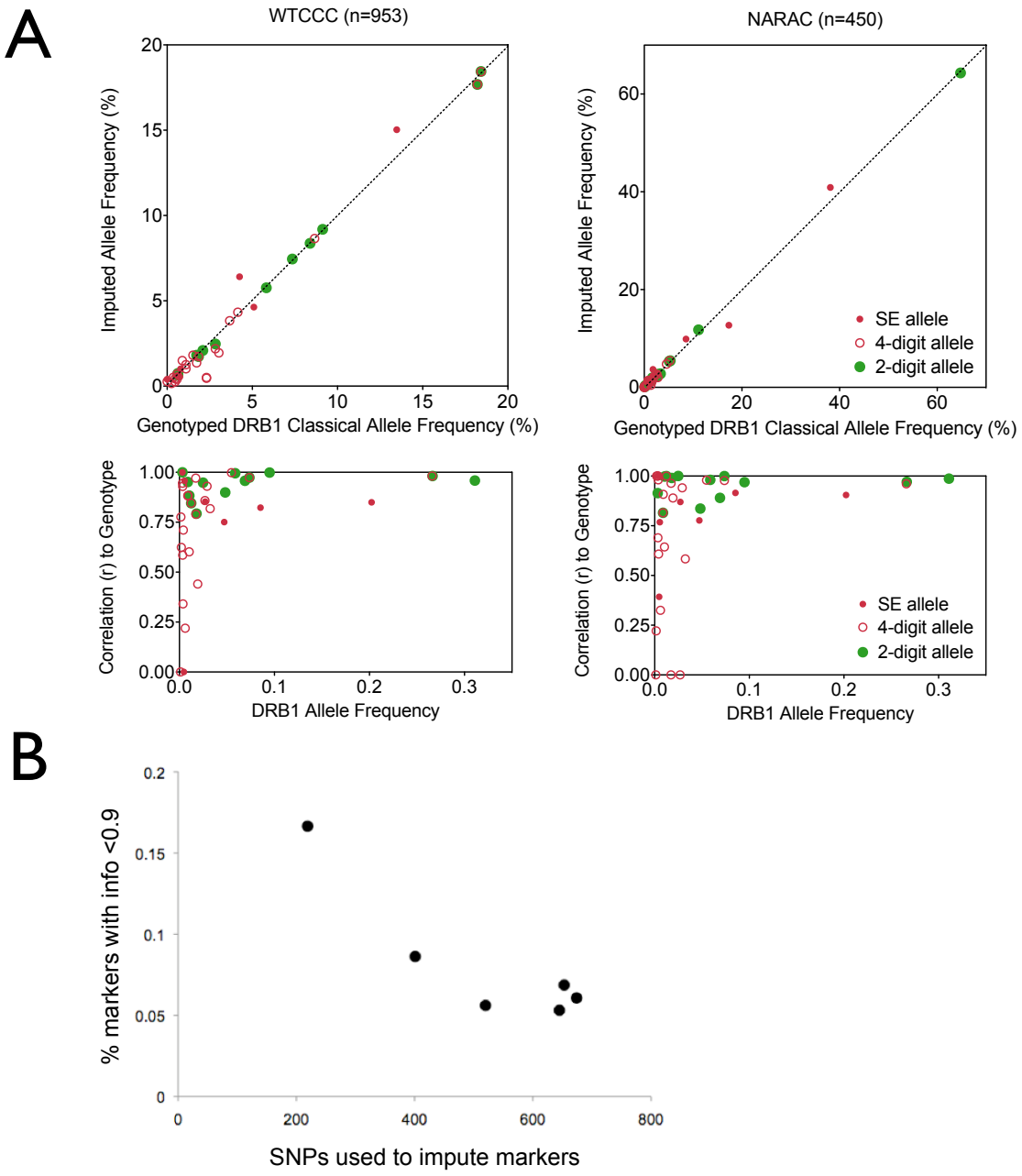
Five amino acids in three HLA proteins explain most of the association between MHC and seropositive rheumatoid arthritis

Supplementary Materials

Soumya Raychaudhuri[1-4,†], Cynthia Sandor[1-4], Eli A. Stahl[1,2,4], Jan Freudenberg[5], Hye-Soon Lee[6], Xiaoming Jia[1,4,7], Lars Alfredsson[8], Leonid Padyukov[9], Lars Klareskog[9], Jane Worthington[10], Katherine A. Siminovitch[11], Sang-Cheol Bae[6], Robert M. Plenge[1,2,4], Peter K. Gregersen[5], and Paul I.W. de Bakker[1,4,12,13,†]

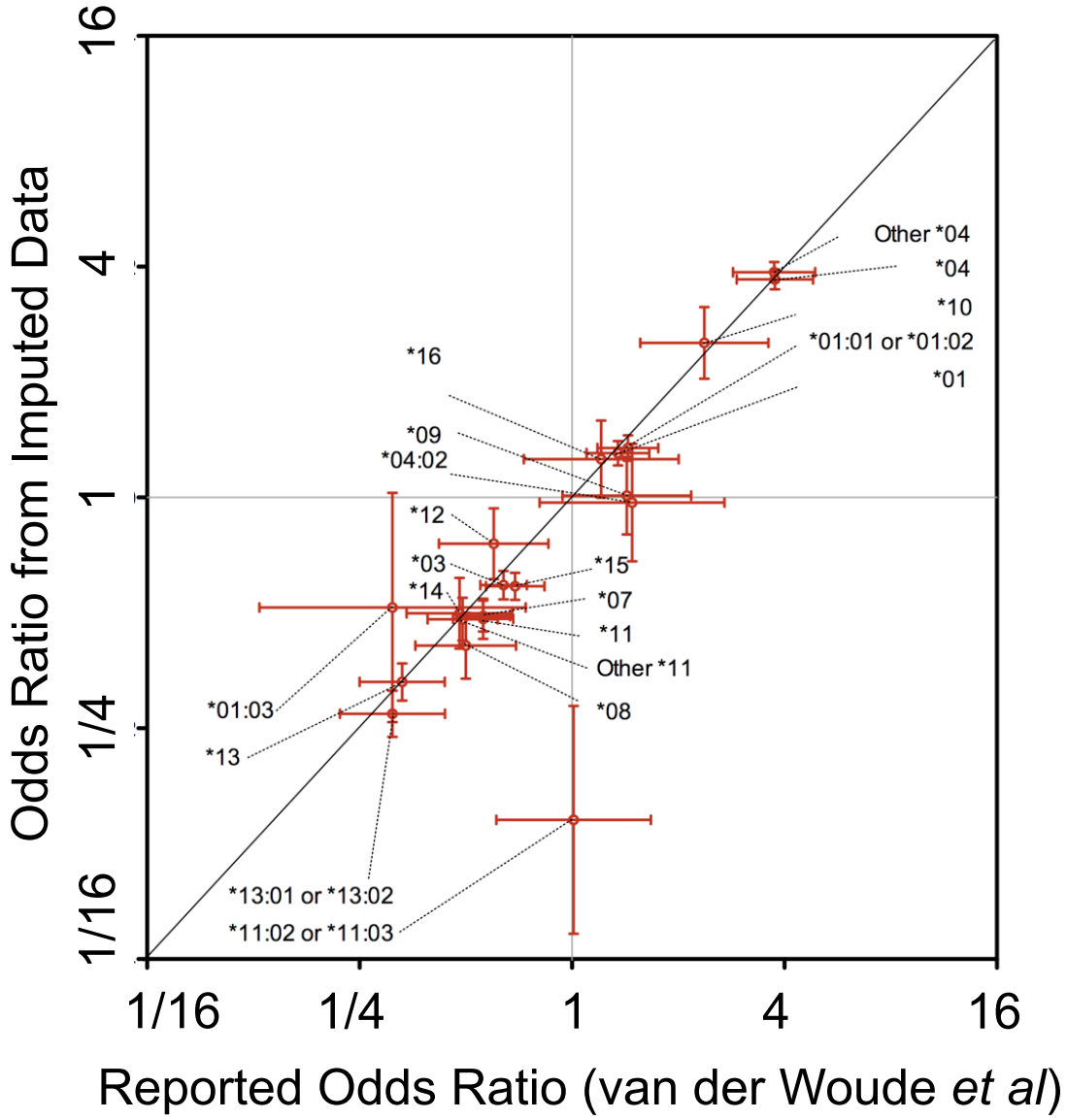
1. Division of Genetics, Brigham & Women's Hospital, Harvard Medical School, Boston, MA, USA
2. Division of Rheumatology, Brigham & Women's Hospital, Harvard Medical School, Boston, MA, USA
3. Partners HealthCare Center for Personalized Genetic Medicine, Boston, MA, USA
4. Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA, USA
5. Robert S. Boas Center for Genomics and Human Genetics, Feinstein Institute for Medical Research, North Shore–Long Island Jewish Health System, Manhasset, NY USA
6. Department of Rheumatology, Hanyang University Hospital for Rheumatic Diseases, Seoul, South Korea
7. Harvard-MIT Division of Health Sciences and Technology, Boston, MA
8. Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden
9. Rheumatology Unit, Department of Medicine, Karolinska Institutet at Karolinska University Hospital Solna, Stockholm, Sweden
10. Arthritis Research UK Epidemiology Unit, Manchester Academic Health Sciences Centre, University of Manchester, Manchester, UK
11. Department of Medicine, University of Toronto, Mount Sinai Hospital and University Health Network, Toronto, Ontario, Canada
12. Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, The Netherlands
13. Department of Medical Genetics, University Medical Center Utrecht, Utrecht, The Netherlands

[†]To whom correspondence should be addressed: soumya@broadinstitute.org (S.R.); pdebakker@rics.bwh.harvard.edu (P.I.W.d.B.)



Supplementary Figure 1. A. Imputation accuracy of classical *HLA-DRB1* alleles. In addition to genome-wide SNP data, we had access to classical *HLA-DRB1* genotypes at 4-digit resolution in 953 individuals from the UK 1958 birth cohort (labeled WTCCC, left panels) and 450 individuals (175 controls and 275 cases) from the NARAC cohort (right panels). Upper panels plot the imputed allele frequency as a function of the actual allele frequency for all 2- and 4-digit alleles. The correlation is 0.989 for the WTCCC four digit alleles and 0.989 for the NARAC four digit alleles. Lower panels plot the correlation between imputed and genotyped allelic dosages (defined as the number of alleles carried by a given individual) as a function of the classical *HLA-DRB1* allele frequency in

the imputation reference panel. For most common alleles, the imputations are of good quality. We have presented results for 4-digit DRB1 alleles, two digit DRB1 alleles, and SE (shared epitope) alleles separately. **B.** Here we plot the proportion of imputed markers with info scores <0.9 (see Supplementary Table 4) as a function of the number of SNPs in each of the six cohorts used to impute markers across the MHC (see Supplementary Table 1). The proportion of imputed markers with lower quality performance metrics is lower when more genotyped SNPs are available for imputation.



Supplementary Figure 2. Comparison of classical *HLA-DRB1* allele effect sizes from our imputation-based study to those from a recently published study. The odds ratios (and 95% confidence intervals) for imputed classical *HLA-DRB1* alleles in our study of 19,992 individuals are plotted as a function of the reported odds ratios (and 95% confidence intervals), taken from a recent association analysis in anti-CCP positive RA cases and controls by van der Woude *et al*[1]. Certain *DRB1* genotypes are grouped together in this figure (ie *13:01 and *13:02) to be consistent with previous reporting in the separate study.

Supplementary Table 1. Overview of genome-wide SNP data sets used in our analysis. We used samples from a recent rheumatoid arthritis meta-analysis of six GWAS data sets[2].

Case-Control Collection	Geographical Origin	Case Antibody Status	# Cases	# Controls	Genotyping Platform (# post-QC SNPs)	# Genotyped SNPs used in MHC Imputation	Case-Control Stratification Correction
Brigham Rheumatoid Arthritis Sequential Study (BRASS)	Boston, USA	100% CCP+	483	1631	Affymetrix 6.0 (682K)	401	Samples matched with PC covariates
CANADA	Toronto, Canada	100% CCP+	589	1554	Illumina 370K (306K)	645	Samples matched with PC covariates
Epidemiological Investigation of Rheumatoid Arthritis (EIRA) *	Sweden	100% CCP+	1122	1060	Illumina 317K (299K)	653	Epidemiologically matched
North American Rheumatoid Arthritis Consortium (NARAC) I	North America	100% CCP+	868	1194	Illumina 550K (503K)	674	Samples matched with PC covariates
NARAC III	North America	100% CCP+	902	6613	Illumina 317K (264K)	520	Samples matched with PC covariates
Wellcome Trust Case Control Consortium (WTCCC)	United Kingdom	100% CCP+	1054	2922	Affymetrix 500K (346K)	219	Geographically matched

Supplementary Table 2. Imputation accuracy. A Classical *HLA-DRB1* alleles. We imputed classical *HLA-DRB1* alleles at 2-digit (upper rows) and 4-digit (lower rows) resolution in two separate cohorts (WTCCC and NARAC) for samples successfully typed to 4-digit resolution at both alleles. In both cohorts we estimated the genotyped allele frequency and the imputed allele frequency; for accurate imputations these values should be equal. We also calculated the Pearson's correlation coefficient (*r*) between genotyped and impute allele dosage in individual; accurate imputations will result in a correlation of 1. Finally, for each allele we calculate and allelic accuracy; the allelic accuracy is the aggregate difference between the actual number of alleles observed and the imputed number of alleles observed. A perfect imputation achieves 100% accuracy. **B. Imputation accuracy for classical *HLA* alleles.** We imputed classical alleles in samples from the WTCCC, and compared to genotyped alleles at *HLA-A*, *HLA-B*, *HLA-C*, *HLA-DQB1*, and *HLA-DRB1* loci. For each locus we calculate and allelic accuracy at two and four digit resolution for samples successfully typed to four digit resolution at both alleles; the allelic accuracy is the aggregate difference between the actual number of alleles observed and the imputed number of alleles observed. **C. Imputation accuracy for individual classical *HLA* alleles.** We imputed classical alleles in samples from the WTCCC, and compared to genotyped alleles at *HLA-A*, *-B*, *-C*, *-DQB1*, and *-DRB1* for samples successfully typed to four-digit resolution at both alleles. Data for *HLA-DRB1* is presented in Supplementary Table 3A. Number of individuals for each locus is presented in Supplementary Table 3B. We estimated the genotyped allele frequency (g) and the imputed allele frequency (i). We also calculated the Pearson's correlation coefficient (r) between genotyped and impute allele dosage in individual. Finally, for each allele we calculate and allelic accuracy (acc).

Supplementary Table 2A

Resolution	Alleles	1958 British Birth Cohort Data (WTCCC, *n*=953)				North American Rheumatoid Arthritis Consortium (NARAC, *n*=450)			
		Frequency		*r*	Accuracy	Frequency		*r*	Accuracy
		Genotyped	Imputed			Genotyped	Imputed		
2 Digit *DRB1* Classical Alleles	*01	7.3%	7.4%	1.00	99.8%	11.1%	11.8%	0.97	99.3%
	*03	18.4%	18.4%	0.98	99.0%	5.3%	5.5%	0.97	99.7%
	*04	22.9%	23.6%	0.96	97.7%	64.8%	64.3%	0.99	99.5%
	*07	18.2%	17.7%	0.97	98.8%	2.7%	2.7%	1.00	100.0%
	*08	2.1%	2.1%	0.95	99.7%	0.1%	0.1%	1.00	100.0%
	*09	1.8%	1.7%	0.85	99.0%	0.4%	0.4%	1.00	100.0%
	*10	0.6%	0.6%	1.00	100.0%	3.3%	2.8%	0.91	99.4%
	*11	5.8%	5.8%	0.90	98.2%	1.7%	1.9%	0.84	99.3%
	*12	1.7%	1.8%	0.95	99.7%	0.2%	0.3%	0.82	99.9%
	*13	8.4%	8.4%	0.96	98.7%	2.7%	2.3%	0.89	99.4%
	*14	2.8%	2.5%	0.88	99.2%	1.3%	1.4%	1.00	100.0%
	*15	9.1%	9.2%	1.00	99.7%	5.1%	5.3%	0.98	99.8%
	*16	0.6%	0.7%	0.79	99.6%	1.1%	1.2%	0.99	99.9%
	Overall				94.6%				98.2%
4 Digit *DRB1* Classical Alleles	*01:01	4.3%	6.4%	0.82	97.5%	8.6%	9.9%	0.92	97.7%
	*01:02	0.7%	0.5%	0.96	99.7%	1.1%	1.4%	0.77	99.1%
	*01:03	2.3%	0.5%	0.71	97.8%	1.3%	0.5%	0.61	98.5%

*03:01	18.4%	18.4%	0.98	99.0%	5.2%	5.4%	0.96	99.5%
*04:01	13.5%	15.0%	0.85	95.5%	38.1%	40.9%	0.90	93.4%
*04:02	0.4%	0.5%	0.44	99.4%	1.0%	1.4%	0.89	99.5%
*04:03	2.3%	0.5%	0.22	97.4%	1.6%	1.2%	0.32	97.7%
*04:04	5.1%	4.6%	0.75	97.0%	17.3%	12.7%	0.78	92.8%
*04:05	0.8%	1.0%	0.85	99.5%	1.8%	2.4%	0.87	99.2%
*04:07	0.9%	1.5%	0.58	98.8%	2.8%	2.0%	0.69	97.8%
*04:08	0.0%	0.4%	0	99.6%	1.8%	3.7%	0.39	96.5%
*07:01	18.2%	17.7%	0.97	98.8%	2.6%	2.7%	0.98	99.9%
*08:01	1.7%	1.4%	0.97	99.5%	0.0%	0.1%	0	99.9%
*08:04	0.3%	0.2%	0.62	99.8%	0.0%	0.0%	1	100.0%
*09:01	1.8%	1.7%	0.85	99.0%	0.4%	0.4%	1.00	100.0%
*10:01	0.6%	0.6%	1.00	100.0%	2.8%	2.8%	1.00	100.0%
*11:01	3.7%	3.8%	0.82	97.6%	0.8%	1.4%	0.58	98.9%
*11:02	0.5%	0.3%	0.78	99.8%	0.1%	0.0%	0.22	99.9%
*11:03	0.6%	0.4%	0.34	99.2%	0.1%	0.1%	0.98	100.0%
*11:04	1.1%	1.3%	0.60	98.6%	0.7%	0.4%	0.64	99.3%
*12:01	1.5%	1.8%	0.88	99.5%	0.2%	0.3%	0.82	99.9%
*13:01	4.1%	4.3%	0.93	98.9%	1.9%	1.9%	0.94	99.7%
*13:02	3.0%	1.9%	0.86	98.8%	0.0%	0.0%	0	100.0%
*13:03	1.1%	1.0%	0.93	99.7%	0.3%	0.3%	1.00	99.9%
*14:01	2.8%	2.2%	0.88	99.0%	1.1%	1.2%	0.91	99.6%
*14:04	0.0%	0.3%	0	99.7%	0.0%	0.2%	0	99.8%
*15:01	8.7%	8.7%	1.00	99.8%	4.6%	4.8%	0.98	99.8%
*15:02	0.5%	0.4%	0.95	99.9%	0.6%	0.5%	1.00	99.9%
*16:01	0.6%	0.7%	0.79	99.6%	1.0%	1.0%	0.96	99.8%
Overall				83.6%				85.0%

Supplementary Table 2B

		2 digit resolution		4 digit resolution	
Locus	#Samples	#Imputed Alleles	Accuracy	#Imputed Alleles	Accuracy
HLA-A	741	15	97.0%	24	94.0%
HLA-B	1289	26	95.0%	41	87.6%
HLA-C	908	14	95.8%	23	92.6%
HLA-DQB1	863	5	95.3%	15	90.0%
HLA-DRB1	953	13	94.6%	29	83.6%

Supplementary Table 2C

	frequency					frequency					frequency					frequency			
Allele	g	i	*r*	acc	Allele	g	i	*r*	acc	Allele	g	i	*r*	acc	Allele	g	i	r	acc
HLA_A*01	27.2%	26.9%	0.98	99.1%	HLA_B*13	1.8%	1.8%	0.97	99.9%	HLA_B*3508	0.2%	0.4%	0.65	99.5%	HLA_C*0303	6.5%	6.5%	0.90	98.5%
HLA_A*02	2.5%	1.8%	0.83	99.1%	HLA_B*14	4.1%	4.2%	0.98	99.8%	HLA_B*3701	1.8%	1.7%	0.95	99.8%	HLA_C*0304	10.4%	11.0%	0.93	98.0%
HLA_A*03	19.9%	20.2%	0.99	99.6%	HLA_B*15	7.6%	7.9%	0.96	99.1%	HLA_B*3801	0.8%	0.7%	0.86	99.6%	HLA_C*0401	10.4%	10.2%	0.99	99.8%
HLA_A*11	8.5%	8.6%	0.98	99.8%	HLA_B*18	3.5%	3.4%	0.98	99.7%	HLA_B*3901	0.8%	0.9%	0.75	99.3%	HLA_C*0501	13.2%	12.7%	0.97	98.9%
HLA_A*23	2.1%	2.2%	0.97	99.9%	HLA_B*27	4.7%	4.8%	0.96	99.4%	HLA_B*3906	0.4%	0.7%	0.85	99.5%	HLA_C*0602	12.7%	12.9%	0.98	99.5%
HLA_A*24	10.4%	10.5%	0.99	99.7%	HLA_B*35	4.9%	4.7%	0.93	98.9%	HLA_B*4001	5.9%	6.5%	0.98	99.2%	HLA_C*0701	1.2%	0.5%	0.65	99.1%
HLA_A*25	2.2%	2.0%	0.92	99.4%	HLA_B*37	1.8%	1.7%	0.95	99.8%	HLA_B*4002	0.8%	0.7%	0.92	99.8%	HLA_C*0702	14.0%	15.6%	0.96	97.8%
HLA_A*26	2.6%	2.8%	0.91	99.1%	HLA_B*38	0.8%	0.7%	0.86	99.6%	HLA_B*4101	0.4%	0.3%	0.76	99.8%	HLA_C*0704	1.9%	2.0%	0.91	99.5%
HLA_A*29	6.1%	6.1%	0.99	99.8%	HLA_B*39	1.7%	1.6%	0.92	99.4%	HLA_B*4102	0.3%	0.3%	1.00	100.0%	HLA_C*0801	0.0%	0.0%	0	100.0%
HLA_A*30	3.1%	3.3%	0.96	99.7%	HLA_B*40	7.2%	7.1%	0.97	99.5%	HLA_B*4402	10.3%	11.6%	0.98	98.4%	HLA_C*0802	4.8%	5.6%	0.95	99.2%
HLA_A*31	3.7%	3.4%	0.97	99.7%	HLA_B*41	0.7%	0.6%	0.87	99.8%	HLA_B*4403	5.3%	5.5%	0.96	99.2%	HLA_C*1202	0.9%	1.0%	0.95	99.9%
HLA_A*32	5.4%	5.3%	0.99	99.9%	HLA_B*44	17.1%	17.3%	0.98	99.2%	HLA_B*4405	0.1%	0.0%	0.71	99.9%	HLA_C*1203	3.5%	3.5%	0.95	99.6%
HLA_A*33	0.1%	0.1%	1.00	100.0%	HLA_B*45	0.7%	0.9%	0.87	99.7%	HLA_B*4501	0.7%	0.9%	0.87	99.7%	HLA_C*1402	1.0%	1.0%	0.88	99.6%
HLA_A*66	0.3%	0.6%	0.73	99.7%	HLA_B*47	0.4%	0.4%	0.95	99.9%	HLA_B*4701	0.4%	0.4%	0.95	99.9%	HLA_C*1502	2.2%	2.2%	0.92	99.6%
HLA_A*68	6.0%	6.1%	0.96	99.6%	HLA_B*49	1.3%	1.2%	0.96	99.8%	HLA_B*4901	1.3%	1.2%	0.96	99.8%	HLA_C*1504	0.0%	0.0%	0	100.0%
HLA_A*0101	27.2%	26.9%	0.98	99.1%	HLA_B*50	1.0%	0.9%	0.88	99.7%	HLA_B*5001	0.9%	0.9%	0.88	99.7%	HLA_C*1505	0.2%	0.2%	0.82	99.9%
HLA_A*0201	0.7%	0.0%	0	99.3%	HLA_B*51	3.1%	3.4%	0.91	99.2%	HLA_B*5101	3.1%	3.4%	0.91	99.1%	HLA_C*1601	4.6%	4.8%	0.98	99.7%
HLA_A*0202	0.0%	0.1%	0.83	99.9%	HLA_B*52	0.8%	0.8%	0.95	99.9%	HLA_B*5201	0.8%	0.8%	0.95	99.9%	HLA_C*1602	0.4%	0.3%	0.85	99.8%
HLA_A*0205	1.6%	1.6%	0.98	99.9%	HLA_B*53	0.2%	0.2%	0.87	99.9%	HLA_B*5301	0.2%	0.2%	0.87	99.9%	HLA_C*1604	0.1%	0.0%	0	99.9%
HLA_A*0206	0.1%	0.1%	1.00	100.0%	HLA_B*55	1.9%	2.0%	0.99	99.9%	HLA_B*5501	1.9%	2.0%	0.99	99.9%	HLA_C*1701	0.6%	0.7%	0.96	99.9%
HLA_A*0301	18.2%	20.1%	0.99	98.0%	HLA_B*56	0.5%	0.5%	0.91	99.9%	HLA_B*5601	0.5%	0.5%	0.91	99.9%	HLA_DQB1*02	5.5%	3.7%	0.85	97.8%
HLA_A*0302	0.3%	0.1%	0.72	99.8%	HLA_B*57	4.5%	4.4%	0.96	99.6%	HLA_B*5701	4.5%	4.4%	0.96	99.6%	HLA_DQB1*03	44.3%	45.0%	0.96	97.3%
HLA_A*1101	8.1%	8.5%	0.98	99.4%	HLA_B*58	0.6%	0.5%	0.90	99.9%	HLA_B*5801	0.6%	0.5%	0.90	99.9%	HLA_DQB1*04	3.1%	3.1%	0.99	99.9%
HLA_A*2301	2.1%	2.2%	0.97	99.9%	HLA_B*73	0.0%	0.0%	0	100.0%	HLA_B*7301	0.0%	0.0%	0	100.0%	HLA_DQB1*05	18.7%	18.8%	0.95	97.5%
HLA_A*2402	9.9%	10.4%	0.98	99.0%	HLA_B*0702	13.2%	14.2%	0.98	98.7%	HLA_C*01	5.0%	5.1%	0.98	99.7%	HLA_DQB1*06	28.5%	29.4%	0.97	98.0%
HLA_A*2403	0.2%	0.1%	0.11	99.7%	HLA_B*0705	0.1%	0.0%	0	99.9%	HLA_C*02	4.0%	4.2%	0.90	99.4%	HLA_DQB1*0201	5.5%	2.0%	0.66	96.2%
HLA_A*2501	2.2%	2.0%	0.92	99.4%	HLA_B*0801	12.5%	14.9%	0.98	97.2%	HLA_C*03	17.1%	17.6%	0.95	98.6%	HLA_DQB1*0301	23.7%	25.7%	0.92	95.0%
HLA_A*2601	2.6%	2.6%	0.87	99.0%	HLA_B*1302	1.8%	1.8%	0.97	99.9%	HLA_C*04	10.4%	10.2%	0.99	99.8%	HLA_DQB1*0302	13.5%	12.1%	0.90	96.3%
HLA_A*2901	0.5%	0.0%	0	99.5%	HLA_B*1401	1.5%	1.6%	0.97	99.9%	HLA_C*05	13.2%	12.7%	0.97	98.9%	HLA_DQB1*0303	6.9%	7.1%	0.92	98.1%
HLA_A*2902	5.6%	6.1%	0.99	99.3%	HLA_B*1402	2.7%	2.6%	0.99	99.9%	HLA_C*06	12.7%	12.9%	0.98	99.5%	HLA_DQB1*0304	0.2%	0.1%	0.07	99.8%
HLA_A*3001	2.0%	2.0%	1.00	100.0%	HLA_B*1501	7.0%	7.2%	0.97	99.2%	HLA_C*07	19.0%	18.1%	0.94	98.4%	HLA_DQB1*0402	3.0%	3.1%	0.99	99.9%
HLA_A*3002	1.0%	1.2%	0.89	99.7%	HLA_B*1503	0.0%	0.1%	0.20	99.9%	HLA_C*08	4.8%	5.6%	0.95	99.2%	HLA_DQB1*0501	15.3%	14.8%	0.97	98.8%
HLA_A*3101	3.7%	3.4%	0.97	99.7%	HLA_B*1510	0.0%	0.0%	0	100.0%	HLA_C*10	0.0%	0.0%	0	100.0%	HLA_DQB1*0502	0.9%	0.9%	0.81	99.5%
HLA_A*3201	5.4%	5.3%	0.99	99.9%	HLA_B*1517	0.1%	0.2%	0.87	100.0%	HLA_C*12	4.4%	4.5%	0.96	99.6%	HLA_DQB1*0503	2.3%	3.0%	0.87	99.1%
HLA_A*3301	0.0%	0.0%	0	100.0%	HLA_B*1518	0.4%	0.3%	0.94	99.9%	HLA_C*14	1.0%	1.0%	0.88	99.6%	HLA_DQB1*0504	0.1%	0.1%	0.71	99.9%
HLA_A*3303	0.1%	0.1%	1.00	100.0%	HLA_B*1801	3.5%	3.4%	0.98	99.7%	HLA_C*15	2.4%	2.4%	0.93	99.6%	HLA_DQB1*0601	0.7%	0.6%	0.97	99.9%
HLA_A*6601	0.3%	0.6%	0.73	99.7%	HLA_B*2702	0.1%	0.1%	0.62	99.8%	HLA_C*16	5.2%	5.1%	0.98	99.6%	HLA_DQB1*0602	16.0%	16.8%	0.97	98.3%
HLA_A*6801	5.4%	5.5%	0.95	99.6%	HLA_B*2705	3.9%	4.7%	0.94	98.7%	HLA_C*17	0.6%	0.7%	0.96	99.9%	HLA_DQB1*0603	6.8%	6.8%	0.91	98.0%
HLA_A*6802	0.6%	0.6%	1.00	100.0%	HLA_B*3501	3.8%	3.5%	0.88	98.6%	HLA_C*0102	5.0%	5.1%	0.98	99.7%	HLA_DQB1*0604	3.8%	3.8%	0.98	99.8%
HLA_B*07	14.2%	14.2%	0.98	99.4%	HLA_B*3502	0.3%	0.2%	0.78	99.7%	HLA_C*0202	4.0%	4.2%	0.90	99.4%	HLA_DQB1*0609	1.1%	1.4%	0.94	99.7%
HLA_B*08	14.8%	14.9%	0.98	99.4%	HLA_B*3503	0.6%	0.5%	0.77	99.4%	HLA_C*0302	0.1%	0.1%	0.98	100.0%					

Supplementary Table 3. Estimated effects for imputed classical *HLA-DRB1* alleles in the European meta-analysis and Korean samples. For each classical *HLA-DRB1* allele we list the odds ratio of RA risk conferred by its presence with the 95% confidence interval and association test statistics. Allele frequencies in cases and controls are also listed. Results are based on imputed alleles in European samples and genotyped alleles in Korean samples.

	European Metanalysis						Korean					
Classical *HLA-DRB1*	Univariate Analysis				Allele Frequency		Univariate Analysis				Allele Frequency	
Allele	Odds Ratio		χ^2	p	Controls	Cases	Odds Ratio		χ^2	p	Controls	Cases
HLA-DRB1*01	1.3	(1.21 - 1.40)	49.4	2.10E-12	11.30%	14.50%	1.16	(0.87 - 1.56)	1	3.20E-01	7.20%	8.20%
HLA-DRB1*01:01	1.38	(1.28 - 1.50)	63.3	1.80E-15	9.70%	13.30%	1.16	(0.87 - 1.56)	1	3.20E-01	7.20%	8.20%
HLA-DRB1*01:02	0.93	(0.66 - 1.31)	0.2	6.70E-01	1.00%	0.60%						
HLA-DRB1*01:03	0.52	(0.26 - 1.03)	3.9	4.80E-02	0.60%	0.50%						
HLA-DRB1*03	0.59	(0.54 - 0.64)	158.5	2.40E-36	12.80%	8.20%	0.29	(0.13 - 0.61)	11.6	6.70E-04	2.20%	0.60%
HLA-DRB1*03:01	0.59	(0.54 - 0.64)	158.5	2.40E-36	12.80%	8.20%	0.29	(0.13 - 0.61)	11.6	6.70E-04	2.20%	0.60%
HLA-DRB1*04	3.71	(3.49 - 3.93)	2133.1	10E-464	17.40%	45.00%	2.48	(2.05 - 3.01)	95.1	1.80E-22	19.00%	35.70%
HLA-DRB1*04:01	4.14	(3.86 - 4.44)	1677.5	10E-366	10.40%	30.90%	2.73	(1.29 - 6.30)	7	8.30E-03	0.70%	1.80%
HLA-DRB1*04:02	0.97	(0.68 - 1.38)	0	8.60E-01	1.10%	0.60%						
HLA-DRB1*04:03	6.24	(3.09 - 12.63)	25	5.70E-07	0.40%	0.60%	0.61	(0.36 - 1.01)	3.7	5.50E-02	3.00%	1.90%
HLA-DRB1*04:04	3.17	(2.83 - 3.54)	397.7	1.70E-88	3.60%	9.10%	2.74	(1.33 - 6.05)	7.7	5.70E-03	0.70%	1.90%
HLA-DRB1*04:05	2.31	(1.77 - 3.01)	36	1.90E-09	0.70%	1.20%	3.93	(3.09 - 5.05)	140	2.60E-32	8.40%	25.20%
HLA-DRB1*04:06							0.55	(0.36 - 0.85)	7.5	6.20E-03	4.70%	2.80%
HLA-DRB1*04:07	0.94	(0.65 - 1.37)	0.1	7.60E-01	0.70%	0.70%	0.13	(0.01 - 0.73)	5.7	1.70E-02	0.60%	0.10%
HLA-DRB1*04:08	5.48	(4.11 - 7.30)	139.5	3.40E-32	0.50%	1.70%	1.12	(0.13 - 9.41)	0	9.10E-01	0.10%	0.20%
HLA-DRB1*04:10							2.87	(1.36 - 6.60)	7.9	5.00E-03	0.70%	1.90%
HLA-DRB1*07	0.49	(0.45 - 0.54)	257.3	6.70E-58	13.30%	6.40%	0.48	(0.33 - 0.70)	15.1	1.00E-04	6.70%	3.40%
HLA-DRB1*07:01	0.49	(0.45 - 0.54)	257.3	6.70E-58	13.30%	6.40%	0.48	(0.33 - 0.70)	15.1	1.00E-04	6.70%	3.40%
HLA-DRB1*08	0.41	(0.34 - 0.50)	91.6	1.10E-21	2.90%	1.30%	0.62	(0.46 - 0.83)	10.3	1.30E-03	9.40%	6.10%
HLA-DRB1*08:01	0.34	(0.26 - 0.44)	72.3	1.90E-17	1.90%	0.90%						
HLA-DRB1*08:02							0.36	(0.18 - 0.68)	10.4	1.30E-03	2.60%	1.00%
HLA-DRB1*08:03							0.73	(0.52 - 1.02)	3.5	6.20E-02	6.80%	5.10%
HLA-DRB1*08:04	0.18	(0.05 - 0.69)	9.2	2.40E-03	0.20%	0.00%						
HLA-DRB1*09	1.01	(0.80 - 1.28)	0	9.30E-01	1.10%	1.30%	1.49	(1.17 - 1.89)	10.6	1.10E-03	9.70%	14.00%
HLA-DRB1*09:01	1.01	(0.80 - 1.28)	0	9.30E-01	1.10%	1.30%	1.49	(1.17 - 1.89)	10.6	1.10E-03	9.70%	14.00%
HLA-DRB1*10	2.53	(2.04 - 3.14)	69.6	7.40E-17	0.80%	2.00%	1.9	(1.16 - 3.18)	6.5	1.10E-02	1.90%	3.40%
HLA-DRB1*10:01	2.53	(2.04 - 3.14)	69.6	7.40E-17	0.80%	2.00%	2.11	(1.26 - 3.60)	8.2	4.10E-03	1.70%	3.40%
HLA-DRB1*11	0.48	(0.43 - 0.54)	158.4	2.60E-36	9.40%	3.90%	0.49	(0.31 - 0.75)	11	9.00E-04	5.00%	2.50%
HLA-DRB1*11:01	0.44	(0.38 - 0.52)	107.8	3.00E-25	6.10%	2.80%	0.48	(0.30 - 0.74)	11	9.00E-04	4.70%	2.40%
HLA-DRB1*11:02	0.19	(0.06 - 0.61)	11.5	7.00E-04	0.20%	0.10%						
HLA-DRB1*11:03	0.13	(0.06 - 0.29)	32.5	1.20E-08	0.60%	0.20%						
HLA-DRB1*11:04	0.15	(0.10 - 0.23)	105.2	1.10E-24	2.40%	0.80%						
HLA-DRB1*11:06							2.28	(0.22 - 49.2)	0.5	4.90E-01	0.10%	0.20%
HLA-DRB1*12	0.76	(0.61 - 0.94)	6.8	8.90E-03	1.70%	1.40%	0.75	(0.54 - 1.02)	3.3	6.80E-02	7.80%	6.00%
HLA-DRB1*12:01	0.76	(0.61 - 0.94)	6.8	8.90E-03	1.70%	1.40%	0.86	(0.57 - 1.27)	0.6	4.40E-01	4.50%	3.90%
HLA-DRB1*12:02							0.62	(0.37 - 1.02)	3.5	6.00E-02	3.30%	2.10%
HLA-DRB1*13	0.33	(0.30 - 0.37)	465	4.00E-103	11.40%	4.40%	0.46	(0.34 - 0.63)	24.7	6.80E-07	10.10%	5.00%
HLA-DRB1*13:01	0.28	(0.24 - 0.33)	302	1.20E-67	6.10%	2.10%	0.32	(0.10 - 0.81)	5.9	1.50E-02	1.30%	0.40%
HLA-DRB1*13:02	0.29	(0.23 - 0.38)	114.7	9.00E-27	2.70%	1.20%	0.48	(0.34 - 0.66)	20.3	6.70E-06	8.80%	4.50%
HLA-DRB1*13:03	0.57	(0.41 - 0.79)	12.3	4.40E-04	1.00%	0.50%						
HLA-DRB1*14	0.5	(0.40 - 0.62)	46.6	8.50E-12	2.50%	1.20%	0.52	(0.37 - 0.73)	15.6	7.70E-05	8.40%	4.50%
HLA-DRB1*14:01	0.46	(0.36 - 0.59)	45.9	1.30E-11	2.20%	1.10%	0.5	(0.28 - 0.86)	6.4	1.10E-02	3.00%	1.50%
HLA-DRB1*14:02							2.19	(0.21 - 47.2)	0.4	5.10E-01	0.10%	0.20%
HLA-DRB1*14:03							0.27	(0.08 - 0.73)	6.8	8.90E-03	1.20%	0.30%
HLA-DRB1*14:04	0	(0.00 - 0.02)	43.1	5.20E-11	0.30%	0.10%						
HLA-DRB1*14:05							0.48	(0.27 - 0.82)	7.5	6.30E-03	3.20%	1.50%
HLA-DRB1*14:06							0.7	(0.21 - 2.13)	0.4	5.40E-01	0.60%	0.40%
HLA-DRB1*14:07							0.65	(0.13 - 2.69)	0.3	5.60E-01	0.40%	0.20%
HLA-DRB1*15	0.59	(0.54 - 0.64)	179.2	7.20E-41	14.20%	9.20%	0.79	(0.61 - 1.01)	3.5	6.30E-02	11.70%	9.50%
HLA-DRB1*15:01	0.57	(0.53 - 0.62)	187.4	1.20E-42	13.60%	8.90%	0.61	(0.45 - 0.82)	10.5	1.20E-03	9.20%	5.80%
HLA-DRB1*15:02	1.13	(0.71 - 1.81)	0.3	6.10E-01	0.50%	0.30%	1.39	(0.90 - 2.19)	2.2	1.40E-01	2.50%	3.60%
HLA-DRB1*16	1.26	(1.00 - 1.59)	3.7	5.50E-02	1.30%	1.20%	1.19	(0.55 - 2.58)	0.2	6.60E-01	1.00%	1.10%
HLA-DRB1*16:01	1.28	(1.01 - 1.62)	4	4.50E-02	1.20%	1.10%						
HLA-DRB1*16:02							1.19	(0.55 - 2.58)	0.2	6.60E-01	1.00%	1.10%

Supplementary Table 5A. Forward conditional haplotype analysis on individual HLA-DRβ1 amino acid residues in the European RA meta-analysis. For each amino acid position in HLA-DRβ1 (column 1), we partitioned the classical alleles into the groups defined by different amino acid residues at that position. In addition to testing amino acid positions within the protein structure, we also tested amino acids in the leader peptide (negative numbers in column 1). Then we assess the omnibus association to RA risk across those haplotypes, where the degree of freedom (df) is determined by the number of defined haplotype groupings (minus 1), and calculate a p-value. Then we conditioned on specific amino acid positions (11 or 13, and 11+71, and 11+71+74), and test if further partitioning classical alleles on other amino acids results in additional strength of association to RA risk. Grey boxes indicate that positions were no longer informative (variable). **B. Improvement in model fit as HLA-DRβ1 amino acids are added.** Following the analysis above in Supplementary Table 6A, we sequentially added amino acids to the model. Adding amino acids sequentially improves the fit by increasing the number of parameters in the model (above the baseline 35 principal component and indicator variables). To assess the improvement in the model fit we calculated the improvement in deviance (-2*log likelihood, column 3, ∆deviance) over the baseline, and then by adding additional parameters. To assess the improvement in the model fit accounting for the increasing number of parameters we also calculated the improvement in the Akaike information criterion (∆AIC, column 4), and also the improvement in the Bayesian information criterion (∆BIC, column 5). Adding the first three amino acids results in clear improvement in model fit in each of the three metrics. But, adding the fourth amino acid does not result in a clear improvement in the model, after accounting for the additional model parameters.

Supplementary Table 5A

Condition:	On Nothing			On Position 11			On Position 13			On Positions 11 and 71			On Positions 11, 71 and 74		
Amino Acid Position	χ^2	df	$\log_{10}$(p)	χ^2	df	p	χ^2	df	p	χ^2	df	p	χ^2	df	p
-25	878.5	1	-192.3	0.1	2	0.95	38.7	3	2.0E-08	5.8	3	0.12	3.0	2	0.23
-24	2111.6	1	-460.3	0.0	1	0.87				1.0	1	0.33	0.0	1	0.99
-17	19.1	1	-4.9				6.0	1	1.4E-02						
-16	878.5	1	-192.3	0.1	2	0.95	38.7	3	2.0E-08	5.8	3	0.12	3.0	2	0.23
-1	183.2	1	-41.0	25.9	1	3.5E-07	25.6	1	4.2E-07						
4	232.0	1	-51.7				5.2	1	2.2E-02						
9	195.5	2	-42.5				34.4	2	3.5E-08						
10	1121.2	2	-243.5	0.0	1	0.87	31.2	1	2.4E-08	1.0	1	0.33	0.0	1	0.99
11	2698.4	5	-581.4				34.4	2	3.5E-08						
12	1085.4	1	-237.3												
13	2665.2	5	-574.2	1.1	2	0.57				1.9	2	0.38	9.6	2	8.3E-03
14	259.1	1	-57.6												
16	81.2	1	-18.7	1.1	1	0.30				1.0	1	0.32	9.6	1	2.0E-03
25	259.1	1	-57.6												
26	176.9	2	-38.4	59.6	3	7.1E-13	63.9	3	8.5E-14	12.9	3	4.9E-03	9.6	2	8.3E-03
28	19.0	2	-4.1	11.6	2	3.1E-03	17.7	2	1.4E-04	12.9	2	1.6E-03	9.6	2	8.3E-03
30	355.8	5	-74.0	11.6	2	3.1E-03	46.9	3	3.7E-10	12.9	2	1.6E-03	9.6	2	8.3E-03
31	120.2	2	-26.1	0.0	1	0.87	31.2	1	2.4E-08	1.0	1	0.33	0.0	1	0.99
32	509.9	1	-112.2	0.8	2	0.67	43.7	3	1.8E-09	9.8	4	4.4E-02	11.8	3	7.9E-03
33	2111.6	1	-460.3	0.0	1	0.87				1.0	1	0.33	0.0	1	0.99

37	1458.3	4	-313.8	11.8	3	8.2E-03	47.0	5	5.7E-09	14.5	4	6.0E-03	11.8	2	2.7E-03
38	75.2	2	-16.3	11.6	2	3.1E-03	43.6	2	3.4E-10	12.9	2	1.6E-03	9.6	2	8.3E-03
40	70.0	1	-16.2	0.0	1	0.87	31.2	1	2.4E-08	1.0	1	0.33	0.0	1	0.99
47	1217.4	1	-266.0	26.1	2	2.1E-06	38.9	3	1.8E-08	2.7	2	0.26			
57	333.3	3	-71.2	13.4	4	9.5E-03	26.2	6	2.1E-04	13.0	5	2.3E-02	16.7	3	8.1E-04
58	156.1	1	-35.1	0.2	1	0.62	0.1	1	0.70	3.0	2	0.22	11.8	2	2.7E-03
60	287.7	2	-62.5	11.7	2	2.9E-03	17.8	3	4.8E-04	12.2	2	2.2E-03	9.6	1	2.0E-03
67	1579.2	2	-342.9	140.7	6	7.2E-28	173.2	8	2.8E-33	12.4	5	3.0E-02	14.9	4	5.0E-03
70	1140.8	2	-247.7	137.2	6	3.9E-27	151.6	6	3.6E-30	1.0	3	0.81	0.0	1	0.99
71	950.3	3	-205.0	188.4	6	5.6E-38	189.0	6	4.1E-38						
73	456.2	1	-100.5	38.4	1	5.7E-10	45.9	1	1.2E-11	0.0	1	0.94			
74	647.5	4	-138.1	104.6	4	1.0E-21	131.9	5	9.2E-27	56.6	4	1.5E-11			
77	151.3	1	-34.0	38.4	1	5.7E-10	45.9	1	1.2E-11	0.0	1	0.94			
78	232.0	1	-51.7				5.2	1	2.2E-02						
85	4.5	1	-1.5	17.5	2	1.6E-04	20.1	2	4.4E-05	20.6	2	3.3E-05	18.3	2	1.0E-04
86	459.7	1	-101.3	58.8	4	5.3E-12	68.2	5	2.4E-13	26.5	6	1.8E-04	34.1	7	1.7E-05
96	2550.8	3	-552.3	0.0	1	0.87	34.4	2	3.5E-08	1.0	1	0.33	0.0	1	0.99
98	1076.4	1	-235.4	0.0	1	0.87	5.2	1	2.2E-02	1.0	1	0.33	0.0	1	0.99
104	1076.4	1	-235.4	0.0	1	0.87	5.2	1	2.2E-02	1.0	1	0.33	0.0	1	0.99
112	43.0	1	-10.3	0.0	1	0.88	0.1	1	0.77	0.0	1	0.94	0.0	0	1.00
120	2237.4	1	-487.6	0.0	0	1.00	31.2	1	2.4E-08	0.0	0	1.00	0.0	0	1.00
133	152.7	1	-34.4	0.0	0	1.00	0.0	0	1.00	0.0	0	1.00	0.0	0	1.00
140	191.9	1	-42.9	1.6	1	0.20	32.8	2	7.6E-08	1.1	1	0.29	1.2	1	0.28
142	152.7	1	-34.4	0.0	0	1.00	0.0	0	1.00	0.0	0	1.00	0.0	0	1.00
149	1083.3	1	-236.9	1.6	1	0.20	1.5	1	0.23	1.1	1	0.29	1.2	1	0.28
166	70.0	1	-16.2	0.0	1	0.87	31.2	1	2.4E-08	1.0	1	0.33	0.0	1	0.99
180	2111.6	1	-460.3	0.0	1	0.87	0.0	0	1.00	1.0	1	0.33	0.0	1	0.99
181	136.8	1	-30.9	0.0	1	0.87	6.0	1	1.4E-02	1.0	1	0.33	0.0	1	0.99
189	4.8	1	-1.5	0.5	1	0.49	0.7	1	0.40	0.5	1	0.49	0.2	1	0.68
231	70.0	1	-16.2	0.0	1	0.87	31.2	1	2.4E-08	1.0	1	0.33	0.0	1	0.99
233	1077.0	1	-235.5	1.1	1	0.29	2.2	2	0.34	1.6	2	0.44	1.4	2	0.51

Supplementary Table 5B

	Parameters	∆deviance	∆AIC	∆BIC
DRB1-11	5	2698.4	2688.4	2648.9
DRB1-11,71	11	188.4	176.4	129.0
DRB1-11,71,74	15	56.6	48.6	17.0
DRB1-11,71,74,86	22	34.1	20.1	-35.2

Supplementary Table 6. Conditional haplotype analysis on individual DRβ1 amino acid residues in South Korean samples. We used classical *HLA-DRB1* allele genotype data in a data set of 616 anti-CCP positive RA cases and 675 controls. Then, as in Table S6, for each amino acid position in HLA-DRβ1 (column 1), we partitioned the classical alleles into the groups defined by different amino acid residues at that position. Then we assess the omnibus association to RA risk across those haplotypes, where the degree of freedom (df) is determined by the number of defined haplotype groupings (minus 1), and calculate a p-value for significance of association.

Amino Acid Position	χ^2	df	p
-25	80.4	1	3.1E-19
-24	94.2	1	2.9E-22
-17	0.5	1	4.6E-01
-16	80.4	1	3.1E-19
-1	3.5	1	6.3E-02
4	0.3	1	5.6E-01
9	14.7	2	6.3E-04
10	101.0	2	1.2E-22
11	175.0	5	6.1E-36
12	97.6	1	5.0E-23
13	176.4	5	3.1E-36
14	15.1	1	1.0E-04
16	22.7	2	1.2E-05
23	1.5	1	2.3E-01
25	15.1	1	1.0E-04
26	4.8	2	9.3E-02
28	12.1	2	2.3E-03
30	35.3	5	1.3E-06
31	20.5	2	3.5E-05
32	41.1	1	1.5E-10
33	95.1	1	1.8E-22
37	75.4	4	1.7E-15
38	11.0	2	4.0E-03
40	8.2	1	4.1E-03
47	56.2	1	6.5E-14
57	105.7	3	9.0E-23
58	11.0	1	9.0E-04
60	6.5	2	3.9E-02
67	63.0	2	2.1E-14
70	82.6	2	1.1E-18
71	33.7	3	2.3E-07
73	25.4	1	4.8E-07
74	53.6	4	6.5E-11
77	11.6	1	6.7E-04
78	0.3	1	5.6E-01
85	2.9	1	8.6E-02
86	39.0	1	4.2E-10
96	108.3	3	2.6E-23
98	84.3	1	4.2E-20
104	84.3	1	4.2E-20
112	6.4	1	1.1E-02
120	106.4	1	6.2E-25
133	2.7	1	1.0E-01
140	0.0	1	8.7E-01
142	2.7	1	1.0E-01
149	102.2	1	5.1E-24
166	8.2	1	4.1E-03
180	94.2	1	2.9E-22
181	2.8	1	9.2E-02
189	10.3	1	1.3E-03
231	8.2	1	4.1E-03

Supplementary Note

Population stratification. We previously addressed case-control stratification in each of the individual GWAS data sets (see **Supplementary Table 1**, Refs. [2-4] for details). For each data set, cases and controls were either recruited from the same geographic area (geographically matched), or individually matched based on principal components from genome-wide data. For the final association analyses in this study, we included five principal components for each of the individual data sets as covariates based on these SNPs. We calculated principal components with Eigenstrat[5]. In the European study we calculated components from a total of 296,756 genotyped or imputed SNPs with very high quality in all six data sets (info score >0.9 from ref [6]), and excluding the *PTPN22*, *MHC*, and all other regions with known association to RA. In the South Korean study we examined 441,398 genotyped markers. For all studied we excluded SNPs from three regions with known long-distance LD (1) the MHC region (chr6:26-36 Mbp) , (2) the chromosome 8 inversion (chr8:6-16 Mbp), and (3) a chromosome 17 region (chr17:40-45 Mbp). To assess the extent of population stratification we calculated single-SNP association statistics across the genome and the genomic inflation factor (λ_{gc}).

***Assessing imputation accuracy*.** In order to assess imputation accuracy of *HLA-DRB1* classical alleles we used classical *HLA-DRB1* allele genotype data from two separate sources: (1) 953 individuals from the 1958 British Birth Cohort, included as WTCCC controls and (2) 450 individuals from the NARAC-I data set consisting of both controls (n=175) and cases (n=275). We used only samples for which we had four digit resolution genotypes for both *DRB1* alleles. We compared these results to imputed classical *DRB1* alleles based on genome-wide SNP data as described.

We also assessed imputation accuracy at all of the other classical alleles in the 1958 British Birth Cohort samples, that were included as WTCCC controls. Again, we included only samples for which we had genotype data to 4-digit resolution for both alleles.

We estimated the quality of our imputation using three different figures of merit. First, we estimated and compared the frequency of the imputed allele versus the genotyped allele. Second, for each classical allele, we calculated the correlation between the genotyped allele dosages (which can be 0, 1, or 2 alleles) and the imputed probabilistic allele dosages (which can range continuously between 0 to 2) across all individuals. Third, for each classical allele, we calculated the accuracy, defined as

$$1 - \frac{\sum_i |g_i - x_i|}{2n}$$

where g_i is the genotyped allele dosage for individual i, x_i is the imputed allele dosage for individual i, and n is the total number of individuals in the sample set. We also calculated overall accuracy across all individuals (separately for alleles at 2- and 4-digit resolution), defined as the proportion of genotyped classical alleles per individual accurately called by imputation:

$$1 - \frac{\sum_i \sum_a \left[\delta\left(g_{i,a} > x_{i,a}\right)\right]\left(g_{i,a} - x_{i,a}\right)}{2n}$$

where $g_{i,a}$ is the dosage for genotyped allele a for individual i, $x_{i,a}$ is the dosage for the imputed allele a for individual i, and δ is a function that is 1 if the genotyped allele is greater than the imputed allele and otherwise 0.

Calculating epistasis. We tested for statistical interactions between each of 44 HLA alleles (42 2- and 4-digit alleles of HLA-DRB1, Asp-9 in B, and Phe-9 in DPB1) and each

of 35 previously validated RA risk common SNP associations. We tested for interaction using logistic regression dosages in all six RA risk GWAS datasets, using five indicator covariates and five principal components (PC) analysis eigenvectors for each dataset. Interaction between a given HLA allele and RA risk-associated SNP was tested by application of the likelihood ratio test of a model with covariates (see above), HLA allele and SNP, versus the same model plus an HLA allele x SNP interaction term. Since we conducted 1,540 tests altogether, we used a Bonferroni-corrected P-value of 0.05/1540 = 0.00003 for statistical significance.

Percent variance explained. As described previously[4], for each of the loci (*DRB1*, *B*, and *DPB1*) we calculated percent variance explained separately[7]. We used a model based on the biometrical model from Fisher[8] and the liability threshold model from Pearson and Lee[9]. We assume that RA risk is the consequence of an underlying liability score that is normally distributed with a mean of zero and a variance of one, and that individuals with a score above a pre-specified threshold get disease. We assume an additive genetic model, and also that genetic factors alter the liability score threshold. To investigate the percent variance explained by a single locus, we determine the prevalence of disease for individuals with different genotypes. We then determine the corresponding change in the liability score threshold for different genotypes. This distance between the thresholds for homozygotes is taken to be the change in the genotypic means, and thus we can calculate the variance attributable to the locus using Fisher's formula. For this calculation we assumed the prevalence of RA was 0.5%. For each SNP we assumed (1) population frequency was equal to average control frequency, and (2) effect size was the derived odds ratios from our study.

1. van der Woude, D. et al. Protection against anti-citrullinated protein antibody-positive rheumatoid arthritis is predominantly associated with HLA-DRB1*1301: a meta-analysis of HLA-DRB1 associations with anti-citrullinated protein antibody-positive and anti-citrullinated protein antibody-negative rheumatoid arthritis in four European populations. *Arthritis Rheum* **62**, 1236-45 (2010).
2. Stahl, E.A. et al. Genome-wide association study meta-analysis identifies seven new rheumatoid arthritis risk loci. *Nat Genet* **42**, 508-514 (2010).
3. Freudenberg, J. et al. Genome-wide association study of rheumatoid arthritis in Koreans: population-specific loci as well as overlap with European susceptibility loci. *Arthritis Rheum* **63**, 884-93 (2011).
4. Raychaudhuri, S. et al. Common variants at *CD40* and other loci confer risk of rheumatoid arthritis. *Nat Genet* **40**, 1216-23 (2008).
5. Price, A.L. et al. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* **38**, 904-9 (2006).
6. 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).
7. Kwan, S.H., Purcell, S. & Sham, P.C. Biometrical Genetics. in *Statistical Genetics: Genemapping through linkage and association* (eds. Neale, B.M., Ferreira, M.A.R., Medland, S.E. & Posthuma, D.) (Taylor & Francis, London, UK, 2007).
8. Fisher, R.A. The correlation between relatives on the supposition of Mendelian inheritance. *Transactions of the Royal Society of Edinburgh* **52**, 399-433 (1918).
9. Pearson, K. & Lee, A. On the inheritance of characteristics not capable of exact quantitative measurement. *Philosophical Transactions of the Royal Society of London Series A* **195**, 79 –150 (1901).