**Figures**

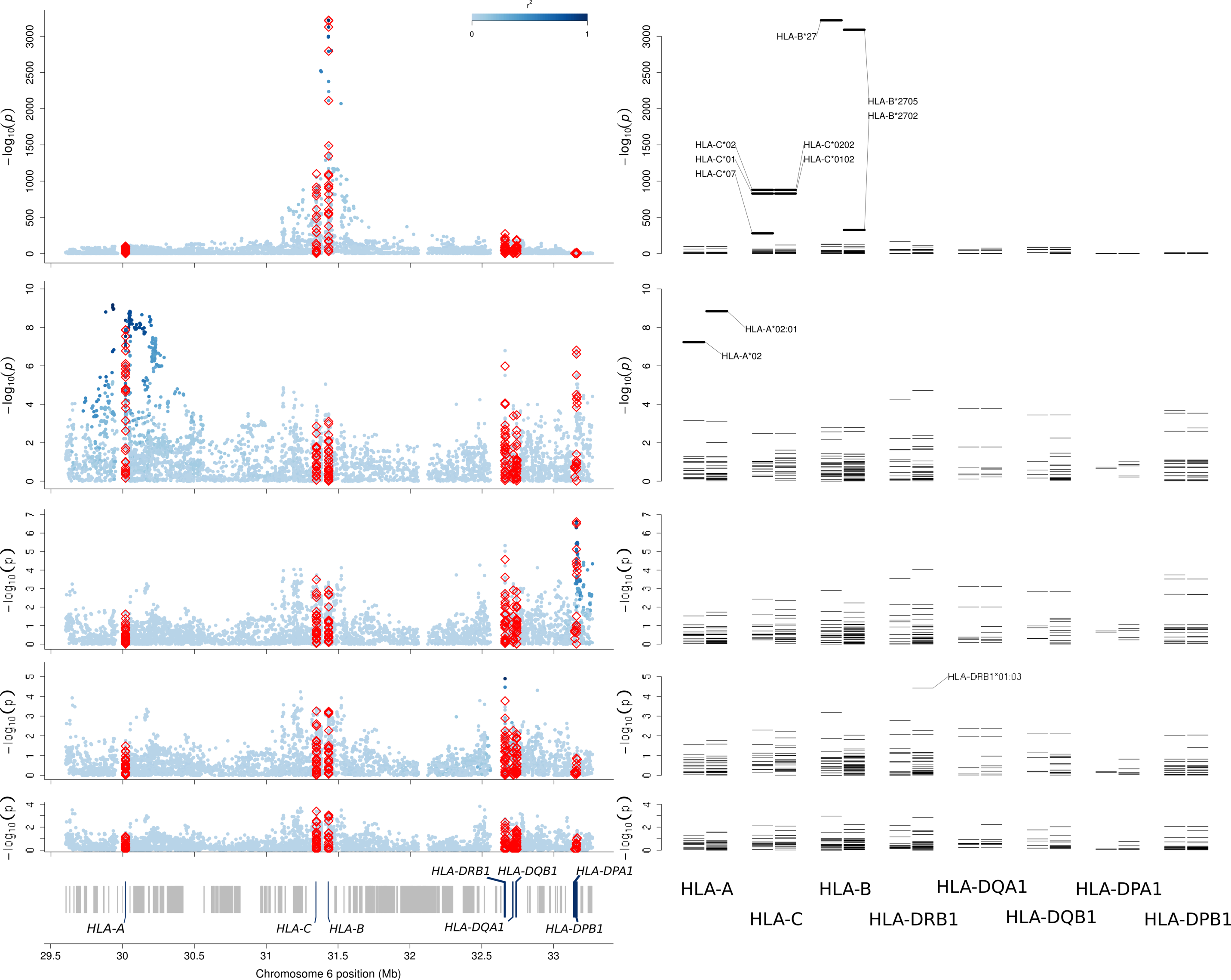
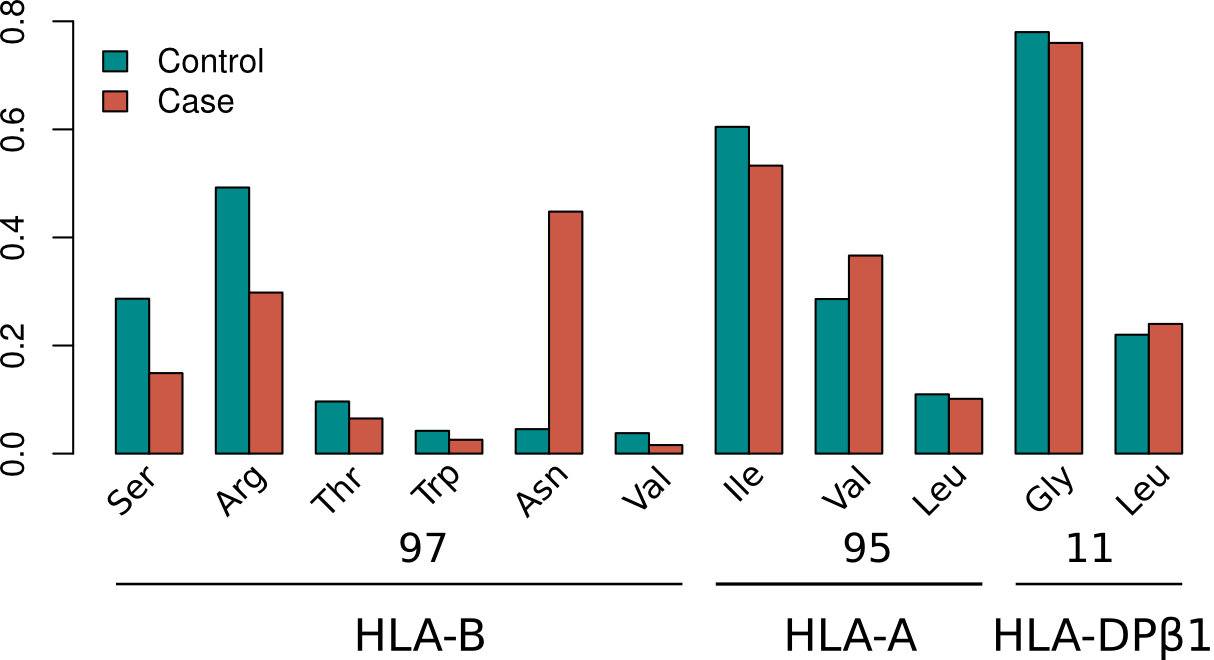
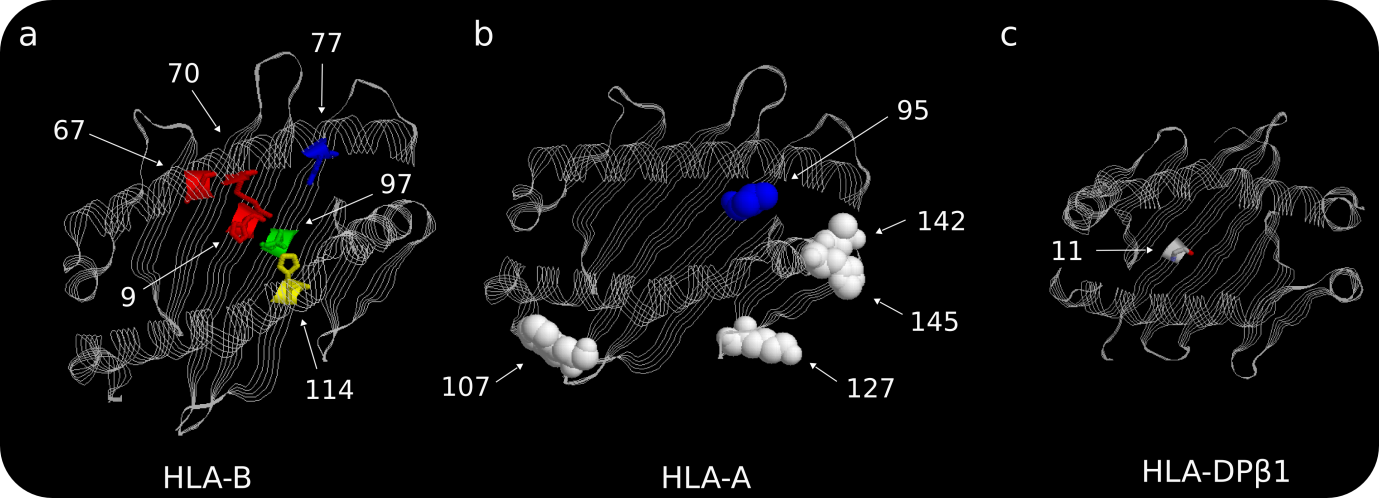


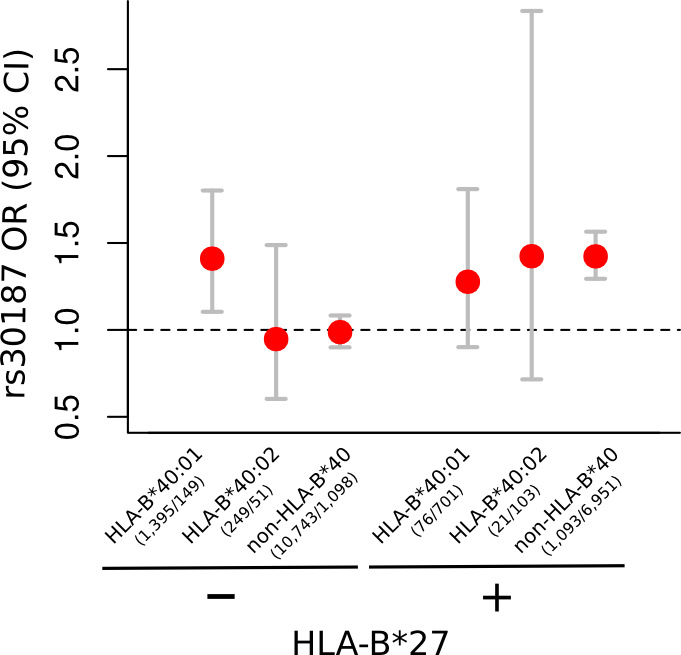
Figure 1



**Figure 2** Amino acid residue frequencies in controls and cases within associated amino acid positions within HLA proteins.



**Figure 3** Three-dimensional models for the HLA-B, HLA-A and HLA-DPβ1 proteins. These structures are based on Protein Data Bank entries 3LV3, 3UTQ and 3LQZ, respectively, with a direct view to the peptide-binding groove.



**Figure 4** Interaction between *ERAP1* and *HLA-B* susceptibility alleles. For each stratified group, effect size for the *ERAP1* variant rs30187 is given. Number of samples in each group (controls/cases) is given below the *HLA-B\*40* genotype.

**Tables**

**Table 1** Most significant polymorphic positions and imputed classical alleles associated with ankylosing spondylitis susceptibility (*P*-value < 1 × 10-2000).

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Position | rs | AA position | Classical allele | χ2 | DF | *P*-value |
| 31,430,829 | rs41558317 | - | - | 14,823 | 1 | < 10-3221 |
| 31,432,180 | rs1071652 | - | - | 14,841 | 3 | < 10-3221 |
| 31,432,180 | - | 97 | - | 14,857 | 5 | < 10-3221 |
| 31,431,272 | - | - | HLA-B\*27 | 14,820 | 1 | < 10-3221 |
| 31,432,179 | rs1140412 | - | - | 14,823 | 2 | < 10-3219 |
| 31,432,506 | - | 70 | - | 14,812 | 3 | < 10-3215 |
| 31,432,129 | - | 114 | - | 14,402 | 2 | < 10-3128 |
| 31,432,130 | rs709055 | - | - | 14,401 | 2 | < 10-3128 |
| 31,432,131 | rs1050628 | - | - | 14,389 | 1 | < 10-3127 |
| 31,431,272 | - | - | HLA-B\*2705 | 14,220 | 1 | < 10-3090 |
| 31,430,834 | rs3819282 | - | - | 13,798 | 1 | < 10-2999 |
| 31,430,345 | rs3819299 | - | - | 13,757 | 1 | < 10-2990 |
| 31,430,346 | rs3819299 | - | - | 13,757 | 1 | < 10-2990 |
| 31,451,646 | rs4463302 | - | - | 12,898 | 1 | < 10-2803 |
| 31,432,485 | - | 77 | - | 12,871 | 2 | < 10-2795 |
| 31,432,486 | rs1131217 | - | - | 12,849 | 1 | < 10-2793 |
| 31,377,108 | rs2394967 | - | - | 11,613 | 1 | < 10-2524 |
| 31,381,125 | rs6905036 | - | - | 11,552 | 1 | < 10-2511 |
| 31,432,208 | rs41556113 | - | - | 10,929 | 1 | < 10-2376 |
| 31,432,843 | rs41553720 | - | - | 10,299 | 2 | < 10-2237 |
| 31,432,515 | - | 67 | - | 9,741 | 4 | < 10-2112 |
| 31,432,515 | rs1071816 | - | - | 9,725 | 3 | < 10-2110 |
| 31,518,387 | rs2844510 | - | - | 9,525 | 1 | < 10-2071 |

**Table 2** Evidence for association of HLA-B alleles with susceptibility to ankylosing spondylitis.

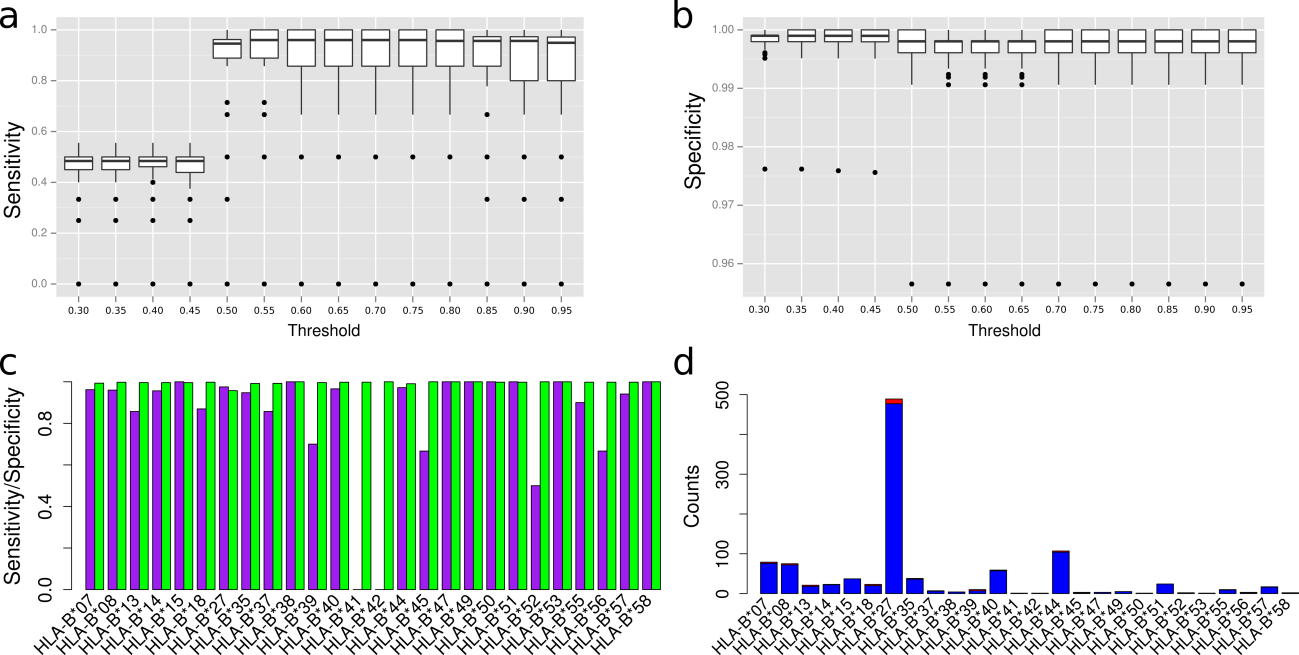
|  |  |  |  |
| --- | --- | --- | --- |
| Round | Allele | OR (95%) | PVAL |
| 1 | 27:05 | 62.41 (56.90-68.45) | < 1 × 10-321 |
| 2 | 27:02 | 43.41 (29.80-63.23) | 1.07 × 10-122 |
| 3 | 07:02 | 0.82 (0.74-0.91) | 5.04 × 10-6 |
| 4 | 57:01 | 0.75 (0.61-0.92) | 5.13 × 10-4 |
| 5 | 51:01 | 1.33 (1.14- 1.56) | 2.14 × 10-3 |
| 6 | 47:01 | 2.35 (1.43-3.86) | 2.25 × 10-3 |
| 7 | 40:02 | 1.59 (1.19-2.14) | 4.65 × 10-3 |
| 8 | 13:02 | 1.43 (1.14-1.80) | 4.29 × 10-3 |
| 9 | 40:01 | 1.22 (1.06-1.40) | 4.93 × 10-3 |
| All other alleles | | | > 0.05 |

**Table 3** Haplotype analysis of SNPs encoding the amino acid 97 of HLA-B

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | *HLA-B* codon 97 position | | |  |  | | | | | | |
| 1 | 2 | 3 |
| SNP | rs41558317 | rs1140412 | rs1071652 | rs41556417 | Amino acid residue |
| Position (HG18) | 31,430,829 | 31,432,179 | 31,432,180 | 31,432,181 |
| Reference allele or amino acid | A | G | C | T | Serine (S) |
| Alternate allele(s) | G | C/A | G/A/T | A/C |  |
| Allele frequency in controls (ref/alt(s)) | 0.95/0.05 | 0.29/0.67/0.05 | 0.82/0.10/0.04/0.05 | 0.92/0.04/0.04 |
| Single locus univariate *P*-value | < 1 × 10-3221 | < 1 × 10-3219 | < 1 × 10-3221 | 2.10 × 10-65 | Multivariate OR | 95% CI | Haplotype counts | | Unadjusted haplotype frequency | | *P*-value |
| Risk allele univariate OR (95% CI) | 60.36 (55.47-65.74) | 59.99 (55.13-65.33) | 59.99 (55.14-65.34) | 2.03 (1.86-2.21) | Controls | Cases | Controls | Cases |
| Haplotype | G | A | T | T | Asparagine (N) | 16.55 | 15.47-17.74 | 1,221 | 8,134 | 0.045 | 0.449 | < 1 × 10-300 |
| A | A | T | T | Asparagine (N) | 3.31 | 0.56-22.63 | 3 | 4 | 1.00 × 10-04 | 2.00 × 10-04 | 0.110 |
| G | C | C | T | Arginine (R) | 2.49 | 0.18-34.27 | 2 | 2 | 1.00 × 10-04 | 1.00 × 10-04 | 0.325 |
| A | C | G | T | Threonine (T) | 1.12 | 1.03-1.2 | 2,620 | 1,177 | 0.097 | 0.065 | 4.50 × 10-3 |
| A | C | C | T | Arginine (R) | 1.00 | Reference | 13,375 | 5,382 | 0.493 | 0.297 | 1 |
| A | C | C | A | Tryptophan (W) | 1.00 | 0.89-1.12 | 1,137 | 458 | 0.042 | 0.025 | 1 |
| A | G | C | T | Serine (S) | 0.86 | 0.82-0.91 | 7,769 | 2,691 | 0.286 | 0.148 | 5.17 × 10-8 |
| A | C | A | C | Valine (V) | 0.68 | 0.59-0.78 | 1,024 | 281 | 0.038 | 0.016 | 1.41 × 10-8 |

**SUPPLEMENTARY MATERIAL**

**Supplementary Figures**



**Supplementary Figure 1** Performance of classical allele imputation at two digit resolution in the *HLA-B* locus. HLA typing at the *HLA-B* locus was available for 536 AS patients, which also had imputed genotype data. The optimal threshold dosage for genotype calling was assessed by computing the sensitivity (a) and specificity (b) for all HLA-B alleles at different dosage thresholds. For a dosage threshold of 0.7 the median sensitivity (c; purple bars) and specificity (c; green bars) for all HLA-B alleles was 0.958 and 0.998, respectively. (d) Proportion of false negatives in the imputed alleles for each allele at a 0.7 dosage threshold, blue sections are proportional to the number of true positives and red sections are proportional to the number of false negatives.

**Supplementary Tables**

**Supplementary Table 1** Level of conditional association significance at the five most significant amino acid positions in the HLA-B protein. For each position we conditioned on it and tested for association at the other four positions.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | | Omnibus amino acid tests | | | | | |
| 97 | 70 | 114 | 77 | 67 | 9 |
| Conditional amino acid position | 97 | -- | 0.014 | 1.08 × 10-3 | 2.28 × 10-4 | 0.040 | 0.016 |
| 70 | 6.28 × 10-11 | -- | 6.84 × 10-5 | 6.05 × 10-3 | 0.882 | 7.28 × 10-3 |
| 114 | 3.46 × 10-99 | 1.32 × 10-92 | -- | 2.10 × 10-9 | 4.16 × 10-17 | 1.12 × 10-13 |
| 77 | < 1 × 10-300 | < 1 × 10-300 | < 1 × 10-300 | -- | 1.21 × 10-144 | 1.32 × 10-33 |
| 67 | < 1 × 10-300 | < 1 × 10-300 | < 1 × 10-300 | < 1 × 10-300 | -- | < 1 × 10-300 |
| 9 | < 1 × 10-300 | < 1 × 10-300 | < 1 × 10-300 | < 1 × 10-300 | < 1 × 10-300 | -- |

**Supplementary Table 2** No evidence of interaction between *HLA-B\*27* and *HLA-B\*40* risk alleles. Odds of disease for each genotype class was computed against the reference (*HLA-B\*27*-negative and *\*40*-negative). The risk effect for the genotype *HLA-B\*27*-/*HLA-B\*40*+ was estimated to be 1.35 and for the class *HLA-B\*27*+/*HLA-B\*40*- the effect was 47.76. Under no interaction between these two alleles we would expect the effect size of the genotype class *HLA-B\*27*+/*HLA-B\*40*+ to be 1.35\*47.76 = 64.48 (assuming a model for risk on the multiplicative scale), which is close to what we observed, 65.19.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Locus Genotypes | Counts | | Frequency | | OR | CI (95%) | *P*-value |
| Controls | Cases | Controls | Cases |
| HLA-B\*27/ HLA-B\*27 | 33 | 363 | 0.002 | 0.038 | 84.78 | 58.9-125.93 | 3.23 × 10-279 |
| HLA-B\*27/ HLA-B\*40 | 97 | 822 | 0.007 | 0.087 | 65.19 | 52.42-81.85 | < 1 × 10-300 |
| HLA-B\*27/non-HLA-B\*40 | 1,069 | 6,626 | 0.079 | 0.698 | 47.76 | 43.8-52.14 | < 1 × 10-300 |
| HLA-B\*40/ HLA-B\*40 | 61 | 12 | 0.004 | 0.001 | 1.52 | 0.74-2.85 | 0.20 |
| HLA-B\*40/non-HLA-B\*27 | 1,587 | 278 | 0.117 | 0.029 | 1.35 | 1.17-1.55 | 3.38 × 10-5 |
| other | 10,723 | 1,391 | 0.790 | 0.147 | 1.00 | Reference | |