Supporting Information

Rioux et al. 10.1073/pnas.0909307106

SI Text

Cohort Descriptions. The entire set of DNA samples consisted of 10,576 individuals derived from 13 cohorts, specifically from 7 different immune-mediated diseases and 3 shared control cohorts. The population from which each cohort was derived, as well as the distribution of DNAs for each disease, were as follows: 654 SLE cases from the United Kingdom, 486 SLE cases from the United States, 427 CD cases from the United States and Italy, 688 UC cases from Italy, 1,343 RA cases from Sweden, 674 RA cases from the United States, 453 MG cases from Sweden, 270 IGAD cases from Sweden, 502 MS mother-father-affected child trios from the United Kingdom, and 531 MS mother-father-affected child trios from the United States for a total of 3099 DNAs, 1,056 controls from the United States, 673 controls from Sweden, and 753 controls from the United Kingdom. Of the 1,343 Swedish RA cases, only 657 ACPA-positive individuals were used in the analysis. After quality control, there remained 643 SLE cases from the United Kingdom, 483 SLE cases from the United States, 396 CD cases from the United States and Italy, 667 UC cases from Italy, 1,308 RA cases from Sweden, 604 RA cases from the United States, 438 MG cases from Sweden, 267 IGAD cases from Sweden, 494 MS mother-father-affected child trios from the United Kingdom, and 518 MS mother-father-affected child trios from the United States for a total of 3,036 DNAs, 1,049 controls from the United States, 672 controls from Sweden, and 746 controls from the United Kingdom.

DNA Handling and SNP Genotyping. All DNAs were received at the Broad Institute Center for Genotyping and Analysis (CGA) in 96-well plates. The concentration of double-stranded DNA was assessed using PicoGreen (Molecular Probes), and concentrations were normalized to 50–100 ng/μL. Some samples were native DNA, and others underwent whole-genome amplification (WGA) before receipt. Native and WGA DNAs were never arrayed on the same plate for Illumina processing. Overall, 120 96-well plates were processed; of these, 17 contained WGA DNA. DNAs from HapMap CEU cell lines (Coriell Cell Repositories) served as process controls. We genotyped 96-well plates of DNA on the CGA's Illumina GoldenGate BeadLab platform as described previously.

Genotype calls were performed using the BeadStudio program (Illumina) to define genotype clusters based on signal intensity. The 103 native DNA plates and the 17 WGA DNA plates were clustered in a stepwise fashion, the rationale being that native DNA samples exhibit better defined and separated (tighter) clusters than WGA DNA. First, for all native DNA samples, the genotype cluster positions (centers and sizes) were determined automatically using the BeadStudio clustering algorithm, followed by manual review. During the manual review process, advanced user modifications were applied to reflect optimally the distribution of sample set genotypes for each SNP. These adjusted user calls received a second round of manual review. After the final clusters were defined for the native DNAs, the modified static cluster definitions (as an .egt file) were applied independently to the WGA DNA plate intensities. The manual review process of the WGA DNAs followed the same iterative workflow as the review of native DNAs; advanced user modifications were applied to reflect optimally the sample set per SNP. Certain SNP assays were marked as failed within BeadStudio because of either poor cluster separation or low signal intensity. The genotypes for the remaining, passing SNPs were exported from BeadStudio and analyzed as described in the following sections.

HLA Typing. Previous 2- or 4-digit typing of *HLA-A*, *HLA-B*, *HLA-C*, *HLA-DRB1*, and *HLA-DQB1* was available for 15%, 18%, 9%, 24%, and 71% of the patient datasets, respectively. HLA typing was performed by different methodologies available to the investigators, including PCR-based sequence-specific oligonucleotide probe reverse-line blot assay, sequence-specific oligonucleotide (LABType) typing, and exons 2/3 sequence-based typing.

Imputation of Genotypes at HLA Genes. For individuals for whom classic HLA alleles were not available, HLA genotypes were imputed from SNPs in the MHC using a recently developed statistical approach (1). Briefly, this method utilizes a database of SNP genotypes and classic HLA alleles for chromosomes when the haplotype phase is known (or has been estimated) and uses a population genetic model to impute HLA alleles for additional individuals for whom only SNP data are available. The SNPs used for HLA allele imputation were selected from the intersection of the SNPs in the database and those genotyped on individuals in the present study. The training database used was from a previously created map of 7,500 SNPs, deletion insertion polymorphisms, and HLA alleles for 182 Utah residents (29 extended families containing 45 unrelated parent-offspring trios) of European ancestry in the Centre d"Etude du Polymorphisme Humain collection (2). We used a forward-selection, backwards-elimination approach to search the space of possible SNP sets efficiently. For each set of SNPs, leave-one-out cross-validation was used to assess prediction accuracy in the training data, averaging across all chromosomes in the database. For each HLA locus the best set of up to 40 prediction SNPs was chosen to be used for imputation. SNPs chosen for predicting each allele are available on request.

To validate the imputed classic HLA type data used in our analyses, we compared our imputed data with available 4-digit classic HLA data. Specifically, we calculated the sensitivity, specificity, and positive predictive value of the imputed HLA data for the *HLA-B*, *HLA-C*, and *HLA-DRB1* loci in the CD dataset (450 samples with *HLA-B* and *HLA-C*), SLE dataset (313 samples with *HLA-DRB1*), and MS dataset (2,257 samples with *HLA-DRB1*), because these disease cohorts had with the most complete datasets for the classic typing. Results are shown in Fig. S1.

Discrepancies between imputed and typed classic HLA alleles could result from errors in the SNP typing, in the imputation process itself, or in the actual HLA typing. Although molecular HLA typing is considered the reference standard, its high cost often limits the number of loci typed, the level of resolution (2 digit vs. 4 digit), and the number of samples that can be typed in a given study. In addition, even in a single dataset, the classic HLA typing often is performed with a variety of laboratory methods and allele-calling algorithms. For example, the HLA typing in the IgAD cohort in the current dataset had been carried out over a period of 20 years using a variety of different methods (including serology and DNA-based typing). In this dataset there were discrepancies between the classic HLA and the imputed HLA genotypes in 94 individuals (64 *HLA-A*, 59 *HLA-B*, 55 *HLA-DRB1*, and 34 *HLA-DQB1* alleles). These samples therefore were retyped, using PCR sequence-specific primer (SSP) kits (Dynal Biotech). We then recalculated the sensitivity and

specificity of the imputed alleles against the typed alleles in both the original and the retyped data. When discrepancies were detected after retyping of the classic HLA loci, the sensitivity increased for many HLA alleles (e.g., from 72.7% to 90% for *HLA-A*11* and from 89.7% to 95.4% for *HLA-B*07*). The increased sensitivity observed in the retyped dataset showed that some original HLA genotypes were inaccurate and that some of the errors were actively corrected by the imputation algorithm. The final results showed an overall concordance of 95% between the SSP-based typing and the SNP-based imputation for most HLA alleles: close to 100% for the *HLA-A* alleles (with the exception of *HLA-A33* and *-A66*); 95% to 100% for most *HLA-B* alleles (notable exceptions being *HLA-B27*, *-B47*, and *-B55*); 90% to 100% for most *HLA-DR* alleles (exceptions being *HLA-DRB1*01* and *HLA-DRB1*09*); and 96% to 100% for the *HLA-DQ* alleles. For the noted exceptions, the imputation did not perform well because the training data did not contain samples (haplotypes) representing these alleles.

Statistical Analyses. We performed quality filtering of both samples and SNPs to ensure robust association testing. To determine the appropriate thresholds, we examined sample heterozygosity, Hardy-Weinberg equilibrium, Mendelian inheritance errors (ME), and inflation. We applied 5 filters iteratively: SNPs less than 70%, samples less than 90%, SNPs less than 95%, families with > 50 ME, and SNPs > 15 ME. After filtering, SNP and sample outliers disappear, and remaining samples and SNPs fall under normal distribution. Overall, 83.85% of the SNPs (1288/1536) and 97.48% of samples (10,309/10,576) passed quality control and were included in analysis; the average call rate after quality control was 99.0%.

Association testing of all SNPs and imputed HLA alleles was performed by the transmission disequilibrium test for the MS trios and by a standard χ^2 test carried out on a 2×2 contingency table for case/control cohorts, as implemented in the PLINK analysis software (http://pngu.mgh.harvard.edu/purcell/plink/) (3). For case/control analyses, each disease cohort was paired with a matching population control cohort when available; otherwise, pairing was performed with the best-matching control cohort. Matching of the case/control cohorts was evaluated by calculating a genomic control coefficient (GCC) using the non-MHC SNPs that were included in the genotyping panel (see earlier sections). When more than a single-source population was available for a given disease, the case/control (e.g., the RA cohorts from the United States and Sweden) or trio cohorts (e.g., the MS cohorts from the United States and the United Kingdom) were combined into a single analysis cohort for increased power. Because an evaluation of the combined case/control cohorts by Cochran-Mantel-Haenszel analysis did not alter the association results significantly, all analyses were performed simply by combining cohorts. An overall GCC was calculated for each combined disease case/control population and was used to correct the association χ^2 results. Association results from the combined disease cohorts are reported as 2-tailed nominal significance p-values (GCC corrected for case-control cohorts).

Conditional logistic regression analyses were performed in the different disease cohorts for the top associated SNP and HLA alleles using the WHAP analysis software (http://pngu.mgh.harvard.edu/purcell/whap/). Briefly, each SNP and HLA allele was evaluated for independence from the top SNP or HLA allele in a pairwise fashion, and independent association signal results are reported as 2-tailed nominal significance p-values after GCC correction.

1. Leslie S, Donnelly P, McVean G (2008) A statistical method for predicting classical HLA alleles from SNP data. *Am J Hum Genet* 82:48–56.
2. de Bakker PI, et al. (2006) A high-resolution HLA and SNP haplotype map for disease association studies in the extended human MHC. *Nat Genet* 38:1166–1172.
3. Purell S, et al. (2007) PLINK: A toolset for whole-genome association and population-based linkage analysis. *Am J Hum Genet* 81:559–575.

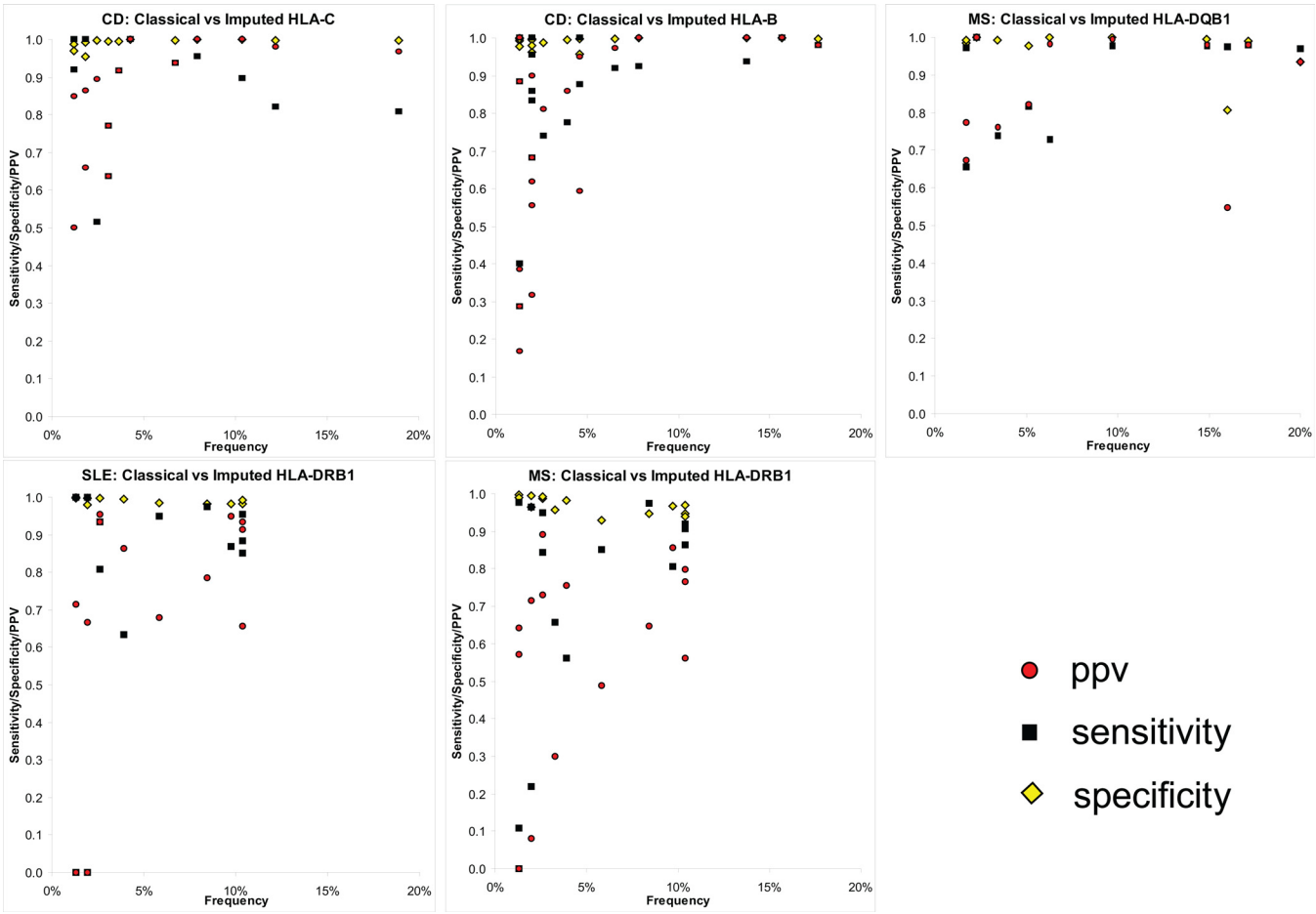

Fig. S1. Assessment of the quality of HLA allele imputation. Sensitivity, specificity, and positive predictive value were calculated for imputed HLA alleles in the datasets for which classic HLA typing data also were available (see *Materials and Methods*). These quality metrics are plotted as a function of allele frequencies in the extended HapMap CEU population (2). HLA alleles that are observed only once (or not at all) in the imputation training set were excluded.

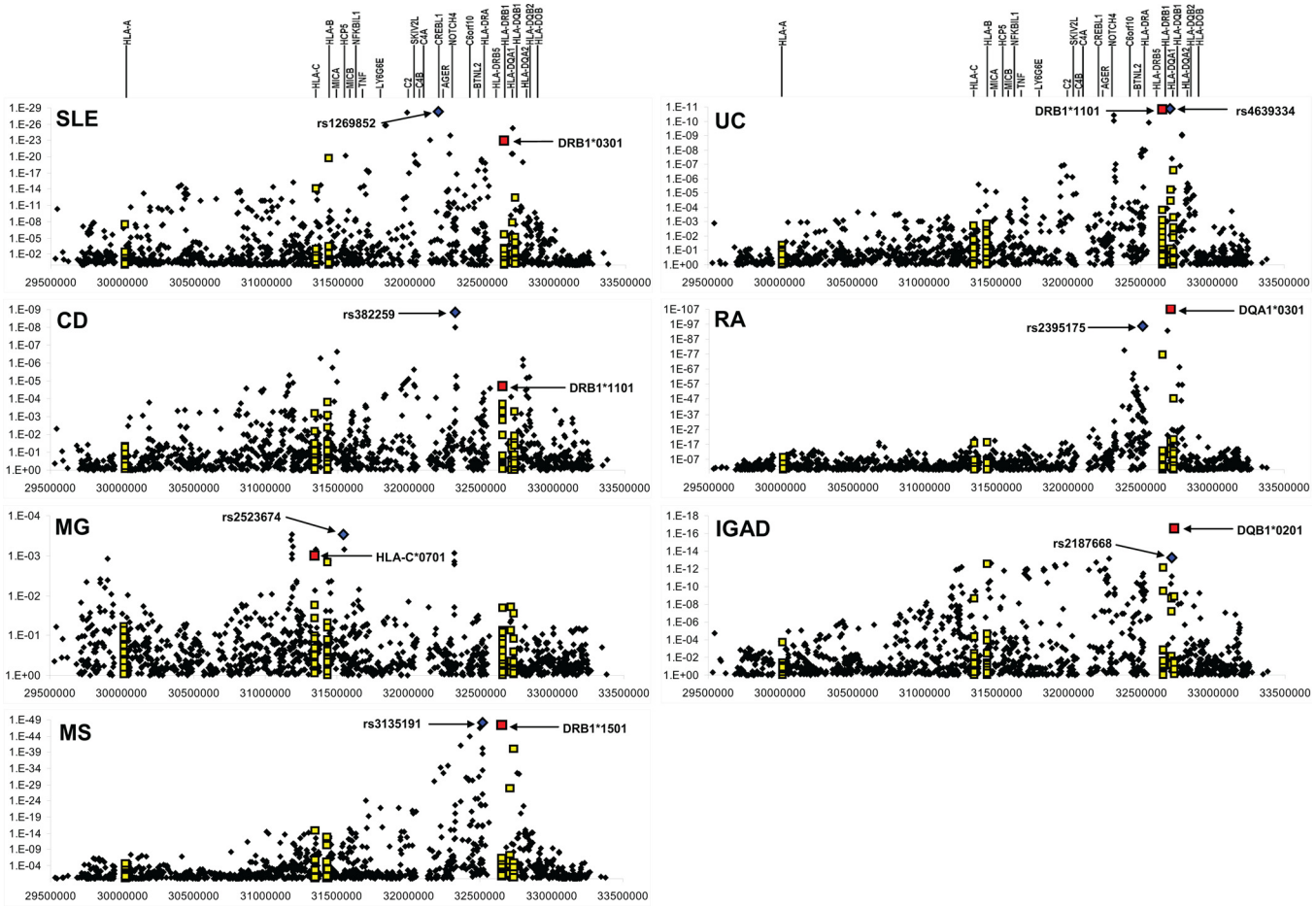

Fig. S2. Results of association and logistic regression analysis for all 7 diseases. Results of allelic tests of association (*Top*) for SNPs (*black diamonds*) and imputed HLA alleles (*yellow boxes*). All association results are represented as the −log10 of the p-values (*y*-axis). The most highly associated SNPs and HLAs are highlighted in blue and red, respectively.

Table S1. Variants that are statistically equivalent to each top (primary) association

Disease	SNP	Position	A1	F_A	F_U	A2	OR	CHISQ	Association P-value	r^2 to primary signal	LRT	Reciprocal conditional P-value
SLE	rs1269852	32151660	G	0.22	0.11	C	2.35	124.80	5.63E-29	1	0	1
	rs558702	31974755	A	0.23	0.11	G	2.34	124.44	6.75E-29	0.96	0	1
	rs3130484	31820160	C	0.22	0.11	T	2.25	113.61	1.59E-26	0.93	0	1
	rs3131378	31829561	G	0.22	0.11	A	2.24	113.25	1.90E-26	0.93	0	1
	rs3131379	31825312	A	0.22	0.11	G	2.24	113.25	1.90E-26	0.93	0	1
UC	rs4639334	32653366	T	0.39	0.13	C	4.04	45.8	1.30E-11	1	0	1
	DQB1*0301	32676000	Y	0.40	0.20	N	2.69	26.5	2.64E-07	0.75	3.1	0.2116
	DRB1*1101	32605000	Y	0.32	0.09	N	4.62	45.4	1.59E-11	0.72	2.3	0.3163
	DQA1*0501	32660500	Y	0.44	0.26	N	2.30	20.4	6.37E-06	0.59	1.5	0.4699
	rs3129763	32642431	T	0.42	0.25	C	2.18	17.4	3.04E-05	0.52	1.9	0.3799
CD	rs382259	32280470	G	0.46	0.27	A	2.30	36.67	1.40E-09	1	0	1
	rs419132	32282252	G	0.46	0.28	A	2.18	32.81	1.02E-08	0.94	1.8	0.18
RA	DQA1*0301	32660500	Y	0.36	0.17	N	2.81	333.52	1.64E-74	1	0	1
MG	rs2523674	31541152	G	0.53	0.43	A	1.5	13.1	0.0003	1	0	1
IGAD	DQB1*0201	32676000	Y	0.40	0.19	N	2.8	71.3	3.04E-17	1	0	1
MS	rs3135391	32482210	T	558	163	C	3.4	216.4	5.51E-49	1	0	1
	rs3135352	32464116	G	560	170	T	3.3	208.4	3.14E-47	0.99	0	1
	DRB1*1501	32605000	Y	557	166	N	3.4	211.5	6.62E-48	0.99	0	1
	rs3135388	32484240	A	561	168	G	3.3	211.9	5.39E-48	1	0.7	0.7093

The SNPs listed for each disease are part of the same equivalence class; they can explain the association observed for top association signal or can be said to be statistically equivalent to top signal. A SNP was defined as equivalent to the top signal if it showed a correlation of $r2 > 0.5$ and caused the top signal to lose significance (p-value > 0.05) following reciprocal conditional analysis.

Table S2. Full set of secondary signals

SLE	Position	A1	F_A	F_U	A2	OR	CHISQ	Association P-value	LRT	rs1269852 Conditional P-value	Signal ID*
rs3135391	32482210	T	0.18	0.14	C	1.3	11.7	0.0006	24.9	3.90E-06	2
rs3135388	32484240	A	0.17	0.14	G	1.3	10.8	0.0010	23.4	8.30E-06	2
rs3135352	32464116	G	0.17	0.14	T	1.3	10.6	0.0012	23.0	1.02E-05	2
rs6932517	32724561	G	0.54	0.41	C	1.7	82.6	1.03E-19	22.7	1.16E-05	3
DRB1*1501	32605000	Y	0.17	0.14	N	1.3	10.2	0.0014	22.3	1.46E-05	2
DQB1*0602	32676000	Y	0.18	0.15	N	1.2	7.9	0.0051	19.8	4.91E-05	2
rs396960	32263048	A	0.19	0.27	T	0.6	47.8	4.66E-12	18.4	0.0001	4
rs3129888	32482949	G	0.22	0.20	A	1.2	5.3	0.0215	17.9	0.0001	2
rs1264708	30163379	G	0.22	0.22	A	1.0	0.2	0.6957	17.5	0.0002	5
rs2227139	32484648	G	0.51	0.39	A	1.6	75.5	3.58E-18	15.9	0.0004	3
rs3135006	32713527	T	0.25	0.23	C	1.1	3.3	0.0699	15.4	0.0005	2
rs3134975	32699001	T	0.25	0.23	C	1.1	3.1	0.0760	15.0	0.0005	2
rs9267948	32283686	G	0.40	0.40	A	1.0	0.1	0.7068	14.4	0.0007	-

UC	Position	A1	F_A	F_U	A2	OR	CHISQ	Association P-value	LRT	rs463933 Conditional P-vaule	Signal ID*
rs382259	32280470	G	0.55	0.27	A	3.334	43.8	3.65E-11	24.5	4.82E-06	7
rs419132	32282252	G	0.56	0.28	A	3.234	42.0	9.26E-11	24.5	4.91E-06	7
rs659445	31968733	G	0.13	0.33	A	0.2988	28.2	1.08E-07	16.3	0.00028	8
rs605203	31951441	G	0.13	0.34	T	0.3015	27.7	1.40E-07	16.0	0.00034	8

CD	Position	A1	F_A	F_U	A2	OR	CHISQ	Association P-value	LRT	rs382259 Conditional P-value	Signal ID*
rs4713436	31190882	T	0.29	0.17	C	2.0	19.0	1.28E-05	16.0	0.00033	10
rs2844511	31493983	A	0.26	0.43	G	0.5	26.7	2.43E-07	15.8	0.00036	11
rs3916766	32728023	A	0.45	0.29	T	2.0	24.8	6.48E-07	14.7	0.00063	12

RA	Position	A1	F_A	F_U	A2	OR	CHISQ	Association P-value	LRT	DQA1*0301 Conditional P-value	Signal ID*
DQB1*0501	32676000	Y	0.16	0.12	N	1.5	22.15	2.52E-06	95.9	1.48E-21	14
DRB1*0101	32605000	Y	0.14	0.10	N	1.5	21.04	4.49E-06	94.6	2.88E-21	14
DQA1*0101	32660500	Y	0.17	0.14	N	1.2	7.23	0.0072	78.5	8.79E-18	14
rs6457614	32759877	C	0.15	0.12	A	1.3	10.77	0.0010	77.9	1.21E-17	14
rs3817969	32469365	A	0.16	0.14	G	1.1	1.46	0.2263	55.1	1.11E-12	14
rs1555115	32462497	C	0.15	0.14	G	1.1	2.54	0.1108	54.6	1.42E-12	14
rs6457617	32771828	G	0.26	0.50	A	0.4	309.91	2.28E-69	54.4	1.51E-12	15
rs2294884	32475236	C	0.18	0.17	A	1.0	0.23	0.6283	49.7	1.63E-11	14
rs2294883	32475428	A	0.18	0.18	T	1.0	0.04	0.8445	49.3	1.96E-11	14
rs2621326	32891873	A	0.49	0.41	G	1.4	30.82	2.83E-08	47.8	4.18E-11	16
rs3129878	32516712	G	0.28	0.32	T	0.8	13.34	0.0003	39.5	2.62E-09	17
rs3129859	32508916	G	0.27	0.32	C	0.8	10.86	9.83E-04	38.8	3.67E-09	17
rs2064476	33181299	C	0.21	0.30	T	0.6	58.42	2.11E-14	38.4	4.50E-09	-
rs3129845	32504254	G	0.27	0.31	A	0.8	9.55	2.00E-03	38.2	5.01E-09	17
rs3135340	32506849	C	0.26	0.30	A	0.8	9.06	0.0026	38.1	5.31E-09	17
rs3128930	33183643	A	0.17	0.25	G	0.6	49.04	2.50E-12	37.9	5.88E-09	-
rs3129876	32515989	A	0.26	0.30	G	0.8	9.67	1.87E-03	37.6	6.97E-09	17
rs3091282	33165175	G	0.21	0.30	C	0.6	57.48	3.42E-14	37.5	7.34E-09	-
rs9277378	33158256	G	0.20	0.29	A	0.6	54.51	1.54E-13	35.8	1.65E-08	-
rs1894408	32894810	C	0.42	0.37	G	1.3	16.88	3.97E-05	35.6	1.88E-08	16
rs3135024	33155443	G	0.16	0.24	A	0.6	45.31	1.68E-11	33.5	5.40E-08	-
rs2621413	32850508	G	0.24	0.31	T	0.7	29.19	6.58E-08	33.2	6.08E-08	-
rs2621416	32849845	G	0.35	0.27	A	1.5	41.37	1.26E-10	32.5	8.82E-08	16
rs2857207	32851150	A	0.24	0.30	G	0.7	22.39	2.23E-06	30.6	2.28E-07	-
rs2071473	32890582	A	0.41	0.35	G	1.3	19.38	1.07E-05	30.3	2.62E-07	16
rs3130071	31702606	T	0.10	0.15	A	0.6	27.02	2.02E-07	29.5	3.94E-07	-
rs3117213	33172582	T	0.19	0.26	G	0.6	40.66	1.81E-10	29.2	4.61E-07	-
rs3117234	33181961	G	0.17	0.23	A	0.7	35.24	2.92E-09	29.0	5.17E-07	-
rs532098	32686029	T	0.58	0.44	C	1.8	101.50	7.13E-24	28.8	5.67E-07	-
rs2857177	32860817	A	0.29	0.37	G	0.7	35.42	2.66E-09	28.6	6.09E-07	-
rs6907322	32432922	A	0.18	0.19	G	1.0	0.29	0.5901	28.6	6.13E-07	-
rs2395314	33170650	T	0.19	0.26	G	0.6	40.50	1.97E-10	28.0	8.16E-07	-
rs2621384	32867250	G	0.29	0.37	A	0.7	35.35	2.75E-09	27.9	8.70E-07	-
rs2076536	32447325	G	0.19	0.34	A	0.5	145.84	1.41E-33	27.6	9.96E-07	-

SLE	Position	A1	F_A	F_U	A2	OR	CHISQ	Association P-value	LRT	rs1269852 Conditional P-value	Signal ID*
rs3117226	33165636	A	0.18	0.26	G	0.7	38.20	6.37E-10	27.3	1.15E-06	-
rs3128968	33164230	A	0.17	0.24	T	0.7	35.75	2.24E-09	27.3	1.19E-06	-
rs3093662	31652167	G	0.03	0.08	A	0.4	51.34	7.78E-13	26.9	1.46E-06	-
rs2071475	32890364	T	0.23	0.21	C	1.1	2.36	0.1248	26.5	1.77E-06	-
rs1264419	30684759	G	0.39	0.51	C	0.6	78.60	7.60E-19	26.4	1.82E-06	-
rs2075801	31836245	A	0.20	0.15	G	1.425	22.29	2.34E-06	26.0	2.26E-06	-
rs2621421	32848648	G	0.35	0.35	C	1.011	0.03	0.8548	25.9	2.36E-06	-
rs910320	33183420	A	0.16	0.22	G	0.6813	27.89	1.28E-07	25.0	3.81E-06	-
rs3097671	33155589	C	0.12	0.17	G	0.6883	21.20	4.13E-06	24.8	4.13E-06	-
rs419132	32318776	G	0.16	0.26	A	0.536	76.87	1.83E-18	24.1	5.78E-06	-
rs2239709	31615425	A	0.09	0.07	G	1.382	9.79	0.0018	24.0	6.02E-06	-
rs2857150	32871798	A	0.34	0.34	G	0.9897	0.03	0.8597	23.9	6.42E-06	-
rs2857154	32870593	T	0.34	0.34	C	0.9902	0.03	0.8672	23.9	6.62E-06	-
rs6926737	32483722	C	0.29	0.49	T	0.4297	208.58	2.80E-47	23.8	6.84E-06	-
rs3830076	32204221	A	0.15	0.07	G	2.4	93.72	3.64E-22	23.7	7.01E-06	-
rs2050191	32446878	A	0.12	0.23	T	0.4526	107.52	3.42E-25	23.5	7.71E-06	-
rs2395153	32453572	C	0.62	0.38	G	2.621	287.70	1.58E-64	23.3	8.54E-06	-
rs9268507	32485516	T	0.30	0.49	C	0.4312	206.64	7.44E-47	23.3	8.67E-06	-
rs2853926	31371029	G	0.21	0.26	A	0.7725	15.18	9.79E-05	23.0	1.00E-05	-
rs6932517	32786159	G	0.25	0.44	C	0.4187	206.55	7.78E-47	22.6	1.24E-05	15
rs2395175	32513003	T	0.43	0.16	C	3.746	434.96	1.36E-96	22.5	1.30E-05	-
rs439852	33113185	T	0.17	0.26	C	0.5874	59.40	1.29E-14	22.5	1.31E-05	-
rs2621323	32896684	C	0.35	0.32	T	1.169	7.01	0.0081	21.9	1.73E-05	16
rs544167	31998136	G	0.02	0.06	T	0.3657	44.58	2.45E-11	21.8	1.89E-05	-
rs4576294	32906525	A	0.12	0.10	G	1.119	1.59	0.2075	21.5	2.15E-05	-
rs9332730	32019987	G	0.06	0.04	C	1.707	18.04	2.17E-05	21.4	2.21E-05	-
rs1264423	30679449	G	0.40	0.51	A	0.6267	68.62	1.19E-16	21.0	2.74E-05	-
rs2395178	32513339	G	0.18	0.34	C	0.4437	152.21	5.69E-35	21.0	2.78E-05	-
DRB1*0404	32654526	Y	0.13	0.07	N	1.964	50.54	1.17E-12	21.0	2.82E-05	-
rs1383266	32942709	A	0.28	0.26	G	1.113	2.82	0.0929	20.9	2.92E-05	-
rs241438	32905597	A	0.29	0.35	G	0.7781	17.28	3.22E-05	20.8	3.04E-05	-
rs382259	32317004	G	0.15	0.25	A	0.5581	66.16	4.16E-16	20.5	3.51E-05	-
rs3132958	32405878	A	0.12	0.23	G	0.4659	100.18	1.39E-23	20.5	3.56E-05	-
rs3135338	32509194	G	0.18	0.33	A	0.4469	148.32	4.04E-34	20.4	3.75E-05	-
rs984778	32508065	C	0.18	0.34	T	0.4453	151.06	1.02E-34	20.2	4.04E-05	-
rs3093553	31657534	G	0.03	0.07	T	0.4083	41.36	1.26E-10	20.2	4.19E-05	-
rs6903608	32536262	G	0.16	0.30	A	0.4385	141.68	1.14E-32	20.0	4.48E-05	-
rs430188	33095519	C	0.01	0.03	G	0.2646	29.67	5.12E-08	19.9	4.67E-05	-
rs3130299	32311514	C	0.16	0.25	T	0.5892	55.79	8.07E-14	19.8	5.05E-05	-
rs3129904	32418373	A	0.12	0.23	G	0.4661	99.73	1.74E-23	19.6	5.52E-05	-
rs10484565	32903009	A	0.18	0.09	G	2.232	94.60	2.33E-22	19.2	6.94E-05	-
rs365053	32303965	G	0.16	0.25	T	0.5951	53.64	2.41E-13	19.1	6.95E-05	-
rs3817973	32469088	C	0.36	0.59	T	0.4063	249.91	2.71E-56	18.9	7.69E-05	-
rs2856993	32899380	C	0.17	0.17	G	0.9598	0.30	0.5829	18.9	7.97E-05	-
rs507778	32317838	A	0.23	0.36	G	0.5413	96.55	8.71E-23	18.8	8.37E-05	-
rs1140809	30719654	T	0.40	0.51	G	0.6395	62.81	2.28E-15	18.7	8.68E-05	-
rs416352	32315370	A	0.22	0.35	C	0.5312	100.27	1.33E-23	18.6	9.25E-05	-

MG	Position	A1	T	U	A2	OR	CHISQ	Association P-value	LRT	rs1269852 Conditional P-value	Signal ID*
No secondary signal detected											

IGAD	Position	A1	F_A	F_U	A2	OR	CHISQ	Association P-value	LRT	DQB1*0201 Conditional P-value	Signal ID*
DRB1*1501	32605000	Y	0.03	0.14	N	0.2	39.2	3.77E-10	28.4	6.65E-07	20
rs3135352	32464116	G	0.03	0.15	T	0.2	39.3	3.68E-10	28.2	7.70E-07	20
rs3135388	32484240	A	0.03	0.15	G	0.2	39.2	3.88E-10	28.1	8.10E-07	20
DQB1*602	32676000	Y	0.04	0.15	N	0.2	36.7	1.37E-09	25.0	3.68E-06	20
rs3135391	32482210	T	0.04	0.15	C	0.2	34.9	3.55E-09	23.6	7.66E-06	20
rs2395165	32459325	G	0.11	0.19	A	0.5	15.1	9.97E-05	22.6	1.25E-05	20
DQA1*0101	32660500	Y	0.19	0.15	N	1.4	5.1	0.0234	21.9	1.79E-05	21
rs3129878	32479959	G	0.55	0.35	T	2.2	48.6	3.09E-12	21.7	1.92E-05	22
rs6940467	31546500	C	0.14	0.08	T	1.8	9.9	0.0016	21.6	2.01E-05	23
rs2596501	31425403	A	0.42	0.41	G	1.0	0.1	0.7376	20.9	2.84E-05	-

SLE	Position	A1	F_A	F_U	A2	OR	CHISQ	Association P-value	LRT	rs1269852 Conditional P-value	Signal ID*
rs6457614	32698320	C	0.16	0.11	A	1.5	5.3	0.0217	19.2	6.80E-05	21
rs3134954	32143362	C	0.05	0.15	T	0.3	26.8	2.28E-07	19.0	7.48E-05	20
rs1044506	32243509	T	0.04	0.14	G	0.3	29.5	5.72E-08	18.9	7.91E-05	20
rs6903608	32499465	G	0.12	0.27	A	0.4	39.3	3.60E-10	18.7	8.54E-05	-
rs2242660	31702031	A	0.59	0.39	G	2.2	47.3	6.08E-12	18.5	9.85E-05	-
rs3131283	32191361	T	0.04	0.14	C	0.3	29.4	6.00E-08	18.4	9.96E-05	20
DQB1*0501	32676000	Y	0.16	0.11	N	1.4	4.5	0.0330	18.2	0.0001	21
rs3130283	32209987	T	0.04	0.15	G	0.3	29.3	6.35E-08	18.0	0.0001	20
rs2395182	32484506	G	0.10	0.24	T	0.3	37.6	8.85E-10	18.0	0.0001	20
rs1345274	31296875	C	0.04	0.01	G	4.5	16.1	6.08E-05	17.1	0.0002	-
rs805301	31722400	C	0.59	0.39	A	2.2	46.1	1.11E-11	17.1	0.0002	-
rs3129888	32482949	G	0.09	0.23	A	0.3	35.2	2.97E-09	16.8	0.0002	20
rs879882	31245720	A	0.24	0.38	G	0.5	27.8	1.33E-07	16.6	0.0003	-
rs2442752	31455945	G	0.54	0.33	A	2.3	53.7	2.30E-13	16.5	0.0003	-
rs1265078	31218862	C	0.47	0.28	G	2.3	50.2	1.40E-12	16.4	0.0003	-
rs1265087	31216070	A	0.47	0.28	G	2.3	50.2	1.40E-12	16.4	0.0003	-
rs3132571	31011552	G	0.50	0.33	A	2.0	34.0	5.62E-09	15.8	0.0004	-
rs3129886	32481799	T	0.12	0.27	C	0.4	37.7	8.17E-10	15.7	0.0004	20
rs130073	31217440	T	0.18	0.32	C	0.4	32.5	1.17E-08	15.3	0.0005	-
rs130078	31224818	C	0.16	0.30	G	0.4	30.1	4.18E-08	15.3	0.0005	-
rs707929	31846347	G	0.54	0.35	A	2.2	43.6	3.93E-11	15.3	0.0005	-
rs3131786	31008111	C	0.51	0.35	G	1.9	32.6	1.15E-08	15.2	0.0005	-
rs1065461	31236748	A	0.16	0.29	G	0.4	29.8	4.82E-08	15.1	0.0005	-
rs3130501	31242718	T	0.16	0.29	C	0.4	29.8	4.82E-08	15.1	0.0005	-
rs2256583	31363578	G	0.15	0.12	A	1.3	2.9	0.0900	14.7	0.0006	-
rs3130071	31698906	T	0.07	0.16	A	0.4	21.7	3.11E-06	14.5	0.0007	-
rs3094225	31219312	G	0.18	0.32	A	0.5	29.0	7.26E-08	14.3	0.0008	-

MS	Position	A1	T	U	A2	OR	CHISQ	Association P-value	LRT	rs1269852 Conditional P-value	Signal ID*
B*4402	31428700	Y	81	188	N	0.4	42.6	6.85E-11	36.7	1.08E-08	25
rs2743951	29817197	T	418	528	C	0.8	12.8	0.00035	17.2	0.00019	26
C*0501	31344180	Y	94	172	N	0.5	22.9	1.73E-06	16.5	0.00026	25
rs3823342	30018907	G	386	503	A	0.8	15.4	8.71E-05	15.3	0.00047	26
rs2523393	29813622	C	431	532	T	0.8	10.6	0.0011	14.9	0.00058	26
rs2270190	31186832	G	121	205	A	0.6	21.6	3.28E-06	14.6	0.00067	25
rs2213567	32758660	C	350	597	G	0.6	64.4	1.00E-15	14.5	0.00070	27
rs7451258	31280850	G	173	271	C	0.6	21.6	3.31E-06	14.3	0.00080	25

*Signal ID. Additional signals that show significant association to the trait following conditioning on primary signal and pairwise $r^2 > 0.5$

Table S3. Summary of top primary and secondary associations

					Correlation neighborhood ($r^2 > 0.8$)**					
					Telomeric boundary		Centromeric boundary		Interval***	
Disease	Signal	#*	Marker	positiion	Marker	position	Marker	position	(bp)	
SLE	Primary	1	rs1269852	32151660	rs497309	31996913	rs3117103	32420761	423848	
	Secondary	2	rs3135391	32482210	rs9268148	32331090	DQB1*0602	32676000	344910	A
		3	rs6932517	32724561	rs2647012	32710879	rs9275572	32725378	14499	
		4	rs396960	32263048	rs396960	32263048	rs396960	32263048	0	
		5	rs1264708	30163379	rs1264708	30163379	rs1264701	30172567	9188	
UC	Primary	6	rs4639334	32653366	rs4639334	32653366	rs4639334	32653366	0	
	Secondary	7	rs382259	32280470	rs382259	32280470	rs419132	32282252	1782	B
		8	rs659445	31968733	rs605203	31951441	rs659445	31968733	17292	
CD	Primary	9	rs382259	32280470	rs382259	32280470	rs419132	32282252	1782	B
	Secondary	10	rs4713436	31190882	rs4713436	31190882	rs2284177	31195839	4957	
		11	rs2844511	31493983	rs2853977	31483504	rs2516448	31494609	11105	
		12	rs3916766	32728023	rs2647087	32727427	rs3916766	32728023	596	
RA	Primary	13	DQA1*0301	32660500	rs660895	32628856	DQA1*0301	32660500	31644	
	Secondary	13	DQB1*0501	32676000	DRB1*0101	32605000	DQB*0501	32676000	71000	
		15	rs6457617	32710272	rs6457617	32710272	rs6457617	32710272	0	
		16	rs2621326	32830860	rs2856997	32828750	rs1894408	32833797	5047	
		17	rs3129878	32479959	rs3129845	32467488	rs3129881	32480707	13219	C
MG	Primary	18	rs2523674	31541152	rs2523674	31541152	rs2523674	31541152	0	
IGAD	Primary	19	DQB1*0201	32676000	DQB1*0201	32676000	DQB*0201	32676000	0	
	Secondary	20	DRB1*1501	32605000	rs9268148	32331090	DQB1*0602	32676000	344910	A
		21	DQA1*0101	32660500	DRB1*0101	32654526	DQB1*0501	32735652	81126	
		22	rs3129878	32479959	rs3129845	32467488	rs3129881	32480707	13219	C
		23	rs6940467	31546500	rs6940467	31546500	rs6940467	31546500	0	
MS	Primary	24	rs3135391	32482210	rs9268148	32331090	DQB1*0602	32676000	344910	A
	Secondary	25	B*4402	31428700	C*0501	31344180	B*4402	31428700	84520	
		26	rs2743951	29817197	rs3998799	29798645	rs2844846	29819999	21354	
		27	rs2213567	32758660	rs2859071	32750375	rs2051549	32777093	26718	

*Association signal reference numbers as described in Figure 2.

**The correlation neighborhood, as defined by the furthest markers showing the indicated correlation coefficients (0.8 or 0.5) or greater to the associated marker, were evaluated from the extended HapMap dataset described in Bakker et al. 2006.

***Size of region defined by the correlation neighborhood. ''-'' absence of correlated neighbor.

Letters on left side of table indicate shared signals.

Table S4. Performance of associated HLA alleles on validation datasets

Alleles Associated		Disease							Data Description			Performance in Validation Data		
Locus	Allele 4-Digit Type	SLE	UC	CD	RA	MG	IgAD	MS	Training Data (Number)	Validation DataSet	Validation Data (Number)	Sensitivity	Specificity	PPV
HLA-B	4402							x, s3	21	58BC	197	93%	100%	99%
										CD		94%	100%	100%
HLA-C	0501							s3	20	58BC	0	NA	NA	NA
										CD		82%	100%	98%
	0701					xx			31	58BC	0	NA	NA	NA
										CD		81%	100%	97%
HLA-DRB1	0101				s3				13	58BC	144	100%	98%	83%
										MS		98%	95%	65%
										SLE		98%	98%	78%
	0301	xxx, xx							15	58BC	231	92%	100%	98%
										MS		81%	97%	86%
										SLE		87%	98%	95%
	0404				s3				16	58BC	102	70%	99%	77%
										MS		86%	95%	56%
										SLE		95%	98%	66%
	1101		xx, s2	xx					9	58BC	86	98%	99%	81%
										MS		85%	93%	49%
										SLE		95%	99%	68%
	1501	x, s3					x, s3	xxx, xx, s2	31	58BC	228	100%	100%	100%
										MS		98%	100%	100%
										SLE		97%	100%	98%
HLA-DQA1	0101				s3		s3		23	58BC	0	NA	NA	NA
	0301				xxx, xx, s2				41	58BC	0	NA	NA	NA
	0501		s2						32	58BC	0	NA	NA	NA
HLA-DQB1	0201						xx, s2		28	58BC	329	100%	99%	97%
										MS		97%	81%	55%
	0301		s2						30	58BC	342	90%	99%	95%
										MS		98%	99%	98%
	0501				x, s3		s3		17	58BC	212	85%	100%	97%
										MS		98%	100%	99%
	0502		x						3	58BC	12	92%	100%	100%
										MS		97%	98%	67%
	0602	s3					s3		35	58BC	241	89%	99%	93%
										MS		97%	94%	94%

Note:
For each locus the type of association is indicated (see legend below). The number of individuals that carry the allele in the training data is shown, and the performance of the method on the various training datasets available is shown.
Legend:
xxx Top Association in Screening and Replication Datasets (Table 1).
xx Top Disease Specific Association Signals for the MHC in Entire Datasets (Table 2).
x Most Significant Secondary Association (Table 3).
s2 Variants that are statistically equivalent to primary association signal (Supporting Table 1).
s3 Appears in Full Set of Secondary Signals (Supporting Table 1).