The American Journal of Human Genetics, Volume *93*

Supplemental Data

Coding Variants at Hexa-allelic Amino Acid 13 of HLA-DRB1 Explain Independent SNP Associations with Follicular Lymphoma Risk

Jia Nee Foo, Karin E. Smedby, Nicholas K. Akers, Mattias Berglund, Ishak D. Irwan, Xiaoming Jia, Yi Li, Lucia Conde, Hatef Darabi, Paige M. Bracci, Mads Melbye, Hans-Olov Adami, Bengt Glimelius, Chiea Chuen Khor, Henrik Hjalgrim, Leonid Padyukov, Keith Humphreys, Gunilla Enblad, Christine F. Skibola, Paul I.W. de Bakker, and Jianjun Liu

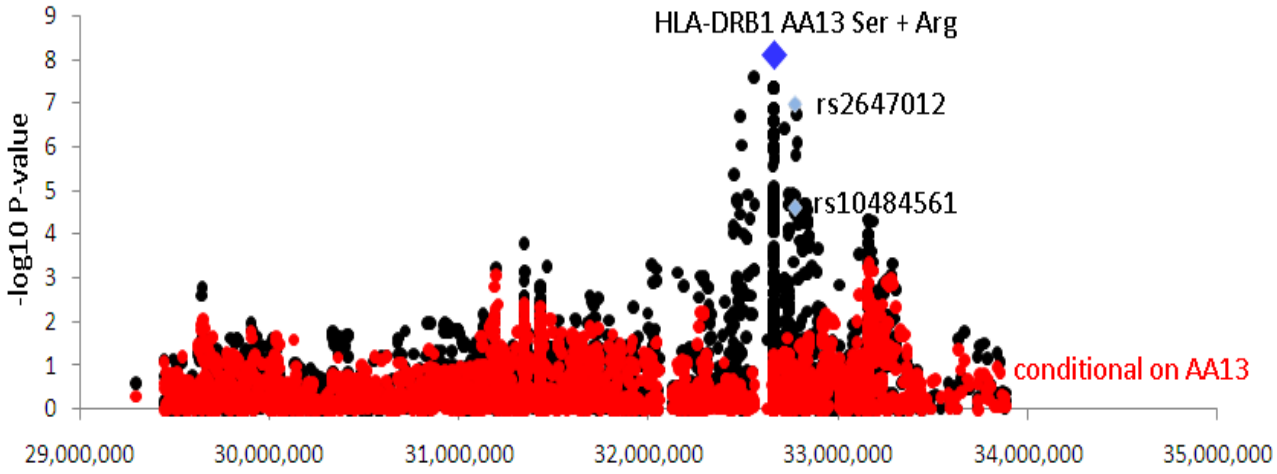

Figure S1. Region Association Plot before and after Conditioning on Alleles at Position 13

Regional association plot of bi-allelic logistic regression test P-values done on imputed amino acid alleles and SNPs surrounding rs2647012 and rs10484561 before (black) and after (red) conditioning on alleles at position 13 in SCALE GWAS samples

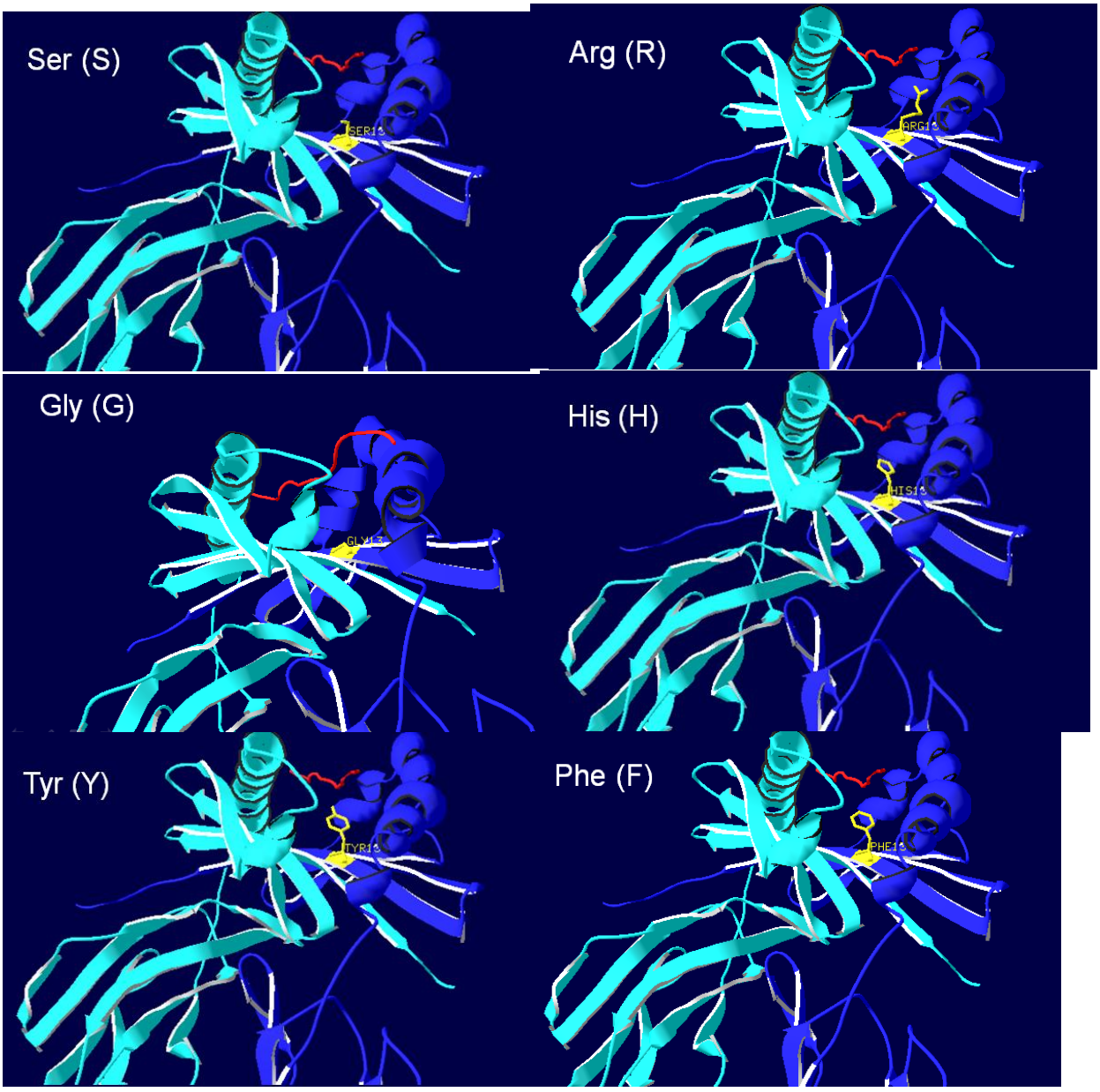

Figure S2. Six Possible Encoded Amino Acids at Position 13 in the HLA-DR Binding Groove

Amino acid 13 (yellow) in the binding groove of structure (PDB ID: 1AQD) of HLA-DR1 (DRA, DRB1 0101, Phe at position 13) protein (extracellular domain) complexed with endogenous peptide (red). Other amino acid alleles were mutated *in silico* using the DeepView/Swiss-PdbViewer program (http://spdbv.vital-it.ch/).

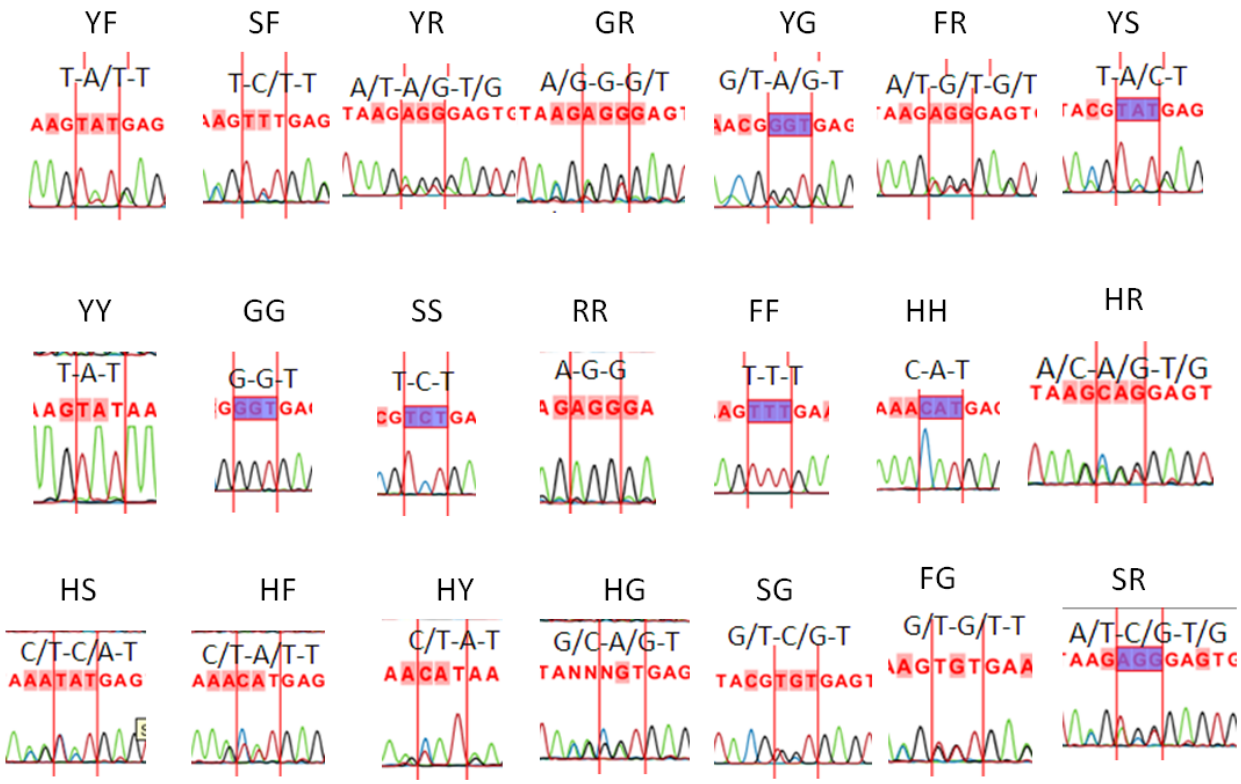

Figure S3. Genotyping of Amino Acid 13 Based on Sequence Chromatograms

All 21 genotypes (the two residues present in each individual, each coded as Y=Tyr, F=Phe, G=Gly, S=Ser, R=Arg, H=His) could be called and distinguished by visual inspection of the chromatograms.

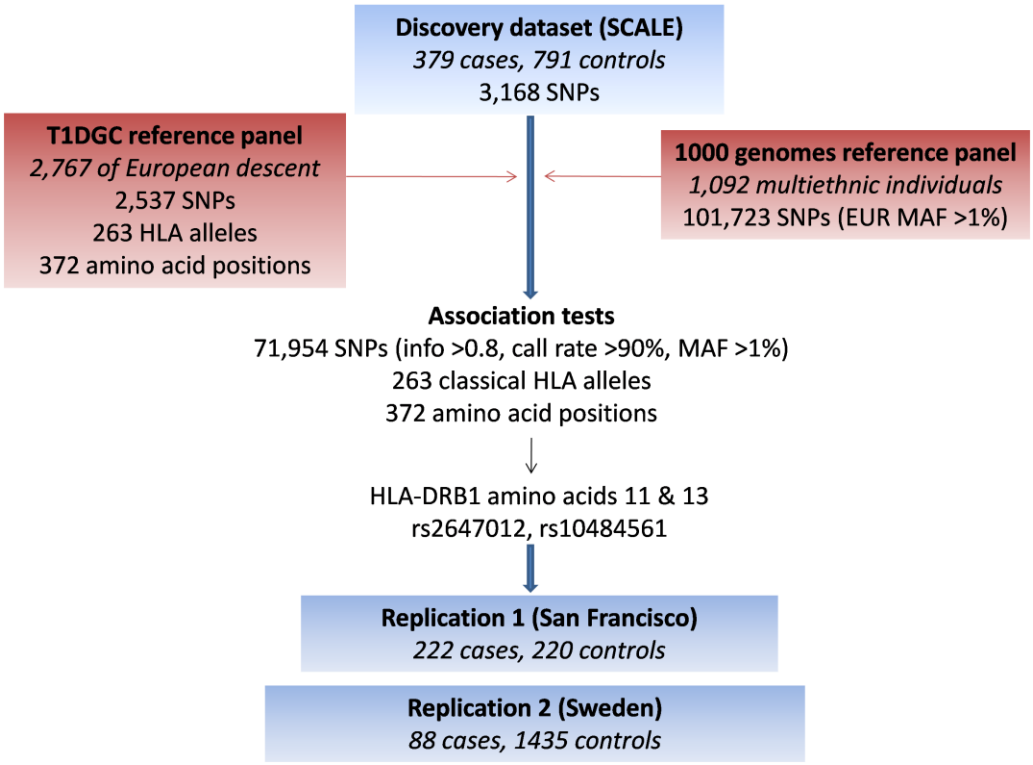

Figure S4. Flowchart on Study Design and Data Acquisition

Flowchart on the imputation, association testing (showing filters for SNPs imputed by IMPUTEv2) and selection of variants for validation in additional samples. The datasets (and sample sizes) used as reference panels for imputation are shown in red boxes and the discovery or replication datasets used for association testing are shown in blue boxes.

Table S1. Conditional Analysis on FL-Associated SNPs in the HLA Class II Region

SNP	Position (b37) chr6	A1	OR (95% CI)	P	OR condition rs2647012 + rs10484561	P condition rs2647012 + rs10484561	r^2 / D' rs2647012	r^2 / D' rs10484561	r^2 / D' rs9378212
rs4530903	32581889	T	1.545 (1.183-2.018)	0.00139	0.509 (0.186-1.392)	0.188	0.085/0.979	0.92/0.991	0.154/0.989
rs9268853	32429643	C	1.302 (1.087-1.559)	0.00409	1.311 (0.995-1.727)	0.054	0.324/0.991	0.07/0.962	0.577/1
rs2621416	32741868	C	1.419 (1.163-1.732)	0.00058	1.056 (0.840-1.328)	0.642	0.207/0.843	0.147/0.641	0.187/0.611
rs2647046	32668336	A	0.617(0.510-0.747)	6.84E-07	NA	NA	0.991/0.996	0.09/0.934	0.558/0.986
rs9276490	32718681	A	0.695 (0.576-0.840)	0.000161	0.871 (0.703-1.080)	0.208	0.263/0.603	0.077/0.735	0.082/0.322
rs7453920	32730012	A	0.689 (0.570-0.833)	0.000116	0.854 (0.687-1.060)	0.152	0.27/0.605	0.075/0.738	0.079/0.319

Linkage disequilibrium and conditional analyses on rs2647012, rs10484561 and rs9378212 and the HLA SNPs reported in a recent FL GWAS by Vijai et al.[4] that were either genotyped (only rs7453920) or imputed in the SCALE discovery GWAS dataset.

Table S2. Association Results at Amino Acids 11 and 13 in the Discovery and Replication Data Sets

	Discovery GWAS (379 cases/791 controls)			Replication 1 (222 cases/220 controls)			Replication 2 (88 cases/1,435 controls)			Meta-analysis		
	Freq Cases/ Controls	OR (95% CI)	P	Freq Cases/ Controls	OR (95% CI)	P	Freq Cases/ Controls	OR (95% CI)	P	OR	P_{het}	I^2
Asp11	0.022/ 0.017	1.311 (0.716-2.402)	0.380	0.009/ 0.016	0.558 (0.161-1.935)	0.358	0.040/ 0.021	1.928 (0.871-4.270)	0.105	1.326	0.218	26.3
Gly11	0.128/ 0.097	1.358 (1.039-1.773)	0.025	0.162/ 0.121	1.388 (0.955-2.017)	0.085	0.091/ 0.077	1.180 (0.705-1.974)	0.528	1.338	4.36×10^{-3}	0
Leu11	0.168/ 0.106	1.724 (1.338-2.222)	2.59×10^{-5}	0.182/ 0.093	2.211 (1.465-3.337)	1.58×10^{-4}	0.227/ 0.115	2.149 (1.498-3.082)	3.22×10^{-5}	1.922	4.82×10^{-12}	0
Pro11	0.124/ 0.151	0.803 (0.624-1.032)	0.087	0.124/ 0.198	0.568 (0.391-0.826)	3.03×10^{-3}	0.114/ 0.165	0.654 (0.408-1.050)	0.079	0.710	4.26×10^{-4}	16.4
Ser11	0.330/ 0.426	0.666 (0.555-0.799)	1.17×10^{-5}	0.360/ 0.421	0.775 (0.590-1.018)	0.067	0.358/ 0.411	0.805 (0.588-1.100)	0.173	0.717	1.67×10^{-6}	0
Val11	0.228/ 0.204	1.150 (0.937-1.411)	0.181	0.158/ 0.152	1.042 (0.723-1.504)	0.824	0.165/ 0.211	0.740 (0.493-1.112)	0.147	1.050	0.558	44.3
Ser13	0.255/ 0.358	0.615 (0.506-0.747)	9.30×10^{-7}	0.315/ 0.375	0.763 (0.576-1.012)	0.061	0.290/ 0.350	0.761 (0.546-1.062)	0.108	0.677	1.18×10^{-7}	5.9
Arg13	0.124/ 0.151	0.803 (0.624-1.032)	0.087	0.124/ 0.198	0.568 (0.391-0.826)	3.03×10^{-3}	0.114/ 0.165	0.654 (0.408-1.050)	0.079	0.710	4.26×10^{-4}	16.4
Gly13	0.075/ 0.068	1.111 (0.795-1.553)	0.538	0.045/ 0.045	0.991 (0.526-1.865)	0.977	0.057/ 0.057	1.000 (0.521-1.922)	0.999	1.069	0.629	0
His13	0.220/ 0.195	1.161 (0.943-1.428)	0.159	0.146/ 0.143	1.026 (0.705-1.494)	0.892	0.165/ 0.203	0.774 (0.513-1.166)	0.221	1.060	0.489	34.0
Tyr13	0.128/ 0.097	1.358 (1.039-1.773)	0.025	0.162/ 0.121	1.388 (0.955-2.017)	0.085	0.091/ 0.077	1.180 (0.705-1.974)	0.528	1.338	4.36×10^{-3}	0
Phe13	0.198/ 0.132	1.649 (1.303-2.087)	3.09×10^{-5}	0.203/ 0.118	1.881 (1.296-2.731)	8.90×10^{-4}	0.284/ 0.148	2.160 (1.547-3.016)	6.11×10^{-6}	1.820	6.71×10^{-12}	0

P_{het}: Cochrane's Q test P-value; I^2: inconsistency

Table S3. Association Statistics at rs2647012, rs10484561, and rs9268839/rs9378212 before and after Adjustment for Alleles at Amino Acid Position 13

Sample collection Cases/Controls	Before adjustment				After adjustment			
rs2647012	**OR (95% CI)**			**P**	**OR (95% CI)**			**P**
Discovery 379 / 791	0.603 (0.500-0.727)			1.10×10^{-7}	0.851 (0.582-1.244)			0.405
Replication 1 222 / 220	0.551 (0.418-0.726)			2.33×10^{-5}	0.636 (0.432-0.936)			0.0216
Replication 2 88 /1435	0.831 (0.609-1.135)			0.245	1.771 (0.965-3.250)			0.065
Meta-analysis	**P**	**OR**	**P_{het}**	**I^2**	**P**	**OR**	**P_{het}**	**I^2**
Fixed effects	4.72×10^{-11}	0.628	0.125	51.99	0.207	0.853	0.021	74.29
Random effects	3.33×10^{-5}	0.639			0.804	0.938		
rs10484561	**OR (95% CI)**			**P**	**OR (95% CI)**			**P**
Discovery 379 / 791	1.686 (1.324-2.148)			2.33×10^{-5}	1.558 (0.864-2.808)			0.140
Replication 1 222 / 220	2.570 (1.701-3.883)			7.41×10^{-6}	6.006 (2.226-16.2)			4.00×10^{-4}
Replication 2 88 /1435	1.791 (1.237-2.595)			2.05×10^{-3}	0.614 (0.311-1.212)			0.160
Meta-analysis	**P**	**OR**	**P_{het}**	**I^2**	**P**	**OR**	**P_{het}**	**I^2**
Fixed effects	2.61×10^{-11}	1.857	0.220	33.97	0.104	1.402	0.0009	85.75
Random effects	9.14×10^{-8}	1.900			0.356	1.698		
rs9268839/rs9378212	**OR (95% CI)**			**P**	**OR (95% CI)**			**P**
Discovery 379 / 791	1.637 (1.372-1.953)			4.57×10^{-8}	NA			NA

Given the high heterogeneity observed across the three datasets, both the fixed effects and random effects models were considered in the meta-analysis. P_{het}: Cochrane's Q test P-value; I^2: inconsistency.

Table S4. Haplotype Analysis of rs10484561 and rs2647012 with the Six Amino Acid Residues at Position 13 in a Meta-analysis across All Three Data Sets

Phased haplotype: AA13 – rs2647012 – rs10484561	Discovery		Replication 1		Replication 2		Meta-analysis		
	MAF Cases/ Controls	OR (95% CI) P	MAF Cases/ Controls	OR (95% CI) P	MAF Cases/ Controls	OR (95% CI) P	OR (95% CI)	P	P_{het}/ I^2
Tyr – C – A	0.128/ 0.097	1.36 (1.04-1.77) 0.025	0.151/ 0.121	1.28 (0.88-1.87) 0.200	0.090/ 0.076	1.19 (0.71-1.99) 0.510	1.31 (1.07-1.60)	8.58E-03	0.896/ 0
Phe – C – A	0.024/ 0.018	1.34 (0.74-2.42) 0.332	0.011/ 0.021	0.54 (0.18-1.64) 0.277	0.056/ 0.018	3.37 (1.65-6.87) 8.58E-04	1.62 (1.06-2.47)	0.025	0.017/ 75.6
Phe – C – C	0.174/ 0.113	1.66 (1.30-2.12) 5.86E-05	0.189/ 0.091	2.396 (1.58-3.63) 3.64E-05	0.208/ 0.127	1.77 (1.22-2.57) 2.64E-03	1.81 (1.51-2.18)	2.55E-10	0.322/ 11.7
Gly – C – A	0.074/ 0.068	1.09 (0.78-1.53) 0.617	0.045/ 0.046	0.99 (0.53-1.87) 0.977	0.056/ 0.055	1.02 (0.53-1.96) 0.965	1.06 (0.81-1.39)	0.682	0.957/ 0
His – C – A	0.220/ 0.193	1.17 (0.95-1.44) 0.139	0.146/ 0.143	1.03 (0.71-1.49) 0.892	0.152/ 0.201	0.71 (0.46-1.08) 0.108	1.05 (0.89-1.25)	0.538	0.111/ 54.6
Arg – C – A	0.008/ 0.007	1.13 (0.44-2.90) 0.808	0.025/ 0.023	1.10 (0.46-2.63) 0.840	0.023/ 0.012	1.79 (0.67-4.80) 0.245	1.28 (0.75-2.19)	0.368	0.725/ 0
Arg – T – A	0.115/ 0.144	0.78 (0.60-1.01) 0.061	0.099/ 0.173	0.52 (0.35-0.78) 0.002	0.090/ 0.153	0.55 (0.33-0.93) 0.025	0.67 (0.55-0.82)	9.80E-05	0.193/ 39.1
Ser – C – A	0.050/ 0.068	0.72 (0.49-1.06) 0.098	0.119/ 0.111	1.08 (0.72-1.60 0.722	0.028/ 0.077	0.35 (0.14-0.86) 0.022	0.81 (0.62-1.05)	0.117	0.059/ 64.7
Ser – T – A	0.201/ 0.290	0.62 (0.50-0.76) 6.62E-06	0.194/ 0.264	0.69 (0.51-0.94) 0.018	0.253/ 0.266	0.94 (0.66-1.32) 0.706	0.69 (0.59-0.81)	2.69E-06	0.132/ 50.6

P_{het}: Cochrane's Q test P-value; I^2: inconsistency

Table S5. Haplotype Analysis of Alleles at Amino Acids 11 and 13 in the Meta-analysis across All Three Data Sets

Haplotype	Average Frequency in controls	P	OR (95% CI)	P_{het}	I^2	LD (r^2) in SCALE
Leu11-Phe13	0.105	4.82×10^{-12}	1.922 (1.597-2.313)	0.468	0	0.79
Asp11-Phe13	0.018	0.218	1.326 (0.847-2.078)	0.257	26.3	0.095
Gly11-Tyr13	0.098	0.004	1.338 (1.095-1.634)	0.870	0	1
Val11-His13	0.180	0.489	1.060 (0.898-1.252)	0.220	34.0	0.95
Ser11-Gly13	0.057	0.629	1.069 (0.816-1.400)	0.929	0	0.12
Pro11-Arg13	0.171	4.26×10^{-4}	0.710 (0.586-0.859)	0.302	16.4	1
Ser11-Ser13	0.361	1.18×10^{-7}	0.677 (0.586-0.782)	0.345	5.9	0.74

P_{het}: Cochrane's Q test P-value; I^2: inconsistency; LD: linkage disequilibrium between the encoded residues at positions 11 and 13.

Table S6. Amino Acid Residue Found at Positions 11 and 13 of Each Classical *HLA-DRB1* Allele

HLA type	11	13
DRB1*01:01	Leu	Phe
DRB1*01:02	Leu	Phe
DRB1*01:03	Leu	Phe
DRB1*01:04	Leu	Phe
DRB1*03:01	Ser	Ser
DRB1*04:01	Val	His
DRB1*04:02	Val	His
DRB1*04:03	Val	His
DRB1*04:04	Val	His
DRB1*04:05	Val	His
DRB1*04:06	Val	His
DRB1*04:07	Val	His
DRB1*04:08	Val	His
DRB1*07	Gly	Tyr
DRB1*08	Ser	Gly
DRB1*09	Asp	Phe
DRB1*10:01	Val	Phe
DRB1*11:01	Ser	Ser
DRB1*11:02	Ser	Ser
DRB1*11:03	Ser	Ser
DRB1*11:04	Ser	Ser
DRB1*12:01	Ser	Gly
DRB1*13:01	Ser	Ser
DRB1*13:02	Ser	Ser
DRB1*13:03	Ser	Ser
DRB1*13:04	Ser	Ser
DRB1*14:01	Ser	Ser
DRB1*14:02	Ser	Ser
DRB1*14:03	Ser	Ser
DRB1*14:04	Ser	Gly
DRB1*15:01	Pro	Arg
DRB1*16:01	Pro	Arg

Data from the EMBL-EBI Immunogenetics HLA Database (http://www.ebi.ac.uk/imgt/hla/)