The American Journal of Human Genetics, Volume *94*

Supplemental Data

Fine Mapping Seronegative and Seropositive Rheumatoid Arthritis to Shared and Distinct HLA Alleles by Adjusting for the Effects of Heterogeneity

Buhm Han, Dorothée Diogo, Steve Eyre, Henrik Kallberg, Alexandra Zhernakova, John Bowes, Leonid Padyukov, Yukinori Okada, Miguel A. González-Gay, Solbritt Rantapää-Dahlqvist, Javier Martin, Tom W.J. Huizinga, Robert M. Plenge, Jane Worthington, Peter K. Gregersen, Lars Klareskog, Paul I.W. de Bakker, and Soumya Raychaudhuri

Supplemental Figures

Figure S1: Frequency of HLA-DRβ1 Val11 in seronegative RA cases. In the US and UK cohorts, we measured the seronegative case frequency of HLA-DRβ1 Val11, a major risk factor for seropositive RA, as we increased the level of stringency of anti-CCP cutoff (i.e. as we reduced the cut-off). As we reduced the cut-off from the default values (5.0 in UK and 20.0 in US), we observed decreasing trend in the Val-11 frequency (Spearman P=6.9×10^{-5}). This suggested that uncertainties in anti-CCP testing might have caused possible confounding from seropositive RA.

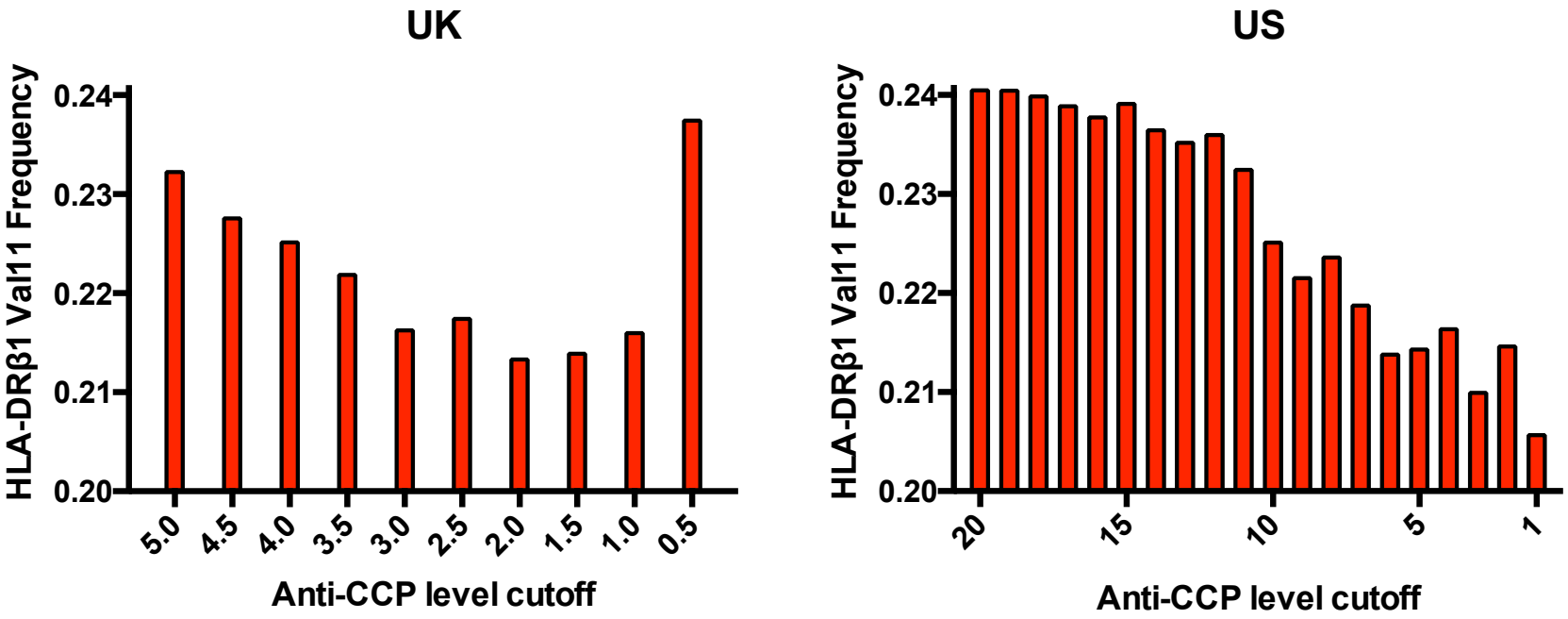


Figure S2: Effects of individual amino acids within HLA proteins on seronegative and seropositive RA. For each amino acid position, we listed the allele frequencies in cases (red) and controls (blue) along with univariate odds ratios (odds ratio of a residue taking the other residues as reference). Newly identified positions are in bold faces. For seronegative RA, we estimated odds ratios of Ser11+Leu11 of HLA-DRβ1 and Asp9 of HLA-B by conditioning on each other. For seropositive RA, we estimated odds ratios at each position by conditioning on previous positions in forward search; the effects of HLA-B are conditioned on the classical *HLA-DRB1* alleles, the effects of HLA-DPβ1 are conditioned on the *HLA-DRB1* alleles and HLA-B position 9, and the effects of HLA-A are conditioned on the *HLA-DRB1* alleles, HLA-B position 9, and HLA-DPβ1 position 9.

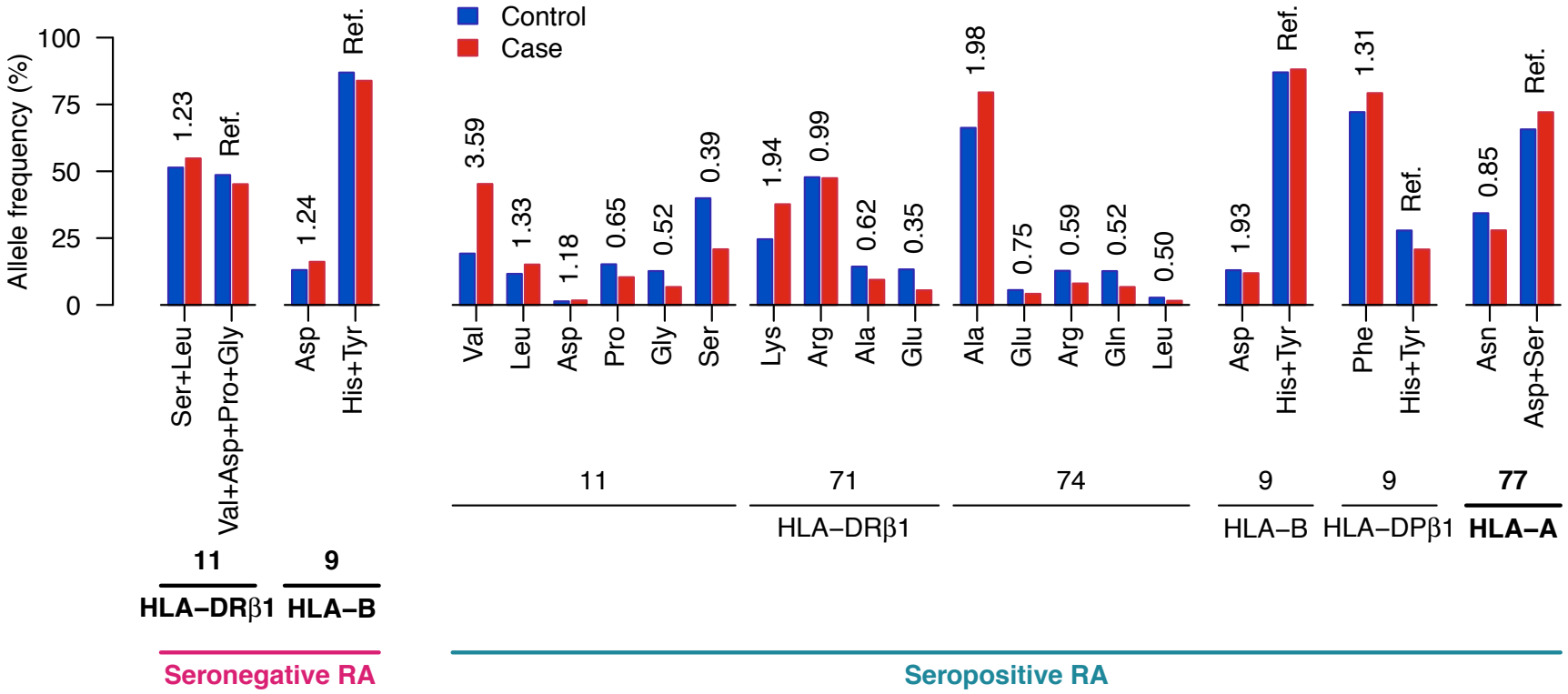


Figure S3: Comparison of confounding proportion estimates when using full GRS and non-overlapping loci GRS. For each cohort, we estimated the confounding proportion from ACPA+ RA (the proportion of samples that actually have ACPA+ RA) using genetic risk scores (GRS). First, we used GRS from the full list of known risk loci for ACPA+ RA (MHC loci in addition to 47 non-MHC loci). Then we used GRS from the selected list of loci that approximates non-overlapping loci between ACPA+ RA and ACPA- RA (38 loci that are not associated to ACPA- RA) (See Table S2). The mean estimate over the five cohorts were 26.3% for full GRS and 28.3% for non-overlapping loci GRS. Vertical lines denote 95% C.I..

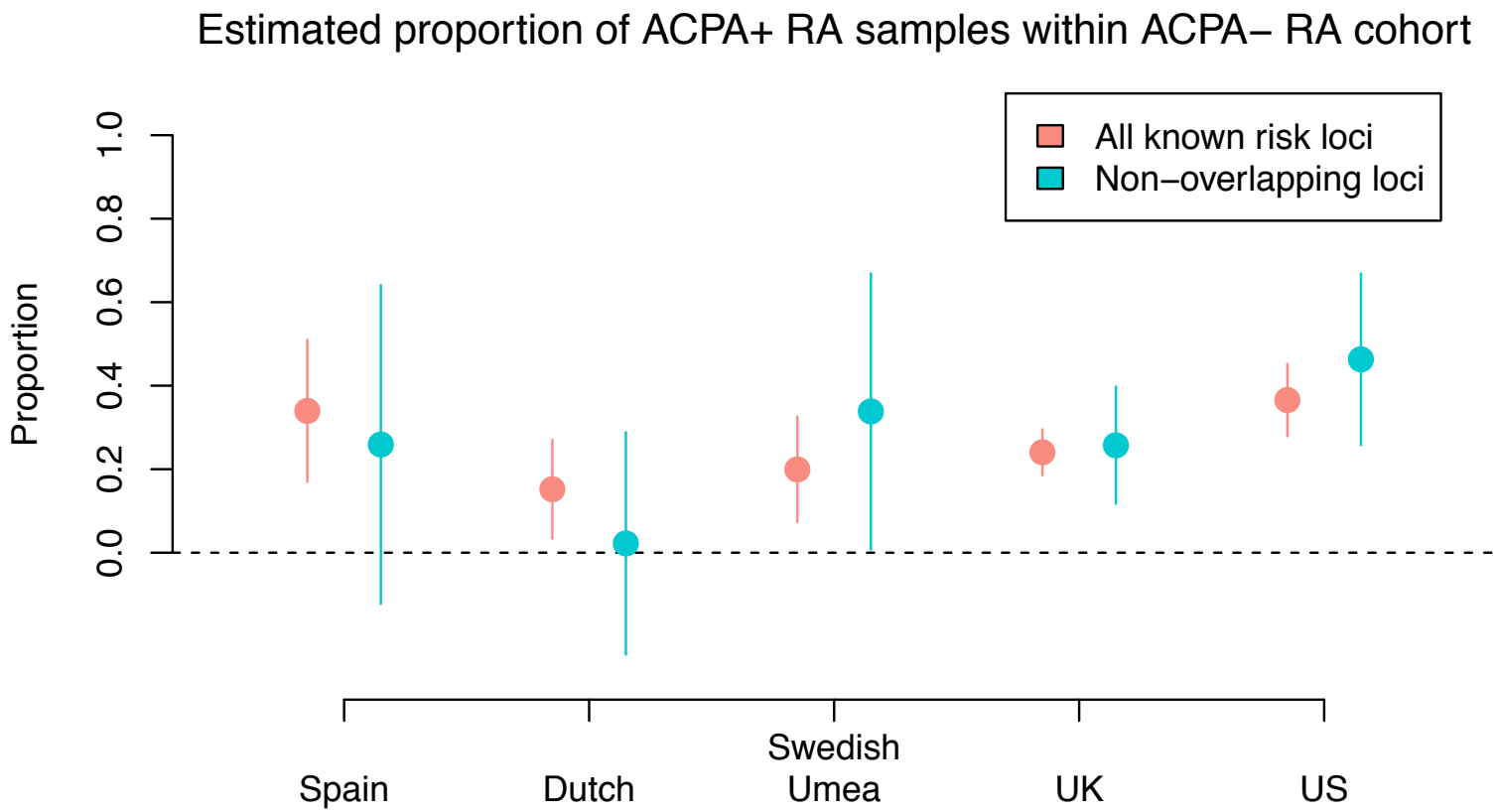


Figure S4. Association results within the MHC to seropositive rheumatoid arthritis. (**A**) We observed the most significant association in *HLA-DRB1*. (**B**) Conditioning on all *HLA-DRB1* alleles revealed an independent association at HLA-B Asp9. (**C**) Conditioning on all *HLA-DRB1* alleles and amino acids at HLA-B position 9 revealed an independent association at HLA-DPβ1 Phe9. (**D**) Conditioning on *HLA-DRB1* alleles and amino acids at HLA-B position 9 and HLA-DPβ1 position 9 revealed an independent association at HLA-A Asn77. (**E**) Conditioning on *HLA-DRB1* alleles and amino acids at HLA-B position 9, HLA-DPβ1 position 9, and HLA-A position 77 did not reveal any convincingly significant association within MHC (P>1.9×10^{-6}).

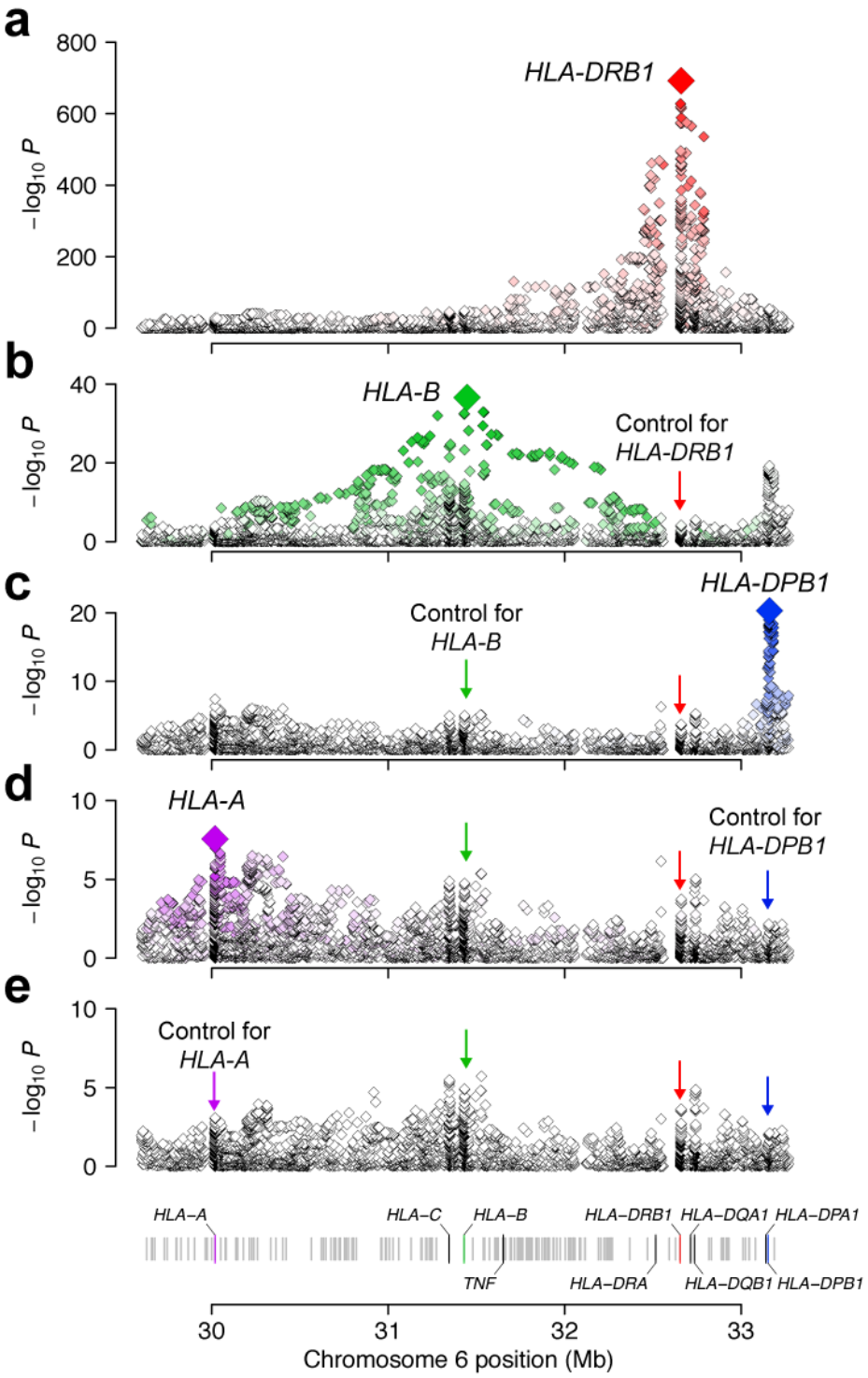


Figure S5: Overview of associated amino acid positions to seronegative and seropositive RA in three dimensional models. All associated positions are in binding grooves. Cyan/Green colors indicate associated positions to both diseases but having distinct effects depending on residues. Orange colors indicate associated positions to both diseases with shared effect size direction. Magenta colors indicate associated positions uniquely to seropositive RA. We used Protein Data Bank (PBD) entries 3pdo (HLA-DR), 3lqz (HLA-DP), 2bvp (HLA-B), and 1x7q (HLA-A) with UCSF Chimera to prepare the figure.

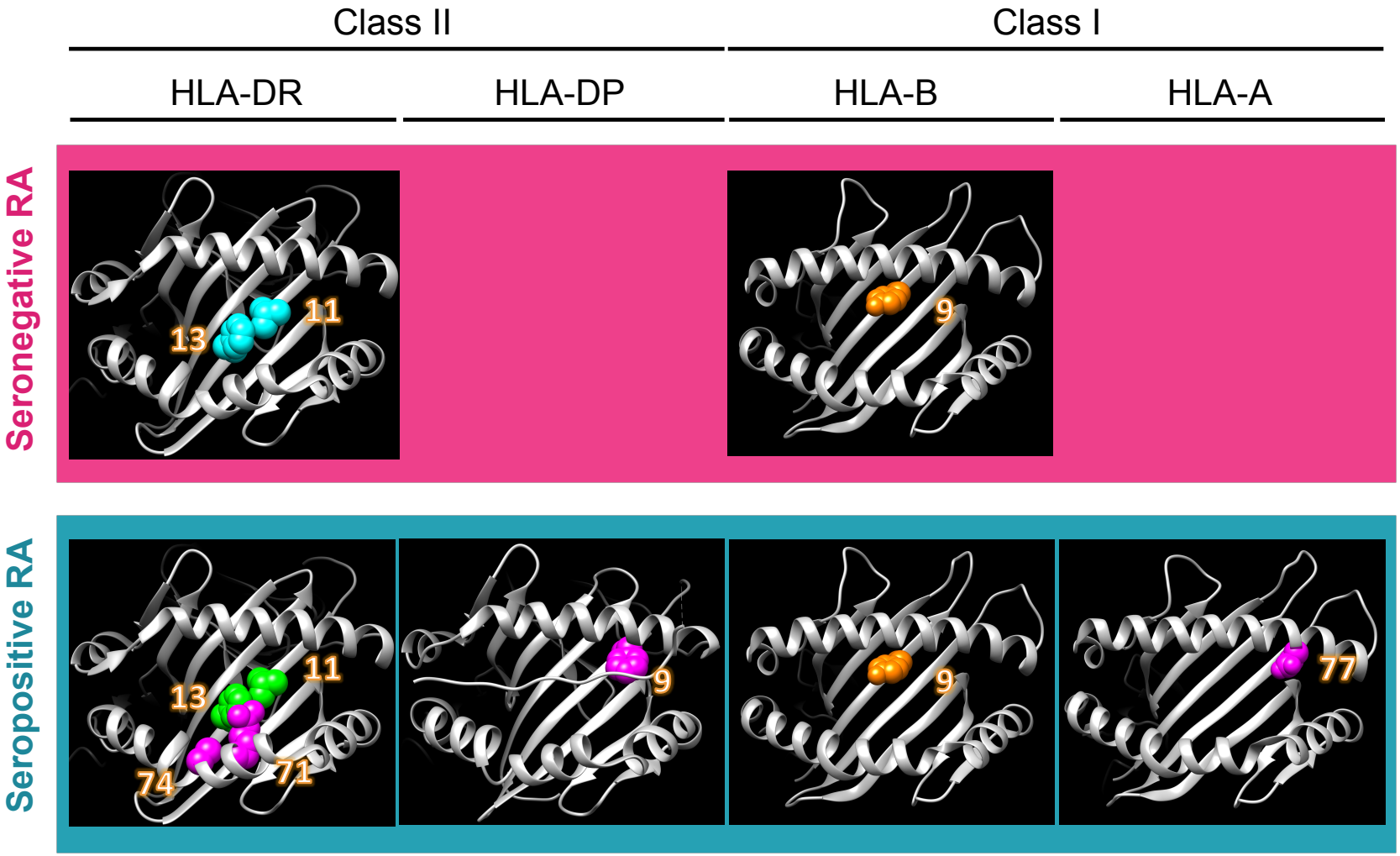


Figure S6: Replication of individual effect sizes of amino acid residues at HLA-DRβ1 position 11. We plot the univariate odds ratio (odds ratio with respect to the other residues as reference) of six residues along with 95% confidence interval, in discovery analysis versus replication analysis. **(A)** When we accounted for possible heterogeneity using risk score corrections in the discovery analysis (See **Methods**), the individual effects were well replicated. The p-value for discordance test was not significant (*P*=0.44). Green line indicates the fitted regression line, which ideally should follow the diagonal line. **(B)** If we do not adjust for risk scores in the discovery analysis, the individual effects were much less concordant to replication (discordance *P*=0.0045).

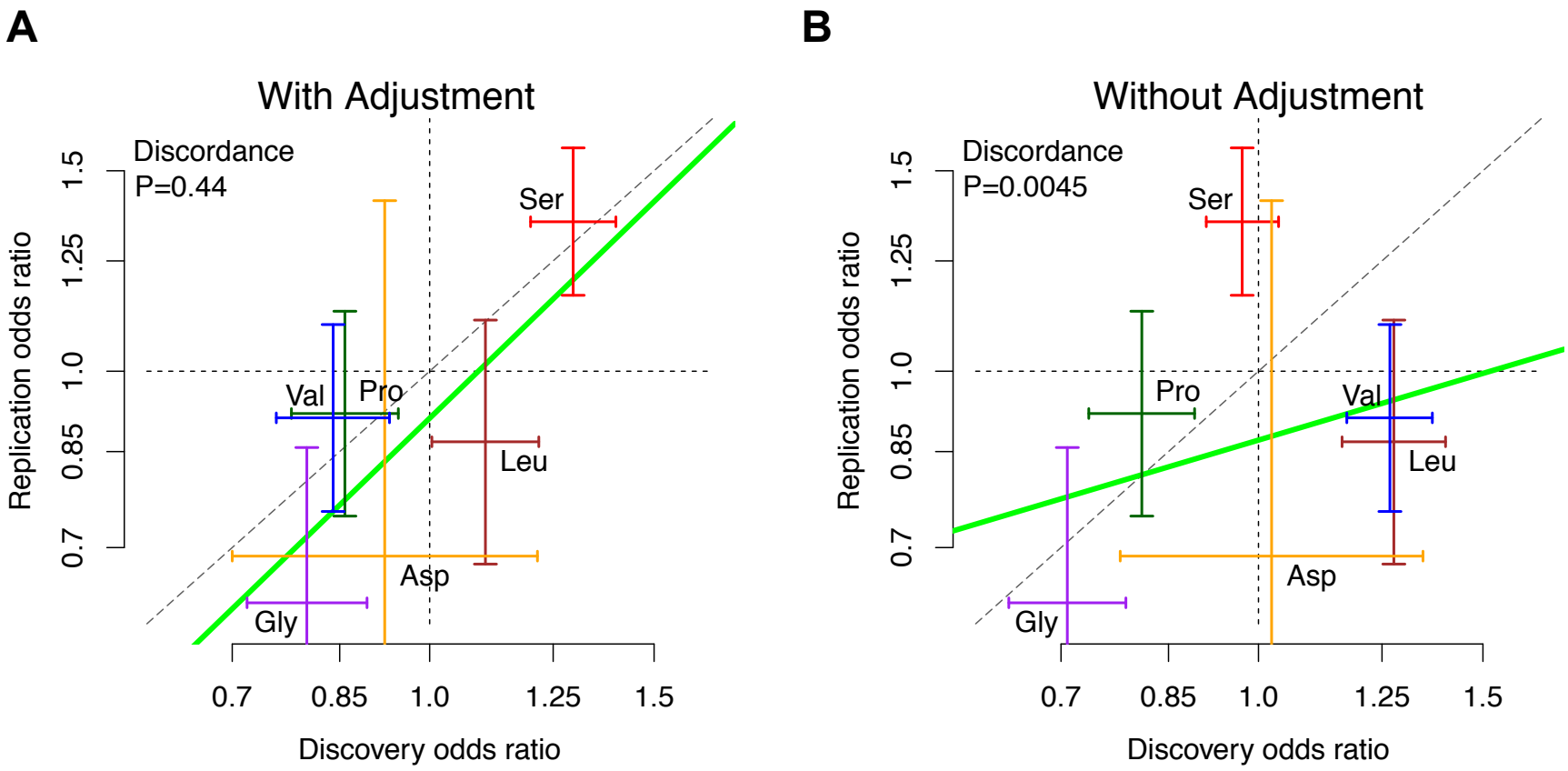
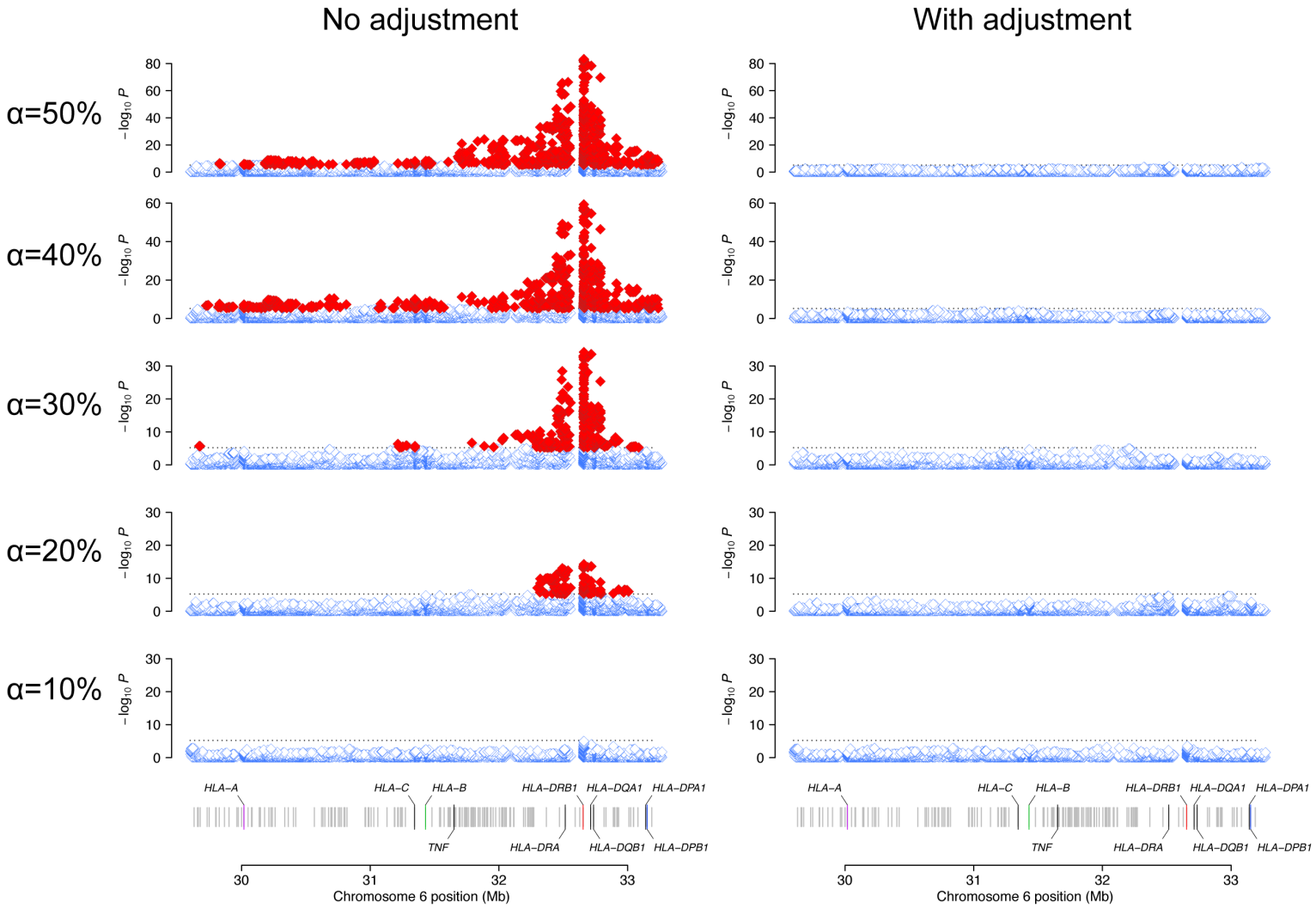


Figure S7: Simulations under the null disease model. We performed a simple simulation that splits UK controls to half and half as null cases and controls, and replaces *α*% of null cases with randomly sampled ACPA+ RA cases to simulate confounding. **(A)** Spurious associations due to the confounding exacerbated with increasing *α* (Left pane). Red diamonds are spurious associations with *P*<6E-6. After we adjust for risk scores, spurious associations disappeared (Right pane). The dotted horizontal line is the threshold 6E-6. **(B)** We approximated the sample proportion of confounding disease using risk score. Vertical lines denote 95% C.I..

A.



B.

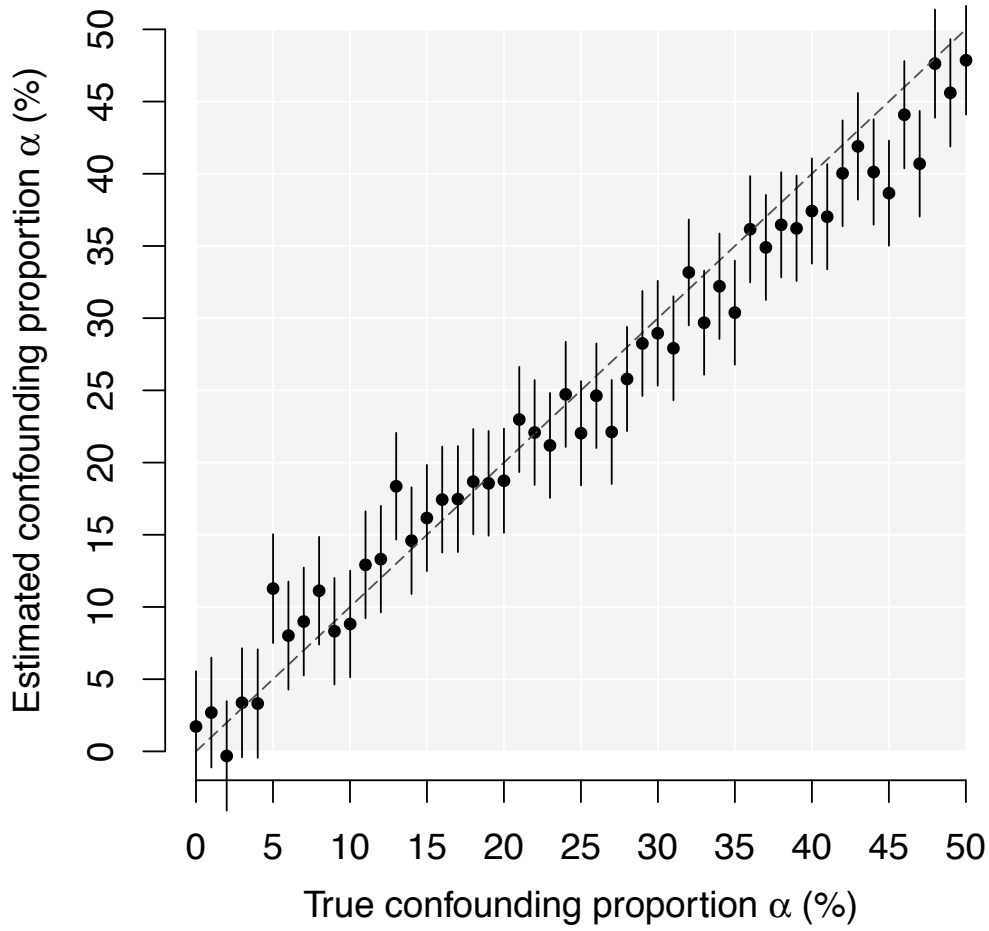
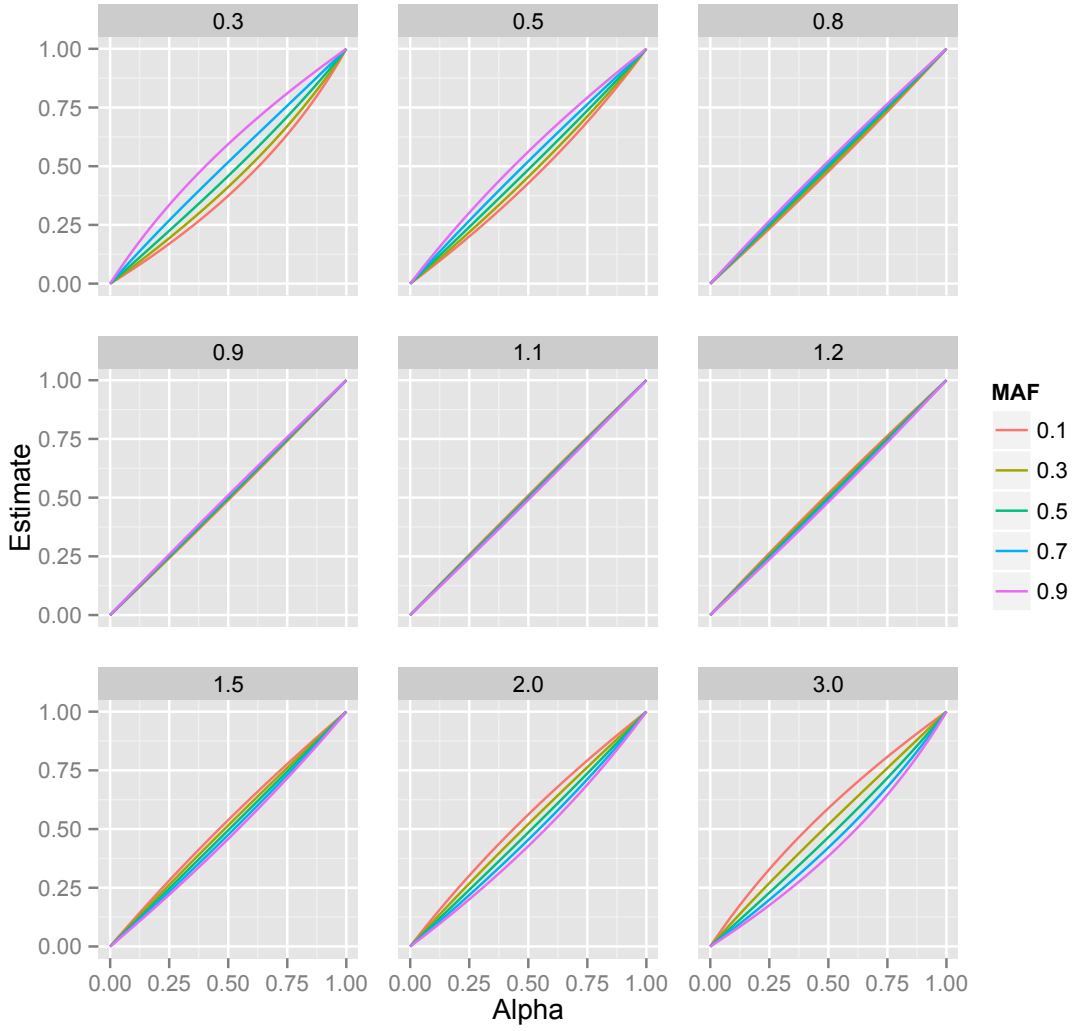

Estimated confounding proportion α (%)
0
5
10
15
20
25
30
35
40
45
50
True confounding proportion α (%)

Figure S8: Approximate relationship between confounding proportion and estimated proportion in logistic regression. We assume one locus with MAF and specific odds ratio. Then we plot for each confounding proportion α (x-axis), the expected value of the estimated proportion in logistic regression (y-axis). The numbers in grey boxes are odds ratios. Unless the odds ratio is large and the MAF is very low or high, the estimated α approximates true α well.



Supplemental Tables

Table S1: Sample Collections. (A) For seronegative RA analysis, we collected five cohorts for the discovery analysis and an independent cohort for replication. Within five cohorts, we confirmed the seronegative status of samples using the conventional anti-CCP testing (yellow colors). In replication cohort, in order to stringently define seronegative samples, we additionally applied newly developed ACPA-specific sensitive testing (blue color) (**Material and Methods**). (B) For seropositive RA analysis, we collected six cohorts. We confirmed seropositive status of these samples using the conventional anti-CCP testing (yellow colors). We re-used the control samples for both seronegative and seropositive analyses.

A. Sample collections for seronegative RA analysis

Cohorts	Case	Control
Discovery analysis		
UK	1096	8430
US	551	2134
Dutch	301	2004
Swedish Umea	242	963
Spanish	216	399
Discovery analysis Total	2406	13930
Replication		
Swedish EIRA	427	1691
Discovery+replication Total	2833	15621

B. Sample collections for seropositive RA analysis

Cohorts	Case	Control
UK	2463	8430
US	1803	2134
Dutch	330	2004
Swedish Umea	524	963
Spanish	397	399
Swedish EIRA	1762	1940
Total	7279	15870

Table S2. List of known associated SNPs used for defining risk scores for ACPA+ RA and AS. (A) We used 47 RA associated loci reported in Eyre et al. 2012 to build genetic risk scores (GRS). We estimated the effect sizes of these loci with respect to ACPA+ RA using leave-one-out approach. In two-step approach, we used loci not associated to ACPA- RA (P>0.01, last column) to build GRS. **(B)** We obtained the list of 24 AS associated loci from the Cortes et al. 2013. We used 19 loci that passed QC in our collections (right most column) in addition to HLA-B*27 to build GRS. We used odds ratios reported in Cortes et al. 2013.

A. 47 RA associated loci (Eyre et al. 2012)

SNP	Gene	Chromosome	Position of proxy	Proxy in Immunochip	r^2 to proxy	ACPA+ association P	ACPA- association P	ACPA- P > 0.01?
rs2843401	TNFRSF14	1	2528133	rs2843401	1	6.57E-09	6.02E-01	TRUE
rs2240336	PADI4	1	17674402	rs2240336	1	5.98E-09	2.83E-02	TRUE
rs883220	INPP5B	1	38616871	rs883220	1	1.01E-04	6.66E-02	TRUE
rs2476601	PTPN22	1	114377568	rs2476601	1	6.99E-77	1.74E-04	FALSE
rs11586238	IGSF2	1	117263138	rs11586238	1	3.42E-03	6.83E-01	TRUE
rs2228145	IL6R	1	154426970	rs2228145	1	1.58E-07	2.44E-02	TRUE
rs12746613	FCGR2B	1	161467042	rs12746613	1	6.91E-05	5.01E-02	TRUE
rs10919563	PTPRC	1	198700442	rs10919563	1	2.88E-04	5.90E-01	TRUE
rs34695944	REL	2	61124850	rs34695944	1	2.75E-08	2.89E-01	TRUE
rs1858036	SPRED2	2	65598241	rs1858036	1	1.04E-06	7.81E-01	TRUE
rs11676922	AFF3	2	100806940	rs11676922	1	2.27E-08	1.64E-02	TRUE
rs13426947	STAT4	2	191933254	rs13426947	1	7.44E-09	2.67E-03	FALSE
rs1980422	CD28	2	204610396	rs1980422	1	2.64E-07	4.72E-01	TRUE
rs11571302	ICOS	2	204742934	rs11571302	1	4.48E-08	1.21E-01	TRUE
rs13315591	PXK	3	58555895	rs9813011	1	1.72E-01	5.00E-01	TRUE
rs12506688	RBPJ	4	26104113	rs12506688	1	2.55E-10	4.18E-02	TRUE
rs6822844	IL21	4	123509421	rs6822844	1	4.04E-02	9.90E-02	TRUE
rs71624119	ANKRD55	5	55440730	rs71624119	1	1.20E-11	5.21E-12	FALSE

rs2561477	PAM	5	102608924	rs2561477	1	2.74E-05	6.20E-05	FALSE
rs548234	PRDM1	6	106568034	rs548234	1	1.57E-02	5.18E-01	TRUE
rs10499194	TNFAIP3	6	138002637	rs10499194	1	8.15E-07	3.36E-01	TRUE
rs6920220	TNFAIP3	6	138006504	rs6920220	1	2.30E-13	3.76E-02	TRUE
rs58721818	TNFAIP3	6	138243739	rs58721818	1	5.99E-12	1.14E-01	TRUE
rs212389	TAGAP	6	159489791	rs212389	1	2.95E-06	7.15E-01	TRUE
rs59466457	CCR6	6	167537754	rs59466457	1	2.91E-10	6.41E-01	TRUE
rs3807306	IRF5	7	128580680	rs3807306	1	1.90E-07	2.22E-02	TRUE
rs10488631	IRF5	7	128594183	rs10488631	1	2.02E-03	4.44E-04	FALSE
rs2736340	BLK	8	11343973	rs2736340	1	1.94E-04	4.92E-04	FALSE
rs951005	CCL21	9	34743681	rs951005	1	3.82E-02	3.73E-01	TRUE
rs2269060	TRAF1	9	123683569	rs2269060	1	5.58E-06	9.46E-02	TRUE
rs10795791	IL2RA	10	6108340	rs10795791	1	4.75E-06	6.24E-02	TRUE
rs4750316	PRKCQ	10	6393260	rs4750316	1	4.54E-04	1.00E-01	TRUE
rs2275806	GATA3	10	8095340	rs2275806	1	1.45E-05	3.50E-02	TRUE
rs12764378	ARID5B	10	63800004	rs12764378	1	1.68E-06	6.93E-01	TRUE
rs540386	TRAF6	11	36509189	rs5030485	0.93	4.47E-02	3.16E-01	TRUE
rs595158	CD5	11	60909581	rs595158	1	3.88E-05	4.07E-03	FALSE
rs10892279	DDX6	11	118611781	rs10892279	1	2.13E-06	8.05E-01	TRUE
rs1678542	KIF5A	12	57968715	rs1678542	1	1.04E-03	6.17E-01	TRUE
rs8043085	RASGRP1	15	38828140	rs8043085	1	1.36E-10	3.70E-01	TRUE
rs8026898	TLE3	15	69991417	rs8026898	1	1.27E-10	3.64E-03	FALSE
rs13330176	IRF8	16	86019087	rs13330176	1	3.85E-08	7.76E-01	TRUE
rs2872507	IKZF3	17	38040763	rs2872507	1	1.28E-06	2.15E-01	TRUE
rs34536443	TYK2	19	10463118	rs34536443	1	2.24E-14	1.16E-02	TRUE
rs4810485	CD40	20	44747947	rs4810485	1	1.45E-09	9.19E-01	TRUE

rs2834512	RCAN1	21	35911599	rs2834512	1	2.16E-04	5.82E-01	TRUE
rs9979383	RUNX1	21	36715761	rs9979383	1	3.76E-05	1.02E-04	FALSE
rs3218253	IL2RB	22	37544810	rs3218253	1	2.55E-07	3.16E-01	TRUE

B. 24 AS associated loci (Cortes et al. 2013)

SNP	Gene	Chromosome	Position	Risk allele/non-Risk allele	Odds ratio	QC passed in our dataset
rs11209026	IL23R	1p31	67478546	G/A	1.62	O
rs1801274	FCGR2A	1q23	159746369	T/C	1.11	O
rs4129267	IL6R	1q21	152692888	C/T	1.14	X
rs41299637	GPR25-KIF21B	1q32	199144473	T/G	1.19	O
rs6600247	RUNX3	1p36	25177701	C/T	1.15	O
rs12615545	UBE2E3	2q31	181756697	C/T	1.12	O
rs4676410	GPR35	2q37	241212412	T/C	1.13	X
rs6759298	Intergenic	2p15	62421949	C/G	1.29	O
rs12186979	PTGER4	5p13	40560617	G/A	1.08	O
rs30187	ERAP1	5q15	96150086	T/C	1.29	O
rs6871626	IL12B	5q33	158759370	A/C	1.1	O
rs17765610	BACH2	6q15	90722494	G/A	1.15	O
rs1128905	CARD9	9q34	138373660	C/T	1.1	X
rs11190133	NKX2-3	10q24	101268715	C/T	1.15	O
rs1250550	ZMIZ1	10q22	80730323	G/T	1.11	O
rs11065898	SH2B3	12q24	110346958	T/C	1.11	O
rs1860545	LTBR-TNFRSF1A	12p13	6317038	C/T	1.13	O
rs11624293	GPR65	14q31	87558574	C/T	1.2	O
imm_16_28525386	IL27-SULT1A1	16p11	28525386	A/G	1.11	X
rs2531875	NOS2	17q11	23172294	G/T	1.12	O
rs9901869	NPEPPS-TBKBP1-TBX21	17q21	42930205	A/G	1.14	O
rs35164067	TYK2	19p13	10386181	G/A	1.14	O
rs2836883	Intergenic	21q22	39388614	G/A	1.18	O

rs7282490	ICOSLG	21q22	44440169	G/A	1.11	X

Table S3. Imputation accuracy in current dataset and previous dataset (Raychaudhuri et al. 2012). To measure imputation accuracy, we typed *HLA-A, HLA-B, HLA-C, HLA-DQB1,* and *HLA-DRB1* in 918 individuals in UK cohort. Then we calculated imputation accuracy as the proportion of the alleles correctly imputed (Online Methods). For each gene, we only used individuals whose four-digit typing was successful.

Gene	Two digit alleles			Four digit alleles		
	Accuracy in previous dataset, A_{prev}	Accuracy in current dataset, A_{new}	Error reduction ratio, $(1-A_{prev})/(1-A_{new})$	Accuracy in previous dataset, A_{prev}	Accuracy in current dataset, A_{new}	Error reduction ratio, $(1-A_{prev})/(1-A_{new})$
HLA-A	0.972	0.988	2.43	0.967	0.983	2.00
HLA-B	0.945	0.982	3.07	0.936	0.972	2.30
HLA-C	0.968	0.985	2.07	0.957	0.978	1.92
HLA-DQB1	0.964	0.994	6.20	0.829	0.987	13.32
HLA-DRB1	0.943	0.974	2.16	0.870	0.926	1.76
Average	**0.959**	**0.985**	**2.70**	**0.912**	**0.969**	**2.87**

Table S4: List of binary markers defined within MHC and the association results at these markers.

Described in a separate Excel file.

Table S5. Exhaustive pairwise search results for associations in *HLA-DRB1* and *HLA-B*. In order to find the best pair of markers explaining the associations to seronegative RA, we tested every possible pair of binary markers between *HLA-DRB1* and *HLA-B*. The total number of tests was 383,130 (495 markers in DRB1 × 774 markers in B). We tested each pair by including them in the logistic regression and comparing the deviance to chi-square distribution with 2 degrees of freedom. We show the top 20 pairs. The best pair was Ser11+Leu11 from *HLA-DRB1,* and from *HLA-B*, one of B*0801, B*08, HLA-B Asp9, and rs2596492 at the first base position of the codon at position 9. The four markers at *HLA-B* are in almost perfect LD ($r^2 \geq 0.997$) and statistically indistinguishable.

	HLA-DRB1		***HLA-B***			
Rank	**SNP2HLA ID**	**Marker**	**SNP2HLA ID**	**Marker**	**Deviance**	**P-value**
1	AA_DRB1_11_32660115_SL	Ser11+Leu11	HLA_B_0801	HLA-B*0801	92.255	9.30E-21
2	AA_DRB1_11_32660115_SL	Ser11+Leu11	HLA_B_08	HLA-B*08	92.251	9.30E-21
3	AA_DRB1_11_32660115_SL	Ser11+Leu11	AA_B_9_31432689_D	Asp9	92.007	1.00E-20
4	AA_DRB1_11_32660115_SL	Ser11+Leu11	SNP_B_31432690_C	rs2596492(C_vs_G+T)	92.007	1.00E-20
5	AA_DRB1_11_32660115_SLD	Ser11+Leu11+Asp11	HLA_B_0801	HLA-B*0801	89.967	2.90E-20
6	AA_DRB1_11_32660115_SLD	Ser11+Leu11+Asp11	HLA_B_08	HLA-B*08	89.962	2.90E-20
7	AA_DRB1_13_32660109_SFG	Ser13+Phe13+Gly13	HLA_B_0801	HLA-B*0801	89.921	3.00E-20
8	AA_DRB1_13_32660109_SFG	Ser13+Phe13+Gly13	HLA_B_08	HLA-B*08	89.916	3.00E-20
9	AA_DRB1_11_32660115_SLD	Ser11+Leu11+Asp11	AA_B_9_31432689_D	Asp9	89.713	3.30E-20
10	AA_DRB1_11_32660115_SLD	Ser11+Leu11+Asp11	SNP_B_31432690_C	rs2596492(C_vs_G+T)	89.713	3.30E-20
11	AA_DRB1_13_32660109_SFG	Ser13+Phe13+Gly13	AA_B_9_31432689_D	Asp9	89.663	3.40E-20
12	AA_DRB1_13_32660109_SFG	Ser13+Phe13+Gly13	SNP_B_31432690_C	rs2596492(C_vs_G+T)	89.663	3.40E-20
13	AA_DRB1_11_32660115_SL	Ser11+Leu11	SNP_B_31430769	rs2523607	86.743	1.50E-19
14	AA_DRB1_11_32660115_SL	Ser11+Leu11	SNP_B_31431395	rs2596495	86.688	1.50E-19
15	AA_DRB1_11_32660115_SL	Ser11+Leu11	SNP_B_31431485	rs4990036	86.619	1.60E-19
16	AA_DRB1_11_32660115_SL	Ser11+Leu11	AA_B_97_31432180_SNV	Ser97+Val97+Asn97	84.906	3.70E-19
17	AA_DRB1_11_32660115_SL	Ser11+Leu11	AA_B_97_31432180_SV	Ser97+Val97	84.652	4.10E-19
18	AA_DRB1_11_32660115_SL	Ser11+Leu11	SNP_B_31432582	rs9266178	84.514	4.40E-19
19	AA_DRB1_11_32660115_SL	Ser11+Leu11	SNP_B_31432583	rs9266179	84.514	4.40E-19
20	AA_DRB1_11_32660115_SL	Ser11+Leu11	AA_B_45_31432581_EG	Glu45+Gly45	84.486	4.50E-19

Table S6. RF status-stratified analysis results in the UK cohort. We obtained rheumatoid factor (RF) data for the cases in the UK cohort. We stratified the cases into two groups based on the RF status and examined association results in each group, controlling for heterogeneity due to possible confounding from ACPA+ RA and AS.

CCP and RF status	# Cases	HLA-DRβ1 Ser11+Leu11		HLA-B Asp9		Estimated confounding proportion	
		P-value	OR (CI95)	P-value	OR (CI95)	ACPA+ RA	AS
All CCP-	1096	5.7E-11	1.38 (1.26-1.53)	2.3E-5	1.31 (1.16-1.48)	0.241	0.099
CCP- / RF+	470	6.2E-9	1.53 (1.32-1.77)	8.1E-7	1.56 (1.32-1.85)	0.255	0.090
CCP- / RF-	546	1.8E-4	1.29 (1.13-1.47)	0.08	1.17 (0.99-1.39)	0.233	0.118

Table S7. Forward conditional haplotype analysis on individual HLA-DRβ1 amino acid residues for ACPA+ RA. For each amino acid position in HLA-DRβ1 (column 1), we partitioned the classical alleles into groups based on the amino acid residues and performed omnibus association testing where the degree of freedom (**df**) is the number of partitions minus one. We included the signal peptide in the test (negative positions). If multiple amino acid positions are statistically the same (give the exactly same partitioning all the time), we only kept the position with the lowest position number. The most significant amino acid positions were 11 and 13 (highlighted), which were statistically indistinguishable ($P > 0.03$). Then we performed conditional analysis where given the partitioning defined on position 11, if further partitioning by additional amino acid gives significant p-value. Conditioning on 11, we found 71 is significant (highlighted), and conditioning on 11 and 71, we found 74 was significant (highlighted). Conditioning on 11, 71, and 74, the most significant was position 70 (highlighted).

Condition:	**On Nothing**			**On Position 11**			**On Positions 11 and 71**			**On positions 11, 71 and 74**		
Amino acid position	**df**	χ^2	$\log_{10}P$	**df**	χ^2	$\log_{10}P$	**df**	χ^2	$\log_{10}P$	**df**	χ^2	$\log_{10}P$
-29	1	124.26	-28.13	1	0.84	-0.44	1	3.30	-1.16	1	10.34	-2.89
-25	2	1274.28	-276.71	2	5.31	-1.15	2	15.67	-3.40	2	19.79	-4.30
-24	2	2669.56	-579.69	2	5.31	-1.15	2	15.67	-3.40	2	19.79	-4.30
-17	2	149.22	-32.40	1	0.84	-0.44	1	3.30	-1.16	1	10.34	-2.89
-16	2	1274.28	-276.71	2	5.31	-1.15	2	15.67	-3.40	2	19.79	-4.30
-1	2	355.64	-77.23	2	15.15	-3.29	1	3.30	-1.16	1	10.34	-2.89
1	1	1.64	-0.70	1	5.12	-1.63	1	4.95	-1.58	1	4.14	-1.38
4	2	251.49	-54.61	1	5.12	-1.63	1	4.95	-1.58	1	4.14	-1.38
9	2	207.59	-45.08	0	0.00	0.00	0	0.00	0.00	0	0.00	0.00
10	2	1572.99	-341.57	1	4.48	-1.47	1	12.67	-3.43	1	8.35	-2.41
11	5	3551.51	-766.45	0	0.00	0.00	0	0.00	0.00	0	0.00	0.00
12	1	1504.66	-328.42	0	0.00	0.00	0	0.00	0.00	0	0.00	0.00
13	5	3489.63	-753.02	2	7.29	-1.58	2	15.01	-3.26	2	9.65	-2.10
14	1	313.74	-69.48	0	0.00	0.00	0	0.00	0.00	0	0.00	0.00
16	1	111.92	-25.43	1	2.81	-1.03	1	1.68	-0.71	1	1.17	-0.55
26	2	224.95	-48.85	2	75.19	-16.33	2	24.89	-5.41	2	10.57	-2.29

28	2	15.31	-3.33	1	15.64	-4.12	1	23.79	-5.97	1	9.60	-2.71
30	5	498.85	-104.85	2	14.94	-3.24	2	22.79	-4.95	2	10.28	-2.23
31	2	218.07	-47.35	1	4.48	-1.47	1	12.67	-3.43	1	8.35	-2.41
32	1	708.24	-155.32	1	4.62	-1.50	1	11.02	-3.04	1	56.73	-13.30
33	1	2639.72	-575.02	1	4.48	-1.47	1	12.67	-3.43	1	8.35	-2.41
37	4	1864.09	-401.81	4	11.68	-1.70	4	12.36	-1.83	4	73.74	-14.43
38	2	144.08	-31.29	2	14.13	-3.07	2	21.95	-4.77	2	9.88	-2.15
40	1	125.01	-28.30	1	4.48	-1.47	1	12.67	-3.43	1	8.35	-2.41
47	1	1642.11	-358.28	1	1.99	-0.80	1	2.60	-0.97	1	9.29	-2.64
57	3	425.20	-91.11	3	10.40	-1.81	3	9.87	-1.71	3	43.16	-8.64
58	1	241.01	-53.63	1	2.57	-0.96	1	1.47	-0.65	1	33.10	-8.06
60	2	374.12	-81.24	2	10.40	-2.26	2	9.56	-2.08	2	41.65	-9.04
67	2	2107.81	-457.70	2	143.52	-31.16	2	2.11	-0.46	2	84.35	-18.32
70	2	1610.74	-349.77	2	85.29	-18.52	2	2.33	-0.51	2	93.66	-20.34
71	3	1279.84	-276.46	3	222.78	-47.30	0	0.00	0.00	0	0.00	0.00
73	1	574.88	-126.31	1	52.92	-12.46	1	4.10	-1.37	0	0.00	0.00
74	4	782.79	-167.39	3	139.68	-29.35	3	97.96	-20.37	0	0.00	0.00
77	1	203.84	-45.52	1	52.92	-12.46	1	4.10	-1.37	0	0.00	0.00
78	1	250.09	-55.61	0	0.00	0.00	0	0.00	0.00	0	0.00	0.00
85	1	17.88	-4.63	1	0.05	-0.08	1	0.06	-0.09	1	2.89	-1.05
86	1	605.20	-132.91	1	35.85	-8.67	1	18.20	-4.70	1	17.32	-4.50
95	1	2.93	-1.06	1	4.81	-1.55	1	4.76	-1.54	1	4.84	-1.56
96	4	3272.65	-707.43	2	9.29	-2.02	2	17.47	-3.79	2	13.18	-2.86
98	2	1381.39	-299.97	2	9.29	-2.02	2	17.47	-3.79	2	13.18	-2.86
102	1	0.03	-0.07	1	0.25	-0.21	1	0.59	-0.36	1	0.58	-0.35
104	2	1377.57	-299.14	2	4.74	-1.03	2	13.28	-2.88	2	8.93	-1.94
112	2	64.54	-14.01	2	0.51	-0.11	2	1.22	-0.26	2	39.31	-8.54

120	2	2858.50	-620.72	1	0.25	-0.21	1	0.59	-0.36	1	0.58	-0.35
133	2	192.39	-41.78	1	0.25	-0.21	1	0.59	-0.36	1	0.58	-0.35
140	2	190.04	-41.27	1	0.25	-0.21	1	0.59	-0.36	1	0.58	-0.35
142	2	192.39	-41.78	1	0.25	-0.21	1	0.59	-0.36	1	0.58	-0.35
149	2	1504.92	-326.79	1	0.25	-0.21	1	0.59	-0.36	1	0.58	-0.35
166	2	125.04	-27.15	2	4.74	-1.03	2	13.28	-2.88	2	8.93	-1.94
180	2	2658.12	-577.20	2	4.79	-1.04	2	13.21	-2.87	2	10.53	-2.29
181	2	220.16	-47.81	2	4.79	-1.04	2	13.21	-2.87	2	10.53	-2.29
182	1	80.78	-18.60	1	0.31	-0.24	1	0.77	-0.42	1	2.18	-0.85
188	1	105.05	-23.92	1	0.12	-0.14	1	2.80	-1.03	1	9.62	-2.72
189	2	105.46	-22.90	2	3.19	-0.69	2	5.08	-1.10	2	8.98	-1.95
190	1	104.05	-23.70	1	0.07	-0.10	1	2.49	-0.94	1	8.74	-2.51
231	2	224.61	-48.77	2	4.55	-0.99	2	14.90	-3.24	2	18.19	-3.95
233	2	1502.31	-326.22	2	3.19	-0.69	2	5.08	-1.10	2	8.98	-1.95
234	1	103.53	-23.59	1	0.04	-0.07	1	2.29	-0.89	1	8.30	-2.40
236	1	99.34	-22.67	1	0.01	-0.04	1	1.46	-0.64	1	5.74	-1.78

Table S8. Effect estimates for the amino acids associated with risk of ACPA- and ACPA+ rheumatoid arthritis.
(A) Effect estimates for ACPA- RA. We estimated the effect size of HLA-DRβ1 Ser11+Leu11 taking Val11+Asp11+Pro11+Gly11 as reference and HLA-B Asp9 taking His9+Tyr9 as reference. The effect size at HLA-DRβ1 Ser11+Leu11 is conditioned on HLA-B Asp9, and the effect size at HLA-B Asp9 is conditioned on HLA-DRβ1 Ser11+Leu11. We also show the unadjusted case/control allele frequencies and the classical alleles of *HLA-DRB1* and *HLA-B* corresponding to the amino acids. **(**B) Effect estimates for ACPA+ RA. For HLA-DRβ1, We defined haplotypes based on the amino acid residues present at position 11, 13, 71, and 74. For each haplotype, the multivariate effect is given as an odds ratio (OR), taking the most frequent haplotype (ProArgAlaAla) in the control samples as the reference (that is, giving that haplotype an OR of 1). The effects are conditioned on the remaining associated loci: position 9 in HLA-B, position 9 in HLA-DPβ1, and position 77 in HLA-A. We show the unadjusted allele frequencies and the classical alleles corresponding to each haplotype. We also list the effect sizes, allele frequencies and classical alleles corresponding to HLA-B Asp9, HLA-DPβ1 Phe9, and HLA-A Asn77. The effects of each of these positions were estimated conditioned on the remaining loci; e.g. the effects in HLA-B were conditioned on *HLA-DRB1* alleles, position 9 in HLA-DPβ1, and position 77 in HLA-A. 95% CI, 95% confidence interval.

A. ACPA- RA					
HLA-DRβ1 amino acid at position 11	OR	95% CI	Unadjusted allele frequencies		Classical *HLA-DRB1* alleles
			Controls	Cases	
Ser+Leu	1.23	1.15-1.32	0.514	0.548	**01, *03, *08, *11, *12, *13, *14*
Val+Asp+Pro+Gly	Reference		0.486	0.452	**04, *07, *09, *10, *15, *16*
HLA-B amino acid at position 9					Classical *HLA-B* alleles
Asp	1.24	1.14-1.36	0.131	0.161	**08*
His, Tyr	Reference		0.869	0.839	**07, *13, *14, *15, *18, *27, *35, *37, *38, *39, *40, *41, *42, *44, *45, *46, *47, *48, *49, *50, *51, *52, *53, *54, *55, *56, *57, *58, *73, *81*

B. ACPA+ RA								
HLA-DRβ1 amino acid at position				Multivariate OR	95% CI	Unadjusted allele frequencies		Classical *HLA-DRB1* alleles
11	13	71	74			Controls	Cases	
Val	Phe	Arg	Ala	4.65	3.80-5.70	0.007	0.021	**10:01*
Val	His	Lys	Ala	4.03	3.72-4.37	0.110	0.292	**04:01, *04:09*
Val	His	Arg	Ala	3.63	3.29-4.01	0.054	0.123	**04:04, *04:05, *04:08, *04:10*
Leu	Phe	Arg	Ala	2.11	1.94-2.31	0.104	0.146	**01:01, *01:02*
Asp	Phe	Arg	Glu	1.82	1.52-2.18	0.013	0.017	**09:01*
Pro	Arg	Arg	Ala	1.58	1.26-1.99	0.009	0.010	**16:01, *16:02*
Val	His	Arg	Glu	1.29	1.06-1.57	0.016	0.012	**04:03, *04:06, *04:07, *04:11*
Ser	Gly	Arg	Ala	1.04	0.86-1.25	0.018	0.013	**12:01, *12:02*
Val	His	Glu	Ala	1.03	0.71-1.50	0.005	0.003	**04:02, *04:37*
Pro	Arg	Ala	Ala	1.00	Reference	0.143	0.094	**15:01, *15:02, *15:03*
Gly	Tyr	Arg	Gln	0.92	0.83-1.02	0.127	0.067	**07:01*
Ser	Ser	Lys	Ala	0.87	0.66-1.14	0.009	0.005	**13:03*

Ser	Gly	Arg	Leu	0.83	0.70-0.98	0.027	0.016	**08:01, *08:02, *08:03, *08:04, *08:06, *14:15*
Ser	Ser	Arg	Glu	0.77	0.64-0.94	0.023	0.011	**14:01, *14:05, *14:07*
Ser	Ser	Arg	Ala	0.76	0.67-0.86	0.067	0.034	**11:01, *11:04, *11:06, *11:08, *13:05, *14:02, *14:06*
Leu	Phe	Glu	Ala	0.71	0.55-0.93	0.012	0.005	**01:03*
Ser	Ser	Lys	Arg	0.67	0.60-0.76	0.127	0.081	**03:01, *03:02, *03:04*
Ser	Ser	Glu	Ala	0.60	0.54-0.67	0.115	0.046	**11:02, *11:03, *13:01, *13:02, *13:04*
Ser	Gly	Arg	Glu	0.49	0.26-0.91	0.003	0.001	**14:04*
HLA-B amino acid at position 9				Classical *HLA-B* alleles				
Asp				2.13	1.91-2.37	0.130	0.118	**08*
His, Tyr				1.00	Reference	0.870	0.882	**07, *13, *14, *15, *18, *27, *35, *37, *38, *39, *40, *41, *42, *44, *45, *46, *47, *48, *49, *50, *51, *52, *53, *54, *55, *56, *57, *58, *73, *81*
HLA-DPβ1 amino acid at position 9				Classical *HLA-DPB1* alleles				
Phe				1.31	1.24-1.39	0.721	0.793	**02, *04, *05, *16, *19, *23, *34*
His, Tyr				1.00	Reference	0.279	0.207	**01, *03, *06, *09, *10, *11, *13, *14, *15, *17, *18, *20, *21, *26, *30, *35*
HLA-A amino acid at position 77				Classical *HLA-A* alleles				
Asn				0.85	0.81-0.90	0.343	0.279	**01, *23, *24, *26, *29, *30, *36, *80*
Asp, Ser				1.00	Reference	0.657	0.721	**02, *03, *11, *25, *30, *31, *32, *33, *34, *66, *68, *69, *74*

Table S9. Frequency of ancestral haplotype in six cohorts. We calculated the frequency of ancestral 8.1 haplotype using the best guess imputation data that was phased across the MHC region.

Cohort	Ancestral Haplotype Frequency
UK	0.131
US	0.106
Dutch	0.165
Swedish Umea	0.095
Spanish	0.052
Swedish EIRA	0.136

Table S10. Estimated proportions of confounding diseases in seronegative RA dataset. Using risk scores built from the known associated loci to ACPA+ RA and ankylosing spondylitis (AS), we estimated the proportion of ACPA+ RA and AS samples within ACPA- (**Online Methods**). We applied two different approaches; (A) We used logistic regression that includes the ACPA+ risk score and AS risk score, and no candidate associations to ACPA- RA. (B) We used logistic regression that includes not only the ACPA+ risk score and AS risk score but also the two variables that are putatively associated to ACPA- RA, Ser+Leu-11 of HLA-DRβ1 and Asp-9 of HLA-B. 95% C.I., 95% confidence interval.

Cohort	**(A) Regression using risk scores only**				**(B) Regression using risk scores and putative associations in ACPA- RA**			
	ACPA+	95% C.I.	AS	95% C.I.	ACPA+	95% C.I.	AS	95% C.I.
UK	0.241	0.185-0.296	0.099	0.055-0.142	0.282	0.223-0.340	0.100	0.056-0.144
US	0.366	0.279-0.452	0.041	-0.032-0.115	0.409	0.320-0.498	0.048	-0.025-0.122
Dutch	0.152	0.034-0.270	0.082	-0.017-0.181	0.184	0.063-0.304	0.088	-0.011-0.188
Swedish Umea	0.199	0.073-0.326	0.108	0.030-0.185	0.256	0.127-0.384	0.112	0.034-0.190
Spanish	0.340	0.170-0.510	0.079	-0.066-0.225	0.381	0.210-0.552	0.073	-0.073-0.220