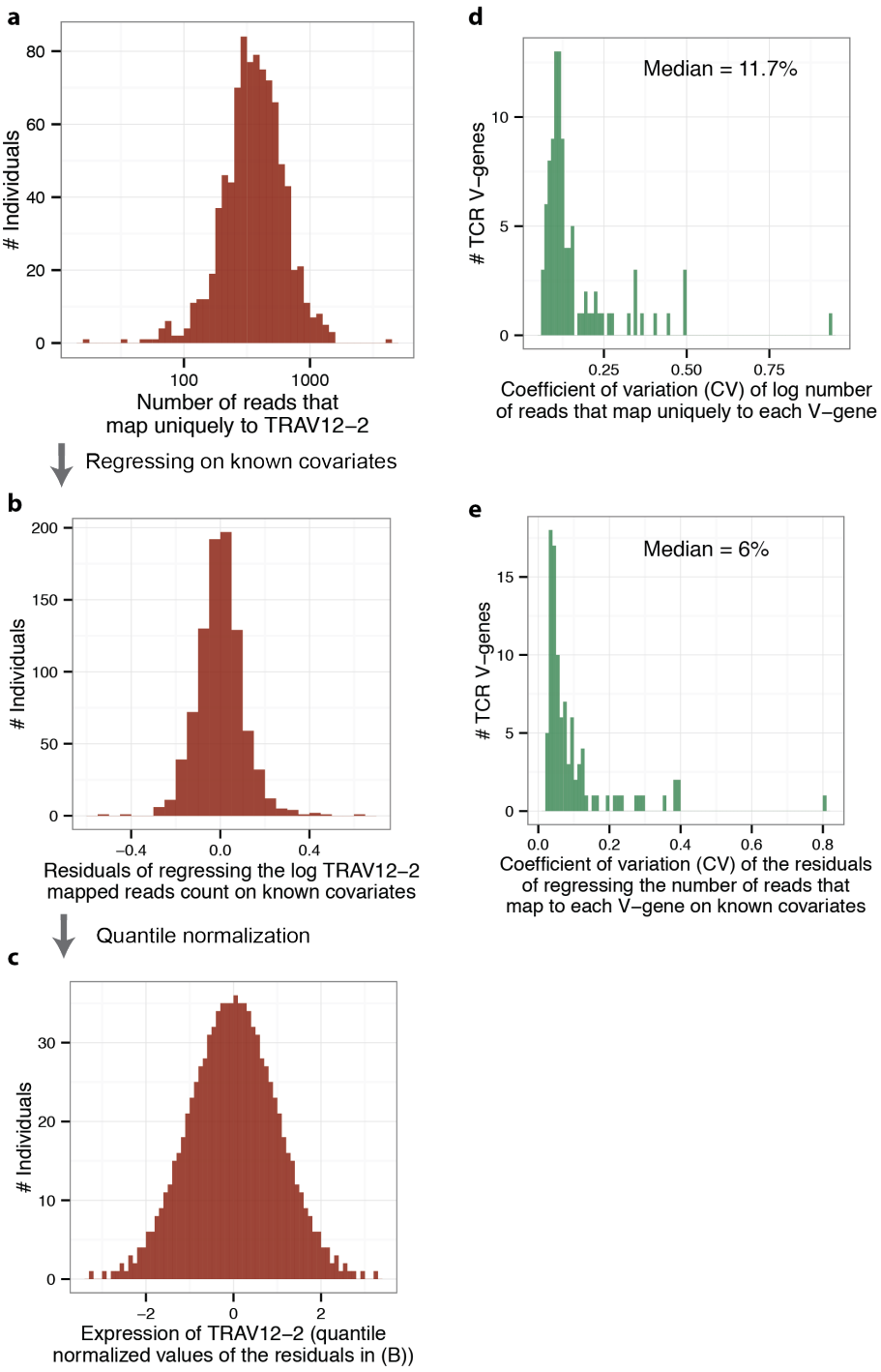
Supplementary Table 10. Probabilities that classical MHC gene amino acid positions influences expression of any TCR Vβ-gene. Posterior probabilities that variable MHC amino acid positions influence the expression of any TCR Vβ-gene, as estimated by the Bayesian model.

Supplementary Table 11. A list of PDB accession codes, MHC alleles and TCR Vα-gene used in the analysis of TCR-pMHC complexes. Data were downloaded on February 23, 2015 from RCSB PDB[2].

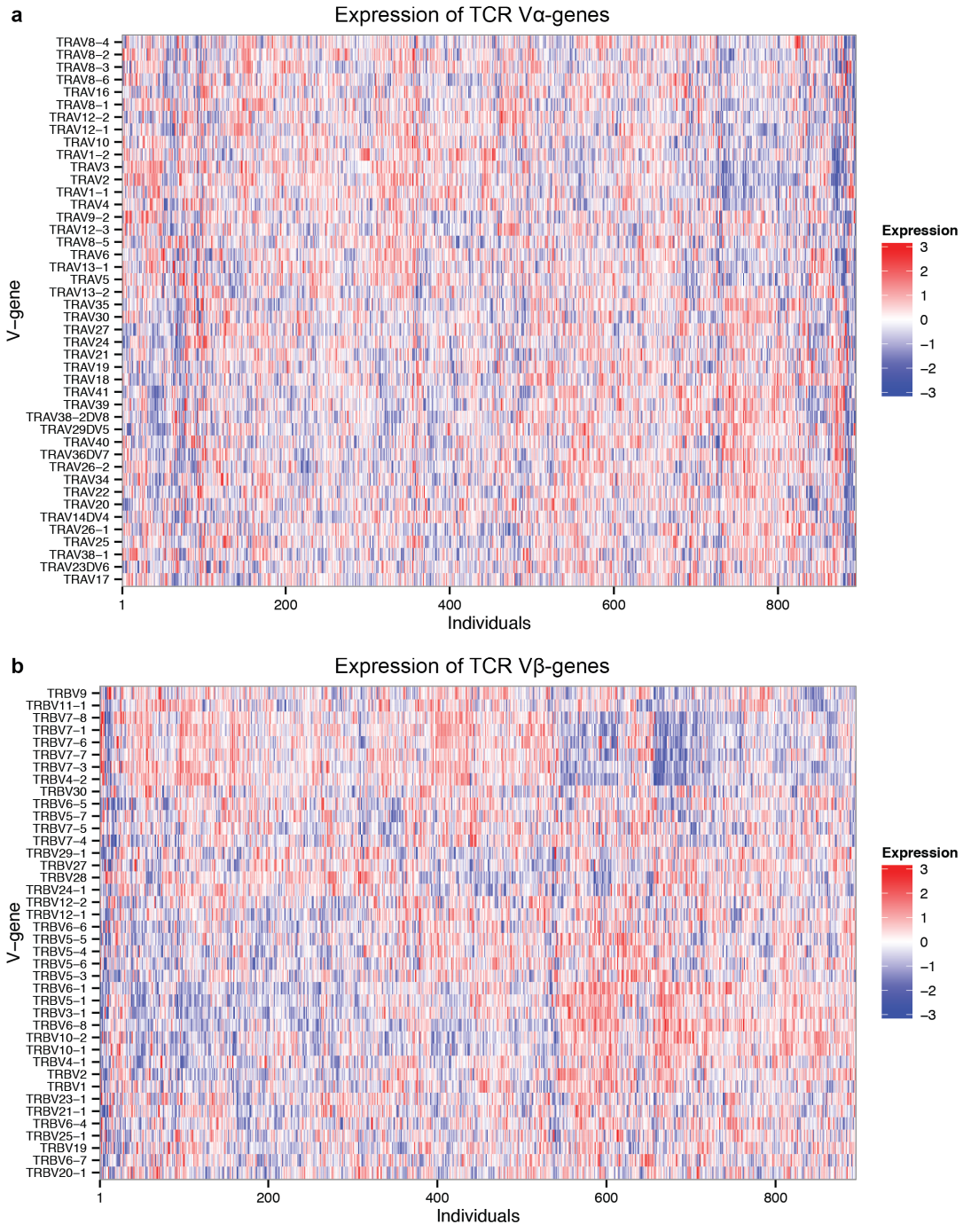
Supplementary figures



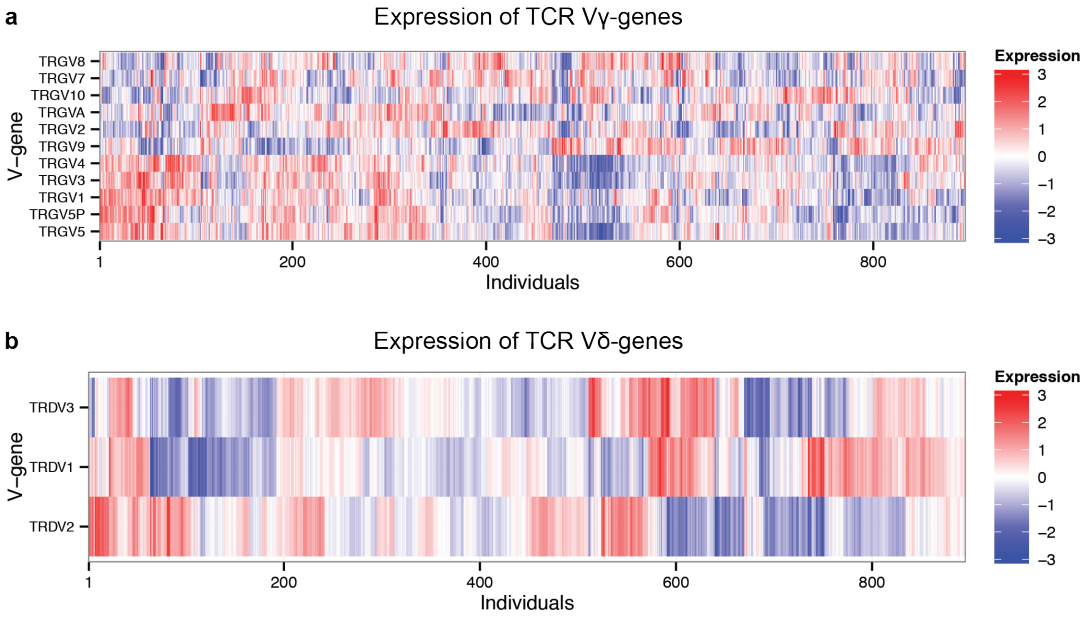
Supplementary Figure1. Read counts per V-gene per individual. Each dot represents the mean over individuals of the number of unique reads that map to a V-gene. The violin illustrates the distribution of these values for the V-genes of each TCR or Ig chain.



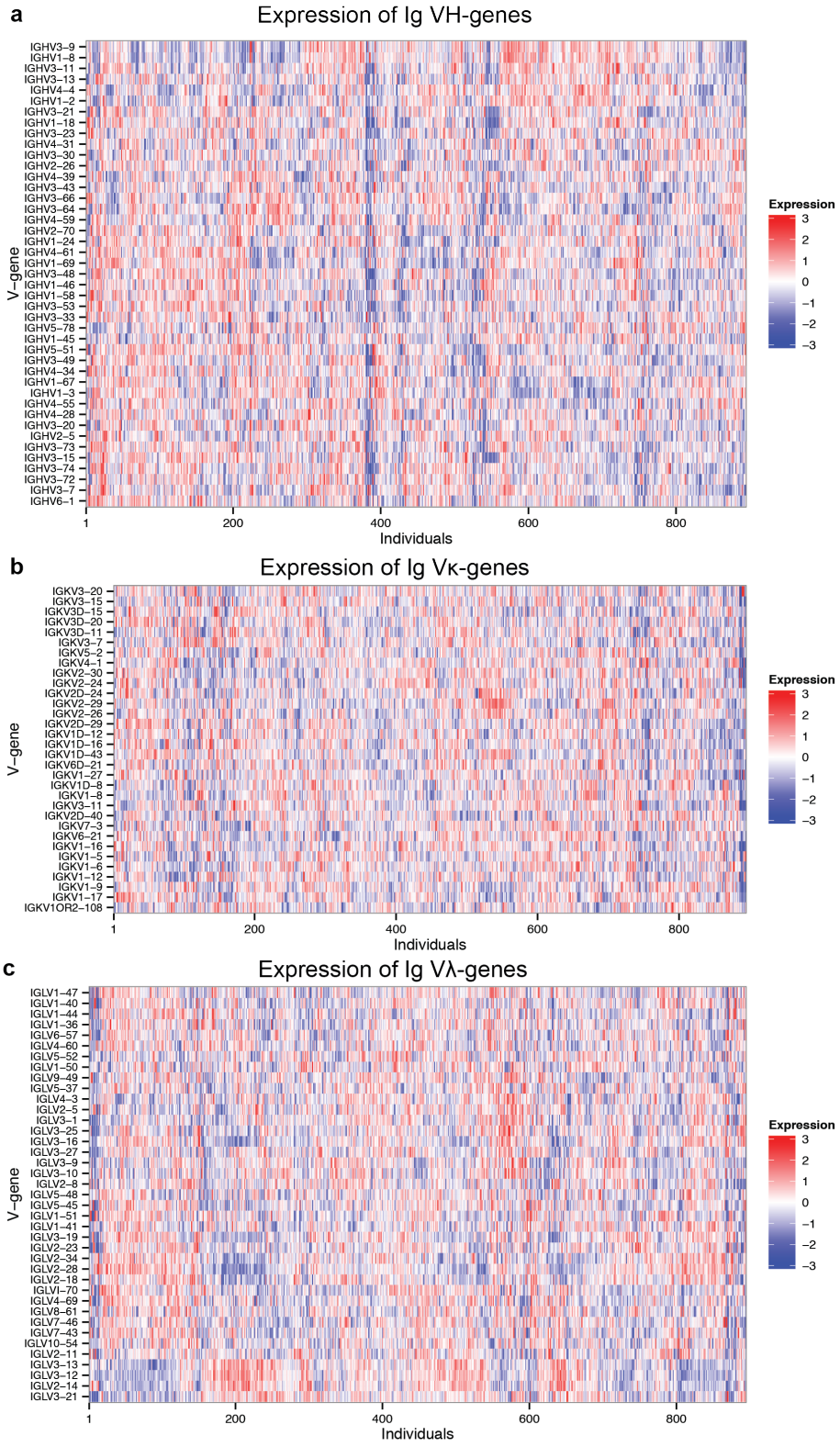
Supplementary Figure 2. Expression variation of TCR V-genes and an example of data normalization. (a-c) An example for expression measurement and data normalization of a TCR V-gene (TRAV12-2). **(a)** The distribution of uniquely mapped read counts over individuals. These counts were regressed on known covariates (see **Supplementary Table 1**) **(b)** and the residuals were quantile-normalized to a normal distribution **(c). (d)** Coefficient of variation (CV) of raw expression measurement ($\log_{10}$ read counts; median = 11.7%) and the residuals of the regression on known covariates (median = 6%) **(e)** of TCR V-genes.



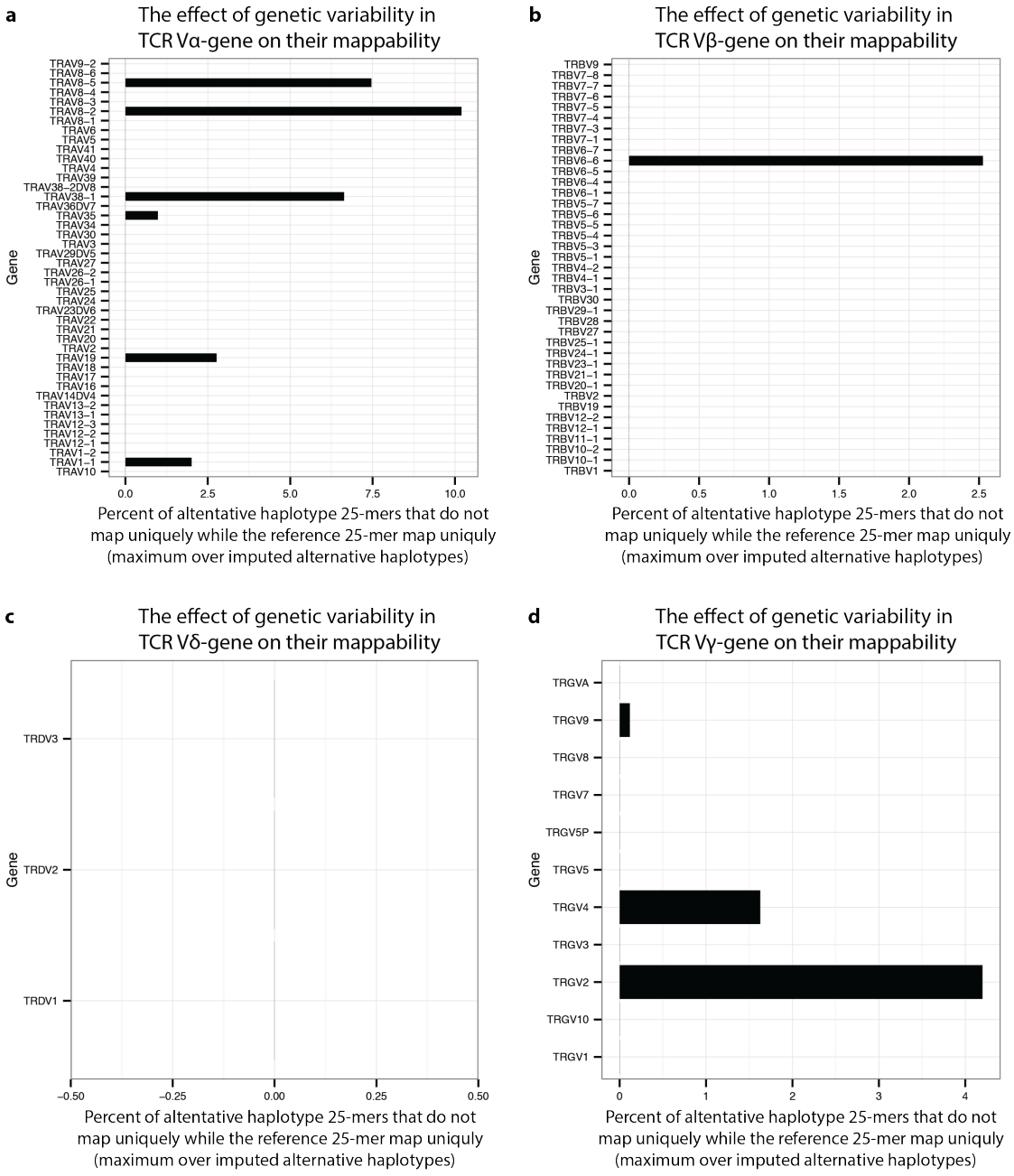
Supplementary Figure 3. Expression of TCR Vα and Vβ-genes. Normalized expression levels of TCR Vα-genes **(a)** and TCR Vβ-genes **(b)**. Rows and columns were clustered using hierarchical clustering.



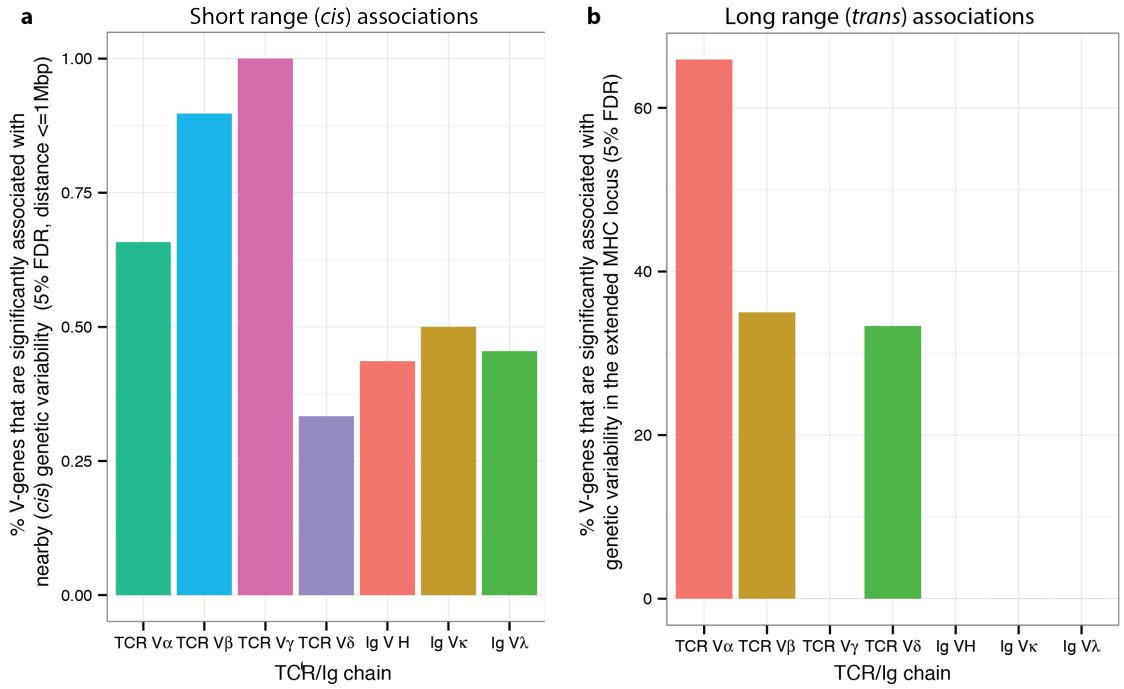
Supplementary Figure 4. Expression of TCR Vγ and Vδ-genes. Normalized expression levels of TCR Vγ-genes **(a)** and TCR Vδ-genes **(b)**. Rows and columns were clustered using hierarchical clustering.



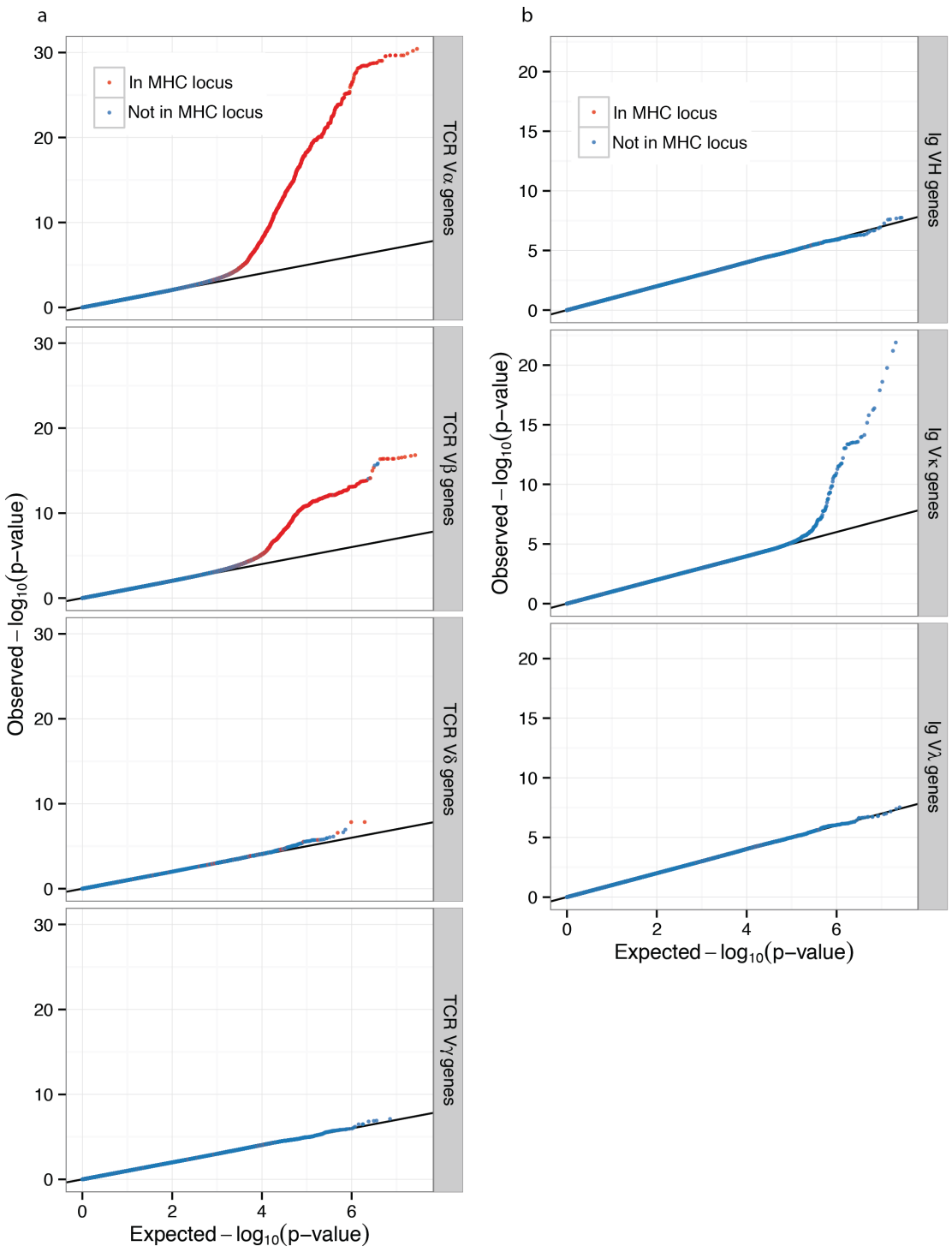
Supplementary Figure 5. Expression of Ig V-genes. Normalized expression levels of Ig VH-genes **(a)**, Vκ-genes **(b)** and Vλ-genes **(c).** Rows and columns were clustered using hierarchical clustering.



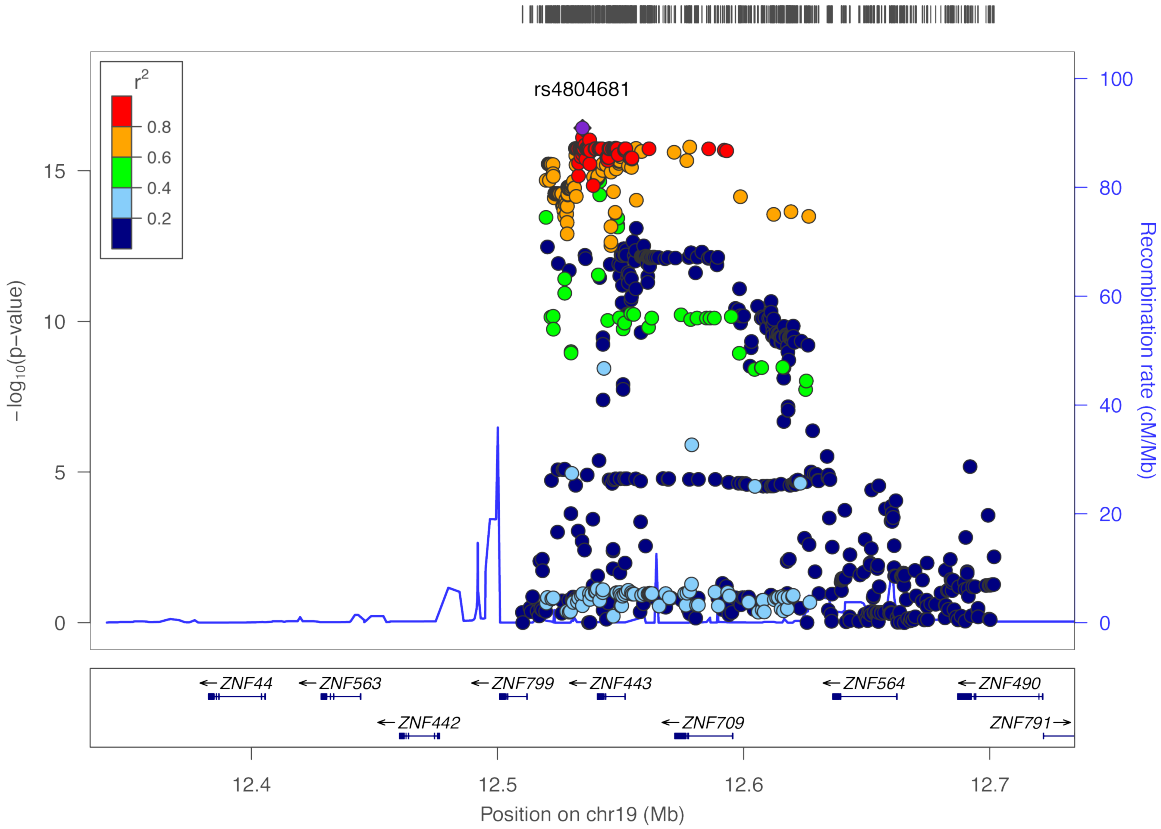
Supplementary Figure 6. Genetic variability affects the mappability of few TCR V-genes. For each TCR V-gene imputed alternative haplotype, the percent of 25-mers that does not map uniquely to its reference haplotype while the equivalent 25-mer extracted from the reference haplotype does map uniquely was computed. Bars show maximum percentage of such 25-mers over all imputed alternative haplotypes for TCR Vα-genes **(a)**, Vβ-genes **(b**) Vδ-genes **(c)** and Vγ-genes. Alternative haplotypes were imputed using SHAPEIT[3] (for pre-phasing) and IMPUTE2[4]; The 1,000 Genomes Phase 1[5] panel was used for imputation with a EUR MAF > 0.01 filter and genotype likelihood > 0.9. Imputed SNPs with MAF < 2.5% were filtered out. 25b is half of the RNA sequencing read length used in this study, when 49-mers are considered only 4 genes show mappability differences. Genes with 25-mers mappability differences were excluded from short-range (*cis*) association analysis.



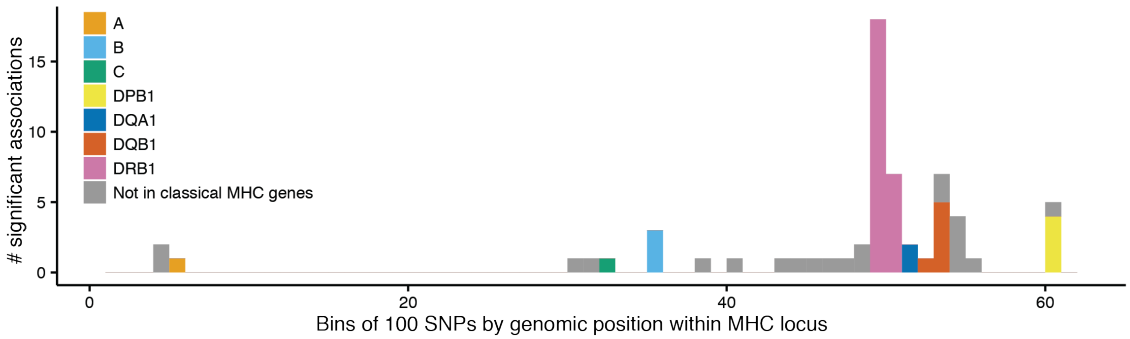
Supplementary Figure 7. Both TCR and Ig genes are associated with *cis* genetic variation, but only TCR V-genes are associated with genetic variability in the MHC locus. Percent of TCR and Ig V-genes that are significantly associated with genetic variation in *cis* **(a)** and in *trans* with variation in the extended MHC locus[6] **(b)** at 5% FDR. Significance of the strongest association for each gene was estimated using 10,000 random permutations of the gene expression values. Genes for which the mappability of the alternative haplotype differ from the reference were excluded from the short range analysis.



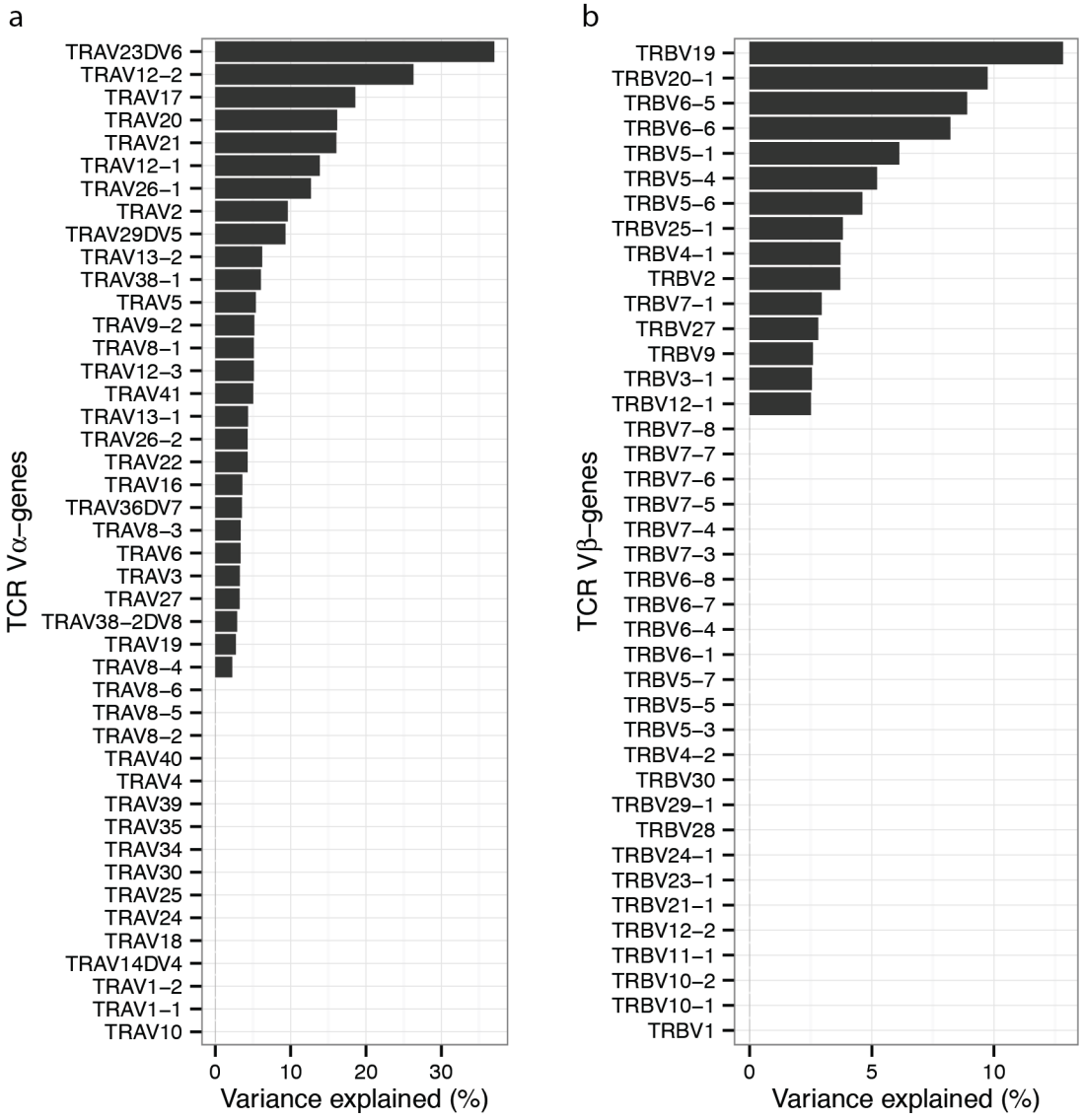
Supplementary Figure 8. QQ plots of genome-wide *trans* association of genetic variation with TCR and Ig V-gene expression. Each panel shows p-values of long-range (>1Mb) associations of SNPs with expression of V-genes from a single TCR **(a)** or Ig **(b)** chain. Red color indicates associations with genetic variation in the MHC locus. All lambda genomic inflation factors are between 0.999 and 1.008.



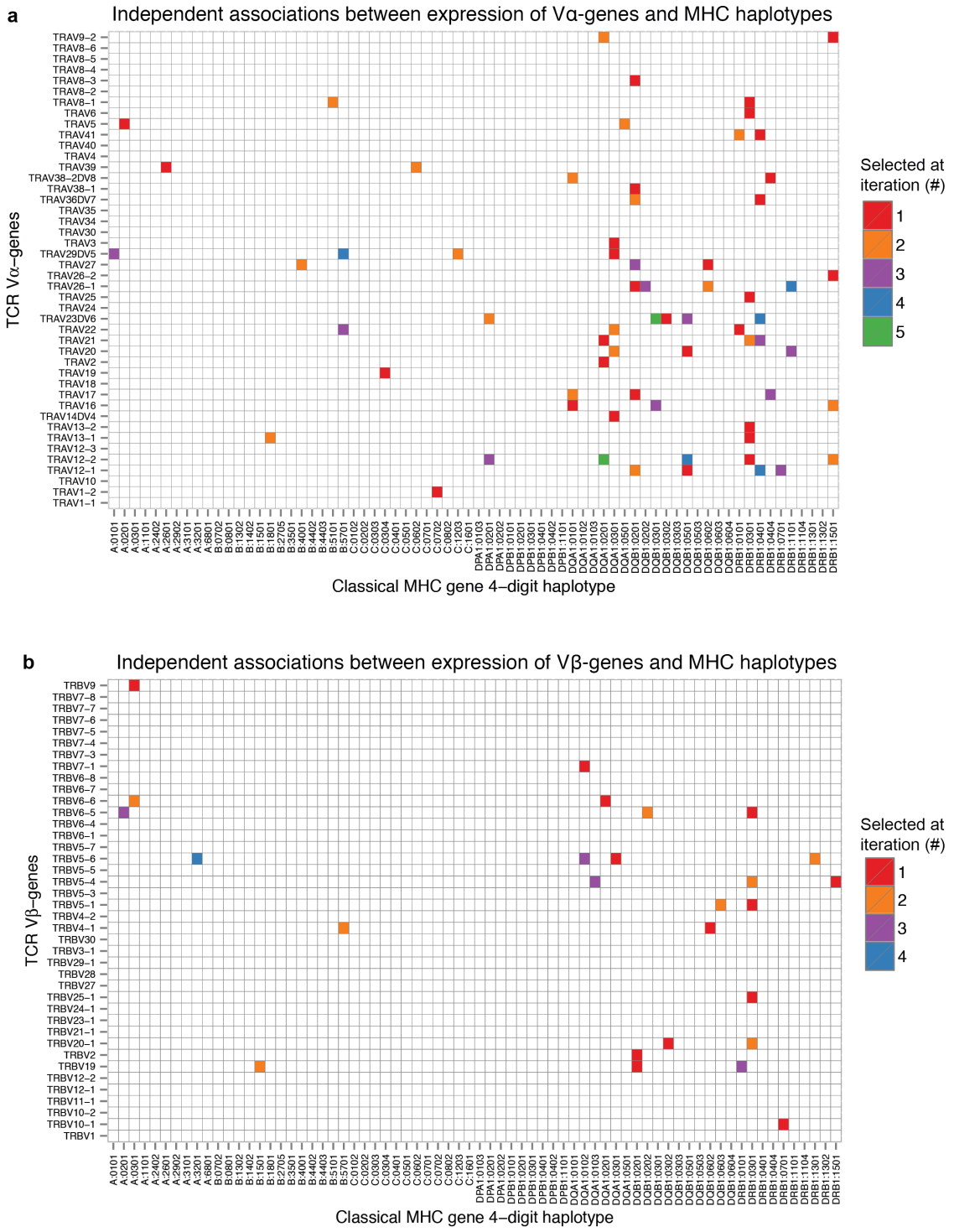
Supplementary Figure 9. A locus in chromosome 19 is associated with TRBV24-1 expression. The plot was generated using Locus Zoom[7]. The most significant SNP is near ZNF443; a gene which is differentially regulated in early Th1 and Th2 cell differentiation[8].



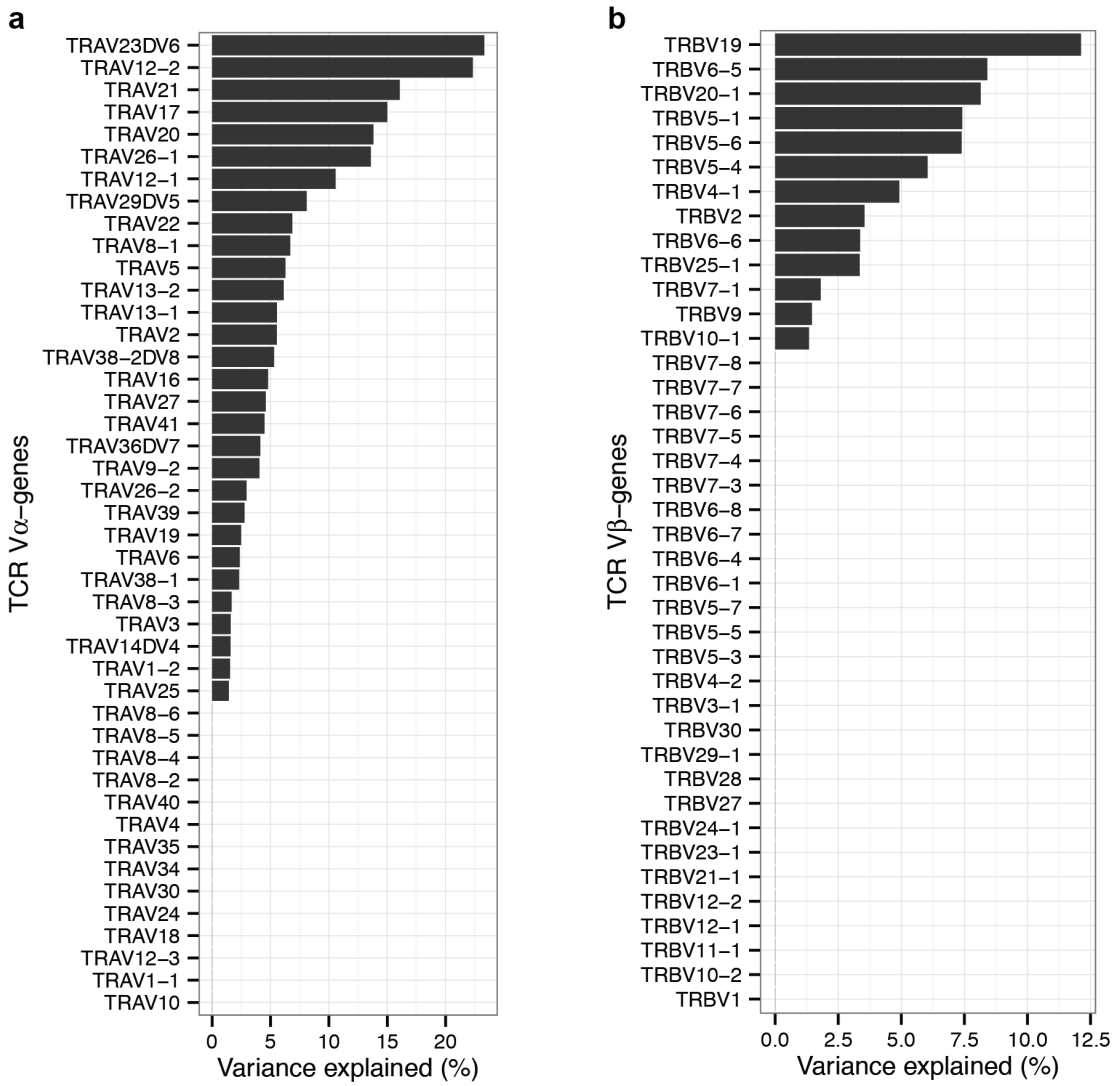
Supplementary Figure 10. Expression of TCR Vα and Vβ genes is associated with amino acid variation in MHC proteins. Independent associations between expression of Vα or Vβ-genes and nucleotide or amino acid variation in the MHC locus were selected using conditional analysis ($p<0.05$ with Bonferroni correction). SNPs are binned according to their genomic positions (100 SNPs per bin).



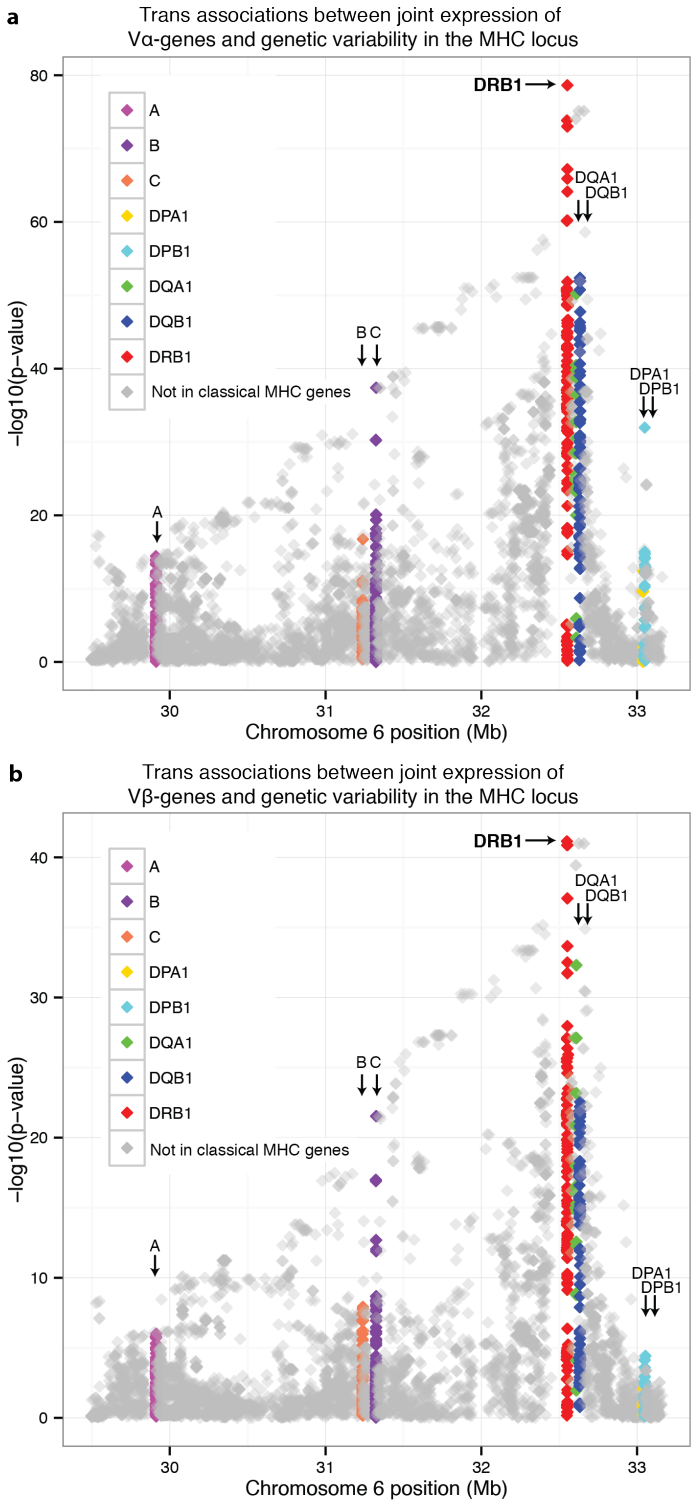
Supplementary Figure 11. Expression variation of TCR Vα and Vβ-genes explained by a linear model of independent association with genetic variation in the MHC locus. Independent associations between Vα **(a)** or Vβ **(b)** expression and nucleotide or amino acid variation in the MHC locus were selected using conditional analysis (i.e. stepwise forward regression procedure; stopping criteria: $p<0.05$ with Bonferroni correction). Genetic variation in the MHC locus explains more expression variation of Vα-genes (which contact MHC II β chain) than of Vβ-genes (which contact MHC II α chain). Part of this may be because MHC II β genes are substantially more variable than MHC II α genes.



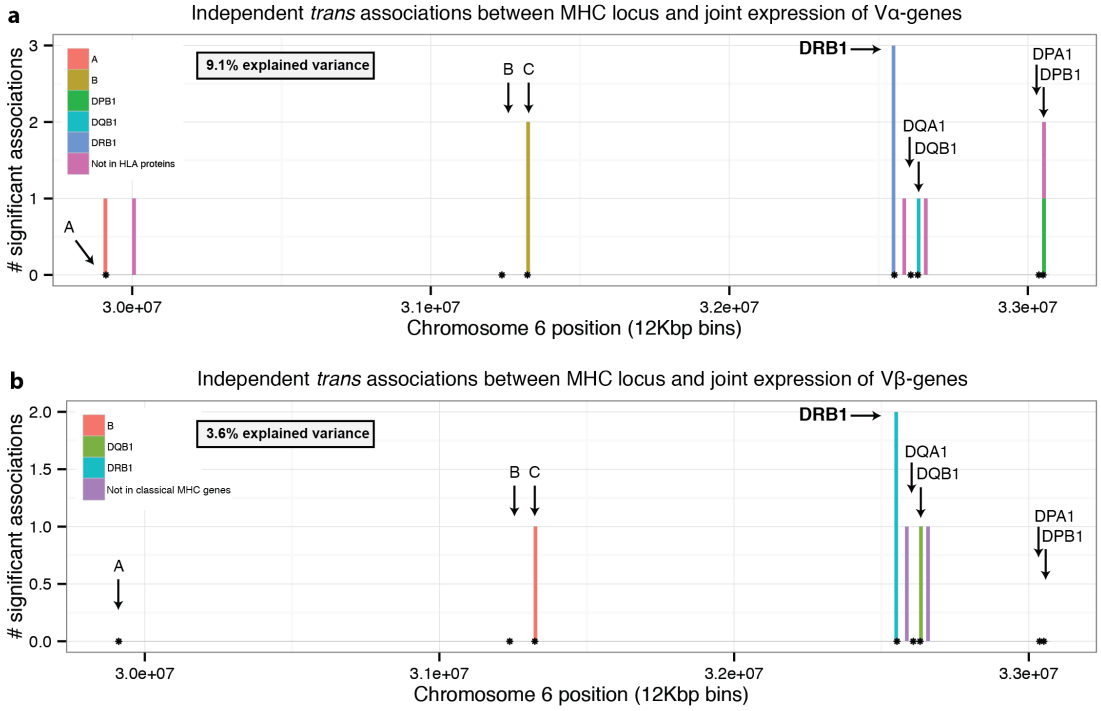
Supplementary Figure 12. Independent association between expression of TCR Vα and Vβ-genes and classical MHC genes 4-digit haplotypes. Independent associations between expression of Vα or Vβ genes and classical MHC 4-digit haplotypes were selected using conditional analysis ($p<0.05$ with Bonferroni correction). Fill color corresponds to the conditional analysis iteration in which the haplotype was selected.



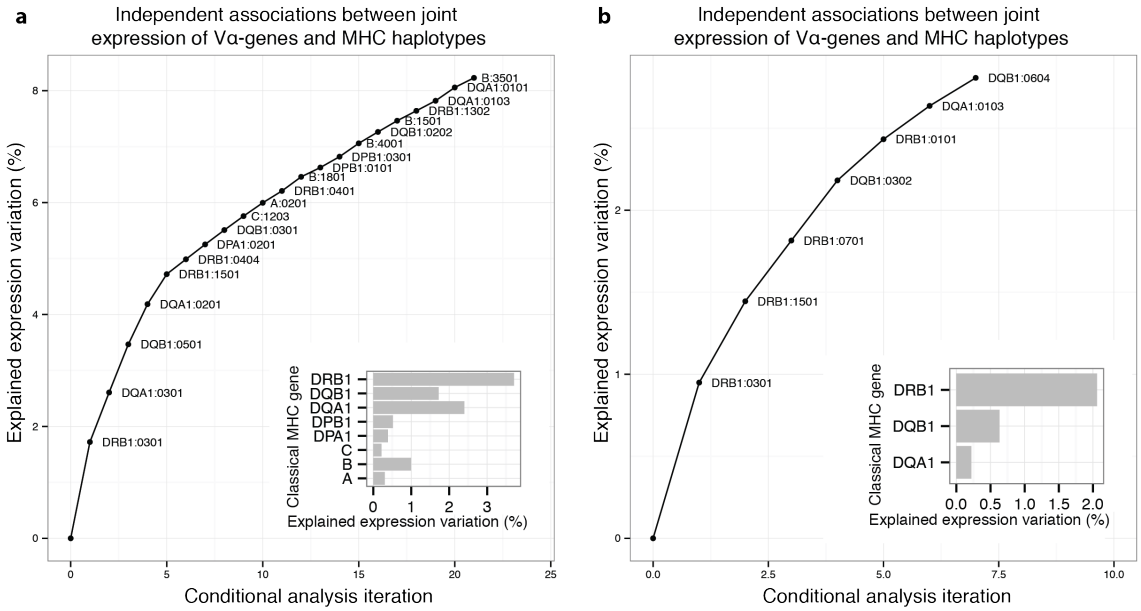
Supplementary Figure 13. Expression variation of TCR Vα and Vβ-genes explained by a linear model of independent association with classical MHC genes 4-digit haplotypes. Independent associations between expression of Vα or Vβ genes and classical MHC 4-digit haplotypes were selected using conditional analysis ($p<0.05$ with Bonferroni correction).



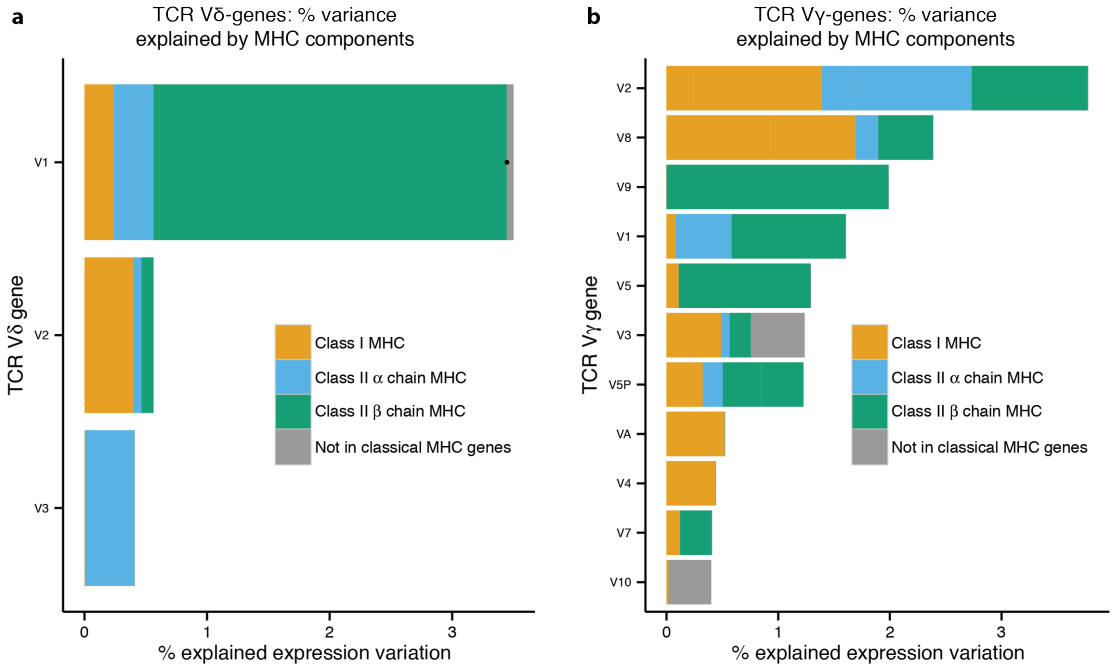
Supplementary Figure 14. Joint expression of TCR Vα and Vβ genes is significantly associated with variation in MHC proteins and particularly in DRB1. A Manhattan plot showing the significance of each binary marker of nucleotide or amino acid variation (imputed by SNP2HLA) association with the joint expression of TCR Vα **(a)** and Vβ **(b)** genes. P-values were computed using a multivariate multiple response regression. The plot shows that the strongest associations with expression of TCR V-genes are with genetic variability in HLA-DRB1.



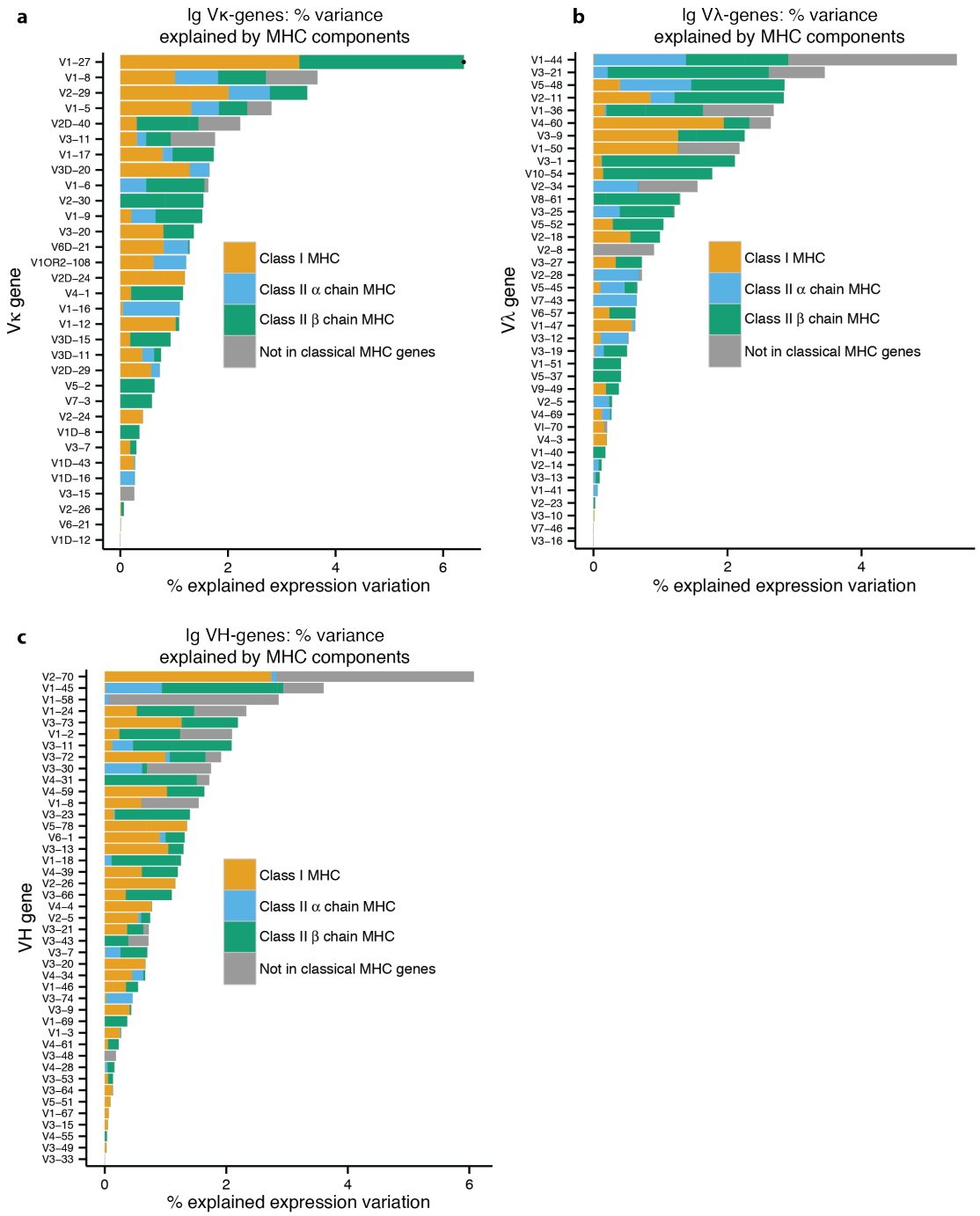
Supplementary Figure 15. Joint expression of TCR Vα and Vβ genes is associated with amino acid variation in classical MHC genes. (a) Independent associations between the joint expression of Vα and Vβ **(b)** genes expression and nucleotide or amino acid variation in the MHC locus ($p<0.05$ with Bonferroni correction). SNPs are binned according to their genomic position (12Kb bins); the points mark the center of the classical MHC genes. The explained variance is the average over the V-genes. Genetic variants, that were selected using a stepwise forward regression as independently associated with joint expression of TCR Vα and Vβ genes, are within or near classical MHC genes and explain 9.1% and 3.6% of Vα and Vβ genes joint expression variation respectively.



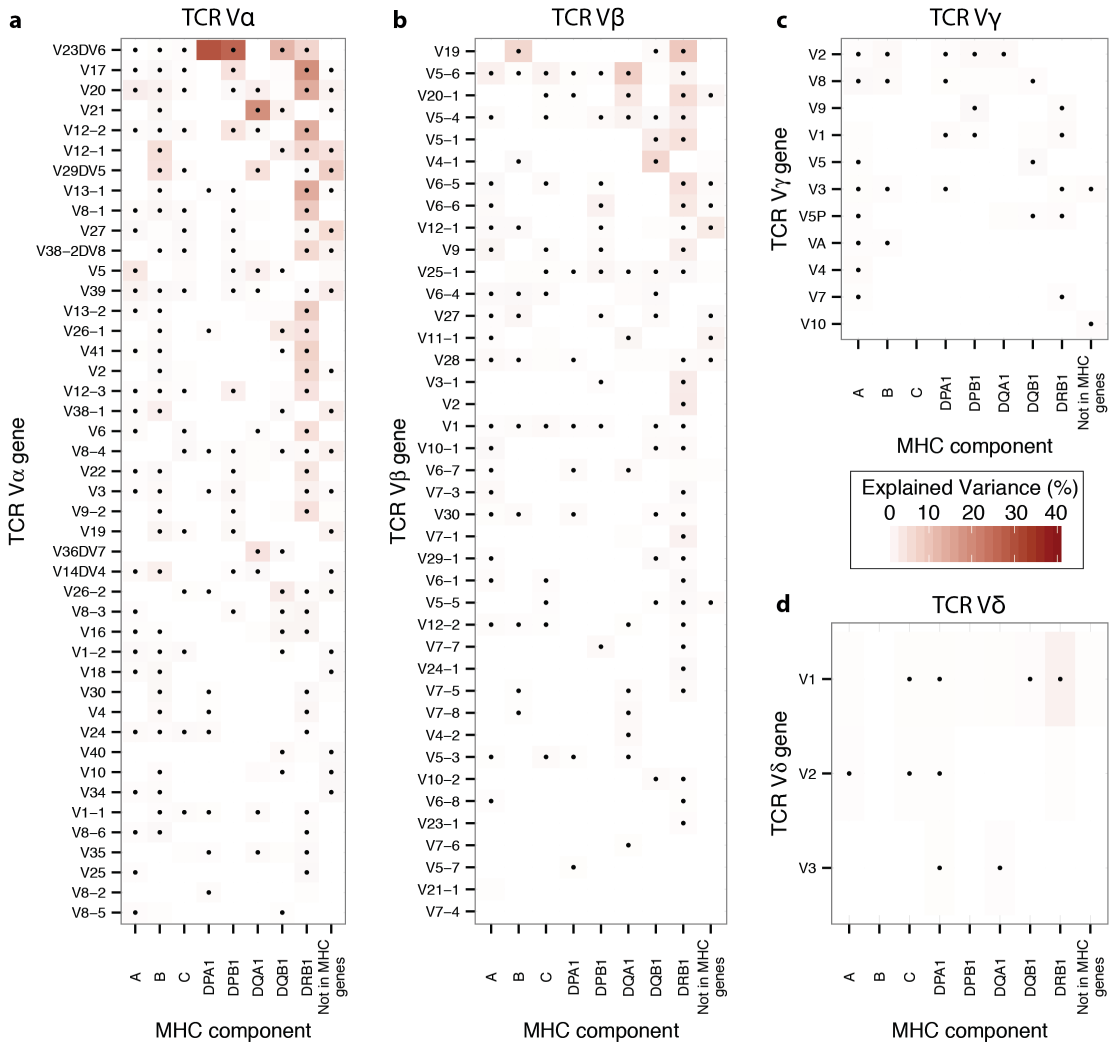
Supplementary Figure 16. Joint expression of TCR Vα and Vβ genes is independently associated with several MHC haplotypes. Line plots are showing TCR Vα-genes **(a)** and Vβ-genes **(b)** joint expression variation explained by a series of linear models selected using stepwise forward regression on MHC 4-digits haplotypes. Annotations show the haplotypes added to the model in each iteration of the forward regression (i.e. condtional analysis). Bar plots show the percent of joint expression variation of Vα-genes **(a)** and Vβ-genes **(b)** explained by models that contain haplotypes of a single MHC gene slecected by the conditional analysis. These plots show that *HLA-DRB1* haplotypes explain the largest percent of joint expression variation of TCR V-genes out of all classical MHC genes.



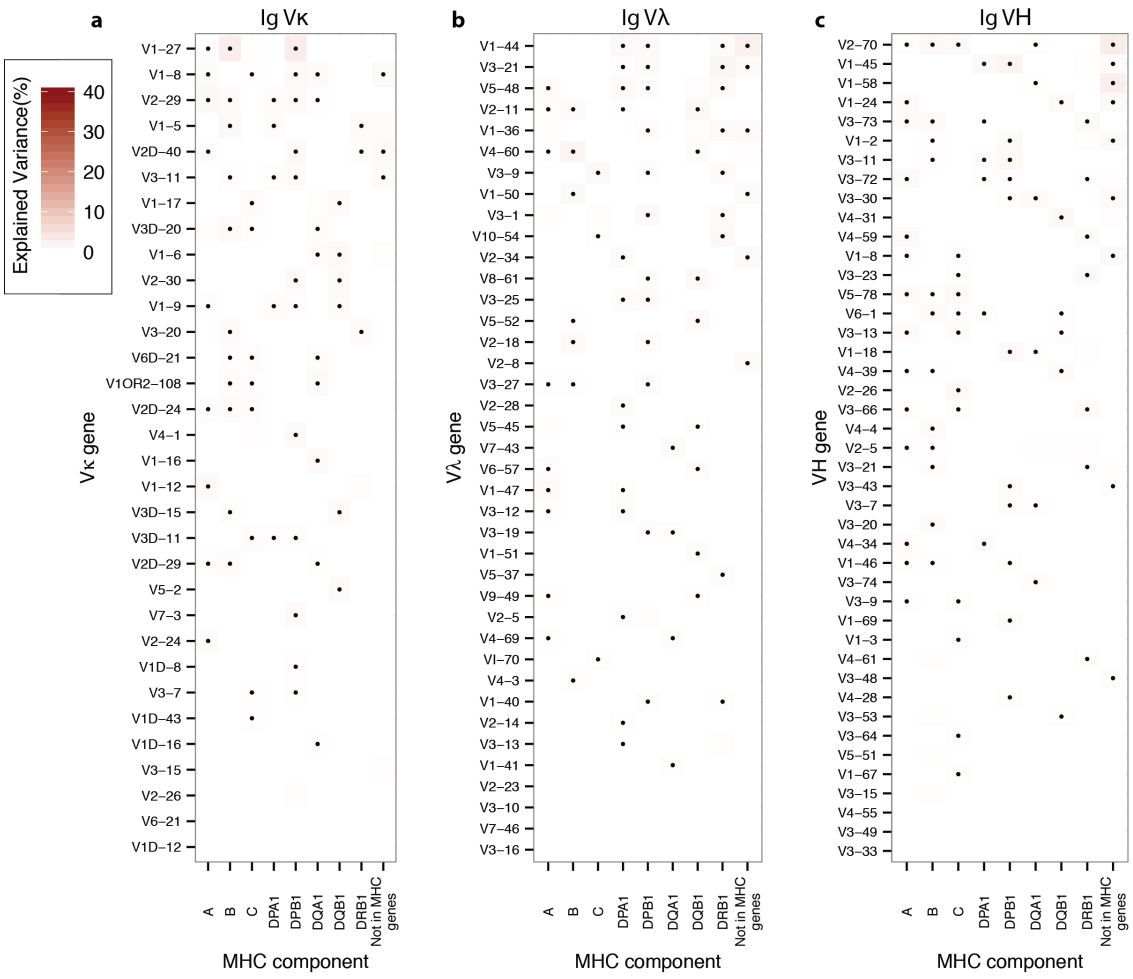
Supplementary Figure 17. Variance of TCR Vγ (a) and Vδ (b) expression explained by MHC variation. Values were computed using GCTA[10]. Dots indicate that the amount of variation explained by the MHC proteins was significant at 5% FDR.



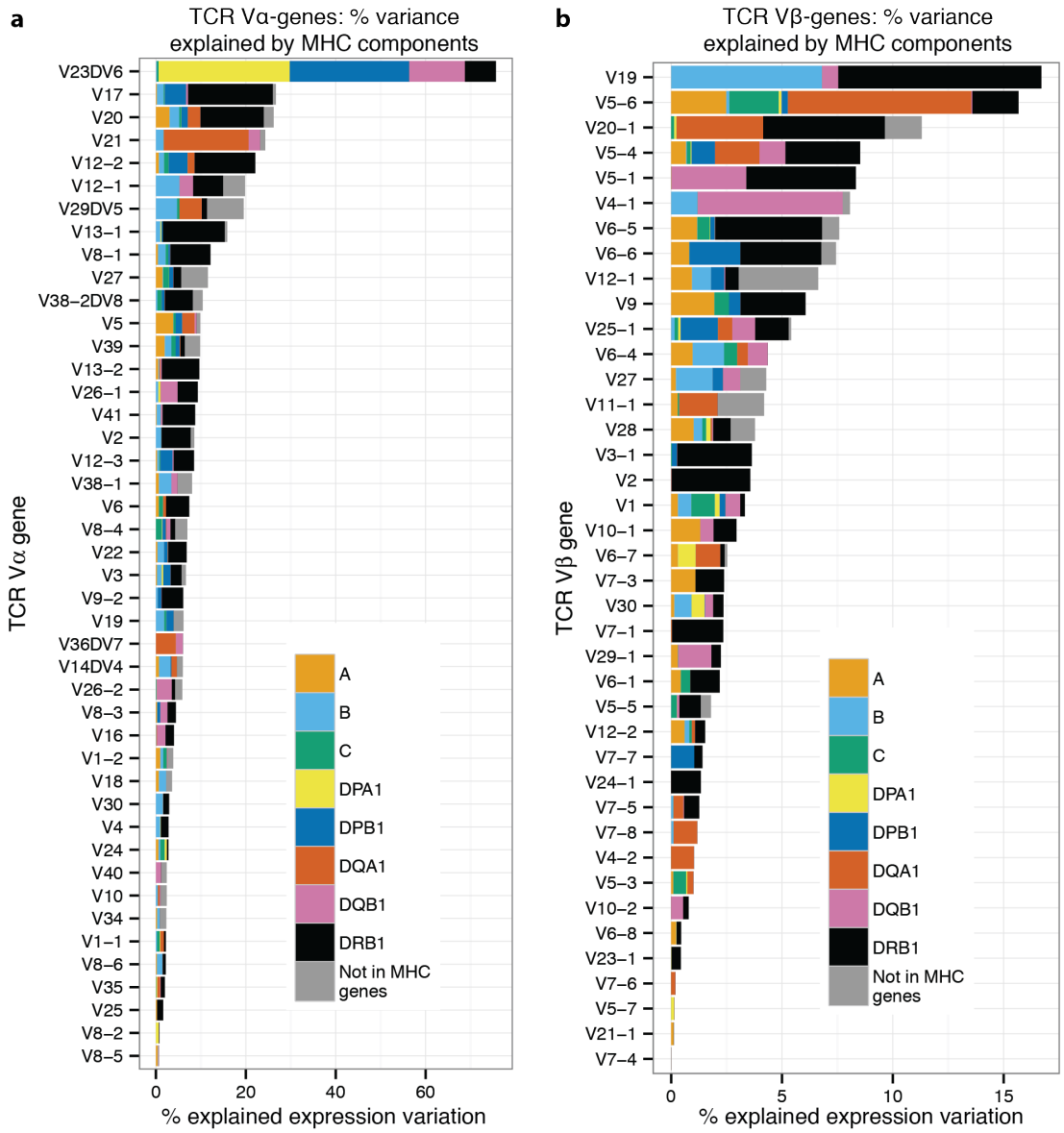
Supplementary Figure 18. Variance of Ig gene expression is not well explained by MHC variation. Variance of Ig Vκ **(a),** Vλ **(b)** and VH **(c)** expression explained that is explained by MHC variation. Values were computed using GCTA[10]. Dots indicate that the amount of variation explained by the MHC proteins was significant at 5% FDR.



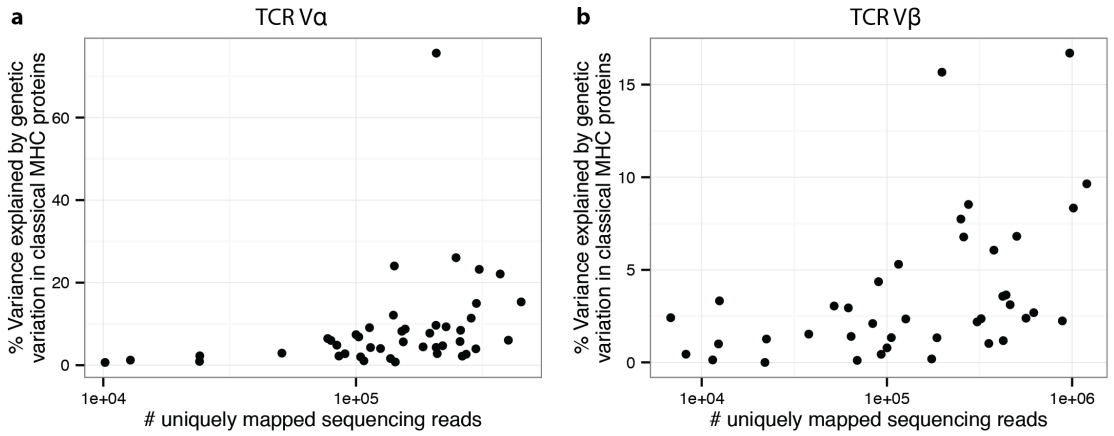
Supplementary Figure 19. Variance in TCR V-gene expression explained by variability in classical MHC proteins and genetic variability in the MHC locus outside the classical MHC genes. The percent of variation of Vα **(a)**, Vβ **(b),** Vγ **(c)** and Vδ **(d)** expression explained by amino acid variation in classical MHC proteins and genetic variability in the MHC locus outside the classical MHC genes ("Not in MHC genes"). V-genes are sorted by the total fraction of explained expression variation (top to bottom). Values were computed using GCTA[10]. Dots mark values that are significantly larger than zero (p-value<0.05 with Bonferroni correction).



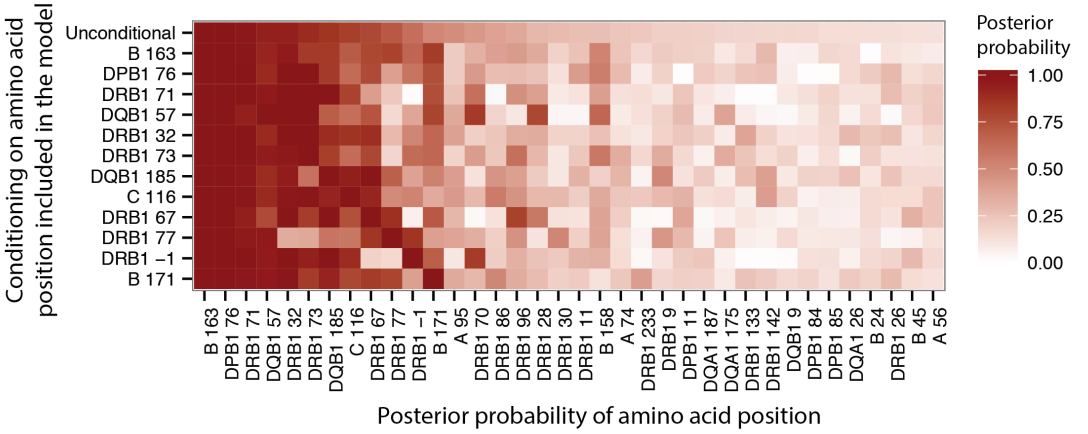
Supplementary Figure 20. Variance in Ig V-gene expression explained by variability in classical MHC proteins and genetic variability in the MHC locus outside the classical MHC genes. The percent of variation of Vκ **(a),** Vλ (**b)** and VH (**c)** expression explained by amino acid variation in classical MHC proteins and genetic variability in the MHC locus outside the classical MHC genes ("Not in MHC genes"). V-genes are sorted by the total fraction of explained expression variation (top to bottom). Values were computed using GCTA[10]. Dots mark values that are significantly larger than zero (p-value<0.05 with Bonferroni correction)



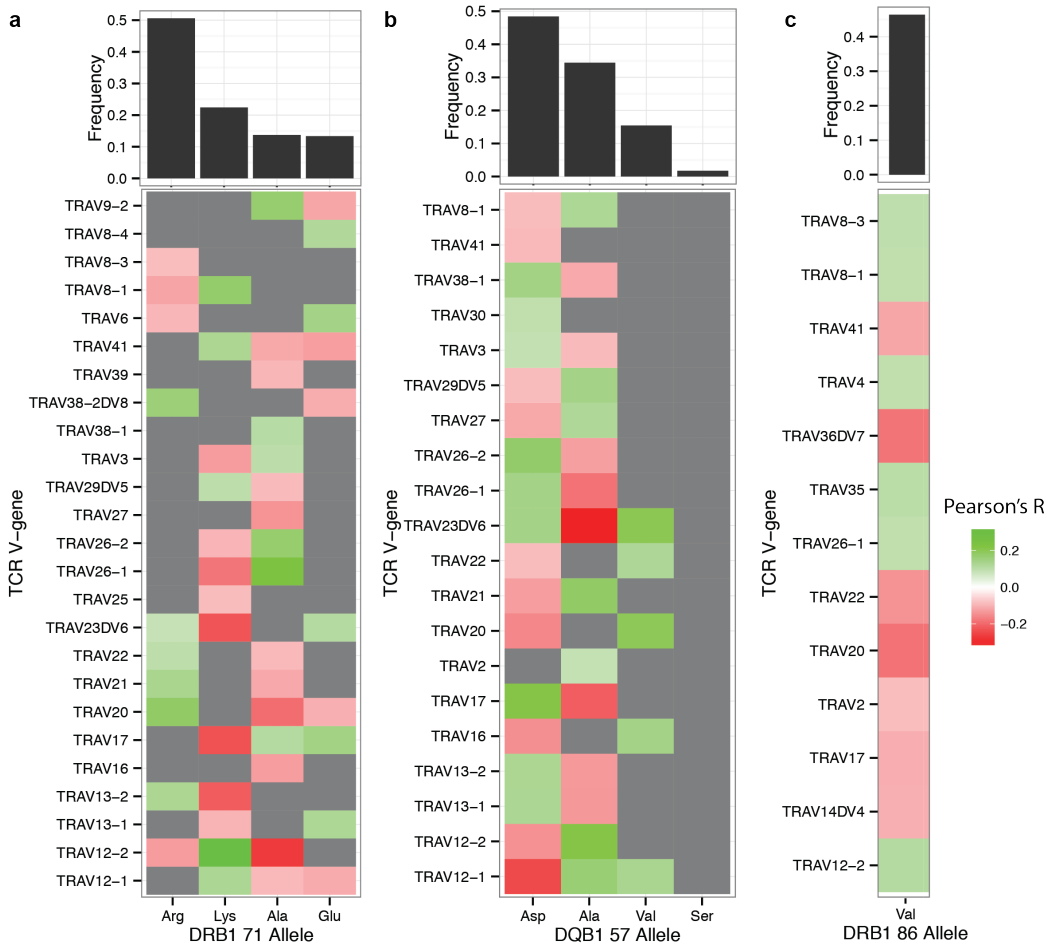
Supplementary Figure 21. Variance in TCR Vα (a) and Vβ (b) expression explained by variability in classical MHC proteins and genetic variability in the MHC locus outside the classical MHC genes. Values were computed using GCTA[10]. This figure is a detailed version of figure 3B–C, which shows the contribution of each classical MHC gene component.



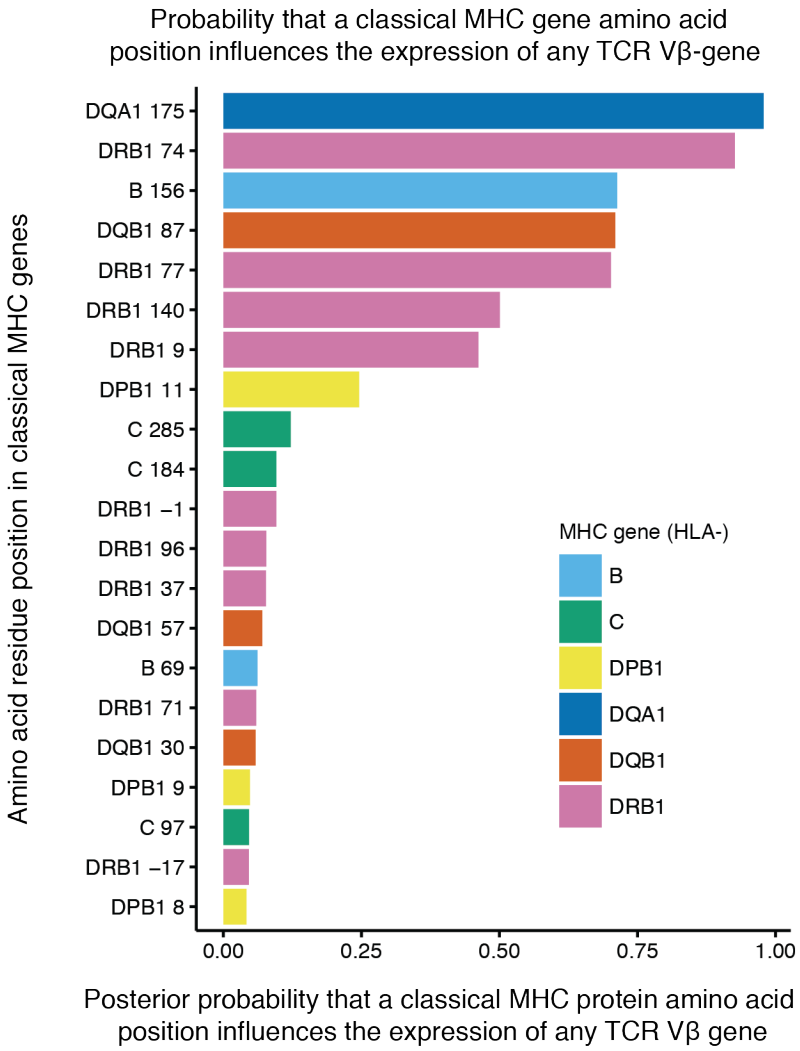
Supplementary Figure 22. Variance explained for V-genes correlates with read depth. The variation of TCR Vα **(a)** and Vβ **(b)** genes for which less than 50,000 reads are mapped is less explainable by genetic variation in the MHC proteins. This may be due to the larger noise of lowly expressed or measured genes. Therefore, GCTA[10] estimates of the genetic components may increase with better measurements due to reduced measurement noise.



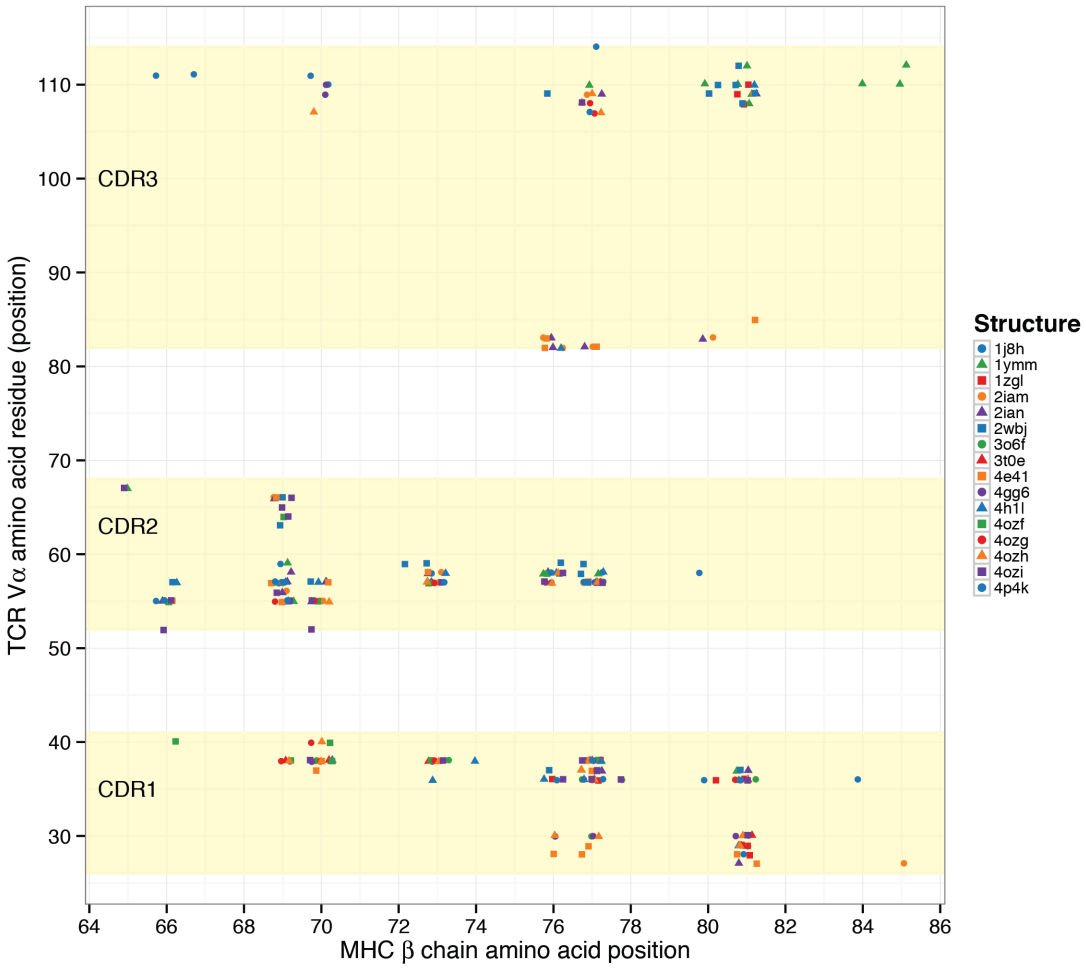
Supplementary Figure 23. Investigating which MHC amino acid positions can explain signals at other positions. For each of the top-ranked amino acid position, we tested the impact it has on the posterior for other positions if we force that position to be in the model. The rows show all positions with marginal posterior ≥50% and columns show all positions with marginal posterior ≥12.5%. Correlations between the genotypes in different positions may lead to uncertainties about which positions drive TCR interactions. In such cases, the posterior may be split between positions. Conditional on one position being causal, the probability of the other correlated positions decreases. An example for this phenomenon is that when position 71 or 67 in DRB1 is used to help explain the expression of TCR Vα genes, the probability of position –1 decreases dramatically, suggesting that support for the -1 position in the model may be due to LD with positions 67 or 71.



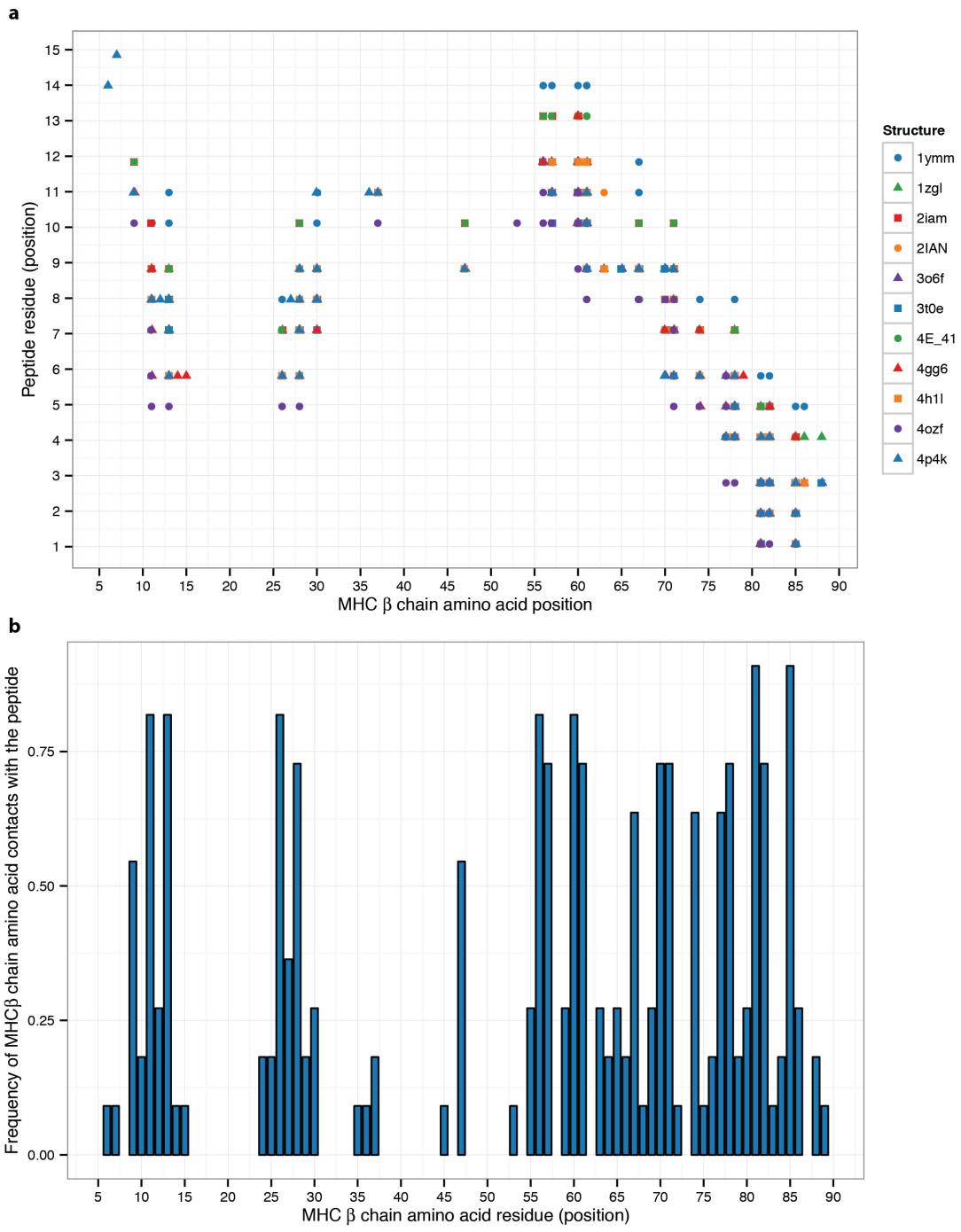
Supplementary Figure 24. MHC amino acid positions that are associated with autoimmune diseases are significantly associated with expression of TCR Vα-genes. Bottom panels show the correlation of each allele of MHC amino acid, which were shown to associate with autoimmune diseases[11–14], with expression of TCR Vα-genes (left - *HLA-DRB1*, center - *HLA-DQB* 57, right - *HLA-DRB1* 86). Only significant associations are shown (5% FDR) and only TCR Vα-genes with at least one significant association are shown. Top panels show alleles frequencies in the study cohort.



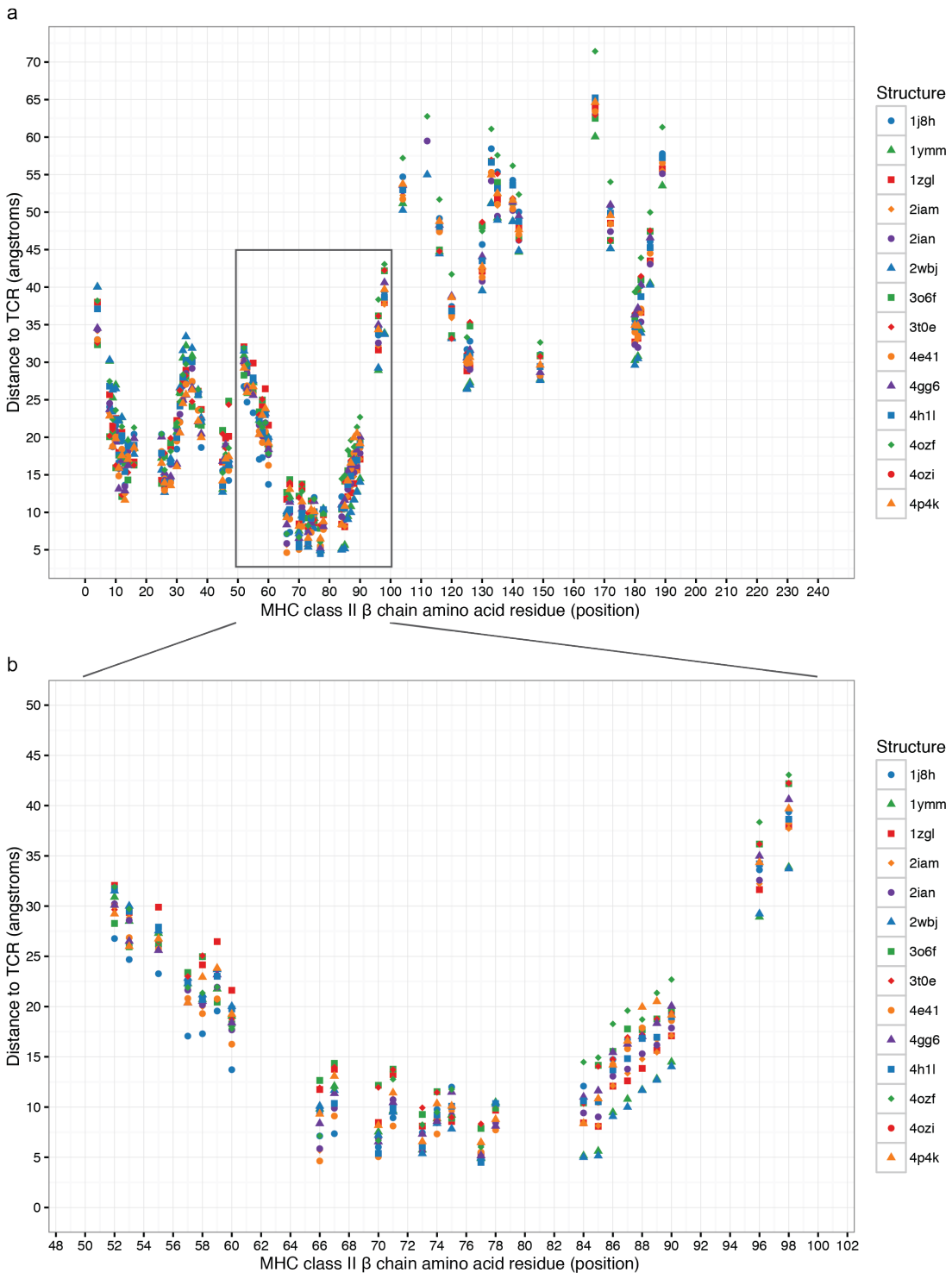
Supplementary Figure 25. Bayesian inference of MHC amino acid residues that are associated with TCR Vβ genes expression biases. Estimated posterior probability that an amino acid residue (y-axis) is influencing the expression of any TCR Vβ-gene.



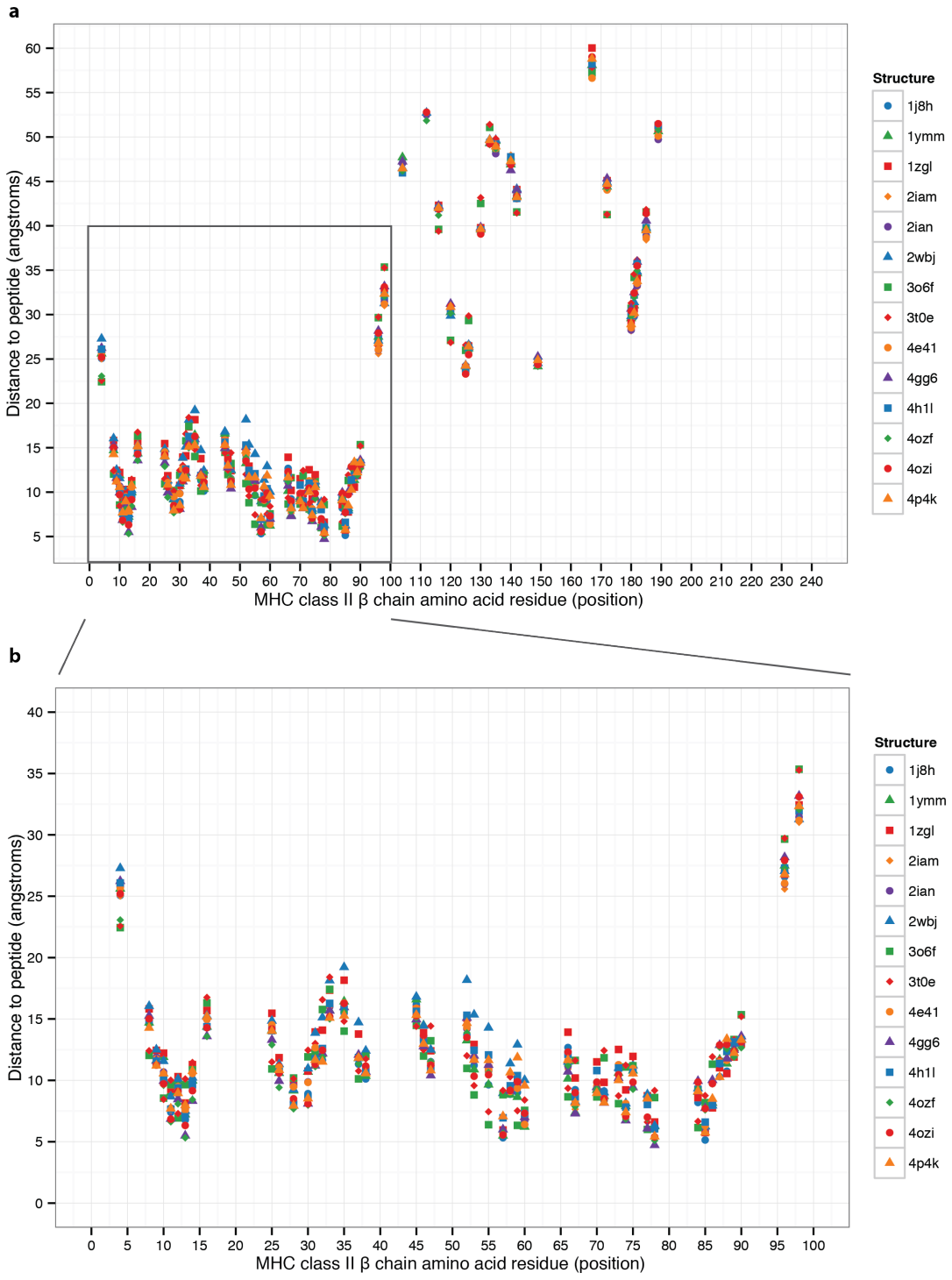
Supplementary Figure 26. Physical contacts between MHC class II β amino acid residues and TCR amino acid residues. Each point represents an intermolecular interaction between an MHC II β amino acid residue and a TCRα residue in a solved TCR-MHC complex. Each RCSB PDB[2] complex is represented by unique symbol type and color. The yellow background coloring indicates positions that are in the CDR 1, 2 & 3 loops in the TCR. TCR-MHC contacts were determined by IMGT.



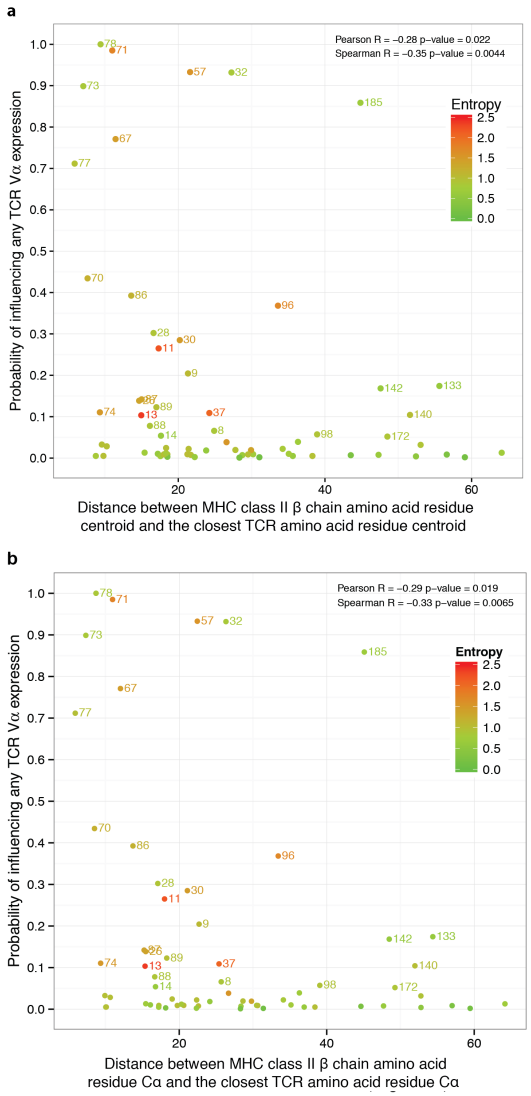
Supplementary Figure 27. Physical contacts between MHC class II β amino acid residues and the peptide amino acid residues. (a) Each point represents an intermolecular interaction between an MHC β-chain residue and a peptide residue in a solved TCR-pMHC complex. Dot type indicates the RCSB PDB ID[2] complex. **(b)** Frequency of peptide contact of MHC amino acid residue in TCR-peptide-MHC complexes presented in **a**.



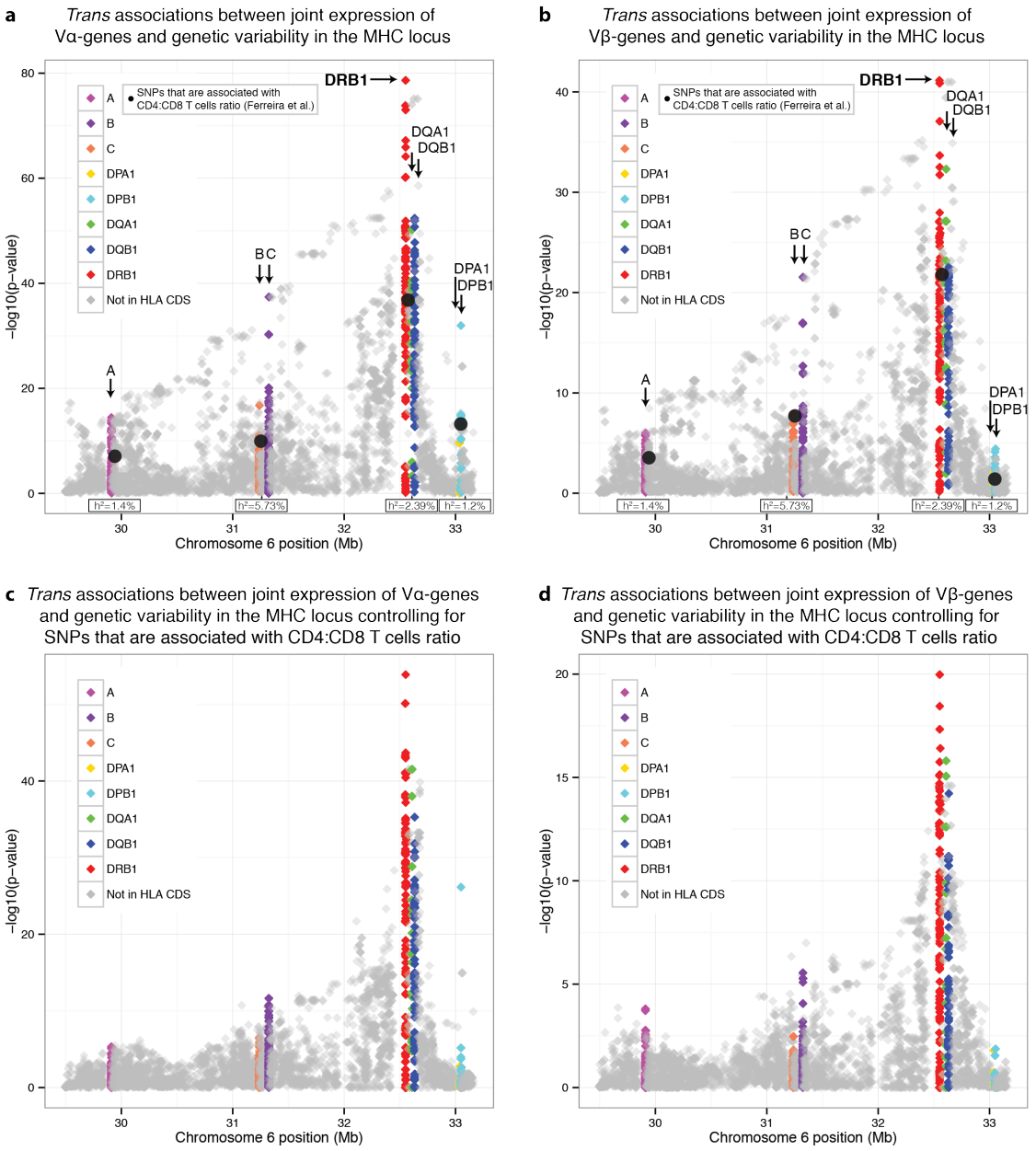
Supplementary Figure 28. Distances between MHC class II β chain amino acid residues and the closest TCR residues in solved TCR-peptide-MHC complexes. (a) Distance between the centroid of an MHC residue and the centroid of the closest TCR residue in solved TCR-peptide-MHC class II protein structures. Data from RCSB PDB[2]. Points are shaped and colored according to TCR-MHC complex PDB ID. **(b)** Zoom in on MHC Class II β residues 50–100. This region contains the residues that encode the MHC β chain α-helix (residues 50–87).



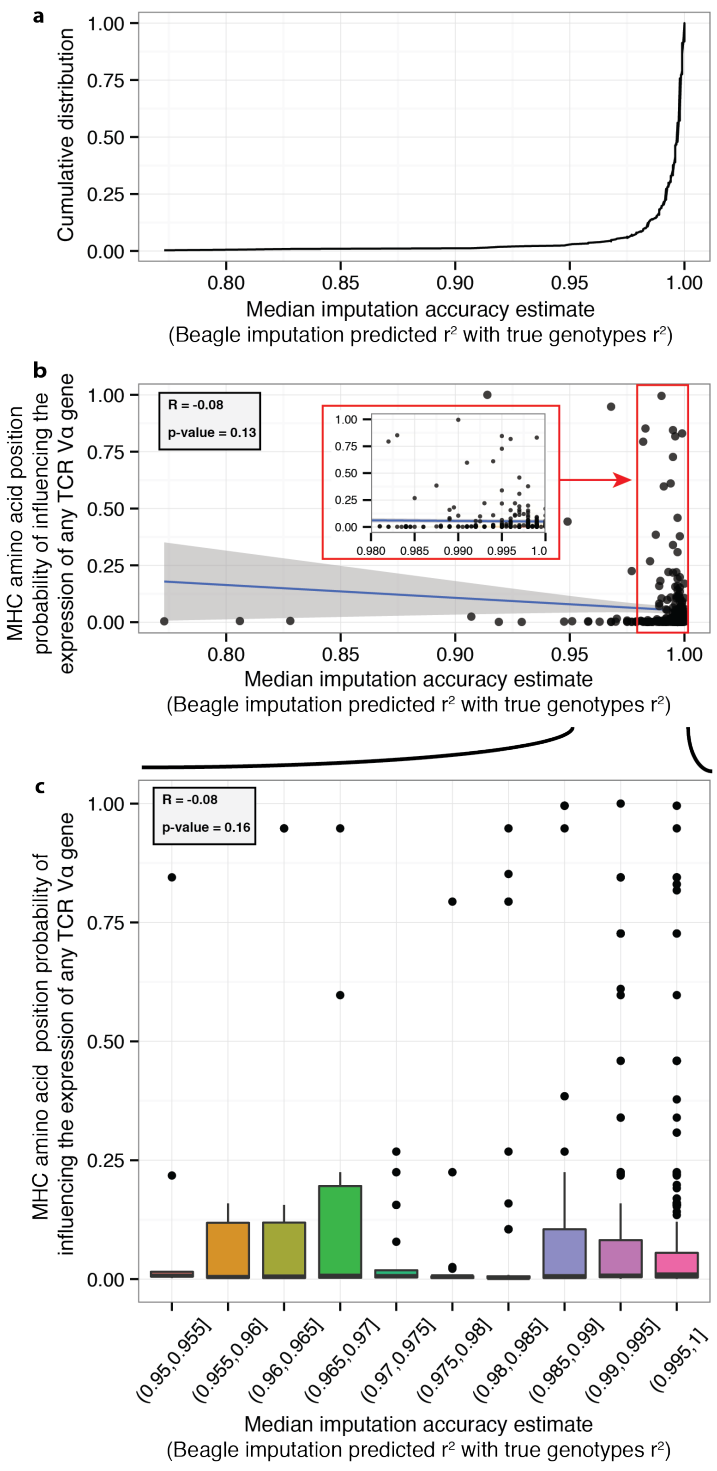
Supplementary Figure 29. Distance between MHC class II β chain amino acid residues and the closest peptide residue in solved TCR-peptide-MHC complexes. (a) Distance between the centroid of an MHC residue and the centroid of the closest peptide residue in solved TCR-peptide-MHC class II protein structures. Data was retrieved from RCSB PDB[2]. Points are shaped and colored by TCR-MHC complex PDB ID**. (b)** Zoom-in on the MHC Class II β residues 1–100 that encompass MHC β chain α-helix (50–87) and β-sheet where the peptides bind (7–42)



Supplementary Figure 30. MHC residues that are associated with TCR Vα genes expression tend to be closer to the TCR. (a) MHC class-II β amino acid residue probabilities of influencing the expression of any TCR Vα (estimated by the Bayesian model) as a function of the distance between the MHC residue centroid and the closest TCR Vα residue centroid. The position of the MHC residue is indicated for those residues with probability >0.0435. The Shannon entropy of the MHC amino acid residue genotype is represented by the color of the points (for amino acid residue with A_h possible variants, Shannon entropy is defined as $E_h = -\sum_{a \in A_h} P(a) \log_2(P(a))$ where $P(a)$ is the abundance of variant a in position h in our dataset). The correlation between the distance to the TCR and the probability of influencing TCR Vα expression was less significant when controlling for the distance to the peptide and the residue genotype entropy (F-test p-value=0.07), suggesting that some of the interactions are mediated by the peptide. The mean distance significantly differs between residues with posterior larger and lower than 0.5 (Wilcoxon rank sum test p-value=0.033) or 0.2 (Wilcoxon rank sum test p-value=0.0056). **(b)** Similar to **(a)** except that the distance is computed between each pair of residues Cα-s instead of the centroids. The mean distance significantly differs between residues with posterior larger and lower than 0.5 (Wilcoxon rank sum test p-value=0.028) or 0.2 (Wilcoxon rank sum test p-value=0.0074).



Supplementary Figure 31. Associations of TCR Vα and Vβ genes with variation in MHC proteins are largely independent of SNPs shown previously to associate with the ratio of CD4:CD8 T cells. Manhattan plots showing the significance of each binary marker of nucleotide or amino acid variation (imputed by SNP2HLA) association with the joint expression of TCR Vα (**a** and **c**) and Vβ (**b** and **d**) genes. Associations were computed using a multivariate multiple response regression. Panels **c** and **d** show the significance of the association controlling for SNPs that were shown by Ferreira et al.[9] to associate with the ratio of CD4:CD8 T cell (excluding one SNP that is not in our imputed set). These SNPs are marked by black dots in panels A and B and the fractions of variation in CD4:CD8 that they explain (h^2) are specified below. The plot shows that the strongest associations with expression of TCR V-genes are with genetic variability in *HLA-DRB1* while the strongest association with CD4:CD8 ratio is in the HLA-B gene. The significance levels are mildly reduced in the conditional analyses (**c** and **d**), presumably due to weak LD with the CD4:CD8-associated SNPs, but the overall pattern of signals remains very similar.



Supplementary Figure 32. MHC amino acid position posterior probability of influencing any TCR Vα expression is not correlated with its imputation quality. (a) The cumulative distribution of imputation accuracy estimate (median beagle r^2 of amino acid residue calculated by SNP2HLA). **(b)** and **(c)** A comparison of the imputation accuracy estimate of MHC amino acids to the probability they influence any TCR Vα gene expression does not show a significant correlation.

References

1. Battle, A. *et al.* Characterizing the genetic basis of transcriptome diversity through RNA-sequencing of 922 individuals. *Genome Res.* **24,** 14–24 (2014).
2. Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Res.* **28,** 235–242 (2000).
3. Delaneau, O. & Marchini, J. Integrating sequence and array data to create an improved 1000 Genomes Project haplotype reference panel. *Nat. Commun.* **5,** 3934 (2014).
4. Howie, B. N., Donnelly, P. & Marchini, J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet.* **5,** e1000529 (2009).
5. Abecasis, G. R. *et al.* An integrated map of genetic variation from 1,092 human genomes. *Nature* **491,** 56–65 (2012).
6. de Bakker, P. I. W. *et al.* A high-resolution HLA and SNP haplotype map for disease association studies in the extended human MHC. *Nat. Genet.* **38,** 1166–72 (2006).
7. Pruim, R. J. *et al.* LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* **26,** 2336–7 (2010).
8. Lund, R. J. *et al.* Genome-wide identification of novel genes involved in early Th1 and Th2 cell differentiation. *J. Immunol.* **178,** 3648–60 (2007).
9. Ferreira, M. A. R. *et al.* Quantitative trait loci for CD4:CD8 lymphocyte ratio are associated with risk of type 1 diabetes and HIV-1 immune control. *Am. J. Hum. Genet.* **86,** 88–92 (2010).
10. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88,** 76–82 (2011).
11. Raychaudhuri, S. *et al.* Five amino acids in three HLA proteins explain most of the association between MHC and seropositive rheumatoid arthritis. *Nat. Genet.* **44,** 291–6 (2012).
12. Hu, X. *et al.* Additive and interaction effects at three amino acid positions in HLA-DQ and HLA-DR molecules drive type 1 diabetes risk. *Nat. Genet.* **advance on,** (2015).
13. Patsopoulos, N. A. *et al.* Fine-mapping the genetic association of the major histocompatibility complex in multiple sclerosis: HLA and non-HLA effects. *PLoS Genet.* **9,** e1003926 (2013).
14. Gutierrez-Achury, J. *et al.* Fine mapping in the MHC region accounts for 18% additional genetic risk for celiac disease. *Nat. Genet.* **47,** 577–8 (2015).