HGG Advances

ARTICLE

Transethnic analysis of psoriasis susceptibility in South Asians and Europeans enhances fine mapping in the MHC and genome wide

Philip E. Stuart,[1,13] Lam C. Tsoi,[1,2,3,13] Rajan P. Nair,[1,13] Manju Ghosh,[4] Madhulika Kabra,[4] Pakeeza A. Shaiq,[5] Ghazala K. Raja,[5] Raheel Qamar,[6] B.K. Thelma,[7] Matthew T. Patrick,[1] Anita Parihar,[8] Sonam Singh,[8] Sujay Khandpur,[8] Uma Kumar,[9] Michael Wittig,[10] Frauke Degenhardt,[10] Trilokraj Tejasvi,[1,12] John J. Voorhees,[1] Stephan Weidinger,[11] Andre Franke,[10] Goncalo R. Abecasis,[2] Vinod K. Sharma,[8] and James T. Elder[1,12,*]

Summary

Because transethnic analysis may facilitate prioritization of causal genetic variants, we performed a genome-wide association study (GWAS) of psoriasis in South Asians (SAS), consisting of 2,590 cases and 1,720 controls. Comparison with our existing European-origin (EUR) GWAS showed that effect sizes of known psoriasis signals were highly correlated in SAS and EUR (Spearman $\rho = 0.78$; $p < 2 \times 10^{-14}$). Transethnic meta-analysis identified two non-major histocompatibility complex (non-MHC) psoriasis loci (1p36.22 and 1q24.2) not previously identified in EUR, which may have regulatory roles. For these two loci, the transethnic GWAS provided higher genetic resolution and reduced the number of potential causal variants compared to using the EUR sample alone. We then explored multiple strategies to develop reference panels for accurately imputing MHC genotypes in both SAS and EUR populations and conducted a fine mapping of MHC psoriasis associations in SAS and the largest such effort for EUR. *HLA-C*06* was the top-ranking MHC locus in both populations but was even more prominent in SAS based on odds ratio, disease liability, model fit, and predictive power. Transethnic modeling also substantially boosted the probability that the *HLA-C*06* protein variant is causal. Secondary MHC signals included coding variants of *HLA-C* and *HLA-B*, but also potential regulatory variants of these two genes as well as *HLA-A* and several HLA class II genes, with effects on both chromatin accessibility and gene expression. This study highlights the shared genetic basis of psoriasis in SAS and EUR populations and the value of transethnic meta-analysis for discovery and fine mapping of susceptibility loci.

Introduction

Psoriasis (MIM: 177900) is a common, chronic, immune-mediated disorder of the skin and joints characterized by cutaneous inflammation, epidermal hyperplasia, and increased risk of arthritis as well as cardiovascular morbidity.[1] Substantial evidence indicates that psoriasis is driven by abnormal interactions between the innate and adaptive immune cells, including keratinocytes, neutrophils, macrophages, dendritic cells, and T cells.[2,3]

Psoriasis affects 0.1% to 6.5% of individuals, depending on ethnicity and geographical location, with higher prevalence recorded at increasing latitudes.[4] In European-origin populations, the prevalence of psoriasis is estimated to vary from 2% to 3%,[5,6] making psoriasis a favorable target for genome-wide association studies (GWASs). As a result, GWASs to date have identified 87 independent genetic signals for psoriasis at genome-wide significance, with 11 of the 86 shared by European and Chinese populations, 56 established for Europeans only, and 20 for Chinese only.[1,7] Secondary psoriasis-association signals (independent of the primary variant) have been reported for at least 11 of the 86 susceptibility regions.[1,7,8]

With the exception of a relatively small study in Japan,[9] most GWASs of psoriasis have been carried out in European-origin and Chinese-origin individuals. In all populations, the strongest psoriasis association signals map to the major histocompatibility complex (MHC), comprising approximately 40% of the detectable heritability of psoriasis.[10] Correspondingly, genome-wide significant MHC associations have been reported for Japanese,[11] Korean,[12] Thai,[13] Pakistani,[14,15] and Indian[16–18] populations.

Transethnic GWASs provide both advantages and challenges for the study of genetically complex traits, with the advantage of increased sample size being counterbalanced by the potential challenge posed by differences in

[1]Department of Dermatology, University of Michigan Medical School, Ann Arbor, MI, USA; [2]Department of Biostatistics, Center for Statistical Genetics, University of Michigan, Ann Arbor, MI, USA; [3]Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor MI, USA; [4]Department of Pediatrics Genetics, All India Institute of Medical Sciences, New Delhi, India; [5]Department of Biochemistry, PMASAA University, Rawalpindi, Pakistan; [6]COMSATS Institute of Information Technology, Islamabad, Pakistan; [7]Department of Genetics, University of Delhi South Campus, 110021 New Delhi, India; [8]Department of Dermatology, All India Institute of Medical Sciences, New Delhi, India; [9]Department of Rheumatology, All India Institute of Medical Sciences, New Delhi, India; [10]Institute of Clinical Molecular Biology, University Medical Center Schleswig-Holstein, Campus Kiel, Kiel 24105, Germany; [11]Department of Dermatology, University Medical Center Schleswig-Holstein, Campus Kiel, Kiel 24105, Germany; [12]Ann Arbor Veterans Affairs Hospital, Ann Arbor, MI, USA
[13]These authors contributed equally
*Correspondence: jelder@umich.edu
https://doi.org/10.1016/j.xhgg.2021.100069.

underlying genetic architecture across populations.[19–21] With these factors in mind, we undertook a GWAS of psoriasis in South Asian populations from India and Pakistan, consisting of 2,590 cases and 1,720 controls. We found that effect sizes of the known psoriasis susceptibility regions were highly correlated in the South Asian (SAS) and European-origin (EUR) datasets, leading us to conduct unconditional and conditional transethnic meta-analyses of psoriasis genetic associations in these two populations. We then investigated whether transethnic analysis could refine Bayesian credible sets for psoriasis loci. Because the MHC carries such a large fraction of the genetic burden for psoriasis,[10] much of which appears to map to variation in genes encoding human leukocyte antigens (HLAs) themselves,[22] we also developed an improved algorithm for HLA imputation based on SNP2HLA[23] and assessed the performance of multiple reference panels derived from EUR and SAS samples. We then built and compared SAS, EUR, and transethnic MHC association models and used multiple bioinformatic tools to explore the potential biological consequences of the many coding and non-coding MHC variants we identified. Finally, we assessed linkage disequilibrium (LD) structure of SAS and EUR to see if it benefited our transethnic analysis.

Subjects and methods

Human subjects

Psoriasis cases were diagnosed by a dermatologist according to established criteria.[24] We included individuals ascertained for psoriatic arthritis (PsA [MIM: 6075074]) by a rheumatologist if they manifested joint, skin, scalp, and/or nail lesions consistent with psoriasis. Control individuals were 18 years of age or older, unrelated to affected individuals, and unaffected with psoriasis or PsA. All participating individuals provided written informed consent and were recruited according to the protocols approved by the institutional review boards of each participating institution.

GWAS genotyping

For the SAS GWAS cohort, three batches of genotyping were performed. The first, consisting of 952 cases and 855 unaffected controls after quality control, was typed on an Illumina OmniExpressExome (8v1-1_B) platform. The second and third batches, consisting of a combined total of 1,638 cases and 865 unaffected controls after quality control, were typed on two iterations of the Illumina HumanCoreExome platform (12v1-1_B and 24v1-0_A, respectively). In all, 2,590 cases and 1,720 controls were included after quality control. For quality control, we removed common variants (minor allele frequency [MAF] ≥ 0.05) with a call rate $< 95\%$, rarer variants (MAF < 0.05) with a call rate $< 99\%$, and variants with a Hardy-Weinberg p value in controls $< 1 \times 10^{-6}$. Samples were removed if they had substantial non-South Asian admixture (based on the principal component analysis [PCA] shown in Figure S1), were duplicates or first- or second-degree relatives of other samples (Plink $\widehat{\pi} \geq 0.20$), had a genotype call rate $< 98\%$, or had an outlier heterozygosity value (>1.5 × interquartile range above third quartile or below first quartile).

Genotyping, quality control, phasing, and imputation of six EUR-origin GWAS cohorts have previously been described,[7] including CASP,[25] Kiel,[26] Genizon,[26] PsA GWAS,[27] WTCCC2,[28] and Exomechip[10] (which contains GWAS content). We also included two datasets based on the Immunochip: PAGE and GAPC.[29] Both phase 3 1000 Genomes Project (1KGP)[30] and r.1.1 Haplotype Reference Consortium (HRC)[31] reference panels were used in imputation, and only well-imputed markers (i.e., imputation quality $r^2 \geq 0.7$) were used in subsequent analysis; if a marker was well-imputed by both reference panels, the imputed dosage for the panel with the higher imputation quality was used. In all, these datasets included 15,967 cases and 28,194 controls. Finally, we examined all pairwise combinations of the eight EUR cohorts and the two SAS cohorts and removed samples that were duplicates or first- or second-degree relatives with a sample in a different cohort. Characteristics of the 10 studies analyzed for psoriasis associations are described in Table S1.

Genome-wide meta-analysis

We conducted a transethnic meta-analysis using eight EUR and two SAS cohorts, as well as EUR-only and SAS-only meta-analyses. For the EUR-only and transethnic analyses, we included only those markers that were well-imputed in at least half of the studies; for the SAS-only analysis, markers had to be well-imputed in both cohorts. Meta-analysis was carried out using the inverse variance-weighted approach implemented in METAL.[32] QQ-plots indicated that the PCA and geographic cohort covariates included in our association models did a good job of controlling for population stratification (Figure S2).

For identifying the Bayesian credible set of markers for each locus, we used association results for genetic variants that were well-imputed in at least half of the cohorts (see Supplemental methods for more details). To compare the number of surrogates for known psoriasis loci, we identified markers in strong LD ($r^2 \geq 0.8$) with their lead markers, using the EUR and SAS samples from the 1KGP.

Genome-wide conditional analysis

We conducted conditional analysis for each psoriasis-associated locus that achieved genome-wide significance to reveal independent signals in the transethnic meta-analysis. Employing only markers that are well-imputed in all cohorts, we used stepwise conditional analysis to reveal secondary signals within ±500 kb of the psoriasis-associated signals. For each of the secondary signals identified, we computed the 95% credible interval set as described above. For comparison, we took the same marker(s) utilized for each round of the conditional analysis as covariates in a separate analysis using only the EUR cohorts.

HLA genotyping

Eight classical HLA genes—*HLA-A* (MIM: 142800), *HLA-B* (MIM: 142830), *HLA-C* (MIM: 142840), *HLA-DPA1* (MIM: 142880), *HLA-DPB1* (MIM: 142858), *HLA-DQA1* (MIM: 146880), *HLA-DQB1* (MIM: 604305), and *HLA-DRB1* (MIM: 142857)—were genotyped to 3-field resolution by the Institute of Clinical Molecular Biology at the University of Kiel (IKMB) in Germany. In-solution targeted capturing with an RNA bait panel was designed to accommodate the complete collection of reference sequences for all HLA genes in version 3.09 of the International Immunogenetics information system (IMGT)/HLA database.[33] Target DNA was fragmented to 150–300 bp size segments, enriched for HLA gene sequences by hybridization with the bait, and subjected to high-throughput paired-end sequencing with read lengths of 115–125 bp. Reads were aligned against the cDNA collection of the

IMGT/HLA database, and the most likely genotype was determined by ranking all possible allele combinations by their harmonic mean for five different quality metrics.

Updated SNP2HLA package

We updated several features of v1.0.3 of SNP2HLA for imputing HLA genotypes.[23] Most importantly, we substituted Beagle version 4.1[34] for the older version 3.0.4 as the imputation engine for both the MakeReference and SNP2HLA scripts, which substantially improved accuracy. The HLA amino acid and SNP sequence dictionaries were updated to a more recent release of the IPD-IMGT/HLA database (r3.30.0, October 2017).[35] For the SNP dictionaries we improved handling of polymorphisms for 2-field HLA alleles by employing International Union of Pure and Applied Chemistry (IUPAC) ambiguity codes rather than arbitrary selection of one allele. We also updated the package to accept modern HLA nomenclature.[36]

Construction of SNP2HLA reference panels

We first built a SNP2HLA reference panel of 397 South Asian individuals. Using a method that maximizes represented genetic diversity,[37] 288 individuals of Pakistani or Indian ancestry were selected from batch 3 of our South Asian psoriasis GWAS. An additional 192 individuals from north India were obtained from a preliminary version of the IKMB multiethnic HLA reference panel;[38] these individuals were originally ascertained for a meta-analysis of inflammatory bowel disease (IBD [MIM: 266600]).[39] The 480 selected South Asians were genotyped for eight classical HLA genes; 233 psoriasis GWAS individuals (the University of Michigan [UM] dataset) were successfully genotyped, as were 164 IBD case-control samples, including 141 that were included in the final IKMB panel (IKMB-SAS dataset) and 23 that were not (B.K. Thelma or BKT dataset). We then input into our updated MakeReference script HLA genotypes for the 397 South Asians, along with genotypes for 2,284 SNPs in the classical MHC region common to the microarray used for the UM dataset and the Immunochip platform used for the IKMB-SAS and BKT datasets.

We constructed an additional 18 SNP2HLA reference panels for imputing our South Asian GWAS samples by rebuilding existing HLA panels and by forming various combinations of the UM, IKMB-SAS, and BKT components of the 397-person SAS panel with four other datasets—the non-SAS subset of the IKMB HLA reference panel,[38] the European ancestry Type 1 Diabetes Genetics Consortium (T1DGC) SNP2HLA panel,[23,40,41] the pan-Asian SNP2HLA panel,[42,43] and data from phase 3 of the 1KGP.[30,44] HLA genotypes for 1KGP were combined separately with both microarray-based (v1) and sequence-based (v2) MHC genotypes for both the full 1KGP dataset (1KGP-ALL) and its SAS subset (1KGP-SAS), resulting in four versions of 1KGP data used for construction of South Asian panels (1KGP-ALL-v1, 1KGP-SAS-v1, 1KGP-ALL-v2, and 1KGP-SAS-v2). We also built 20 SNP2HLA reference panels for imputation of HLA variants in people of European ancestry. Datasets used for these panels consisted of many of those used for South Asians (T1DGC, UM, BKT, IKMB, 1KGP-ALL-v1, and 1KGP-ALL-v2) as well as the EUR subset of the 1KGP with HLA and either microarray or sequence-based MHC data (1KGP-EUR-v1 and 1KGP-EUR-v2, respectively). Additional details of panel construction are provided in Supplemental methods.

Validation of SNP2HLA reference panels

The performance of each HLA reference panel was assessed by comparing gold-standard HLA genotypes for a population-specific validation set with HLA allele dosages that were imputed by applying the panel with our updated SNP2HLA script to the same validation set. The validation set for South Asians was the sequence-based 2-field genotypes of eight HLA genes for the 397 people of our original SAS panel (UM+BKT+IKMB-SAS datasets). For people of European descent, four validation sets were used, consisting of subsets of four of our psoriasis case-control studies with independent genotyping for five HLA genes (Table S2). Leave-one-out cross-validation was used to compare imputed and genotyped dosages for people shared between a South Asian panel being assessed and the SAS validation set.

Two measures of imputation accuracy were used to compare gold-standard and imputed HLA genotypes. Per-individual accuracy was computed as proposed previously:[45]

$$1 - \frac{\sum_{a=1}^{m}[\delta(g_{ia} > x_{ia})](g_{ia} - x_{ia})}{2},$$

where m is the number of 1-field or 2-field alleles for a given HLA gene in the imputation reference panel, g_{ia} is the dosage for genotyped allele a for individual i, $x_{i,a}$ is the dosage for imputed allele a for individual i normalized to a sum of 2.0, and $\delta = 1$ if $g_{ia} > x_{ia}$ else $\delta = 0$. The distribution and mean of per-individual accuracy were then examined for all individuals in a validation set. Per-allele accuracy was computed as the squared Pearson correlation (r^2) of vectors of genotyped and imputed dosages for 1-field or 2-field HLA allele a across all individuals in the validation set. Based on these two measures of imputation accuracy, the relative performances of the SNP2HLA reference panels were assessed as described in Supplemental methods.

MHC variant imputation

The best-performing SNP2HLA reference panel for each combination of three groups of HLA genes (*HLA-A*, *-B*, *-C*, *-DQB1*, *-DRB1*; *HLA-DPA1*, *-DPB1*; *HLA-DQA1*) and two ancestries (SAS and EUR) was used to impute 1-field, 2-field, amino acid, SNP, and insertion or deletion (indel) alleles of HLA genes into either the two SAS or the eight EUR-ancestry psoriasis case-control studies. Before imputation, genotypes for all SNPs in a 20 Mb region (chr6: 20–40 Mb) encompassing the MHC were extracted from the genome-wide set of quality-controlled microarray genotypes for the target study to be imputed. SNPs from the target study were matched to those in the reference panel based on chromosomal position and allele identities. We then applied our improved SNP2HLA script and updated HLA sequence dictionaries to impute the reference panel variants into the target study, using 35 total iterations of the Beagle phasing algorithm. Imputed variant dosages were extracted for only those HLA genes for which the particular reference panel used provided optimal imputation accuracy.

For each of the 10 case-control studies, we also extracted imputed dosages for all variants in the chr6: 24–36 Mb region from the two genome-wide datasets that were imputed with the 1KGP and HRC reference panels. Biallelic variants were extracted from the 1KGP and HRC datasets as described earlier. Selection and processing of multiallelic variants is described in Supplemental methods.

For each ethnic dataset to be analyzed for association, we merged imputed HLA variant dosages from the three appropriate SNP2HLA panels with MHC variant dosages from the combined 1KGP and HRC panels. For SNPs in HLA genes that were duplicated because they were imputed by both the 1KGP/HRC and SNP2HLA panels, the variant with the lesser mean imputation

quality for the studies in the analysis was dropped. We restricted MHC variants used for association analysis to those with a predicted imputation quality (Minimac or Beagle r^2) of 0.7 or better for all case-control studies in an analysis.

Association analysis of MHC variants

We tested for association between MHC variants and psoriasis with a logistic regression model. We defined MHC variants to include both biallelic and multiallelic SNPs and indels in the 12 Mb extended MHC region, 1-field and 2-field classical HLA gene alleles, biallelic HLA amino acid polymorphisms for respective residues, and multiallelic HLA amino acid polymorphisms for respective positions. For MHC variants with m alleles ($m = 2$ for biallelic variants and $m > 2$ for multiallelic variants), we included $m - 1$ alleles as independent variables, excluding either the reference genome allele (1KGP/HRC variants) or the most frequent allele (SNP2HLA variants) as a reference, resulting in the following model:

$$\log(\text{odds}) = \beta_0 + \sum_{i=1}^{m-1} \beta_{1i} x_{1i} + \sum_{j=1}^{n} \beta_{2j} x_{2j} + \sum_{k=1}^{K} \left(\sum_{l=1}^{L_k} \beta_{3kl} x_{3kl} + \sum_{m=1}^{M_k - 1} \beta_{4km} x_{4km} \right) + \sum_{k=1}^{K-1} \beta_{5k} x_{5k} + \varepsilon,$$

where β_0 is the overall intercept, β_{1i} is the additive effect of the dosage of allele i for the variant x_{1i} being tested, and β_{2j} is the additive effects of the dosages of n optional conditioning variants x_{2j}. K is the number of case-control studies, and L_k and M_k are the numbers of study-specific principal components (PCs) and geographic indicator covariates for the k^{th} study (Table S3 tallies study-specific covariates used for controlling population stratification). Variable x_{3kl} is the l^{th} PC for the k^{th} study, x_{4km} is the m^{th} geographic cohort for the k^{th} study, and x_{5k} is the indicator variable for the study-specific intercept. β_{3kl}, β_{4km}, and β_{5k} are the effects of x_{3kl}, x_{4km}, and x_{5k}, respectively, and ε is the error term. Note that the inclusion of study-specific indicator variables assumes fixed effects among the case-control studies. The regression model was fitted using the *-glm* command of Plink 2.0,[46] with the *firth-fallback* option, which requests a standard logistic regression, followed by Firth regression whenever the logistic regression fails to converge. For each tested MHC variant, an omnibus p value for the association of its $m - 1$ alleles was determined with a multivariate Wald test, which follows a χ^2 distribution with $m - 1$ degrees of freedom. For multiallelic variants, in addition to the omnibus test, each of the m alleles were tested individually with biallelic Wald tests, including the reference allele. Additional details of association analysis of the MHC region are given in Supplemental methods.

Association model comparison

We used three measures to compare the goodness of fit of pairs of non-nested association regression models: (1) the Akaike Information Criterion (AIC), which was computed by the logistf R package[47] using log-likelihoods for models fitted with ordinary logistic regression and penalized log-likelihoods for models fitted with Firth's bias-reduced logistic regression; (2) the evidence ratio, which quantifies the relative likelihood of one model versus a second and can be computed from the AIC for each model;[48] and (3) Tjur's R^2, which quantifies the explanatory power of a logistic model by computing the difference in the means of the model-predicted probabilities of a binary outcome for cases and controls.[49] Individual contributions of three groups of regressors (PC and geographic cohort covariates, HLA-C*06, and all genetic variants other than HLA-C*06) to the AIC and Tjur's R^2 values of the full model were determined by decomposing the goodness of fit with the Shorrocks-Shapley procedure.[50]

Bayesian credible sets

Using a Bayesian approach, for association signals in the final regression models we identified the credible set of markers that were 95% likely, based on posterior probability (PP), to contain the causal disease-associated variant. This approach requires specification of a prior distribution for β, the effect size of the variant, where β is assumed to follow a normal distribution with a prior mean of 0 and a prior variance of W.[51] For biallelic loci, the original applications of Bayesian credible sets to disease GWASs[52,53] used a value of 0.2^2 for the prior variance, corresponding to a 90% prior probability that β lies in the interval [−0.165, 0.165]. Because of the wide range of expected effect sizes for psoriasis-associated MHC loci, for biallelic variants we instead computed a mean Bayes factor for a vector of $\sqrt{W}$ priors of 0.1, 0.2, 0.4, 0.8, and 1.6, as was first suggested by Wen and Stephens.[54] For multiallelic variants, we modified the approach of Wen[55] for computing Bayes factors for multiple biallelic SNPs, which uses a g-length vector of 0 s for the prior means and a $g \times g$ W_g matrix for the prior variances and cross-covariances for the effect sizes of g SNPs in a multiple regression model. We extended this multivariant approach to a single multiallelic variant with m alleles in a manner analogous to our association model, namely by decomposing the variant into its m biallelic components and dropping the biallelic variant for the most frequent allele (variants imputed with SNP2HLA panels) or the reference genome allele (variants imputed with 1KGP and HRC panels) to avoid complete linear dependency. Priors for the variances of the effect sizes of the remaining $m - 1$ biallelic variants were set to one of the five priors used for biallelic variants. Priors for the cross-covariances were determined empirically. For each variant in the final association models for the South Asian and European datasets, we dropped that variant and then repeatedly refitted the regression model for each neighboring (±500 kb) multiallelic variant. The variance-covariance matrices from these refitted regressions were converted to correlation matrices. Most (93.5%) of the resulting 4,861 upper-diagonal cross-correlations were positive due to inherent residual dependencies among the $(m-1)$ decomposed variants, but the variation in magnitude was large (SD = 0.22). To accommodate this heterogeneity, the first, third, fifth, seventh, and ninth deciles of this set of cross-correlations (0.011, 0.084, 0.177, 0.313, and 0.558), representing the midpoints of each of the five quantiles, were used instead of a single value to convert variances to cross-covariances. We then determined the mean of the approximate Bayes factors computed by formula (11) of Wen[55] for each of the 25 combinations of 5 variance priors and 5 cross-correlation priors.

For loci outside the extended MHC region, inclusion into Bayesian credible sets was restricted to biallelic variants present in both the SAS and EUR meta-analyses, well-imputed ($r^2 \geq 0.7$) for at least half of the participating case-control studies, and within 200 kb of the lead variant for the association signal being analyzed. For both biallelic and multiallelic MHC signals, the window for inclusion was increased to 500 kb to accommodate the unusually long-range LD that characterizes the MHC region. Credible sets for the individual monoethnic and transethnic MHC analyses were also automatically restricted to variants passing the

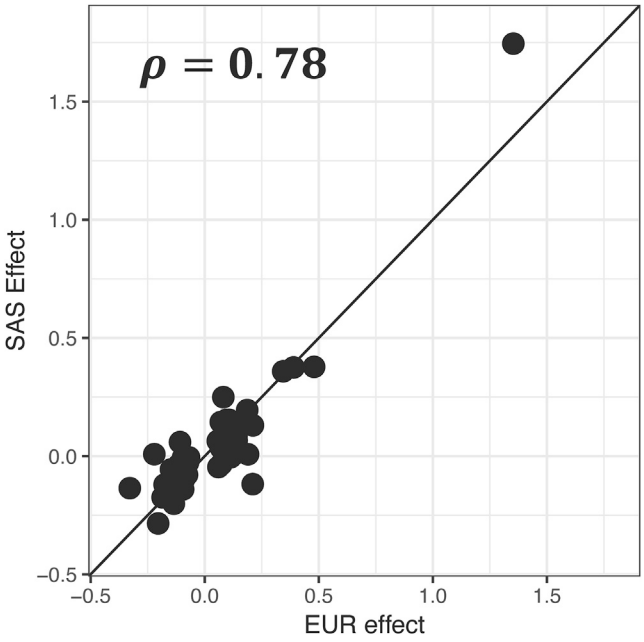

Figure 1. Correlation of effect sizes of psoriasis-associated loci in the European (x axis) and South Asian (y axis) GWASs
Log(OR) effect sizes are plotted; rho is the Spearman correlation coefficient. The plot includes 66 loci identified by past studies having a genome-wide significant ($p \leq 5 \times 10^{-8}$) association with psoriasis in European ancestry populations.

imputation quality threshold of $r^2 \geq 0.7$ imposed upon the studies contributing to a particular ancestry association model. For comparison of credible sets of the four stepwise-selected variants shared among the different MHC models, credible sets were recomputed for all three ancestry models after further restriction to variants with imputation quality of $r^2 \geq 0.7$ in all eight European and both South Asian studies.

Multiallelic linkage disequilibrium

We assessed 12 measures of linkage disequilibrium that can handle multiallelic variants. These metrics included five coefficients either reviewed (D_{hap}, D^*, multiallelic D') or proposed (r^2_{hap}, multiallelic r^2) by Zhao et al.,[56,57] Q^*,[58] W_n^2,[59,60] W_{ab}^2, and W_{ba}^2,[60] r^2 for a multiallelic variant collapsed down to its most common allele versus the rest,[46] ε',[61–63] and r^2_{max}, a metric we devised equal to the maximum biallelic r^2 among all possible pairings of the alleles for two loci. Performance was evaluated empirically for the European and South Asian datasets by computing pairwise LD between all loci in the final full psoriasis association models and their neighboring (±500 kb) variants. Relative magnitudes of the linear (Pearson r) and rank order (Spearman ρ) correlations between measured LD and the $-\log_{10}$ of the p value for each full-model variant were compared across metrics; equality of slopes for the linear fits of LD versus p value for each full-model variant with its neighboring biallelic versus triallelic versus 4^+-allelic variants was also assessed.

Overall, the best-performing measure was W_n^2, which reduces to the familiar r^2 LD coefficient for two biallelic loci. As pointed out by others,[60] W_n^2 is also known as Cramer's V statistic,[64] which is the χ^2 statistic for a contingency table relating two categorical variables, normalized to lie in a [0, 1] interval. Although we used W_n^2 as our primary measure of LD between variants, for annotation of psoriasis-associated variants we also used ε', which was a good performer and is unique among the 12 assessed metrics by being based on differences in the entropy of observed and expected haplotypes rather than differences in their frequencies. Our formulation of ε' multiplies the multiallelic extension[63] of the original multilocus ε statistic[61] by two, because as demonstrated by Liu and Lin,[62] ε can attain a maximum value of only $(n-1)/n$, where n is the number of loci considered.

Before measurement of LD between pairs of variants, fractional imputed dosages were converted to integer hard calls. For most analyses, individuals with poorer-quality imputed genotypes (>0.25 dosage units from an integer) for either variant of the pair were omitted. For analysis of the correlation of imputed genotype quality with association p value for variants in substantial LD with HLA-C*06, as well as for determining the number of strong LD proxies for association signals in the full transethnic MHC model, a stricter hard-call threshold of 0.10 dosage units was used instead.

Details of principal components analysis, determination of phenotypic variance explained, MHC variant annotation, and enrichment analysis are presented in Supplemental methods.

Results

Transethnic analysis reveals additional psoriasis loci

We first compared the European and South Asian GWAS signals, aiming to evaluate shared and unique effects. Despite the lack of power to replicate the established EUR signals in the SAS cohorts using a genome-wide significance threshold, there was a strong correlation between their effect sizes (Figure 1; Spearman's $\rho = 0.78$, $p < 2 \times 10^{-14}$), justifying a transethnic analysis of the two populations.

For the transethnic, EUR, and SAS meta-analyses, we analyzed 8.95 million, 9.01 million, and 9.20 million well-imputed markers, respectively. The meta-analyses revealed 50, 47, and 3 loci, respectively, for the transethnic, EUR and SAS cohorts, that were associated with psoriasis at a genome-wide level of significance (Figure S3). As shown in Table S4, none of the identified psoriasis loci exhibited significant effect size heterogeneity after correction for multiple testing except the primary MHC locus for the EUR and transethnic meta-analyses, which is likely a consequence of differences across the EUR studies in the proportion of purely cutaneous versus PsA cases.[25] The transethnic meta-analysis revealed two psoriasis-associated loci that were previously unreported for either the EUR or SAS populations (Table 1; Figures 2 and S4). Because our signal in 1p36.22 (rs2103876) is close (~200 kb) to two psoriasis-associated missense variants (rs2274976; rs5063) identified from a Chinese study,[65] we investigated pairwise LD among these variants using the EUR samples from the 1KGP. The results indicate $D' > 0.82$ and $r^2 < 0.015$ between rs2103876 and both of the missense SNPs, suggesting that they are in LD but exhibit different population allele frequencies: 1KGP risk allele frequencies for rs2103876 are 0.73 (Gujurati Indians in Huston [GIH]), 0.66 (Utah residents with northern and western European ancestry from CEPH collection [CEU]), and 0.67/0.61 (Southern Han Chinese [CHS]/

Table 1. Psoriasis loci established by the transethnic meta-analysis

Locus	Marker	Location (hg19)	RA/ NRA	SAS		EUR		Transethnic		
				OR (95% CI)	p value	OR (95% CI)	p value	OR (95% CI)	p value	Direction
1p36.22	rs2103876	1:12053100	T/C	1.19 (1.07–1.31)	7.96E−04	1.10 (1.06–1.14)	1.30E−07	1.11 (1.07–1.14)	1.18E−09	++−+++++++
1q24.2	rs12046909	1:168507463	C/T	1.21 (1.09–1.34)	3.85E−04	1.13 (1.08–1.18)	5.56E−07	1.14 (1.09–1.19)	1.68E−09	++++++++++

Abbreviations: Direction, the direction of effect of the risk allele for 10 studies: PsA GWAS, CASP GWAS, Kiel GWAS, Genizon GWAS, WTCCC2 GWAS, Exomechip, PAGE, GAPC, and two South Asian GWASs; EUR, European; NRA, non-risk allele; RA, risk allele; SAS, South Asian.

Han Chinese in Beijing [CHB]); for rs2274976 they are 0.78 (GIH), 0.93 (CEU), and 0.9/0.92 (CHS/CHB); and for rs5063 they are 0.86 (GIH), 0.93 (CEU), and 0.91/0.9 (CHS/CHB). Interestingly, the two Chinese missense variations show no evidence of association in our transethnic meta-analysis. However, we did identify potential regulatory roles for each of our two significant non-MHC signals (see Discussion). For the EUR meta-analysis, we also uncovered a genome-wide significant variation (rs77343625; 5:158208927; $p = 2.03 \times 10^{-8}$) >500 kb upstream of the best signal in the *IL12B* (MIM: 161561) region revealed previously (rs12188300; 5:158829527),[29] but this signal is no longer significant after conditioning on the known *IL12B* signal, indicating that its association results from long-range LD (spreading > 600 kb).

Transethnic Bayesian refinement of primary association signals

To investigate the resolution of localization of causal variants for 65 non-MHC psoriasis loci (63 primary EUR loci from past studies[1,7,8] and 2 transethnic loci established by this study), we compared 95% Bayesian credible sets (BCS) for these susceptibility loci in our EUR and transethnic meta-analyses. Among the 41 loci with stronger association signals in the transethnic than EUR meta-analysis, 19 have fewer markers in the 95% BCS for the transethnic model, 14 have the same number of markers, and 8 have more markers (Table S5). Specifically, for the two psoriasis loci established by the transethnic meta-analysis, the number of markers in the 95% BCS drops from 83 to 23 and from 30 to 10 in the transethnic versus EUR-only meta-analyses.

Conditional analysis identifies 12 secondary non-MHC loci

To identify independent signals for known and previously unreported psoriasis loci, we then conducted a transethnic conditional meta-analysis. Analysis was restricted to 46 non-MHC loci achieving genome-wide significant association in the unconditional transethnic analysis and that also had at least one marker well-imputed ($r^2 \geq 0.7$) for all 10 cohorts, which resulted in an analysis set of 113,745 markers well-imputed in all cohorts mapping to within 500 kb of the lead variant for each of these loci. Altogether, we were able to identify 12 independent signals in nine non-MHC loci (Table S6). Notably, six of the identified independent signals each harbor one genetic variation with a posterior probability ≥ 50% of being causative, with four of the loci (*IL23R* [MIM: 607562], *IFIH1* [MIM: 606951], *TRAF3IP2* [MIM: 607043], and *NFKBIA* [MIM: 164008]) encompassing 10 or fewer variants in their 95% CI set. Overall, all but two of the independent signals harbor fewer variants in the 95% BCS in the transethnic meta-analysis.

Improved imputation of HLA genotypes in both SAS and EUR

Many genetic studies of psoriasis have identified protein and amino acid alleles of HLA genes as potential susceptibility loci.[66] The 1KGP and HRC reference panels used to impute genotypes for the GWAS of this study do include some SNPs and indels within HLA genes but no amino acid or classical protein variants. A reference panel explicitly designed for HLA genes is needed for this purpose. We have had good success using the T1DGC reference panel with the SNP2HLA imputation package[23] to impute HLA genotypes for people of European ancestry.[22] Because no appropriate SNP2HLA panel exists for South Asians, we built a reference panel of 397 individuals of Indian and Pakistani ancestry with our improved version of the SNP2HLA MakeReference script and updated HLA amino acid and DNA sequence dictionaries. This new panel was representative of the population structure of the larger samples of Indians and Pakistanis from which they were drawn, as well as of the various SAS populations represented in the 1KGP (Figure 3). Unfortunately, its imputation accuracy was poor, especially for 2-field alleles that distinguish HLA proteins, where it was under 90% for all but one of eight classical HLA genes (Table S7). As a test, we randomly sampled increasingly large subsets of the T1DGC panel of 5,225 EUR-ancestry individuals, which were then used to build reference panels for HLA imputation of the 397 South Asians. As shown in Figure S5, although the SAS panel did outperform its 397-individual T1DGC counterpart, as the T1DGC subset panels increased in size, mean imputation accuracy increased well beyond that afforded by the SAS panel. We concluded that increasing the sample size of our SAS panel to at least 1,500–2,000 individuals might achieve acceptable imputation accuracies for most HLA genes, with even larger panels needed for imputing HLA-B and HLA-DRB1 genotypes accurately.

The preferred option of adding more South Asians to our panel was not feasible given the prohibitive cost of HLA

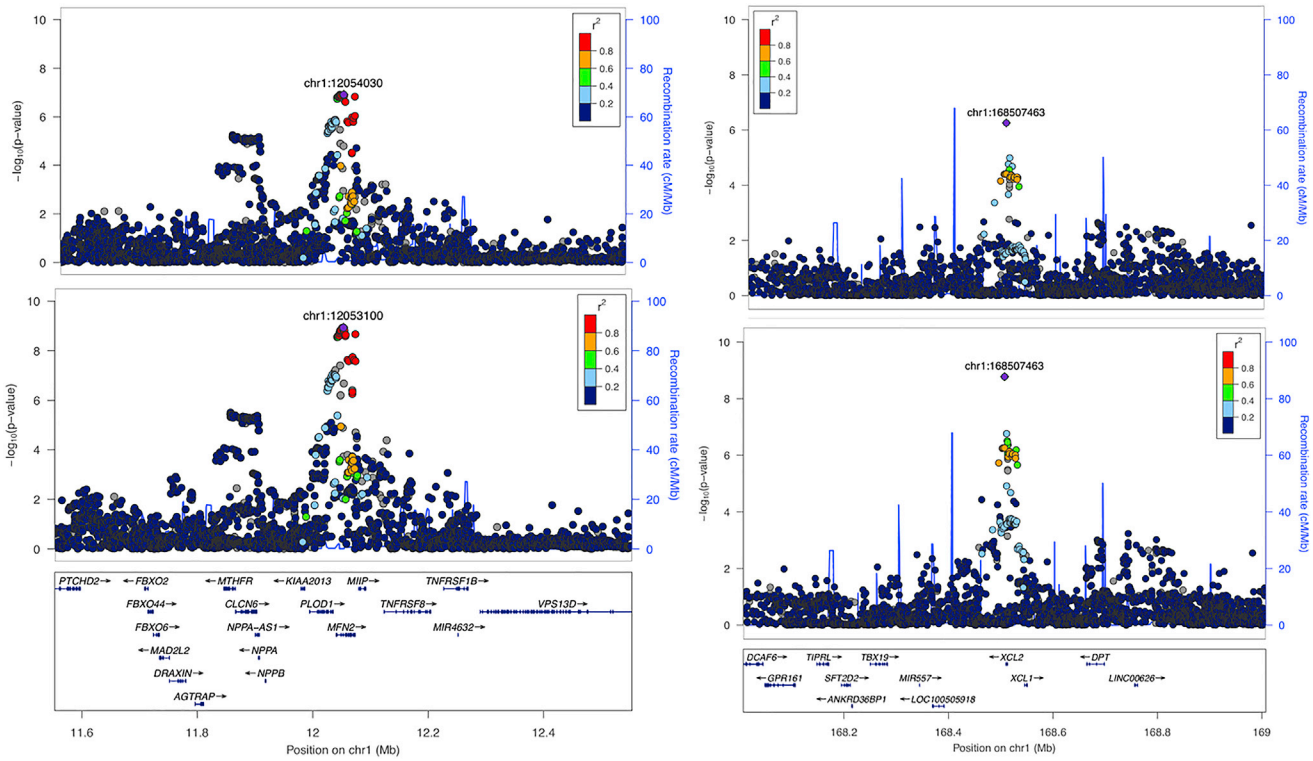

Figure 2. Association plots for psoriasis loci established by the transethnic meta-analysis
The regional association plots for the 1p36.22 and 1p24.2 loci (left and right pairs of panels) are shown for results from the European and transethnic meta-analyses (top and bottom pairs of panels).

genotyping. Past work combining a small population-specific SNP2HLA panel with panels from other populations achieved increased imputation accuracy for the population of interest.[42,43] Similarly, large multi-population reference panels have shown good performance when used with other methods of HLA genotype imputation such as HLA*IMP[67] and HIBAG.[38] Based on these promising findings, we combined data from several sources (Table S8) to build 19 SNP2HLA reference panels for imputing HLA genotypes in South Asians, varying in size from 397–9,343 people of one to several continental ancestries (Table S9). Each panel was then used as a reference with our updated SNP2HLA script to impute 2-field HLA genotypes for the 397 people of the original SAS panel, using leave-one-out cross-validation to generate imputed dosages for any people shared between the panel and target dataset. For the five HLA genes included in all panels, the multiethnic IKMB+BKT+UM+1KGP-ALL-v2 panel gave the best results (Figures S6 and S7), with mean sample accuracies of 93%–97% (Table S10). Among the 11 panels with genotypes for *HLA-DPA1*, *HLA-DPB1*, and *HLA-DQA1*, the IKMB+BKT+UM panel performed best for *HLA-DPA1* and *HLA-DPB1* (Figures S8 and S9), and the IKMB+BKT+UM+T1DGC panel performed best for *HLA-DQA1* (Figures S10 and S11), with mean accuracies ranging from 93%–94% (Table S10). Imputation of *HLA-DQA1* was assessed separately from that for *HLA-DPA1* and *HLA-DPB1* because of *HLA-DQA1* genotyping issues for two of the source datasets (T1DGC and Pan-Asian). As shown in Table S11, the multiethnic IKMB+BKT+UM+1KGP-ALL-v2 reference panel also provided generally excellent imputation results for the global populations of 1KGP, with mean sample accuracies at 2-field HLA resolution of 92%–98% for African, East Asian, EUR, and SAS populations, each constituting 20%–24% of the panel, and 83%–97% accuracy for admixed Americans that constitute only 8% of the panel (Table S8).

We then repeated the procedures used for building South Asian panels to create 20 panels tailored for people of European ancestry (Table S12). Each panel was used to impute 2-field HLA genotypes for four of our psoriasis case-control studies of European ancestry with independent gold-standard HLA genotyping (Table S2). Comparison of imputation accuracy for the five HLA genes in all 20 panels shows that the European-ancestry T1DGC+1KGP-EUR-v2 panel performed best (Figures S12 and S13), with mean sample accuracies of 95%–99% (Table S13). Among the seven panels with *HLA-DPA1*, *HLA-DPB1*, and *HLA-DQA1* genotypes, the IKMB+BKT panel performed best for imputation of *HLA-DQA1* (Figures S14 and S15), achieving a mean sample accuracy of 96% (Table S13). Because we had no independent genotypes for *HLA-DPA1* and *HLA-DPB1*, we assessed panels for imputing these two genes based on imputation accuracies for *HLA-A*, *-B*, *-C*, *-DQB1*, and *-DRB1* as a proxy and found the T1DGC panel performed best (data not shown).

While this study was under review, a large multiethnic HLA reference panel of 21,546 individuals of European,

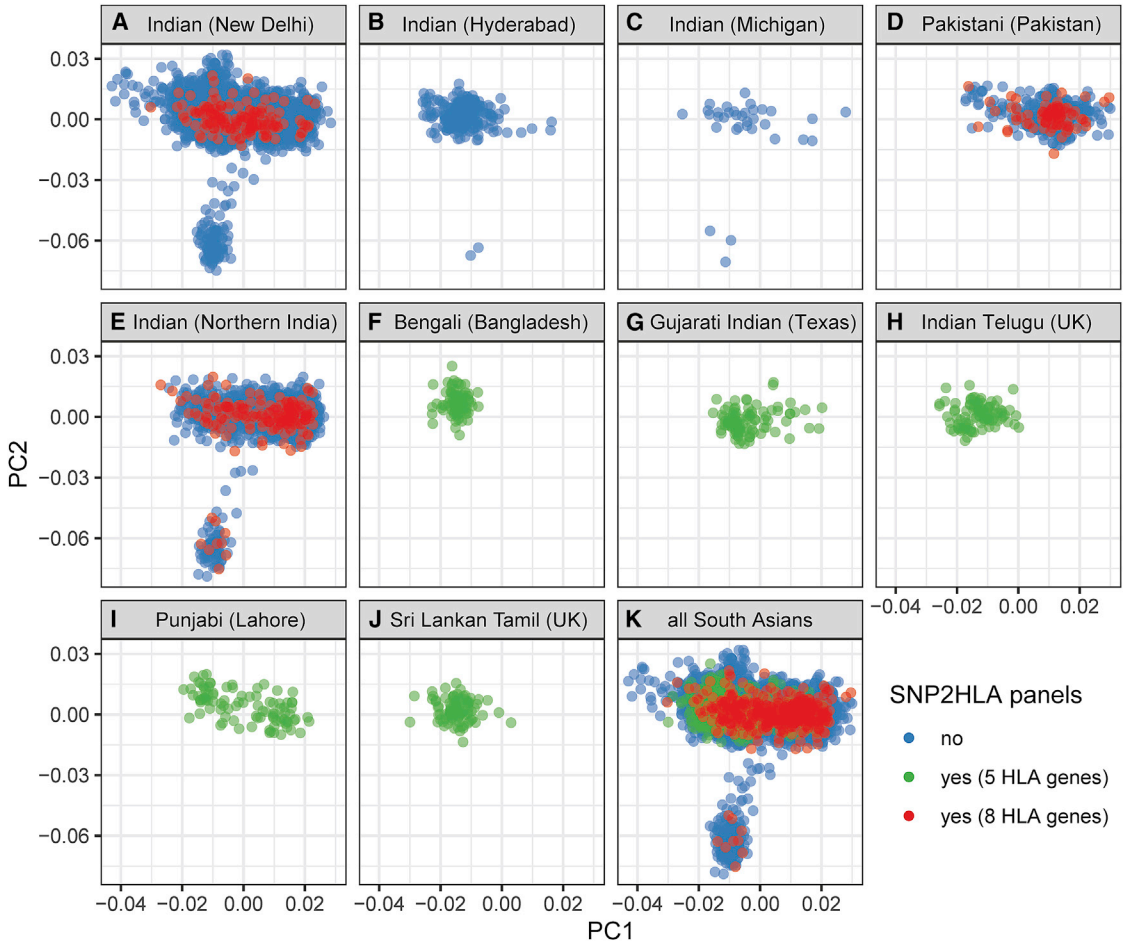

Figure 3. Population structure of South Asians in the SNP2HLA reference panels
The first two principal components from a PCA of 6,421 individuals of South Asian ancestry are shown. (A–J) Results divided by 10 different source populations: populations in (A)–(D) are from the psoriasis GWAS of this study, the population in (E) is from an inflammatory bowel disease case-control cohort of B.K. Thelma, and populations in (F)–(J) constitute the SAS superpopulation of phase 3 of the 1000 Genomes Project. (K) Results for all populations combined. Points are colored by whether they are part of any of the SNP2HLA panels used for imputation of HLA variants in South Asians; blue, green and red colors indicate whether the individual was not part of any panel, was part of panels used for imputation of 5 HLA genes (A, B, C, DQB1, DRB1), or was part of panels used for imputation of 8 HLA genes (A, B, C, DPA1, DPB1, DQA1, DQB1, DRB1), respectively.

admixed African, East Asian, and Latino ancestry[68] became available for use via the Michigan Imputation Server.[69] We compared imputation results for our SAS and EUR sample validation sets obtained with this large panel versus what we achieved with our suite of best-performing EUR and SAS HLA panels. As shown in Figure S16, imputation accuracies with our panels equal or exceed those obtained with the new multiethnic panel for both SAS and EUR target samples and both 1-field and 2-field HLA allele resolutions. The gains in accuracy with our panels were generally larger for class II genes, especially for *HLA-DQA1*, where the conversion of G-group HLA alleles to 2-field alleles that is employed by the newly published panel is most inaccurate.

The three best South Asian reference panels were used with the improved SNP2HLA script to impute HLA genotypes into the two psoriasis case-control studies of SAS ancestry. Similarly, the three best European reference panels were used for imputation into eight case-control studies of EUR ancestry. From the SNP2HLA-imputed genotype datasets, we extracted 1-field, 2-field, amino acid, SNP, and indel variants within those HLA genes for which the panel used was optimal. A comparison of frequencies of imputed 1-field and 2-field HLA alleles reveals many differences between the EUR and SAS populations (Figures S17–S20), which correspond closely to differences in genotyped allele frequencies between these populations published by the National Marrow Donor Program[70] (Figures S21 and S22). We expanded our scope beyond classical HLA genes by extracting imputed genotypes for all 1KGP/HRC panel variants in a 12 Mb region (chr6: 24–36) that encompasses the classical MHC region (chr6: 29.64–33.12 Mb), the extended MHC (chr6: 25.73–33.37 Mb) defined by Horton et al.,[71] as well as flanking sequence. As shown in Table S14, the density of coding, non-coding, and immune-related genes exceeds the genome-wide average for most segments of the 12 Mb extended MHC, peaking within the classical interval. Frequency distributions of the approximately 288,000 imputed MHC variants, cross-

classified by ancestry and MHC region versus reference panel source, MAF and imputation quality, or variant type are shown in Tables S15–S17, respectively.

MHC fine-mapping uncovers multiple independent SAS, EUR, and transethnic loci

Our previous fine-mapping study of the classical MHC[22] identified several HLA gene variants that may be driving the multiple independent psoriasis association signals in the region for people of European ancestry. For this study we extended these fine-mapping efforts to South Asians, built upon past work for people of European ancestry with a dataset that includes more individuals and more variants of generally higher imputation quality, and combined our South Asian and European studies to perform a transethnic association analysis.

Association analysis of the extended MHC was restricted to variants imputed with good accuracy in all participating studies (Table S18). Tables S19–S21 present frequency distributions of the tested MHC variants, cross-classified by ancestry, MHC region, panel source, MAF, imputation quality, and variant type. Figure 4 plots unconditional association with psoriasis in the extended MHC region for all three ancestry analyses. The most strongly associated variant in the SAS analysis is *HLA-C*06*, and the top variant in both the EUR and transethnic analyses is rs12211087, a SNP lying 30 kb upstream of *HLA-C* that is in nearly perfect LD with *HLA-C*06* ($r^2 = 0.986$ and 0.990 in SAS and EUR populations, respectively). The unconditional effect size of the lead variant is very large in all cases—odds ratio (OR) (95% confidence interval [CI]) = 5.80 (5.02–6.71), 3.93 (3.75–4.13), 4.09 (3.90–4.28) for SAS, EUR, and transethnic, respectively—but significantly greater for South Asians than Europeans ($p = 7.4 \times 10^{-8}$).

Stepwise analysis identified five independent South Asian MHC psoriasis susceptibility loci (Figures 5, S23, and S24). Pairwise LD values suggested that the five lead variants are mutually independent (Figure S25). Full model effect sizes and association p values indicated that *HLA-C*06* is much more strongly associated with psoriasis in the SAS population than the other four selected variants, contributing 66% to the total variance in disease liability explained by all five loci (Table 2). The second most strongly associated variant is triallelic SNP rs2428489, whose association is mostly driven by its C allele (OR [95% CI] = 1.64 [1.45–1.85], $p = 1.8 \times 10^{-15}$). All five identified variants are situated near one or more classical HLA genes (Figure 5), but the closest genes for three of the variants are not classical HLA genes (Table S22). Furthermore, only one of the variants other than *HLA-C*06* (indel rs139451799) has a plausible protein-changing surrogate (amino acid 13 or 142 of *HLA-DRB1*) based on two LD measures (Table S22) and comparisons of the magnitude and rank of association p value and Bayesian posterior probability between the variant and its potential surrogate (Tables 2, S22, and S23). Two variants (*HLA-C*06* and rs2442757) have a substantial posterior probability of being causative (0.255 and 0.583, respectively), although the size of the 95% Bayesian credible set for rs2442757 is very large, including 2,868 variants and spanning nearly 1 Mb.

Analysis of the much larger European-ancestry dataset identified 14 independent MHC loci associated with psoriasis (Figures S26–S29). Note that unconditional lead variant rs12211087 (Figure 4) was removed by the backward elimination step after forward selection of *HLA-C*06:02* in the ninth round of the stepwise procedure. Most of the loci in the final EUR regression model are independent of each other, but modest LD ($W_n^2 < 0.5$) is seen between some variants mapping near *HLA-B* and *HLA-C* (Figure S30). The strength of association of top-ranking variant *HLA-C*06:02* is not as dominant as seen for South Asians, contributing 47% and 58% of the variance in disease liability explained for all and the top five ranking MHC loci, respectively (Table S24). As shown in Table S25, five of the 14 lead variants alter an HLA protein (HLA-B amino acid positions 67 and 171, *HLA-C*06:02*, rs41543814, and HLA-DQA1 Arg52), two (rs137854633 and rs371194629) lie within an HLA gene (*HLA-B* and *HLA-G* [MIM: 142871]), and another (rs72866766) is just 468 bp downstream of *HLA-B*. The location within three-dimensional (3D) ribbon models of five HLA-C and HLA-B amino acids altered by EUR risk variants is illustrated in Figure 6. However, none of the nine non-coding variants have any convincing protein-changing surrogates. Notably, three coding variants (amino acids 67 and 171 of HLA-B, rs41543814) and three non-coding variants (rs1655901, rs72866766, and rs4947340) have strong support for being causative, with posterior probabilities ranging from 0.965–1 (Table S26).

Fine-mapping of the transethnic dataset revealed 17 independent psoriasis loci in the extended MHC region; their lead variants all lie within the classical MHC (Figures S31–S35). As was true of the two single ancestry analyses, *HLA-C*06* was the top-ranking variant in the full model (OR [95% CI] = 3.18 [2.95–3.43]; $p = 3.2 \times 10^{-200}$), although position 67 of HLA-B was also very strongly associated (multiallelic $p = 1.3 \times 10^{-134}$; OR [95% CI] = 1.98 [1.86–2.10], 1.30 [1.20–1.40], and 1.25 [1.18–1.33] for its cysteine, methionine, and tyrosine residues, respectively). Most of the transethnic signals are independent of each other, with only moderate LD between *HLA-C*06:02* and rs2844626 in South Asians ($W_n^2 = 0.40$) and modest LD ($0.20 \leq W_n^2 < 0.40$) for seven other pairs of variants in at least one population (Figure S36). LD patterns are broadly similar in EUR and SAS for this set of loci (Figure S37), although there are some substantial differences (e.g., W_n^2 between amino acid 67 of HLA-B and rs2844626 is 0.36 in Europeans and only 0.14 in South Asians). Only two loci in the final transethnic model are protein changing (*HLA-C*06:02* and position 67 of HLA-B), while four others occur in genes: rs1148117870 in the 5′ UTR of *HLA-B*, and rs1736927, rs112540072, and rs559509014/rs147145279 in introns of *HLA-G*, *TSBP1* [MIM: 618151], and *HCP5* [MIM: 604676], respectively

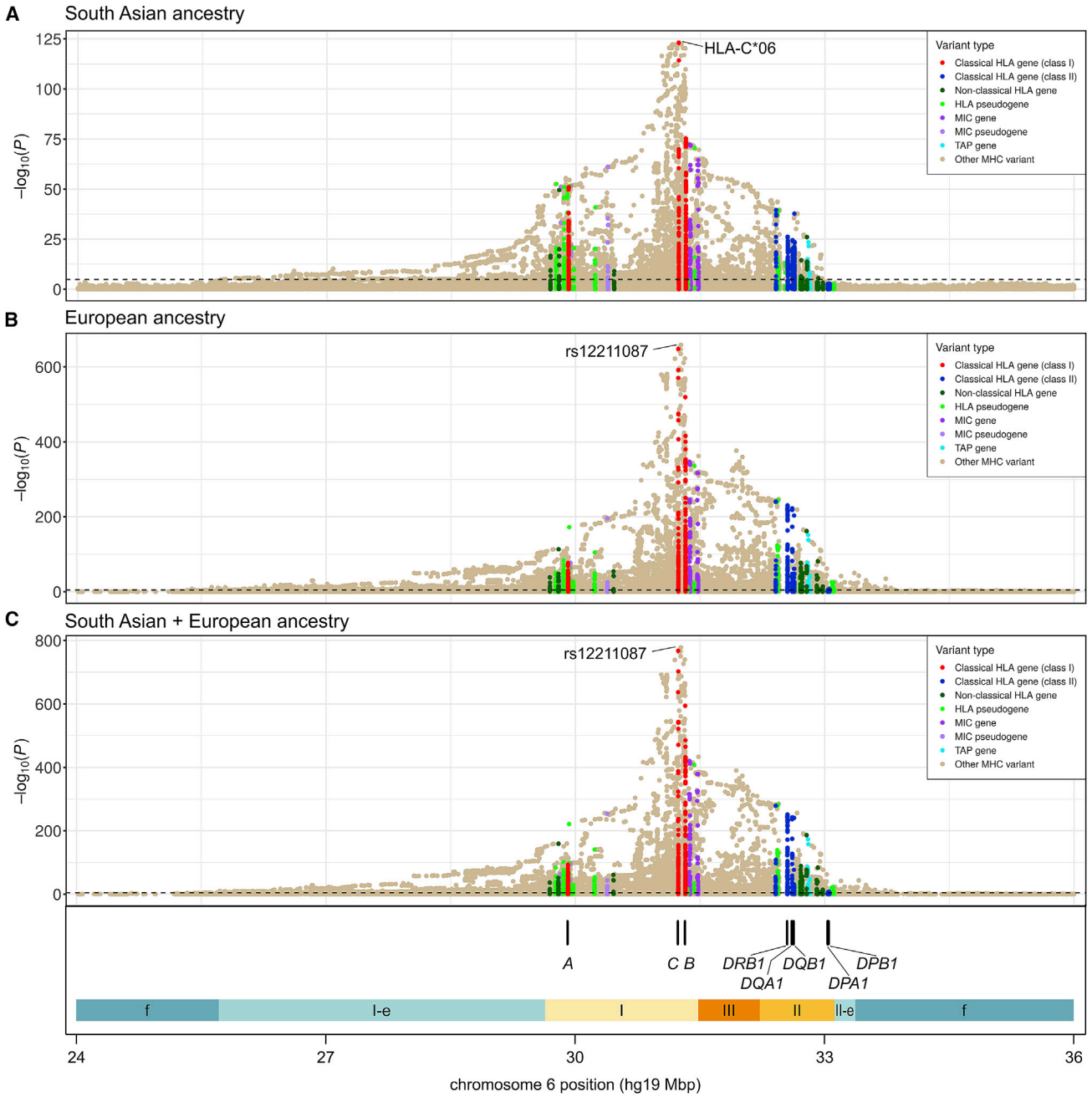

Figure 4. Plots of unconditional psoriasis association for the MHC and flanking regions
Each circle represents the −log(p) of association of an imputed variant, color-coded based on its membership in various categories of MHC genes, as detailed in the keys. The dashed lines denote thresholds of Bonferroni-corrected significance of 0.05. The locations of the eight HLA genes for which amino acid and protein alleles were imputed are shown at the bottom, along with colored segments denoting the boundaries of the classical MHC region (class I, II, and III), the extended MHC class I (I-e) and II (II-e) regions of Horton et al.,[71] and flanking MHC regions (f). We tested association in people of three different ancestries: South Asian (A), European (B), and South Asian and European combined (C).

(Tables S27 and S28). Of the 15 non-coding variants, only one has a plausible protein-changing surrogate (triallelic indel rs147145279 with surrogate biallelic missense SNP rs41556715 in *MICA* [MIM: 600169]) based on examination of two LD metrics in both EUR and SAS (Table S28), as well as a comparison of p values and posterior probabilities of the lead variant with possible surrogates (Tables S27–S29). Bayesian posterior probabilities are very high for three of the variants (amino acid 67 of HLA-B, rs1655901, and rs2853998) and exceed 0.50 for three others (*HLA-C*06:02*, rs2884626, and rs9271539).

HLA protein-changing variants were highly (11.8-fold) and significantly ($p = 4.0 \times 10^{-5}$) enriched in the final EUR model compared to their proportion among tested classical MHC variants (Table S30). Similar albeit non-significant enrichments of HLA coding variants were

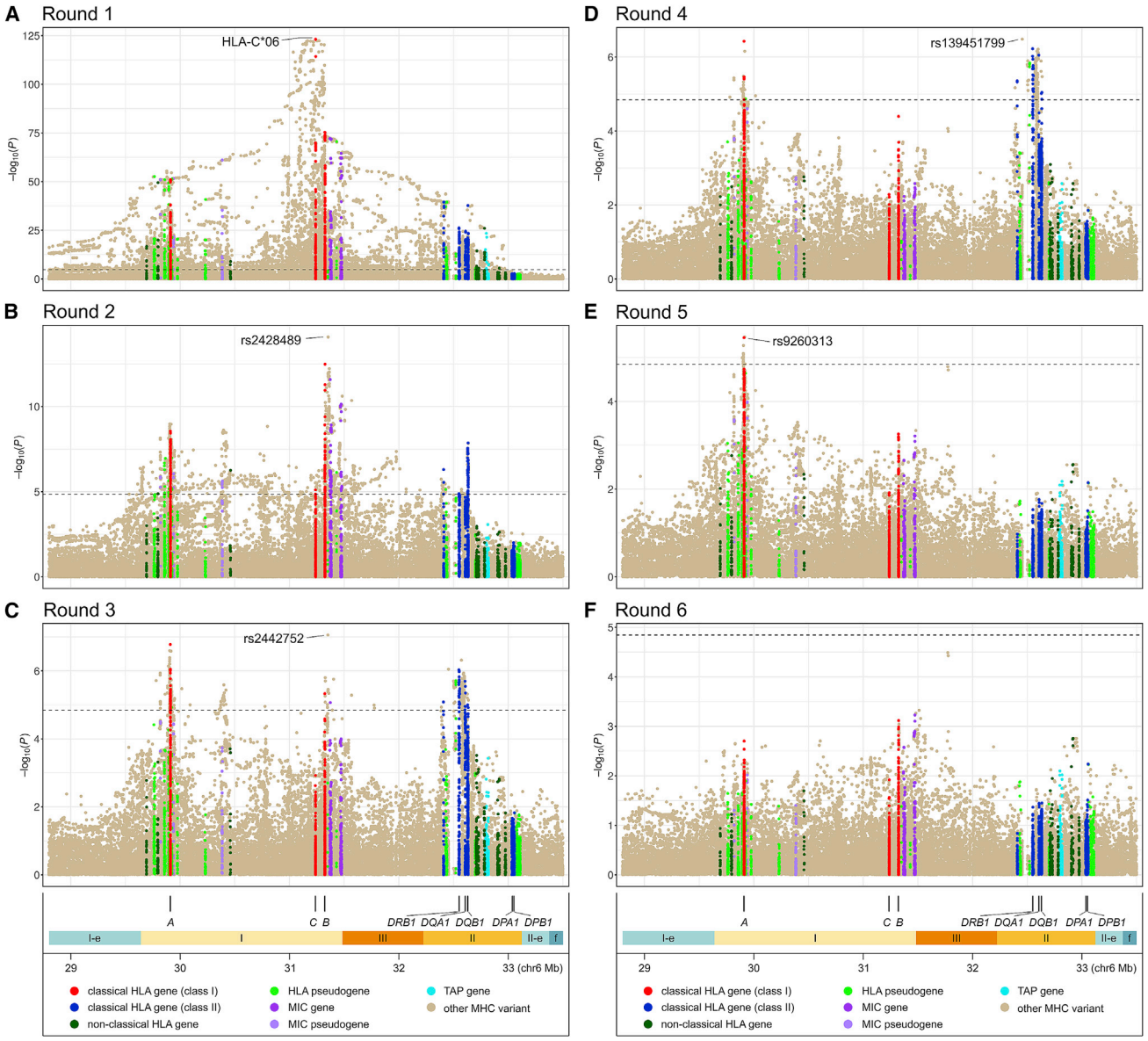

Figure 5. Plots of stepwise analysis of psoriasis association in the extended MHC region in people of South Asian ancestry
(A)–(F) Association results after each of the six rounds of stepwise regression. Each circle represents the −log(p) of association of an imputed variant, color-coded based on its membership in various categories of MHC genes, as detailed in the key at the bottom. Dashed lines denote thresholds of Bonferroni-corrected significance of 0.05. The locations of the eight HLA genes for which amino acid and protein alleles were imputed are shown at the bottom, along with colored segments denoting the boundaries of the classical MHC region (class I, II, and III), the extended MHC class I (I-e) and II (II-e) regions, and flanking MHC regions (f).

observed for the SAS and transethnic models (6.8- and 3.6-fold, respectively). Notably, no such enrichments were observed for protein-changing variants of non-HLA genes. Multiallelic variants were also greatly enriched in all three association models for both the classical and extended MHC regions (Table S30). Compared to the whole genome, both the classical and extended MHC regions show strong enrichment for protein-changing variants and modest enrichment for structural and multiallelic variants (Table S31).

Complete results for the stepwise conditional analysis of SAS, EUR, and transethnic associations in the extended MHC region are provided in Table S32.

SAS and EUR MHC models both similar and dissimilar

We found evidence both for and against the hypothesis that genetic contributions of the MHC region to psoriasis are similar between SAS and EUR.

The top signals in the two association models (*HLA-C*06* in SAS and *HLA-C*06:02* in EUR) are essentially identical (LD $r^2 = 0.9994$ and 0.9998 in SAS and EUR, respectively). There is also good correspondence of rs9260303 in the EUR model with rs1655901 in the SAS model; these two SNPs are only 81 bp apart and 3 kb downstream of HLA-A (Tables S22 and S25) and are in substantial LD ($r^2 = 0.70$ and 0.49 in SAS and EUR, respectively). Furthermore, effect sizes for the five variants in the SAS model are strongly and

Table 2. Psoriasis associations from stepwise meta-analysis of the extended MHC region for two studies of South Asian ancestry

			Alleles		Risk allele frequency		Association at entry into model		Association in final full model		
Step[a]	Variant[b]	chr6 position[c]	Risk[d]	Nonrisk	Cases	Controls	OR (95% CI)	p value	OR (95% CI)	p value	V_g
1	*HLA-C*06*	31238192	*C*06*	other	0.3361	0.1012	5.80 (5.02–6.71)	6.7×10^{-124}	4.68 (3.99–5.50)	1.3×10^{-78}	0.04822
2	rs2428489	31352972	C, T	A	NA	NA	NA	8.2×10^{-15}	NA	9.6×10^{-16}	0.01001
			C	other	0.4094	0.2654	1.41 (1.26–1.58)	2.1×10^{-9}	1.64 (1.45–1.85)	1.8×10^{-15}	NA
			T	other	0.0840	0.1044	0.71 (0.87–1.29)	1.4×10^{-4}	1.06 (0.87–1.29)	0.57	NA
			other	A	0.4934	0.3697	ref	ref	ref	ref	NA
3	rs2442752	31351764	T	C	0.6508	0.5154	1.39 (1.23–1.56)	8.8×10^{-8}	1.39 (1.23–1.57)	8.8×10^{-8}	0.00576
4	rs139451799	32454479	–	other	0.2295	0.2532	1.46 (1.26–1.69)	3.3×10^{-7}	1.43 (1.23–1.65)	1.6×10^{-6}	0.00521
5	rs9260313	29916885	T	C	0.6593	0.5341	1.29 (1.16–1.44)	3.4×10^{-6}	1.29 (1.16–1.44)	3.4×10^{-6}	0.00345

Abbreviations: chr6, chromosome 6; CI, confidence interval; NA, not applicable; OR, odds ratio; ref, reference; V_g, variance in liability explained by the genetic variant.[72]
[a]Round of stepwise regression analysis.
[b]Variant notes: variant ID is build 151 dbSNP rsID when applicable; HLA-C*06 is one biallelic split from a decomposed set of 14 classical 1-field HLA-C alleles; the stepwise-selected variant for triallelic indel rs139451799 is one of its biallelic splits with − versus A+G alleles.
[c]Base pair position in hg19 human reference; for classical HLA proteins the position of the center of the coding unit is given; for indels (all of which are insertions into the reference sequence), the position immediately before the insertion point is given.
[d]Risk allele is based on final full regression model.

significantly correlated with their effect sizes when re-estimated for the EUR dataset; this is true both for full model ($r = 0.96$, $p = 0.0026$) and unconditional model ($r = 0.97$, $p = 0.0012$) coefficients (Figures 7A and 7B). Effect sizes for variants in the EUR model are also significantly correlated with their re-estimated values in the SAS dataset, whether determined for all 14 variants (Figure S38) or for only the top five variants (Figures 7C and 7D). Finally, the explanatory power of the within-population versus cross-population fits of the SAS and EUR models as measured by Tjur's R^2 are similar: $R^2 = 0.268$ for SAS model in SAS versus 0.257 for top 5 of EUR model in SAS, and $R^2 = 0.304$ for top 5 of EUR model in EUR versus 0.296 for SAS in EUR (Table S33).

However, there are also differences between the MHC models for SAS and EUR. First, as was true of the unconditional models, the full model effect size for *HLA-C*06* in SAS is significantly greater than that seen for *HLA-C*06:02* in EUR (OR [95% CI] = 4.68 [3.99–5.50] versus 2.80 [2.59–3.03]; $p = 1.9 \times 10^{-8}$). Second, as shown in Figure 8, there is at best only weak correspondence between three of the five SAS loci with any of the top-ranking EUR loci. Finally, examination of goodness-of-fit measures shows much stronger statistical support ($\Delta AIC > 10$ and evidence ratios > 50)[48] for within-population MHC models than cross-population models (Table S33).

We suspect that the underlying MHC association signals in SAS and EUR are largely the same, but relatively low power (Table S1), coupled with generally lower genotype imputation quality (Table S34), limits accurate identification of lead variants for SAS association signals.

Transethnic MHC model superior to monoethnic models

We used pairwise LD measures to assess the correlation of variants in the monoethnic MHC association models for SAS and EUR with those of the transethnic model. As illustrated in Figure S39, one variant in the SAS model (*HLA-C*06*) has a nearly identical surrogate (*HLA-C*06:02*) in the transethnic model, and two others exhibit moderately

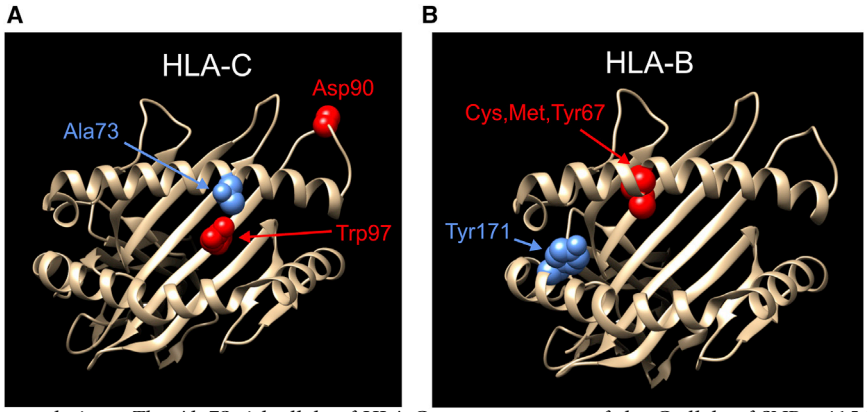

Figure 6. Protein locations of five HLA amino acid variants with a posterior probability ≥ 0.5 of being causative in at least one of the three full MHC association models (EUR, SAS, transethnic)
Three-dimensional ribbon models for the α chains of HLA-C (A) and HLA-B (B) are based on Protein Data Bank entries 4nt6 and 2bvp, respectively, and were created with UCSF Chimera version 1.15.[73] For HLA-C, shown in red is the pairwise combination of Asp90 and Trp97 that makes the HLA-C*06:02 psoriasis risk allele unique among all 2-field HLA alleles in the minimal allele set constituting a total of 99.9% frequency for the SAS and EUR populations. The Ala73 risk allele of HLA-C, a consequence of the C allele of SNP rs41543814, is shown in blue. For HLA-B, the three residues at position 67 that significantly increase risk for psoriasis in both SAS and EUR are shown in red, and the Tyr171 risk allele is shown in blue.

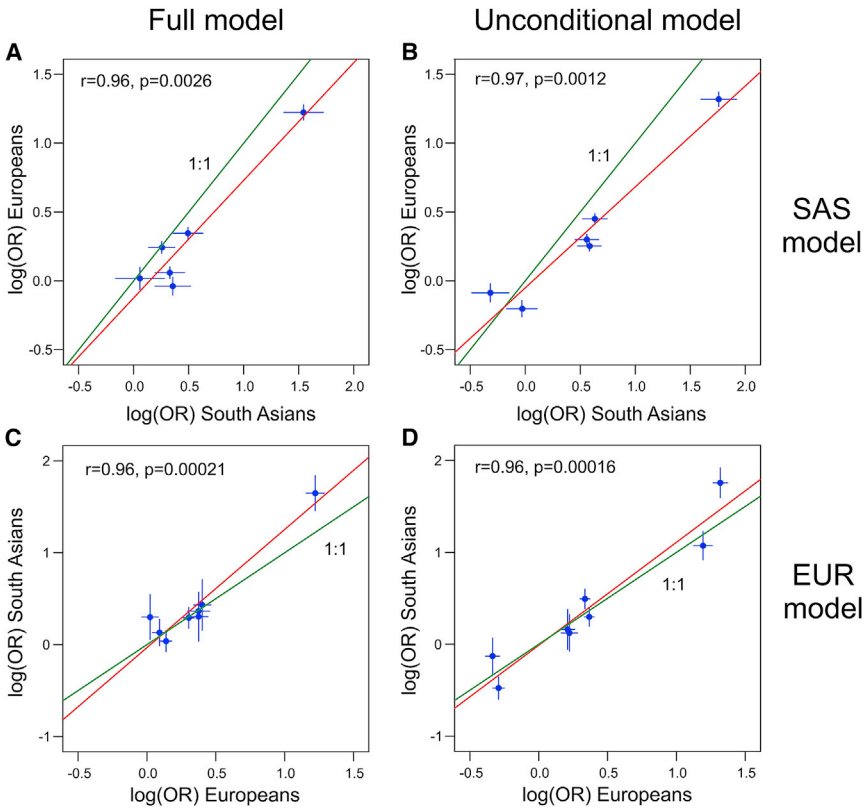

Figure 7. Comparison of association effect sizes in Europeans versus South Asians for the top five variants in the South Asian and European regression models for the MHC region

In (A) and (B), the log(OR) of each variant in the South Asian model as estimated in the European dataset is plotted against their estimates in the South Asian dataset. Conversely, in (C) and (D), the log(OR) of each of the top five variants in the European model as estimated in the South Asian dataset is plotted against their estimates in the European dataset. (A) and (C) plot effect sizes for the final regression model containing all variants; (B) and (D) plot unconditional effect sizes for each variant with no other variants in the regression model. Multiallelic variants with m alleles are represented by the set of $m - 1$ decomposed biallelic variants used for the joint Wald test. The vertical and horizontal bars show the 95% confidence intervals for these estimates in each dataset. Green and red lines depict a 1:1 correspondence and a linear fit, respectively. The Pearson correlation coefficient and its significance are shown in the upper left corner of each plot.

strong LD ($W_n^2 = 0.70$ for rs9260313 and rs1655901) or moderate LD ($W_n^2 = 0.51$ for rs2428489 and rs9266716) with transethnic variants. A similar proportion of EUR model variants have potential counterparts in the transethnic model (Figure S40): four shared loci (rs1655901 near *HLA-A*, *HLA-C*06:02*, position 67 of HLA-B, and rs147145279), one variant in strong LD ($W_n^2 = 0.81$ for rs6935999 and rs28573770), one in moderately strong LD ($W_n^2 = 0.73$ for rs371194629 and rs1736927), and three in moderate LD ($W_n^2 = 0.55$, 0.43, and 0.43 for rs6935999 with rs28573770, rs1128175 and position 67 of HLA-B, and rs41543814 and rs2844626, respectively). We compared 95% BCS for the four variants in the transethnic model that are shared with either of the two monoethnic models, finding a large increase in posterior probability for *HLA-C*06:02* in the transethnic versus the SAS or EUR models (0.664 versus 0.255 and 0.288) (Tables S23, S26, S29, and S37). The 95% credible sets for position 67 of HLA-B and rs1655901 remained unchanged for the transethnic model compared to the European model (i.e., each variant is the sole member of its set with Bayesian posterior probability = 1.000). In contrast, for the fourth shared variant, rs147145279, the size of the credible set increased and the posterior probability of the lead variant decreased for the transethnic model (Table S35). We also assessed the goodness of fit of the transethnic model to that of the two monoethnic models, finding that predictive performance of the transethnic model, as assessed by Tjur's R^2, was in all cases better than its monoethnic counterparts, but that the increase in performance was slight (Table S36). Decomposing the contribution of different regressors to Tjur's R^2 indicated that variants other than *HLA-C*06* were mostly responsible for the slight edge in goodness of fit of the transethnic models (Table S36).

Regulatory effects of non-coding MHC variants

We compiled various potential regulatory features for nine non-coding loci in the final association models that have posterior probabilities exceeding 0.50. These features included *cis*-expression quantitative trait locus (eQTL) effects (Tables 3 and S37), chromatin state (Figure 9; Table S38), as well as conservation metrics, transcription factor binding data, and scores for the overall likelihood of being regulatory or deleterious (Table S38). Allelic variation at all but one of the nine loci was significantly associated with transcription levels of a specific target gene in a majority of the tested psoriasis-relevant tissues, and there was remarkable consistency across tissues in the direction of the effect of the risk allele on transcription for any given eQTL-gene pair. Nearly all the top target genes for these *cis*-eQTL effects have a known role in human immunity, including 10 different HLA genes as well as *AGER* (MIM: 600214), *DDX39B* (MIM: 142560), *HSPA1L* (MIM: 140559), *MICA*, *MICB* [MIM: 602436), *PSMB9* (MIM: 177045), and *NCR3* (MIM: 611550). One of the more interesting of these potential regulatory variants is rs137854633, whose SiPhy score indicates significant evolutionary sequence constraint, and which has modestly strong RegulomeDB and CADD

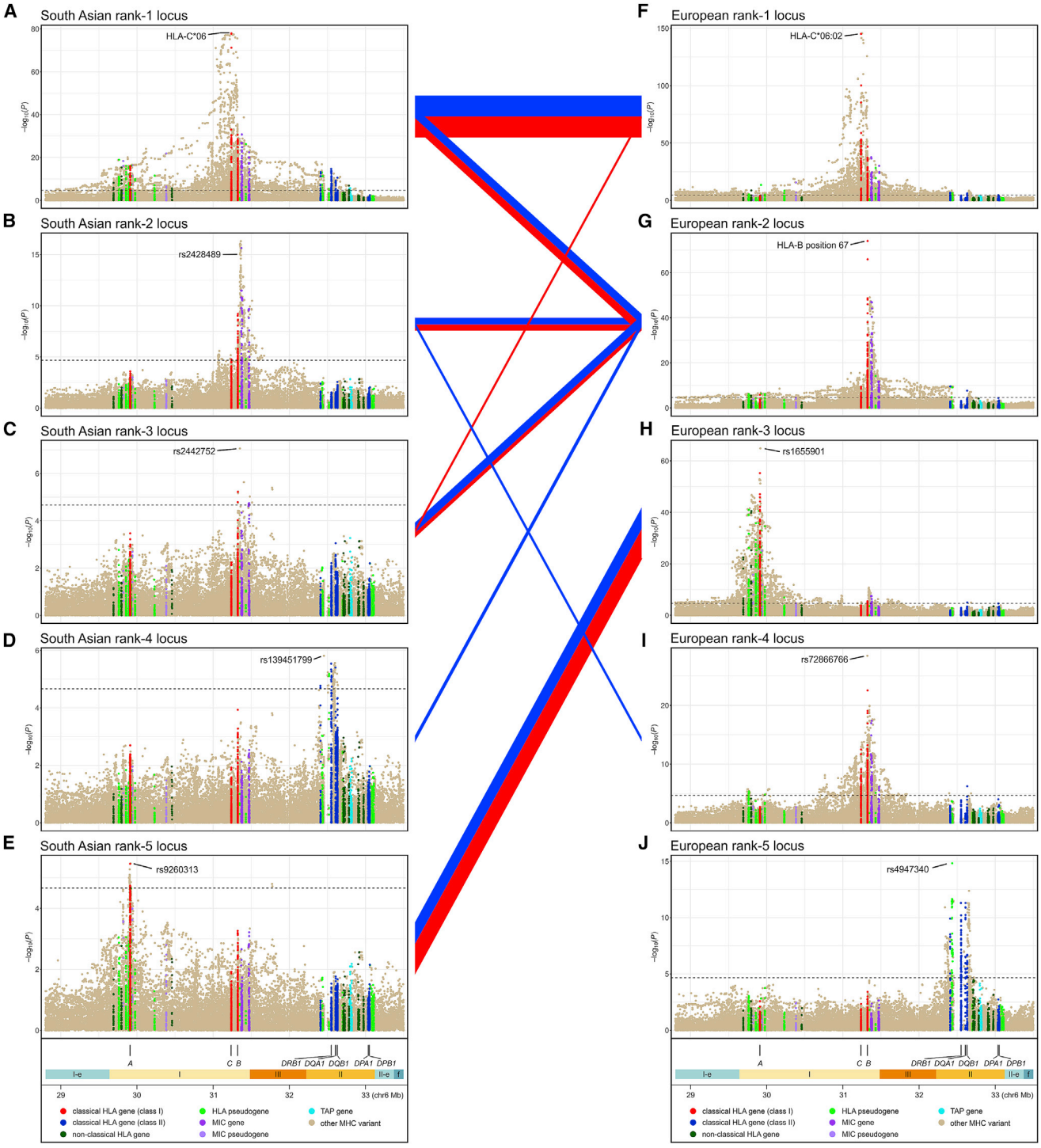

Figure 8. Linkage disequilibrium (LD) between the top five MHC association signals in the European and South Asian models
Association plots for the five most significant psoriasis-associated MHC loci in South Asians (A–E) and Europeans (F–J) are shown in decreasing top to bottom order of their significance in the full regression model. LD (W_n^2 coefficient) is depicted with line segments connecting pairs of South Asian and European loci; the color of the line indicates the population in which the LD was measured (red, South Asian; blue, European), and the thickness of the line is scaled linearly with the magnitude of LD. Only LD values of $W_n^2 \geq 0.1$ are shown.

scores of being regulatory and deleterious, respectively. The risk alleles of rs137854633 are one base insertions into intron 1 of HLA-B that are only 105 bp from the transcription start site and that are significantly and negatively correlated with HLA-B transcript levels in 29 of 31 tested tissues. The interval containing this variant binds at least 10 different transcription factors in chromatin immunoprecipitation sequencing (ChiP-seq) experiments and is in a region

Table 3. Summary of top-ranking eGenes for noncoding psoriasis-associated MHC variants with a Bayesian posterior probability > 0.5

Variant	Ancestry	PP	Nearest gene (position)	Top eGenes[a]	No. tissues[b]	Distance to TSS (kb)	No. tissues with +, − risk allele effect[c]	
							Most significant eQTL	Significant eQTL
rs1655901	EUR	1.000	*HLA-A* (3.1 kb downstream)	*HLA-F-AS1*	46	200.0	13, 0	23, 1
				HLA-A	53	6.5	11, 0	13, 6
				HLA-F	40	225.6	0, 6	0, 15
rs2844626	SAS+EUR	0.541	*HLA-C* (7.0 kb downstream)	*CCHCR1*	58	104.0	14, 0	29, 6
				HLA-C	58	10.3	13, 0	35, 2
				AL645933.2 0.2	43	133.7	0, 16	0, 27
rs72866766	EUR	1.000	*HLA-B* (468 bp downstream)	*MICA*	45	47.3	0, 22	0, 29
				HLA-C	58	81.3	8, 1	10, 7
				DDX39B	58	188.6	5, 0	10, 0
rs137854633	EUR	0.543	*HLA-B* (intron 1)	*HLA-B*	31	0.1	0, 21	0, 29
				HLA-C	44	85.0	0, 10	1, 29
				MICB	31	137.8	3, 0	15, 0
rs2853998	SAS+EUR	0.942	*HLA-B* (2.2 kb upstream)	*AL645933.2*	40	36.1	0, 11	0, 22
				HLA-B	40	2.2	0, 8	0, 35
				HLA-C	53	87.3	0, 7	3, 18
rs2442752	SAS	0.583	*AL671883.3* (7.9 kb downstream)	*HLA-C*	58	111.9	11, 8	13, 18
				HLA-B	45	26.8	5, 0	23, 0
				PSORS1C3	48	173.1	4, 0	20, 0
rs6935999	EUR	0.752	*HLA-DRA* (15 kb upstream)	*HSPA1L*	44	609.3	0, 4	0, 9
				AGER	44	240.7	3, 0	6, 0
				HLA-DRB1	44	164.9	3, 0	5, 1
				NCR3	37	831.2	0, 3	0, 3
rs4947340	EUR	0.965	*HLA-DRA* (23 kb upstream)	*HLA-DQA2*	34	273.8	18, 0	33, 0
				HLA-DRB1	49	122.3	0, 14	0, 48
				HLA-DQB2	33	296.0	6, 0	30, 1
rs9271539	SAS+EUR	0.562	*HLA-DQA1* (15 kb upstream)	*HLA-DRB5*	52	92.0	28, 0	49, 0
				HLA-DRB1	52	32.4	12,0	22, 4
				HLA-DQA1	36	15.2	1, 1	8, 5
				PSMB9	52	231.9	0, 2	0, 24

Abbreviations: PP, posterior probability; eGene, gene whose expression is significantly associated with an eQTL; eQTL, expression quantitative trait locus; EUR, European; SAS, South Asian; TSS, transcription start site.
[a]Expressed genes with a TSS within 1 Mb of the variant were ranked by the number of 58 psoriasis-relevant tissues for which the eQTL effects of that variant upon the gene were the most significant. The top three ranking eGenes are shown for each variant (four eGenes shown for two variants because of ties).
[b]Number of 58 tissues from 16 RNA expression studies for which the eQTL effect of the variant on the eGene was assessed. See Table S37 for citations of the 16 expression studies.

of active or poised transcription in nearly all psoriasis-relevant tissues. Other interesting regulatory candidates include rs2853998 and rs9271539, which lie in an enhancer region in most myeloid and lymphocyte cells but show Polycomb repression in skin-derived and mesenchymal cells. Two other psoriasis risk loci, rs2844626 and rs2442752, are scored by RegulomeDB as likely to affect transcription factor binding and linked to expression of a gene target, which based on *cis*-eQTL studies appear to be *CCHCR1* (MIM: 605310) and *HLA-C* for the former and *HLA-C* and *HLA-B* for the latter.

Reduced extent of LD for SAS versus EUR

We assessed whether there are differences in LD structure between SAS and EUR that may be contributing to the reduction in size of Bayesian credible sets observed for

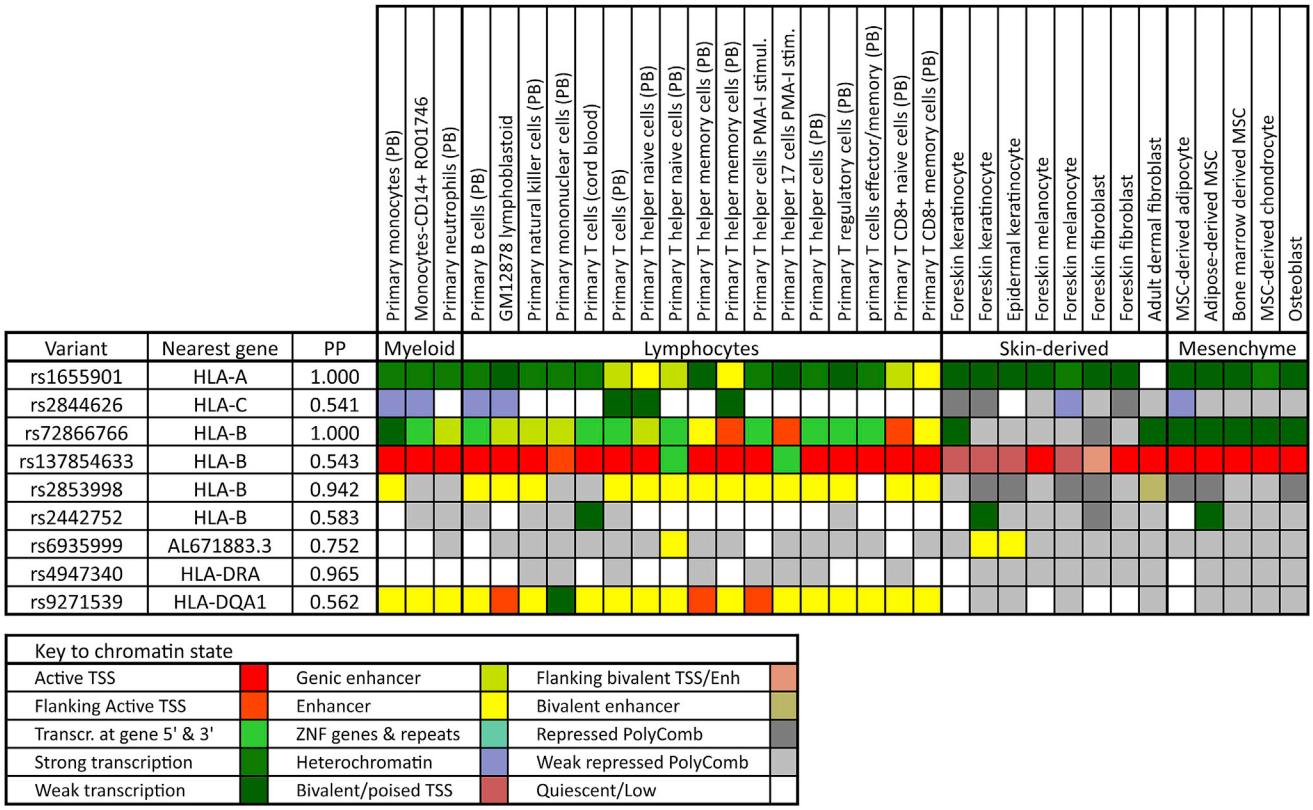

Figure 9. Chromatin states in 33 psoriasis-relevant cell types for psoriasis-associated noncoding variants with a Bayesian posterior probability exceeding 0.50

Chromatin states are derived from the 15-state Roadmap Epigenomics core model based on five chromatin marks.[74] Abbreviations: MSC, mesenchymal stem cell; PB, peripheral blood; PMA-I, PMA-ionomycin treatment; PP, posterior probability (Bayesian); TSS, transcription start site.

some psoriasis loci in our transethnic models. For non-MHC loci, using EUR and SAS samples of the 1KGP, we found fewer LD surrogates on average for the SAS population—44%, 35%, and 21% of the loci had fewer, equal, or more surrogates in SAS compared to EUR (Figure S41A). We were also able to demonstrate positive albeit nonsignificant correlations between the $\log_2$ ratio of the number of LD proxies in EUR versus SAS with both the $-\log_{10}$ (transethnic association p value) (Figure S41B) and with the $\log_2$ EUR versus transethnic ratio of the number of markers in the 95% BCS (Figure S41C). Within the MHC region, as shown in Figure S42, the slope of the linear fit for SAS LD values regressed onto EUR LD values was substantially and significantly different from 1.00 (slope = 0.79 and 0.81, with $p = 9 \times 10^{-29}$ and 3×10^{-25} for W_n^2 and ε' LD coefficients, respectively) when considering all 919 unique pairwise combinations of the MHC loci selected by the three stepwise ethnic analyses.

Discussion

We first performed a GWAS of psoriasis in individuals of South Asian descent, followed by a fine-mapping study of MHC psoriasis signals for South Asians and the largest analysis to date of psoriasis MHC signals for people of European ancestry. Finding that effect sizes for known psoriasis loci were strongly correlated in EUR and SAS ($\rho = 0.78$; $p < 2 \times 10^{-14}$; Figure 1), we conducted a transethnic meta-analysis of non-MHC and MHC signals, identifying two loci that had not previously reached genome-wide significance in EUR populations (Table 1; Figure 2). Inspection of eQTL databases revealed potential regulatory roles for both these signals. For 1p36.22, rs2103876 is a *cis*-eQTL for multiple genes in blood,[75] including *MFN2* (MIM: 608507), *MIIP* (MIM: 608772), *MTHFR* (MIM: 607093), and *CLCN6* (MIM: 602726). *MTHFR*, which encodes methylenetetrahydrofolate reductase, is the only gene whose expression is positively correlated ($p = 4.2 \times 10^{-97}$) with the psoriasis risk allele (T) and significantly upregulated in psoriatic lesional skin ($p = 3.06 \times 10^{-23}$; fold change [FC] = 1.7).[76] Methylenetetrahydrofolate reductase is one of several targets of methotrexate, a highly effective antipsoriatic drug.[77] For the 1q24.2 locus, the only significant eQTL target for rs12046909 in whole blood is *XCL1* (MIM: 600250),[75] whose psoriasis risk allele is negatively associated with *XCL1* expression, with no significant eQTL for this genetic variation in other tissues according to GTEx.[78] *XCL1* encodes a ligand for chemokine receptor XCR1 (MIM: 600552), which is expressed on dendritic

cells, and *XCL1* is overexpressed in lesional psoriatic versus normal skin ($p = 0.011$; FC = 1.95).[76]

Transethnic association studies should improve resolution for fine mapping of genetic variants by leveraging differences in LD architecture among different global populations.[30] Our analysis of both non-MHC and MHC psoriasis associations revealed that on average the range of LD with their lead variants was more limited in South Asians than Europeans. The faster decay of LD in SAS versus EUR, which has been reported previously,[30] may have contributed to the improved fine-mapping resolution our transethnic models achieved for many associated loci, beyond that afforded by a simple increase in sample size.

Based on their known immunologic functions and large observed effect sizes, HLA protein variants are leading causative candidates for psoriasis.[27,79] We improved an existing method to construct high-quality HLA reference panels for South Asians and Europeans and to then impute HLA amino acid, protein, and SNP variants with excellent accuracy. To increase the density and scope of tested MHC variants, we integrated the imputed HLA variants with variants imputed using the 1KGP and HRC reference panels in a 12 Mb interval that encompasses the classical MHC region. For people of EUR ancestry, our MHC fine-mapping analysis surpasses our most recent effort[22] in many respects, including effective sample size (39,335 versus 21,137), MHC interval (12 versus 3.7 Mb), total number of tested variants (83,352 versus 8,739), number of tested classical MHC variants (49,407 versus 8,556), and imputed variant quality ($r^2 \geq 0.70$ versus no r^2 filter). We also included joint association testing of the alleles of multiallelic variants, not only for HLA genes but for the entire 12 Mb extended MHC. To fully incorporate results of multiallelic with biallelic association tests, we assessed 12 metrics for multiallelic LD and utilized the best-performing two, and we also devised a method to compute Bayes factors for multiallelic variants.

Our analysis of the extended MHC region uncovered five independent psoriasis loci in SAS, 14 in EUR, and 17 for the transethnic analysis. Notably, the lead variants for all these risk loci are restricted to the classical MHC. Our efforts to improve HLA variant imputation quality and include multiallelic variants were justified by their significant enrichment in the final regression models. Despite the enrichment of HLA variants, only eight of the 36 identified risk variants change HLA proteins, compared to all six psoriasis loci from our previous MHC fine-mapping study.[22] However, four of the non-coding MHC loci lie within HLA protein-coding genes, and 11 are intergenic with an HLA protein-coding gene as their nearest neighbor. Furthermore, based on the W_n^2 LD measure, the best protein-changing surrogates are in HLA genes for 29 of the 36 identified risk loci, even though HLA proteins represent only 19 of 137 coding genes in the classical MHC. Interestingly, the best coding surrogates for five of the remaining seven risk variants are in *MICA* and *TAP2*, genes with important immune system functions, whose proteins are either structurally akin to class I HLA molecules (*MICA*) or involved in translocation of short peptides from cytosol to endoplasmic reticulum for binding to MHC class I proteins (*TAP2*). Five of the eight HLA coding risk variants have Bayesian posterior probabilities of being causative that exceed 0.50: HLA-C*06:02 for the transethnic analysis, which differs from other common 2-field HLA-C alleles by amino acid combination Asp90 and Trp97 and by underlying two-SNP haplotype variant rs1131123-T/rs1131118-A, the multiallelic amino acid at position 67 of HLA-B for the European and transethnic analyses, and the essentially biallelic amino acids at positions 171 of HLA-B and 73 of HLA-C for the European analysis. Notably, all these amino acid variants except Asp90 are not only in the antigen-binding groove of the HLA protein (Figure 6) but also interact with bound peptide rather than the T cell receptor.[80]

Most (28/36) of the identified MHC risk variants are noncoding. Nine of these noncoding variants have posterior probabilities exceeding 0.50, indicating that some of the psoriasis loci in the MHC may have a regulatory function. Most of these nine variants exhibit strong and consistent *cis*-eQTL effects on target genes with a known function in immunity in a wide variety of psoriasis-relevant tissues. Several of these variants also show evidence of affecting transcription factor binding or correlation with chromatin states indicative of active transcription, enhancers, or regions of Polycomb suppression (Figure 9).

*HLA-C*06* was the top-ranking psoriasis locus for both SAS and EUR, but its strength of association was significantly higher for SAS in both the unconditional and final full regression models. Furthermore, the contribution of *HLA-C*06* to overall psoriasis susceptibility compared to other MHC variants was higher for SAS based on variance in liability explained and goodness-of-fit measures such as AIC and Tjur's R^2, even when the comparison was restricted to the top five ranking loci in the EUR model. Because *HLA-C*06* is more strongly associated with purely cutaneous psoriasis than with PsA in both EUR-origin individuals[27] and our SAS sample (OR [95% CI] = 6.54 [5.50–7.79] versus 5.35 [4.02–7.13]), the stronger *HLA-C*06* association in SAS could be related to the lower prevalence of PsA in SAS (8.7% in the literature[81] and 13.0% in our sample) compared to EUR (25%–30%).[82] The lower prevalence of PsA in SAS may be a consequence of the substantially lower frequency of several known *HLA-B* risk alleles for PsA (*B*08*, *B*27*, *B*38*, and *B*39*)[83,84] in SAS versus EUR that we observed for both our own data (Figures S17B and S19B) and data from the National Marrow Donor Program (NMDP) (Figure S21).

Other than *HLA-C*06*, the only other possibly shared locus between the two monoethnic MHC association models is a signal 3 kb downstream of HLA-A. The similarity of effect sizes for the variants in one ethnic model compared to their refitted values for the other ethnic group argues that MHC associations for the two groups may not be as different as the two final regression models seem to indicate but may stem more from a lack of power in the relatively small SAS study

to accurately identify signals other than *HLA-C*06*. Despite its modest power, combining the SAS dataset with EUR samples in a transethnic analysis did bolster the evidence that the *HLA-C*06* protein variant is causal, increasing its Bayesian posterior probability from 0.255 and 0.288 in the SAS and EUR models to 0.664 in the transethnic model.

Our encouraging findings for the MHC region need to be tempered in light of several limitations. First, many variants in the region were not tested for association. As shown in Table S18, only 42%–52% of the classical MHC variants in the reference panels used for imputation had a high enough imputation quality to qualify for testing, including only 68%–80% and 72%–84% of common (MAF $\geq$ 0.05) and low-frequency (MAF = 0.01–0.05) variants, respectively. Furthermore, amino acid and full-length protein variants were characterized and tested for only eight of the 137 coding genes in the region (Table S14). The 1KGP and HRC reference panels, which are based on short-read sequences aligned to the reference genome, are deficient in their coverage of indels and larger structural variants in the MHC, as was demonstrated by a Danish study that used a *de novo* assembly approach.[85,86] Second, variation in imputation quality among the tested variants is predicted to affect their power to detect association independent of their true association with disease.[87] We demonstrated this for all three of our ancestry datasets, finding a significant positive correlation of full model significance of association with imputation quality for variants in substantial LD with *HLA-C*06* (Figure S43). Third, stepwise regression is unlikely to find the best possible association model, owing to the infeasibility of testing all possible subsets of variants, and to the adverse effects of LD-produced collinearity among genuinely independent susceptibility loci, which can lead to unstable coefficient estimates, inflated standard errors, and a downward bias of their relative importance compared to variants that are uncorrelated with others already in the model.[88] We hypothesize that many of the sizeable differences among our three MHC association models arise from these limitations and the modest power of the SAS dataset, rather than from differences in the true but unknown models. However, until this hypothesis can be validated, we believe it safest to treat each model as the best current summary of MHC psoriasis associations for its corresponding population.

To remedy these issues, methods are needed that can address the unique challenges posed by the MHC region. Genotyping every variant in the MHC with close to 100% accuracy is an essential first step, which may require intensive sequence-based approaches with *de novo* alignment, deep read coverage, very long reads for more accurate haplotype phasing, and multiple libraries with insert sizes ranging from small to very large for proper typing of complex structural variations. Improved association methods are also needed that can better handle large numbers of correlated variants. Notwithstanding these challenges, our results demonstrate the value of genotyping diverse populations both within and beyond the MHC.

Data and code availability

Complete summary statistics for the genome-wide association analyses of the SAS, EUR, and transethnic (SAS and EUR) datasets have been deposited at the NHGRI-EBI GWAS Catalog (accession numbers GCST90019015, GCST90019016, and GCST90019017). Individual-level genotype data for the CASP GWAS, PsA GWAS, and Exomechip case-control studies are available on dbGaP (dbGaP: phs 000019.v1.p1, phs000982.v1.p1, and phs001306.v1.p1). Genotypes for the WTCCC2 psoriasis-control study are archived at the European Genome-Phenome Archive (study ID EGAS00000000108) and can be requested by contacting the data access committee at the Wellcome Trust Sanger Institute (datasharing@sanger.ac.uk). Data sharing restrictions do not allow making genotypes available for the remaining six case-control cohorts analyzed by this study (Kiel GWAS, Genizon GWAS, PAGE and GAPC Immunochip studies, batches 1 and 2+3 of SAS GWAS).

Our updated version of the SNP2HLA imputation package is available on GitHub. Five datasets (UM, 1KGP-ALL-v2, T1DGC, BKT, IKMB; see Table S8) were used to build the best-performing MHC reference panels of this study. Our GitHub repository contains three MHC reference panels built by applying our updated methods to the two datasets (UM, 1KGP-ALL-v2) for which genotype data can be freely shared. Genotype data for the T1DGC dataset can requested from the NIDDK Central Repository. Data restrictions preclude sharing any individual-level genotype data for the BKT and IKMB datasets; however, the MHC reference data for the IKMB dataset is available as a HIBAG model for imputing HLA alleles (IKMB models).

Supplemental information

Supplemental information can be found online at https://doi.org/10.1016/j.xhgg.2021.100069.

Acknowledgments

The authors thank Yang Luo and Soumya Raychaudhuri (Broad Institute of Harvard and MIT) for advice concerning the construction of SNP2HLA reference panels. This work was supported by awards from the National Institutes of Health (R01AR042742, R01AR050511, R01AR054966, R01AR063611, and R01AR065183 to J.T.E.; K01AR072129 to L.C.T.). L.C.T. was also supported by the Dermatology Foundation, the National Psoriasis Foundation, and the Arthritis National Research Foundation. L.C.T., P.E.S., T.T., J.J.V., R.P.N., and J.T.E. are supported by the Dawn and Dudley Holmes Foundation and the Babcock Memorial Trust. J.T.E. is supported by the Ann Arbor Veterans Affairs Hospital. We also wish to acknowledge the Indian Council of Medical Research for support of a foreign collaboration project with the University of Michigan, Ann Arbor, USA funded by National Instititutes of Health, USA (project title Genetic Analysis of Psoriasis and Psoriatic Arthritis in Indians; file no. 50/2/2009-BMS; project code N-1170).

Declaration of interests

The authors declare no competing interests.

Received: January 19, 2021
Accepted: October 24, 2021

Web resources

1000 Genomes, phase 3, Affymetrix 6.0 microarray data, http://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502/supporting/hd_genotype_chip/
1000 Genomes, phase 3, HLA genotypes, http://ftp.1000genomes.ebi.ac.uk/vol1/ftp/data_collections/HLA_types/
1000 Genomes, v5 phase3, full integrated variant call set, ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502/
1000 Genomes, v5 phase3, reduced integrated variant call set (no monomorphic or singleton sites), http://csg.sph.umich.edu/abecasis/MACH/download/1000G.Phase3.v5.html
Beagle 4.1, https://faculty.washington.edu/browning/beagle/b4_1.html
CADD v1.6, https://cadd.gs.washington.edu/
dbGaP, https://www.ncbi.nlm.nih.gov/gap/
eQTLGen Consortium, https://www.eqtlgen.org/
European Genome-Phenome Archive, https://ega-archive.org/
GTEx Portal, https://www.gtexportal.org/home/
HIBAG model for imputing HLA alleles, https://hibag.s3.amazonaws.com/index.html
IPD-IMGT/HLA database, https://www.ebi.ac.uk/ipd/imgt/hla/
logistf R package, https://cran.r-project.org/web/packages/logistf/index.html
Michigan Imputation Server, https://imputationserver.sph.umich.edu/index.html#!
NHGRI-EBI GWAS Catalog, https://www.ebi.ac.uk/gwas/
OMIM, https://www.omim.org/
Pan-Asian SNP2HLA reference panel, http://software.broadinstitute.org/mpg/snp2hla/
Plink 2.0, https://www.cog-genomics.org/plink/
RegulomeDB 2.0, https://regulomedb.org/regulome-search/
SNP2HLA imputation package, https://github.com/CutaneousBioinf/SNP2HLA_UM
SNP2HLA v1.0.3, http://software.broadinstitute.org/mpg/snp2hla/
T1DGC SNP2HLA reference panel, https://repository.niddk.nih.gov/studies/t1dgc-special/

References

1. Gudjonsson, J.E., and Elder, J.T. (2018). Psoriasis. In Dermatology in General Medicine, S. Kang, M. Amagai, A.L. Bruckner, A.H. Enk, A.J. McMichael, and J.S. Orringer, eds. (New York: McGraw-Hill), pp. 457–498.
2. Kim, J., and Krueger, J.G. (2015). The immunopathogenesis of psoriasis. Dermatol. Clin. *33*, 13–23.
3. Greb, J.E., Goldminz, A.M., Elder, J.T., Lebwohl, M.G., Gladman, D.D., Wu, J.J., Mehta, N.N., Finlay, A.Y., and Gottlieb, A.B. (2016). Psoriasis. Nat. Rev. Dis. Primers *2*, 16082.
4. Parisi, R., Symmons, D.P., Griffiths, C.E., Ashcroft, D.M.; and Identification and Management of Psoriasis and Associated ComorbidiTy (IMPACT) project team (2013). Global epidemiology of psoriasis: a systematic review of incidence and prevalence. J. Invest. Dermatol. *133*, 377–385.
5. Schäfer, T. (2006). Epidemiology of psoriasis. Review and the German perspective. Dermatology *212*, 327–337.
6. Naldi, L. (2004). Epidemiology of psoriasis. Curr. Drug Targets Inflamm. Allergy *3*, 121–128.
7. Patrick, M.T., Stuart, P.E., Raja, K., Gudjonsson, J.E., Tejasvi, T., Yang, J., Chandran, V., Das, S., Callis-Duffin, K., Ellinghaus, E., et al. (2018). Genetic signature to provide robust risk assessment of psoriatic arthritis development in psoriasis patients. Nat. Commun. *9*, 4178.
8. Stuart, P.E., Tsoi, L.C., Hambro, C.A., and Elder, J.T. (2018). Genetics of Psoriasis. In Textbook of Psoriatic Arthritis, D.D. Gladman, ed. (New York: Oxford University Press), pp. 35–55.
9. Hirata, J., Hirota, T., Ozeki, T., Kanai, M., Sudo, T., Tanaka, T., Hizawa, N., Nakagawa, H., Sato, S., Mushiroda, T., et al. (2018). Variants at HLA-A, HLA-C, and HLA-DQB1 Confer Risk of Psoriasis Vulgaris in Japanese. J. Invest. Dermatol. *138*, 542–548.
10. Tsoi, L.C., Stuart, P.E., Tian, C., Gudjonsson, J.E., Das, S., Zawistowski, M., Ellinghaus, E., Barker, J.N., Chandran, V., Dand, N., et al. (2017). Large scale meta-analysis characterizes genetic architecture for common psoriasis associated variants. Nat. Commun. *8*, 15382.
11. Ozawa, A., Ohkido, M., Inoko, H., Ando, A., and Tsuji, K. (1988). Specific restriction fragment length polymorphism on the HLA-C region and susceptibility to psoriasis vulgaris. J. Invest. Dermatol. *90*, 402–405.
12. Kim, T.G., Lee, H.J., Youn, J.I., Kim, T.Y., and Han, H. (2000). The association of psoriasis with human leukocyte antigens in Korean population and the influence of age of onset and sex. J. Invest. Dermatol. *114*, 309–313.
13. Choonhakarn, C., Romphruk, A., Puapairoj, C., Jirarattanapochai, K., Romphruk, A., and Leelayuwat, C. (2002). Haplotype associations of the major histocompatibility complex with psoriasis in Northeastern Thais. Int. J. Dermatol. *41*, 330–334.
14. Shaiq, P.A., Stuart, P.E., Latif, A., Schmotzer, C., Kazmi, A.H., Khan, M.S., Azam, M., Tejasvi, T., Voorhees, J.J., Raja, G.K., et al. (2013). Genetic associations of psoriasis in a Pakistani population. Br. J. Dermatol. *169*, 406–411.
15. Munir, S., ber Rahman, S., Rehman, S., Saba, N., Ahmad, W., Nilsson, S., Mazhar, K., and Naluai, Å.T. (2015). Association analysis of GWAS and candidate gene loci in a Pakistani population with psoriasis. Mol. Immunol. *64*, 190–194.
16. Umapathy, S., Pawar, A., Mitra, R., Khuperkar, D., Devaraj, J.P., Ghosh, K., and Khopkar, U. (2011). Hla-a and hla-B alleles associated in psoriasis patients from mumbai, Western India. Indian J. Dermatol. *56*, 497–500.
17. Indhumathi, S., Rajappa, M., Chandrashekar, L., Ananthanarayanan, P.H., Thappa, D.M., and Negi, V.S. (2016). The HLA-C*06 allele as a possible genetic predisposing factor to psoriasis in South Indian Tamils. Arch. Dermatol. Res. *308*, 193–199.
18. Chandra, A., Lahiri, A., Senapati, S., Basu, B., Ghosh, S., Mukhopadhyay, I., Behra, A., Sarkar, S., Chatterjee, G., and Chatterjee, R. (2016). Increased Risk of Psoriasis due to combined effect of HLA-Cw6 and LCE3 risk alleles in Indian population. Sci. Rep. *6*, 24059.
19. Marigorta, U.M., and Navarro, A. (2013). High trans-ethnic replicability of GWAS results implies common causal variants. PLoS Genet. *9*, e1003566.
20. Li, Y.R., and Keating, B.J. (2014). Trans-ethnic genome-wide association studies: advantages and challenges of mapping in diverse populations. Genome Med. *6*, 91.
21. Marigorta, U.M., Rodríguez, J.A., Gibson, G., and Navarro, A. (2018). Replicability and Prediction: Lessons and Challenges from GWAS. Trends Genet. *34*, 504–517.
22. Okada, Y., Han, B., Tsoi, L.C., Stuart, P.E., Ellinghaus, E., Tejasvi, T., Chandran, V., Pellett, F., Pollock, R., Bowcock, A.M., et al. (2014). Fine mapping major histocompatibility complex associations in psoriasis and its clinical subtypes. Am. J. Hum. Genet. *95*, 162–172.

23. Jia, X., Han, B., Onengut-Gumuscu, S., Chen, W.M., Concannon, P.J., Rich, S.S., Raychaudhuri, S., and de Bakker, P.I. (2013). Imputing amino acid polymorphisms in human leukocyte antigens. PLoS ONE *8*, e64683.
24. Nair, R.P., Henseler, T., Jenisch, S., Stuart, P., Bichakjian, C.K., Lenk, W., Westphal, E., Guo, S.W., Christophers, E., Voorhees, J.J., and Elder, J.T. (1997). Evidence for two psoriasis susceptibility loci (HLA and 17q) and two novel candidate regions (16q and 20p) by genome-wide scan. Hum. Mol. Genet. *6*, 1349–1356.
25. Nair, R.P., Duffin, K.C., Helms, C., Ding, J., Stuart, P.E., Goldgar, D., Gudjonsson, J.E., Li, Y., Tejasvi, T., Feng, B.J., et al.; Collaborative Association Study of Psoriasis (2009). Genome-wide scan reveals association of psoriasis with IL-23 and NF-kappaB pathways. Nat. Genet. *41*, 199–204.
26. Ellinghaus, E., Ellinghaus, D., Stuart, P.E., Nair, R.P., Debrus, S., Raelson, J.V., Belouchi, M., Fournier, H., Reinhard, C., Ding, J., et al. (2010). Genome-wide association study identifies a psoriasis susceptibility locus at TRAF3IP2. Nat. Genet. *42*, 991–995.
27. Stuart, P.E., Nair, R.P., Tsoi, L.C., Tejasvi, T., Das, S., Kang, H.M., Ellinghaus, E., Chandran, V., Callis-Duffin, K., Ike, R., et al. (2015). Genome-wide Association Analysis of Psoriatic Arthritis and Cutaneous Psoriasis Reveals Differences in Their Genetic Architecture. Am. J. Hum. Genet. *97*, 816–836.
28. Strange, A., Capon, F., Spencer, C.C., Knight, J., Weale, M.E., Allen, M.H., Barton, A., Band, G., Bellenguez, C., Bergboer, J.G., et al.; Genetic Analysis of Psoriasis Consortium & the Wellcome Trust Case Control Consortium 2 (2010). A genome-wide association study identifies new psoriasis susceptibility loci and an interaction between HLA-C and ERAP1. Nat. Genet. *42*, 985–990.
29. Tsoi, L.C., Spain, S.L., Knight, J., Ellinghaus, E., Stuart, P.E., Capon, F., Ding, J., Li, Y., Tejasvi, T., Gudjonsson, J.E., et al.; Collaborative Association Study of Psoriasis (CASP); Genetic Analysis of Psoriasis Consortium; Psoriasis Association Genetics Extension; and Wellcome Trust Case Control Consortium 2 (2012). Identification of 15 new psoriasis susceptibility loci highlights the role of innate immunity. Nat. Genet. *44*, 1341–1348.
30. Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A., Abecasis, G.R.; and 1000 Genomes Project Consortium (2015). A global reference for human genetic variation. Nature *526*, 68–74.
31. McCarthy, S., Das, S., Kretzschmar, W., Delaneau, O., Wood, A.R., Teumer, A., Kang, H.M., Fuchsberger, C., Danecek, P., Sharp, K., et al.; Haplotype Reference Consortium (2016). A reference panel of 64,976 haplotypes for genotype imputation. Nat. Genet. *48*, 1279–1283.
32. Willer, C.J., Li, Y., and Abecasis, G.R. (2010). METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics *26*, 2190–2191.
33. Wittig, M., Anmarkrud, J.A., Kässens, J.C., Koch, S., Forster, M., Ellinghaus, E., Hov, J.R., Sauer, S., Schimmler, M., Ziemann, M., et al. (2015). Development of a high-resolution NGS-based HLA-typing and analysis pipeline. Nucleic Acids Res. *43*, e70.
34. Browning, B.L., and Browning, S.R. (2016). Genotype Imputation with Millions of Reference Samples. Am. J. Hum. Genet. *98*, 116–126.
35. Robinson, J., Barker, D.J., Georgiou, X., Cooper, M.A., Flicek, P., and Marsh, S.G.E. (2020). IPD-IMGT/HLA Database. Nucleic Acids Res. *48* (D1), D948–D955.
36. Marsh, S.G., Albert, E.D., Bodmer, W.F., Bontrop, R.E., Dupont, B., Erlich, H.A., Fernández-Viña, M., Geraghty, D.E., Holdsworth, R., Hurley, C.K., et al. (2010). An update to HLA nomenclature, 2010. Bone Marrow Transplant. *45*, 846–848.
37. Zhang, P., Zhan, X., Rosenberg, N.A., and Zöllner, S. (2013). Genotype imputation reference panel selection using maximal phylogenetic diversity. Genetics *195*, 319–330.
38. Degenhardt, F., Wendorff, M., Wittig, M., Ellinghaus, E., Datta, L.W., Schembri, J., Ng, S.C., Rosati, E., Hübenthal, M., Ellinghaus, D., et al. (2019). Construction and benchmarking of a multi-ethnic reference panel for the imputation of HLA class I and II alleles. Hum. Mol. Genet. *28*, 2078–2092.
39. Liu, J.Z., van Sommeren, S., Huang, H., Ng, S.C., Alberts, R., Takahashi, A., Ripke, S., Lee, J.C., Jostins, L., Shah, T., et al.; International Multiple Sclerosis Genetics Consortium; and International IBD Genetics Consortium (2015). Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. Nat. Genet. *47*, 979–986.
40. Mychaleckyj, J.C., Noble, J.A., Moonsamy, P.V., Carlson, J.A., Varney, M.D., Post, J., Helmberg, W., Pierce, J.J., Bonella, P., Fear, A.L., et al.; T1DGC (2010). HLA genotyping in the international Type 1 Diabetes Genetics Consortium. Clin. Trials *7* (1, Suppl), S75–S87.
41. Onengut-Gumuscu, S., Chen, W.M., Burren, O., Cooper, N.J., Quinlan, A.R., Mychaleckyj, J.C., Farber, E., Bonnie, J.K., Szpak, M., Schofield, E., et al.; Type 1 Diabetes Genetics Consortium (2015). Fine mapping of type 1 diabetes susceptibility loci and evidence for colocalization of causal variants with lymphoid gene enhancers. Nat. Genet. *47*, 381–386.
42. Pillai, N.E., Okada, Y., Saw, W.Y., Ong, R.T., Wang, X., Tantoso, E., Xu, W., Peterson, T.A., Bielawny, T., Ali, M., et al. (2014). Predicting HLA alleles from high-resolution SNP data in three Southeast Asian populations. Hum. Mol. Genet. *23*, 4443–4451.
43. Okada, Y., Kim, K., Han, B., Pillai, N.E., Ong, R.T., Saw, W.Y., Luo, M., Jiang, L., Yin, J., Bang, S.Y., et al. (2014). Risk for ACPA-positive rheumatoid arthritis is driven by shared HLA amino acid polymorphisms in Asian and European populations. Hum. Mol. Genet. *23*, 6916–6926.
44. Abi-Rached, L., Gouret, P., Yeh, J.H., Di Cristofaro, J., Pontarotti, P., Picard, C., and Paganini, J. (2018). Immune diversity sheds light on missing variation in worldwide genetic diversity panels. PLoS ONE *13*, e0206512.
45. Raychaudhuri, S., Sandor, C., Stahl, E.A., Freudenberg, J., Lee, H.S., Jia, X., Alfredsson, L., Padyukov, L., Klareskog, L., Worthington, J., et al. (2012). Five amino acids in three HLA proteins explain most of the association between MHC and seropositive rheumatoid arthritis. Nat. Genet. *44*, 291–296.
46. Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., and Lee, J.J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience *4*, 7.
47. Heinze, G., and Schemper, M. (2002). A solution to the problem of separation in logistic regression. Stat. Med. *21*, 2409–2419.
48. Burnham, K.P., and Anderson, D.R. (2004). Multimodel inference: Understanding AIC and BIC in model selection. Sociol. Methods Res. *33*, 261–304.

49. Tjur, T. (2009). Coefficients of determination in logistic regression models—a new proposal: the coefficient of discrimination. Am. Stat. *63*, 366–372.
50. Shorrocks, A.F. (2013). Decomposition procedures for distributional analysis: a unified framework based on the Shapley value. J. Econ. Inequal. *11*, 99–126.
51. Wakefield, J. (2009). Bayes factors for genome-wide association studies: comparison with P-values. Genet. Epidemiol. *33*, 79–86.
52. Wellcome Trust Case Control Consortium (2007). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature *447*, 661–678.
53. Maller, J.B., McVean, G., Byrnes, J., Vukcevic, D., Palin, K., Su, Z., Howson, J.M., Auton, A., Myers, S., Morris, A., et al.; Wellcome Trust Case Control Consortium (2012). Bayesian refinement of association signals for 14 loci in 3 common diseases. Nat. Genet. *44*, 1294–1301.
54. Wen, X., and Stephens, M. (2014). Bayesian Methods for Genetic Association Analysis with Heterogeneous Subgroups: From Meta-Analyses to Gene-Environment Interactions. Ann. Appl. Stat. *8*, 176–203.
55. Wen, X. (2014). Bayesian model selection in complex linear systems, as illustrated in genetic association studies. Biometrics *70*, 73–83.
56. Zhao, H., Nettleton, D., Soller, M., and Dekkers, J.C. (2005). Evaluation of linkage disequilibrium measures between multi-allelic markers as predictors of linkage disequilibrium between markers and QTL. Genet. Res. *86*, 77–87.
57. Zhao, H., Nettleton, D., and Dekkers, J.C. (2007). Evaluation of linkage disequilibrium measures between multi-allelic markers as predictors of linkage disequilibrium between single nucleotide polymorphisms. Genet. Res. *89*, 1–6.
58. Hedrick, P.W., and Thomson, G. (1986). A two-locus neutrality test: applications to humans, E. coli and lodgepole pine. Genetics *112*, 135–156.
59. Yamazaki, T. (1977). The effects of overdominance of linkage in a multilocus system. Genetics *86*, 227–236.
60. Thomson, G., and Single, R.M. (2014). Conditional asymmetric linkage disequilibrium (ALD): extending the biallelic r2 measure. Genetics *198*, 321–331.
61. Nothnagel, M., Fürst, R., and Rohde, K. (2002). Entropy as a measure for linkage disequilibrium over multilocus haplotype blocks. Hum. Hered. *54*, 186–198.
62. Liu, Z., and Lin, S. (2005). Multilocus LD measure and tagging SNP selection with generalized mutual information. Genet. Epidemiol. *29*, 353–364.
63. Okada, Y., Momozawa, Y., Ashikawa, K., Kanai, M., Matsuda, K., Kamatani, Y., Takahashi, A., and Kubo, M. (2015). Construction of a population-specific HLA imputation reference panel and its application to Graves' disease risk in Japanese. Nat. Genet. *47*, 798–802.
64. Cramer, H. (1946). Mathematical Models of Statistics (Princeton, NJ: Princeton University Press).
65. Zuo, X., Sun, L., Yin, X., Gao, J., Sheng, Y., Xu, J., Zhang, J., He, C., Qiu, Y., Wen, G., et al. (2015). Whole-exome SNP array identifies 15 new susceptibility loci for psoriasis. Nat. Commun. *6*, 6793.
66. O'Rielly, D.D., Jani, M., Rahman, P., and Elder, J.T. (2019). The Genetics of Psoriasis and Psoriatic Arthritis. J. Rheumatol. Suppl. *95*, 46–50.
67. Motyer, A., Vukcevic, D., D'ilthey, A., Donnelly, P., McVean, G., and Lesie, S. (2016). Practical use of methods for imputation of HLA alleles from SNP genotype data. bioRxiv. https://doi.org/10.1101/091009.
68. Luo, Y., Kanai, M., Choi, W., Li, X., Sakaue, S., Yamamoto, K., Ogawa, K., Gutierrez-Arcelus, M., Gregersen, P.K., Stuart, P.E., et al.; NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium (2021). A high-resolution HLA reference panel capturing global population diversity enables multi-ancestry fine-mapping in HIV host response. Nat. Genet. *53*, 1504–1516.
69. Das, S., Forer, L., Schönherr, S., Sidore, C., Locke, A.E., Kwong, A., Vrieze, S.I., Chew, E.Y., Levy, S., McGue, M., et al. (2016). Next-generation genotype imputation service and methods. Nat. Genet. *48*, 1284–1287.
70. Gragert, L., Madbouly, A., Freeman, J., and Maiers, M. (2013). Six-locus high resolution HLA haplotype frequencies derived from mixed-resolution DNA typing for the entire US donor registry. Hum. Immunol. *74*, 1313–1320.
71. Horton, R., Wilming, L., Rand, V., Lovering, R.C., Bruford, E.A., Khodiyar, V.K., Lush, M.J., Povey, S., Talbot, C.C., Jr., Wright, M.W., et al. (2004). Gene map of the extended human MHC. Nat. Rev. Genet. *5*, 889–899.
72. So, H.C., Gui, A.H., Cherny, S.S., and Sham, P.C. (2011). Evaluating the heritability explained by known susceptibility variants: a survey of ten complex diseases. Genet. Epidemiol. *35*, 310–317.
73. Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera–a visualization system for exploratory research and analysis. J. Comput. Chem. *25*, 1605–1612.
74. Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., Ziller, M.J., et al.; Roadmap Epigenomics Consortium (2015). Integrative analysis of 111 reference human epigenomes. Nature *518*, 317–330.
75. Vosa, U., Claringbould, A., Westra, H.-J., Bonder, M.J., Deelen, P., Zeng, B., Kirsten, H., others, Visscher, P.M., Scholz, M., et al. (2018). Unraveling the polygenic architecture of complex traits using blood eQTL meta-analysis. bioRxiv. https://doi.org/10.1101/447367.
76. Tsoi, L.C., Rodriguez, E., Degenhardt, F., Baurecht, H., Wehkamp, U., Volks, N., Szymczak, S., Swindell, W.R., Sarkar, M.K., Raja, K., et al. (2019). Atopic Dermatitis Is an IL-13-Dominant Disease with Greater Molecular Heterogeneity Compared to Psoriasis. J. Invest. Dermatol. *139*, 1480–1489.
77. Cronstein, B.N., and Aune, T.M. (2020). Methotrexate and its mechanisms of action in inflammatory arthritis. Nat. Rev. Rheumatol. *16*, 145–154.
78. Battle, A., Brown, C.D., Engelhardt, B.E., Montgomery, S.B.; GTEx Consortium; Laboratory, Data Analysis &Coordinating Center (LDACC)—Analysis Working Group; Statistical Methods groups—Analysis Working Group; Enhancing GTEx (eGTEx) groups; NIH Common Fund; NIH/NCI; NIH/NHGRI; NIH/NIMH; NIH/NIDA; Biospecimen Collection Source Site—NDRI; Biospecimen Collection Source Site—RPCI; Biospecimen Core Resource—VARI; Brain Bank Repository—University of Miami Brain Endowment Bank; Leidos Biomedical—Project Management; ELSI Study; Genome Browser Data Integration &Visualization—EBI; Genome Browser Data Integration &Visualization—UCSC Genomics Institute, University of California Santa Cruz; Lead analysts; Laboratory, Data Analysis &Coordinating Center (LDACC); NIH program management; Biospecimen collection; Pathology; and eQTL

manuscript working group (2017). Genetic effects on gene expression across human tissues. Nature *550*, 204–213.
79. Prinz, J.C. (2018). Human Leukocyte Antigen-Class I Alleles and the Autoreactive T Cell Response in Psoriasis Pathogenesis. Front. Immunol. *9*, 954.
80. van Deutekom, H.W., and Keşmir, C. (2015). Zooming into the binding groove of HLA molecules: which positions and which substitutions change peptide binding most? Immunogenetics *67*, 425–436.
81. Kumar, R., Sharma, A., and Dogra, S. (2014). Prevalence and clinical patterns of psoriatic arthritis in Indian patients with psoriasis. Indian J. Dermatol. Venereol. Leprol. *80*, 15–23.
82. Gladman, D.D. (2005). Epidemiology. In Psoriasis and psoriatic arthrits: An integrated approach, G.B. Gordon and E. Ruderman, eds. (Heidelberg: Springer-Verlag), pp. 57–65.
83. Eder, L., Chandran, V., Pellet, F., Shanmugarajah, S., Rosen, C.F., Bull, S.B., and Gladman, D.D. (2011). Human leucocyte antigen risk alleles for psoriatic arthritis among patients with psoriasis. Ann. Rheum. Dis *71*, 50–55, 21900282.
84. Winchester, R., Minevich, G., Steshenko, V., Kirby, B., Kane, D., Greenberg, D.A., and FitzGerald, O. (2012). HLA associations reveal genetic heterogeneity in psoriatic arthritis and in the psoriasis phenotype. Arthritis Rheum. *64*, 1134–1144.
85. Jensen, J.M., Villesen, P., Friborg, R.M., Mailund, T., Besenbacher, S., Schierup, M.H.; and Danish Pan-Genome Consortium (2017). Assembly and analysis of 100 full MHC haplotypes from the Danish population. Genome Res. *27*, 1597–1607.
86. Maretty, L., Jensen, J.M., Petersen, B., Sibbesen, J.A., Liu, S., Villesen, P., Skov, L., Belling, K., Theil Have, C., Izarzugaza, J.M.G., et al. (2017). Sequencing and de novo assembly of 150 genomes from Denmark as a population reference. Nature *548*, 87–91.
87. Zheng, J., Li, Y., Abecasis, G.R., and Scheet, P. (2011). A comparison of approaches to account for uncertainty in analysis of imputed genotypes. Genet. Epidemiol. *35*, 102–110.
88. Harrell, F.E. (2001). Regression Modeling Strategies with Applications to Linear Models, Logistic Regression, and Survival Analysis (New York: Springer-Verlag).

HGGA, Volume *3*

Supplemental information

Transethnic analysis of psoriasis susceptibility in South Asians and Europeans enhances fine mapping in the MHC and genome wide

Philip E. Stuart, Lam C. Tsoi, Rajan P. Nair, Manju Ghosh, Madhulika Kabra, Pakeeza A. Shaiq, Ghazala K. Raja, Raheel Qamar, B.K. Thelma, Matthew T. Patrick, Anita Parihar, Sonam Singh, Sujay Khandpur, Uma Kumar, Michael Wittig, Frauke Degenhardt, Trilokraj Tejasvi, John J. Voorhees, Stephan Weidinger, Andre Franke, Goncalo R. Abecasis, Vinod K. Sharma, and James T. Elder

Table of Contents

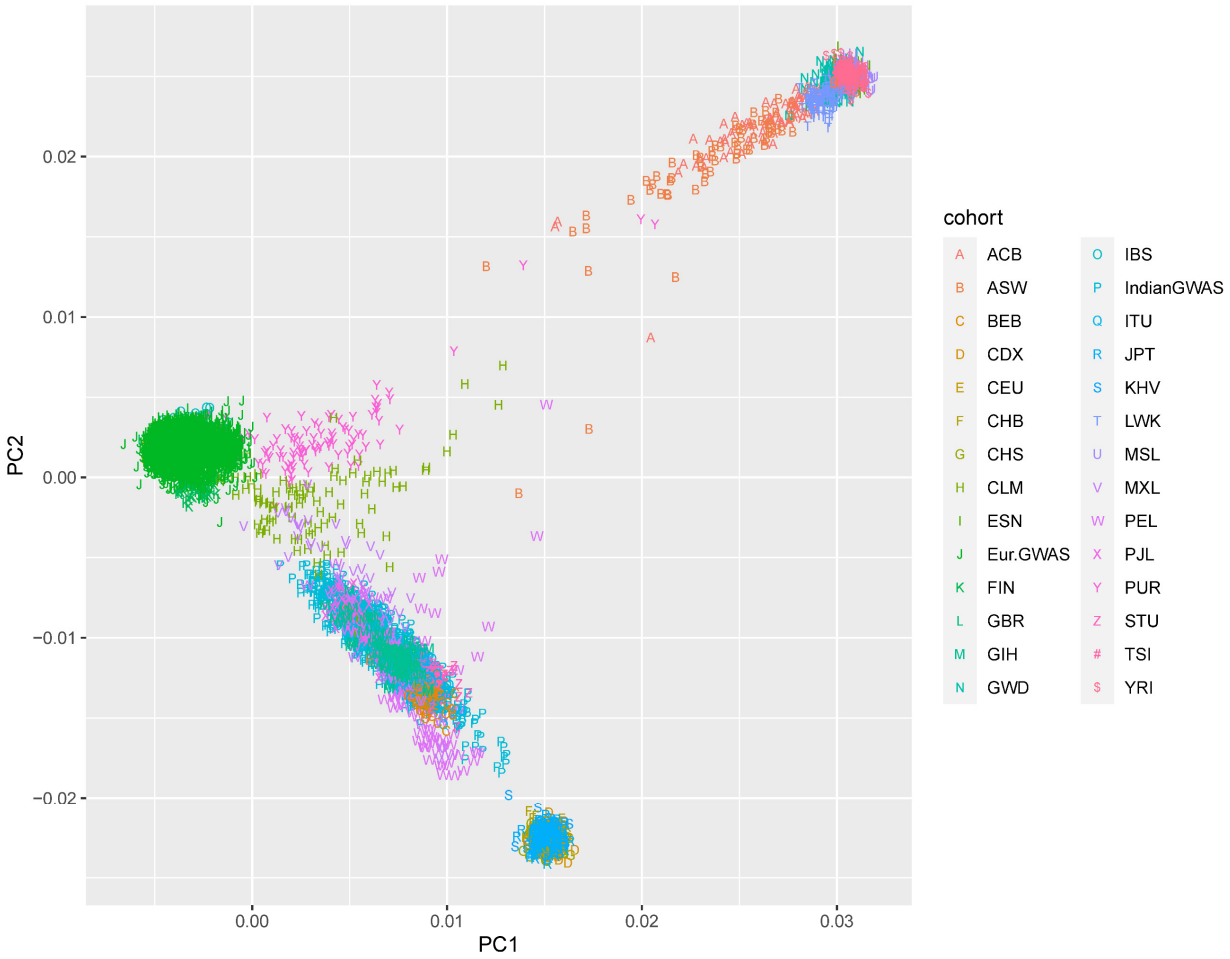

Figure S1. Principal component analysis of the European and South Asian cohorts of this study. Plot of the first two axes of a principal component analysis of the 44,161 individuals of European ancestry (denoted with a green J) and 4310 individuals of South Asian ancestry (denoted with a cyan P) analyzed by this study. The EUR and SAS individuals of this study were analyzed with the 26 global populations of phase 3 of the 1000 Genomes Project. Only those EUR and SAS individuals passing all quality control filters, including removal of population outliers detected with PCA, are plotted here.

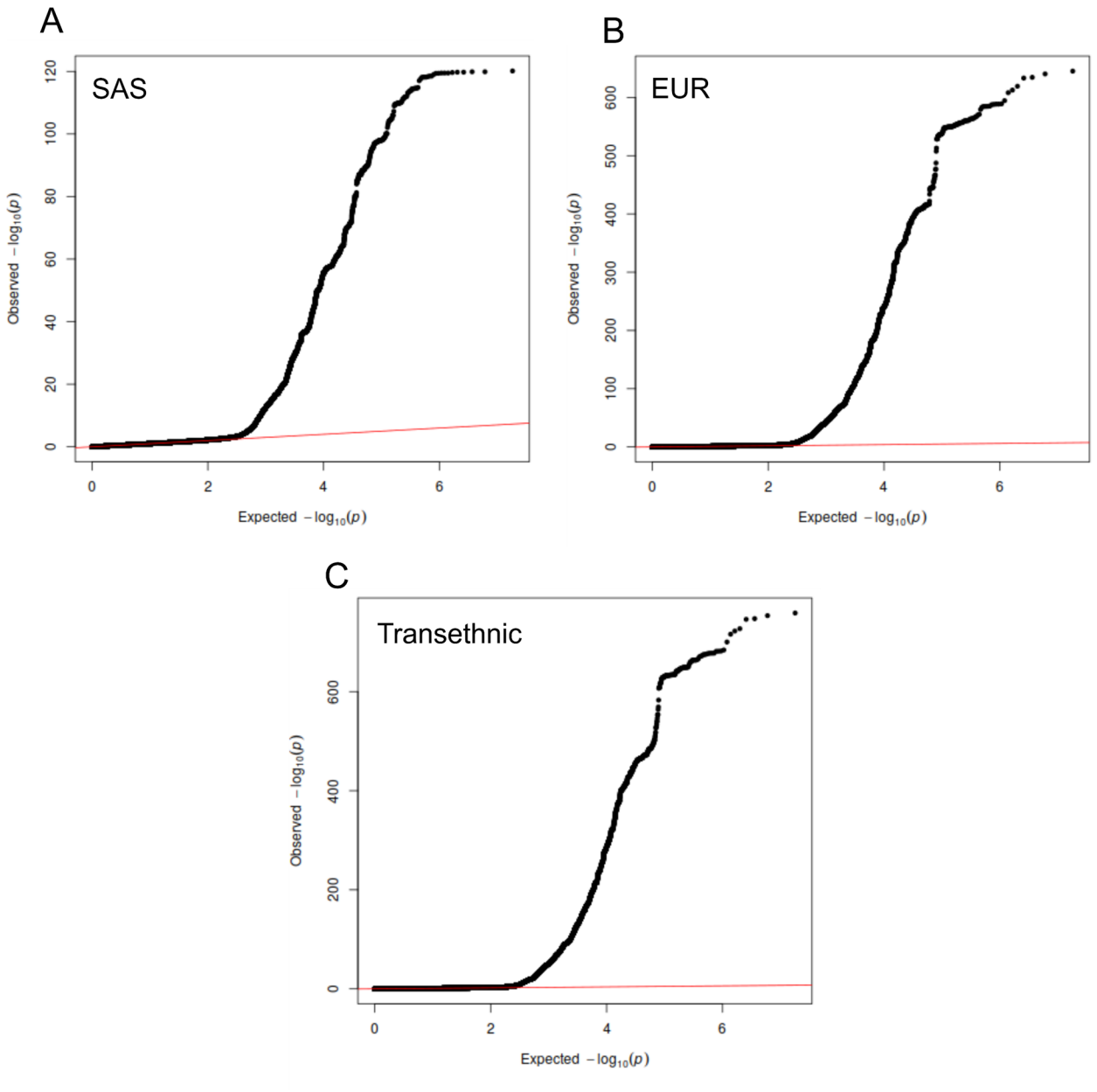

Figure S2. QQ-plots for the three genome-wide meta-analyses of psoriasis associations. Quantile-quantile plots of observed vs. expected $-\log_{10}$ of association p-values are plotted for the South Asian (SAS), European (EUR) and transethnic (SAS+EUR) meta-analyses in panels A, B and C, respectively.

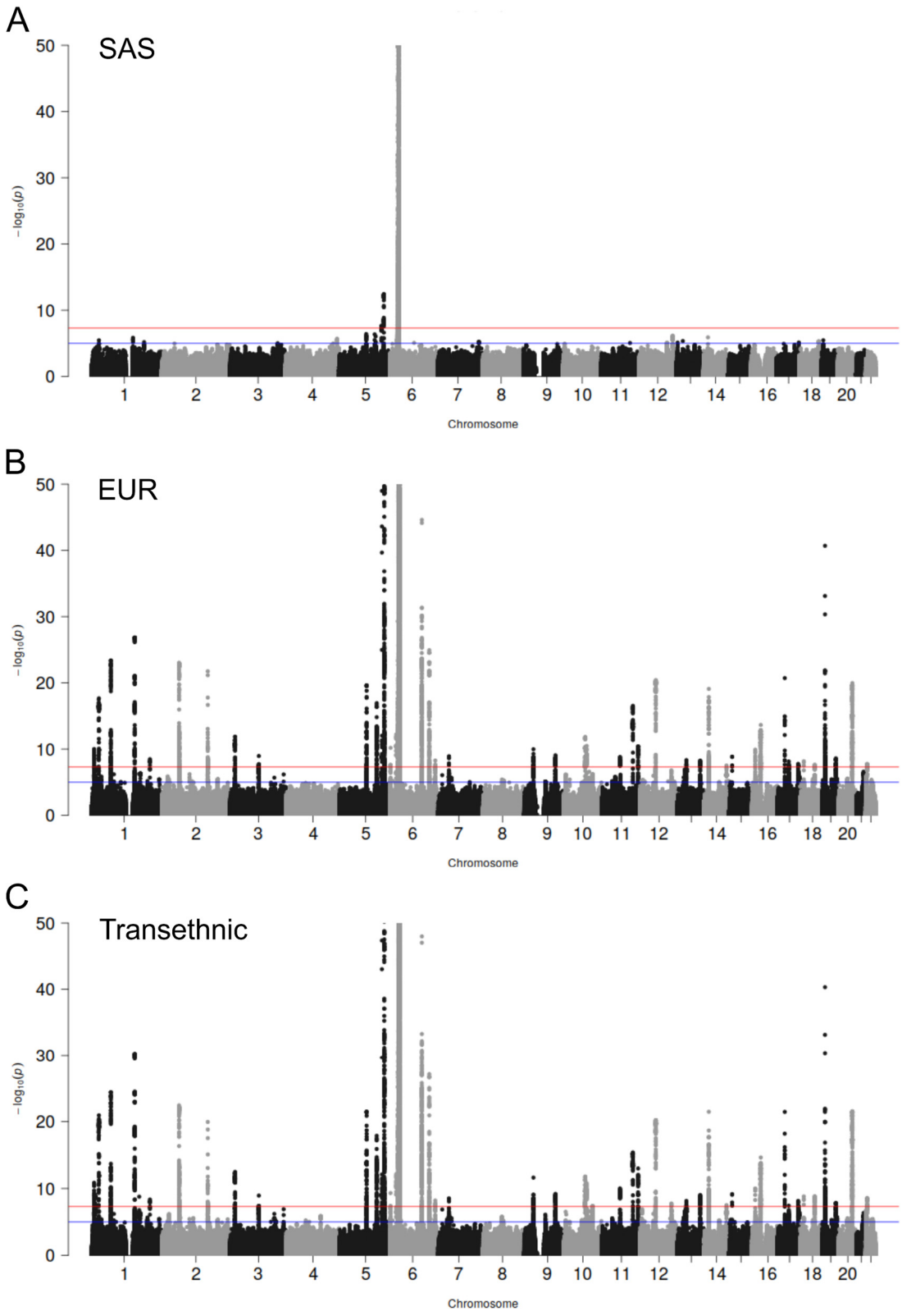

Figure S3. Manhattan plots of psoriasis associations. Negative $\log_{10}$ of association p-values are plotted against chromosomal position for meta-analyses of two South Asian cohorts (panel A), eight European ancestry cohorts (panel B), and a transethnic analysis of all ten cohorts (panel C). For better visualization of less significant peaks, the y-axis of all three panels is truncated at $-\log_{10}(p)$ of 50.

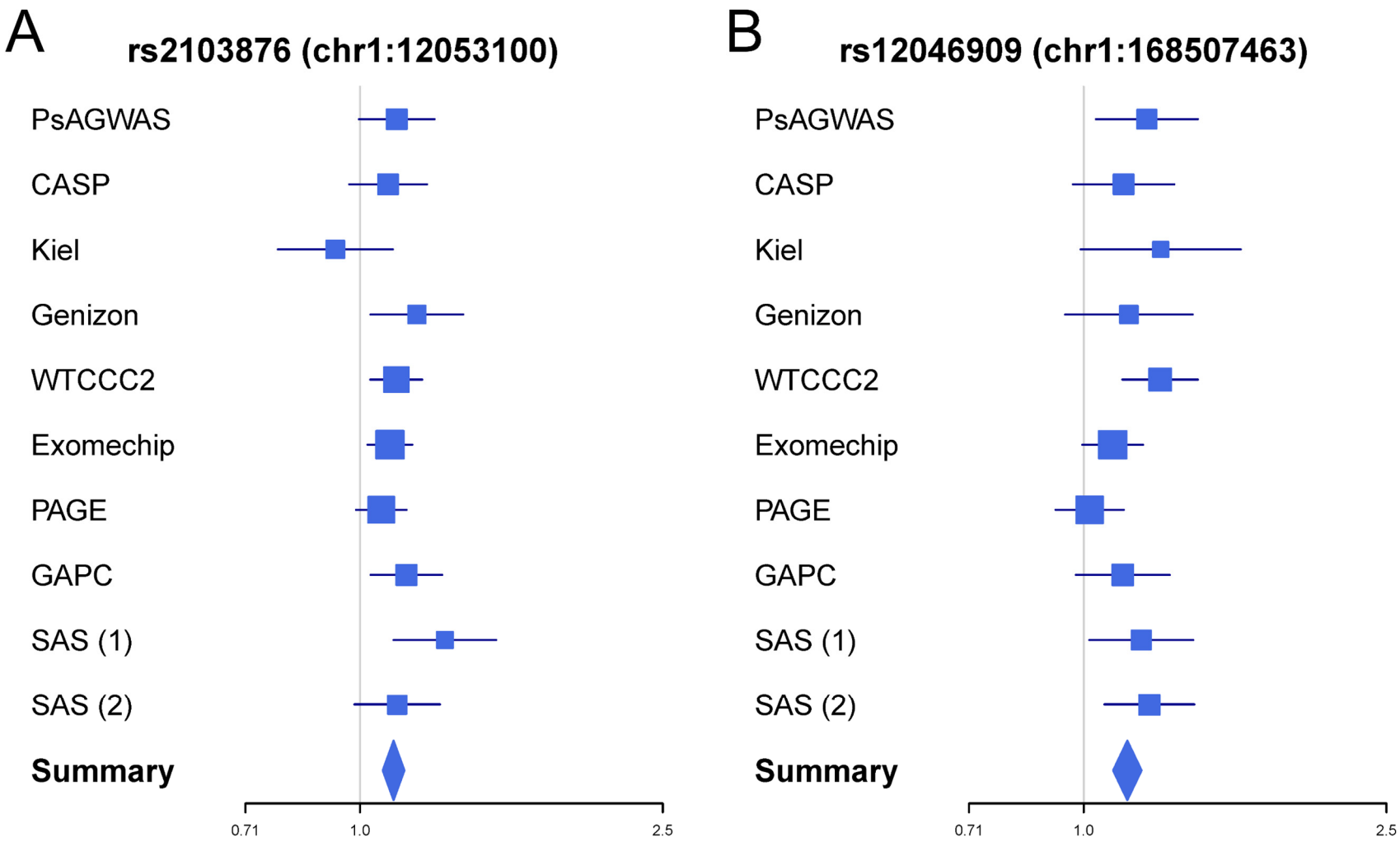

Figure S4. Forest plots for two newly established psoriasis risk loci. Odds ratios and their 95% confidence intervals are plotted for the association of psoriasis with two psoriasis risk variants (rs2103876 in panel A and rs12046909 in panel B) newly established by the transethnic meta-analysis. Results for the eight European ancestry cohorts are shown in the first 8 rows, for the two South Asian (SAS) cohorts in the next two rows, and the overall meta-analysis result is shown at the bottom.

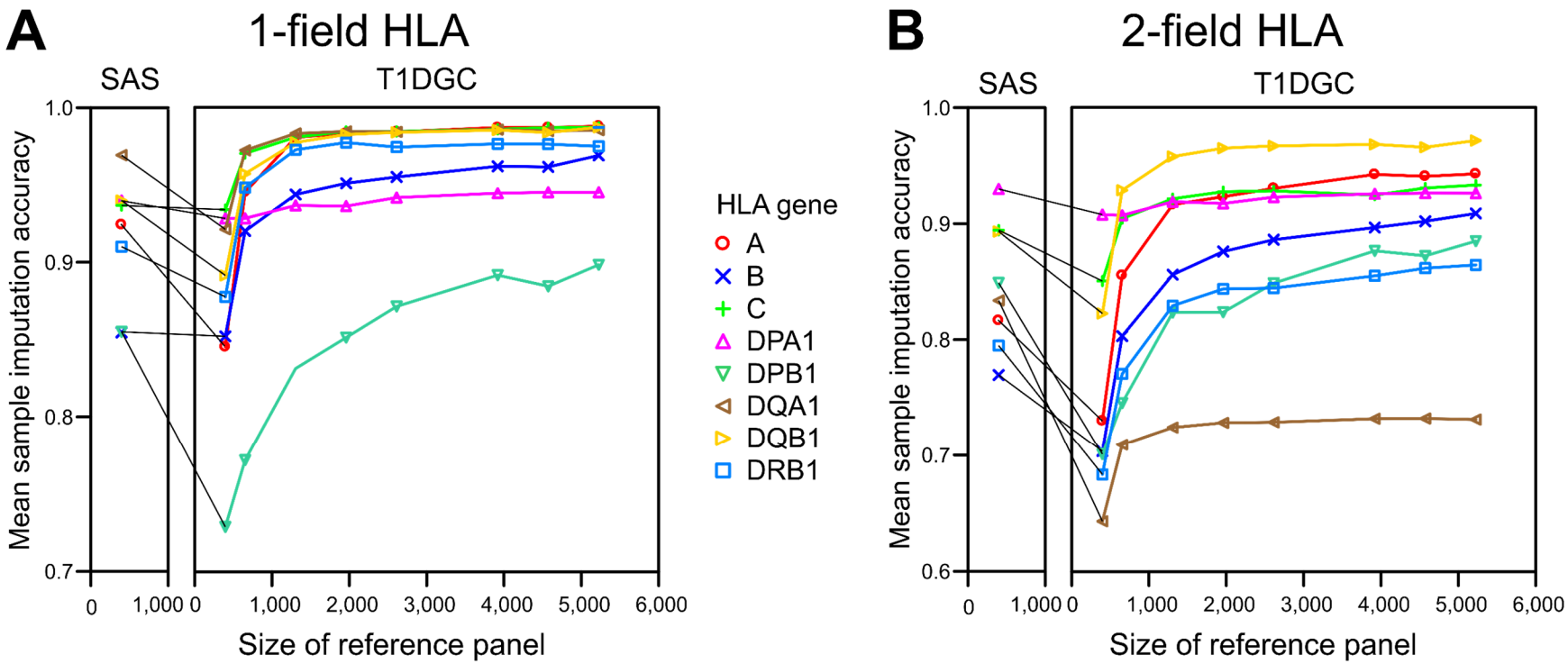

Figure S5. Comparing accuracy of HLA alleles imputed for a set of 397 South Asians when using a SNP2HLA reference panel of the same 397 South Asians vs. increasingly large random subsets of the European-ancestry T1DGC reference panel. Results are shown for 1-field (panel A) and 2-field (panel B) alleles of eight HLA genes. The left and right parts of each panel show accuracies when using the South Asian (SAS) and T1DGC SNP2HLA panels, respectively. Leave-one-out cross-validation was used to estimate imputation accuracies for the South Asian panel. For the T1DGC panel, eight randomly sampled subsets were assessed, with sample sizes of 397, 653, 1306, 1959, 2613, 3919, 4572 and 5225, representing 7.6%, 12.5%, 25%, 37.5%, 50%, 75%, 87.5% and 100% of the full T1DGC panel, respectively. In each panel, lines connect accuracies based on South Asian and European reference panels of equal sizes (n=397).

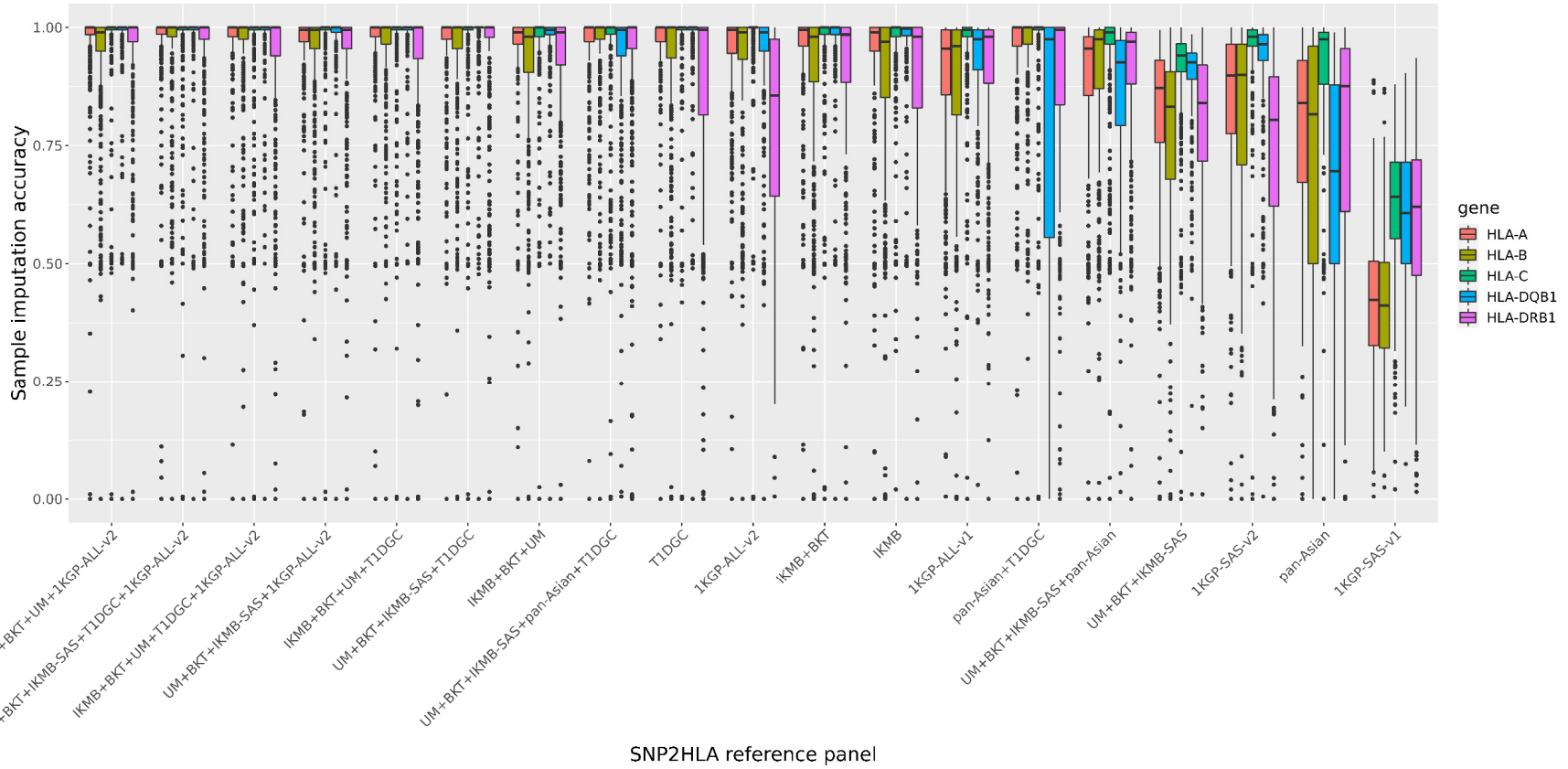

Figure S6. Boxplots comparing sample imputation accuracy of 2-field protein alleles for five HLA genes as a function of 19 different SNP2HLA reference panels for a validation set of 397 people of South Asian ancestry. Panels are listed in decreasing order of their mean rank across 12 performance metrics based on the Wilcoxon signed rank test and the paired t-test. Sample imputation accuracy for a given HLA gene is measured for each individual in the validation set by subtracting, from 1, one-half of the sum for that individual of the positive differences in genotyped vs. imputed dosages of all 2-field alleles of that gene in the reference panel.

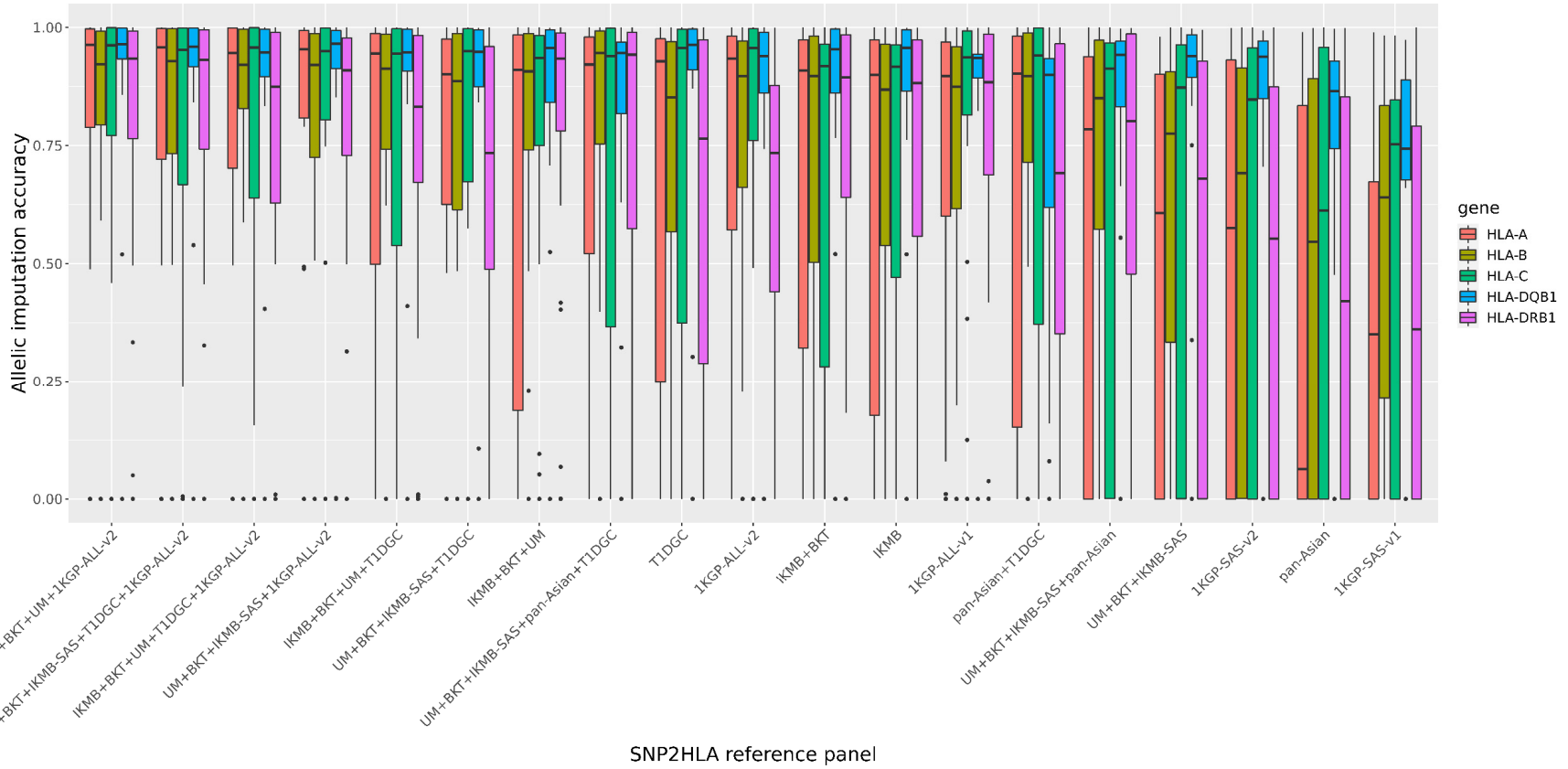

Figure S7. Boxplots comparing allelic imputation accuracy of 2-field protein alleles for five HLA genes as a function of 19 different SNP2HLA reference panels for a validation set of 397 people of South Asian ancestry. Panels are listed in decreasing order of their mean rank across 12 performance metrics based on the Wilcoxon signed rank test and the paired t-test. Allelic imputation accuracy for a given HLA gene is measured for each 2-field allele of that gene in the reference panel by computing the squared Pearson correlation of vectors of genotyped and imputed dosages of that allele for all individuals in the validation set.

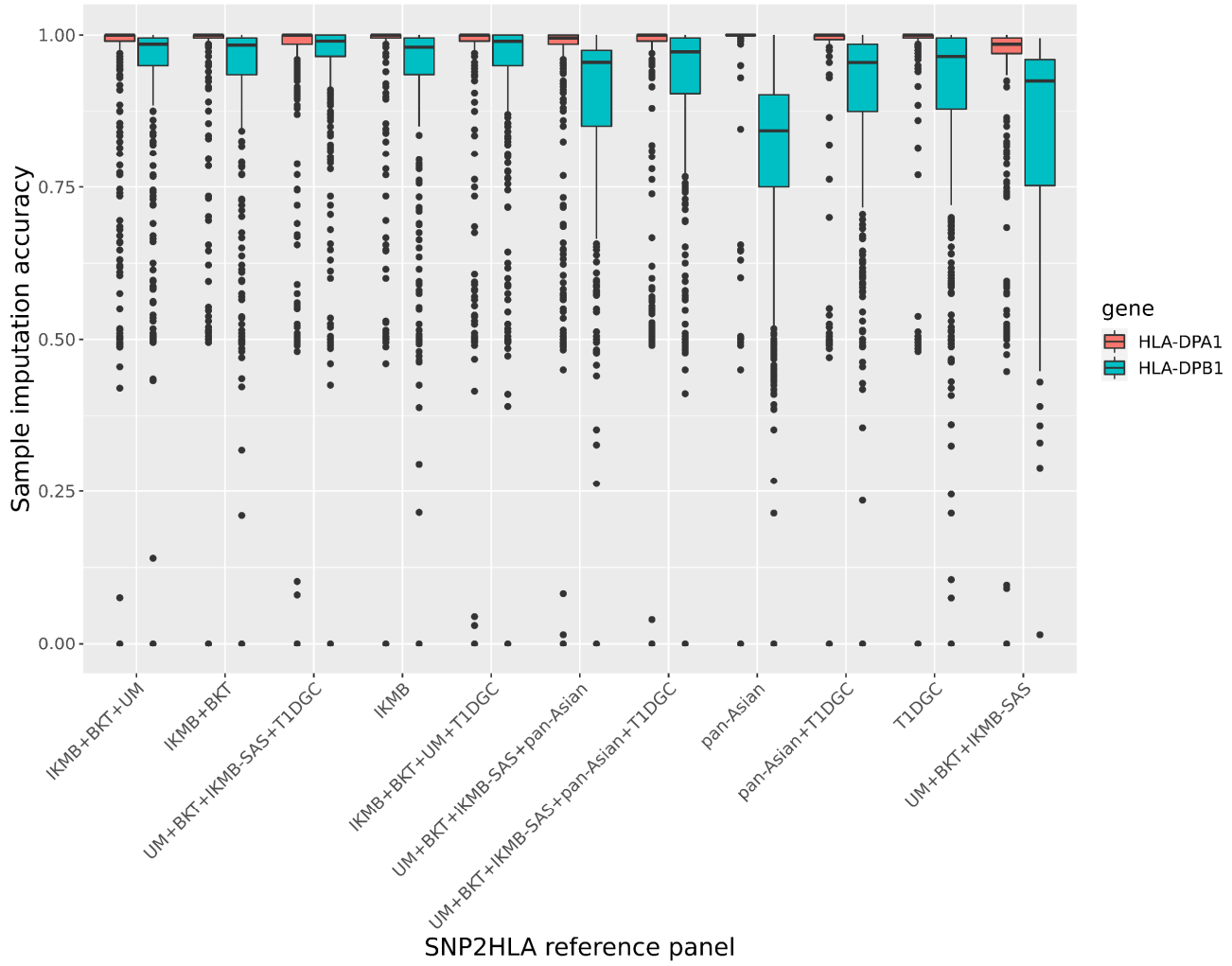

Figure S8. Boxplots comparing sample imputation accuracy of 2-field protein alleles for two HLA genes as a function of 11 different SNP2HLA reference panels for a validation set of 397 people of South Asian ancestry. Panels are listed in decreasing order of their mean rank across 12 performance metrics based on the Wilcoxon signed rank test and the paired t-test. Sample imputation accuracy for a given HLA gene is measured for each individual in the validation set by subtracting, from 1, one-half of the sum for that individual of the positive differences in genotyped vs. imputed dosages of all 2-field alleles of that gene in the reference panel.

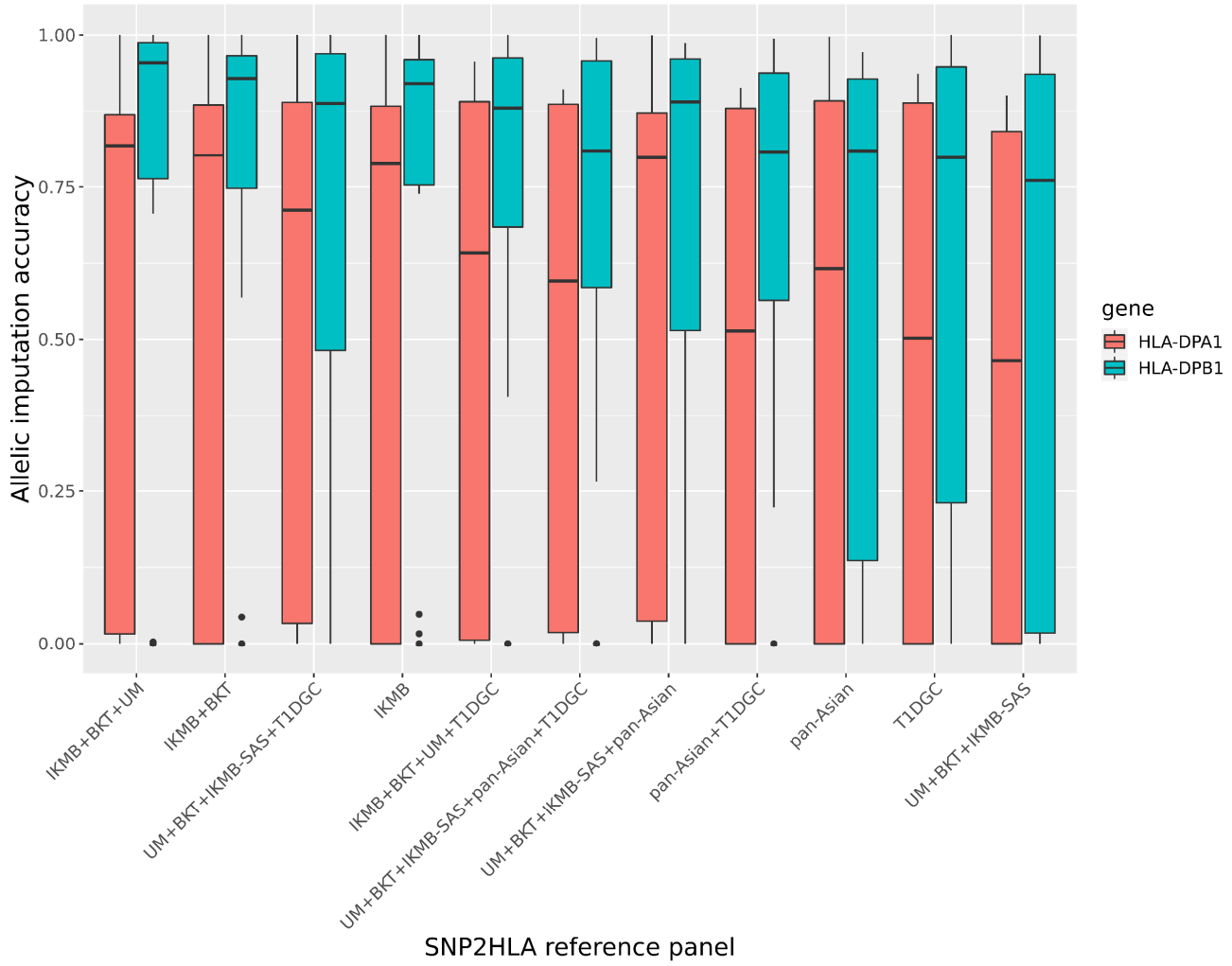

Figure S9. Boxplots comparing allelic imputation accuracy of 2-field protein alleles for two HLA genes as a function of 11 different SNP2HLA reference panels for a validation set of 397 people of South Asian ancestry. Panels are listed in decreasing order of their mean rank across 12 performance metrics based on the Wilcoxon signed rank test and the paired t-test. Allelic imputation accuracy for a given HLA gene is measured for each 2-field allele of that gene in the reference panel by computing the squared Pearson correlation of vectors of genotyped and imputed dosages of that allele for all individuals in the validation set.

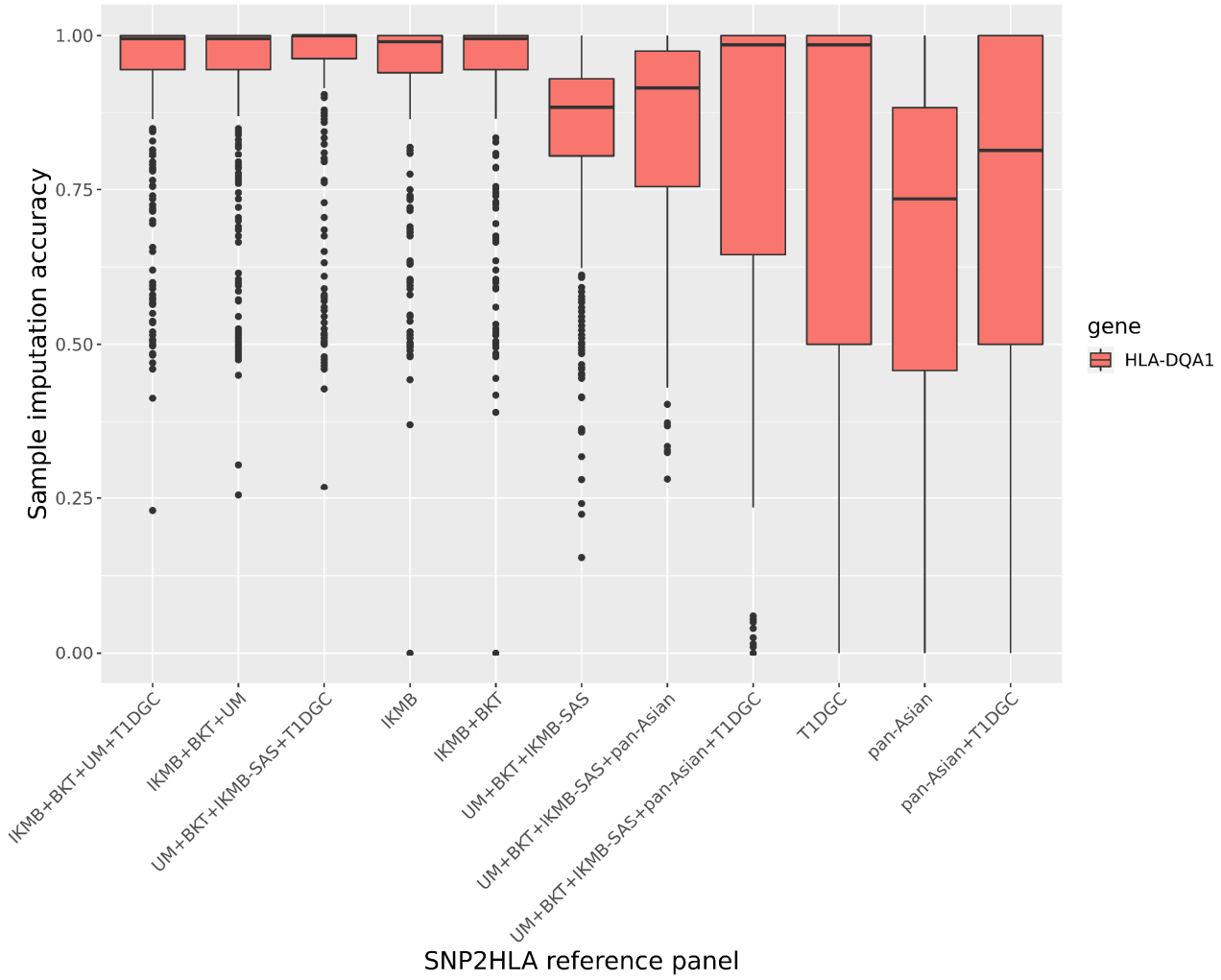

Figure S10. Boxplots comparing sample imputation accuracy of 2-field protein alleles for *HLA-DQA1* as a function of 11 different SNP2HLA reference panels for a validation set of 397 people of South Asian ancestry. Panels are listed in decreasing order of their mean rank across 12 performance metrics based on the Wilcoxon signed rank test and the paired t-test. Sample imputation accuracy is measured for each individual in the validation set by subtracting, from 1, one-half of the sum for that individual of the positive differences in genotyped vs. imputed dosages of all 2-field alleles of *HLA-DQA1* in the reference panel.

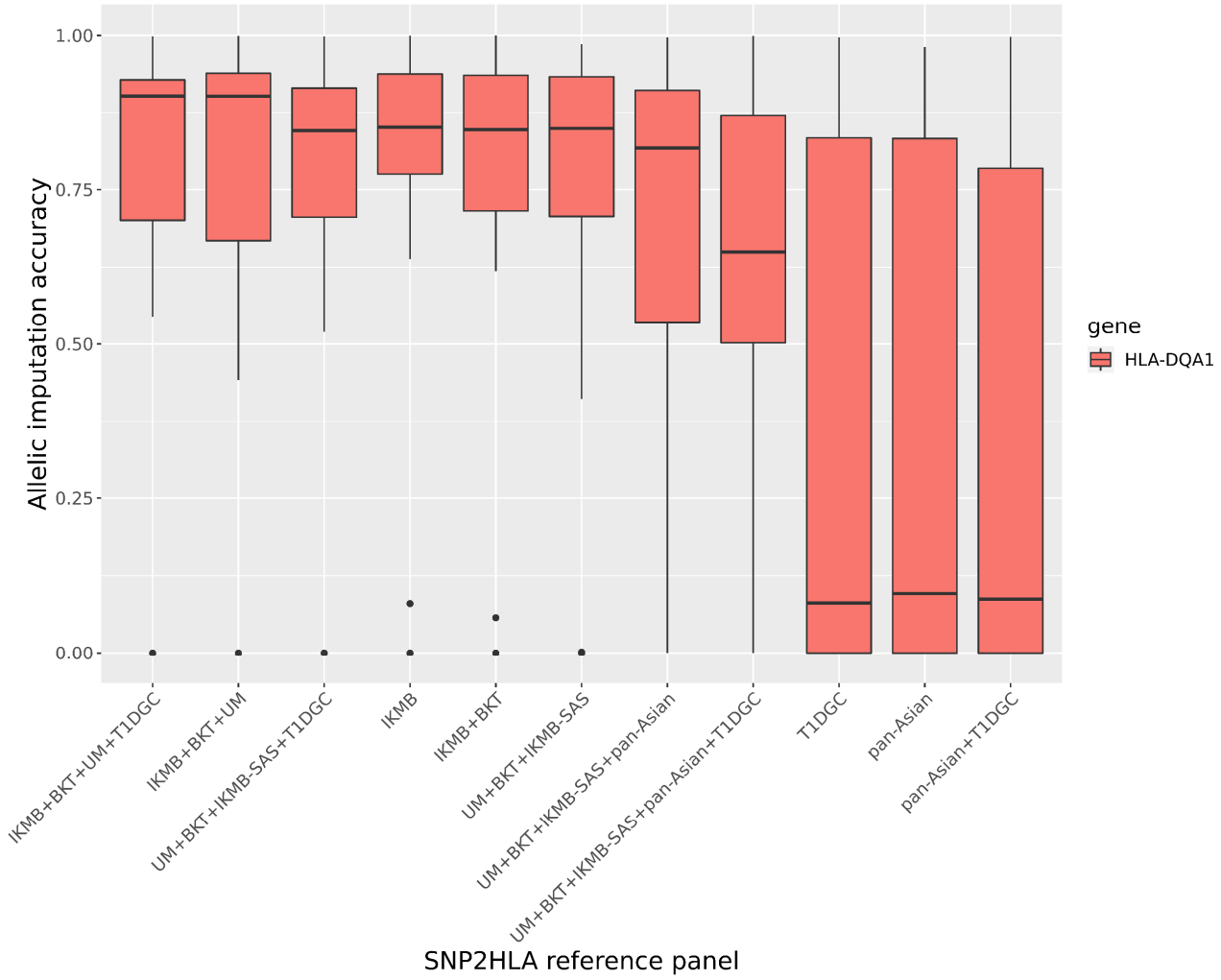

Figure S11. Boxplots comparing allelic imputation accuracy of 2-field protein alleles for *HLA-DQA1* as a function of 11 different SNP2HLA reference panels for a validation set of 397 people of South Asian ancestry. Panels are listed in decreasing order of their mean rank across 12 performance metrics based on the Wilcoxon signed rank test and the paired t-test. Allelic imputation accuracy is measured for each 2-field allele of *HLA-DQA1* in the reference panel by computing the squared Pearson correlation of vectors of genotyped and imputed dosages of that allele for all individuals in the validation set.

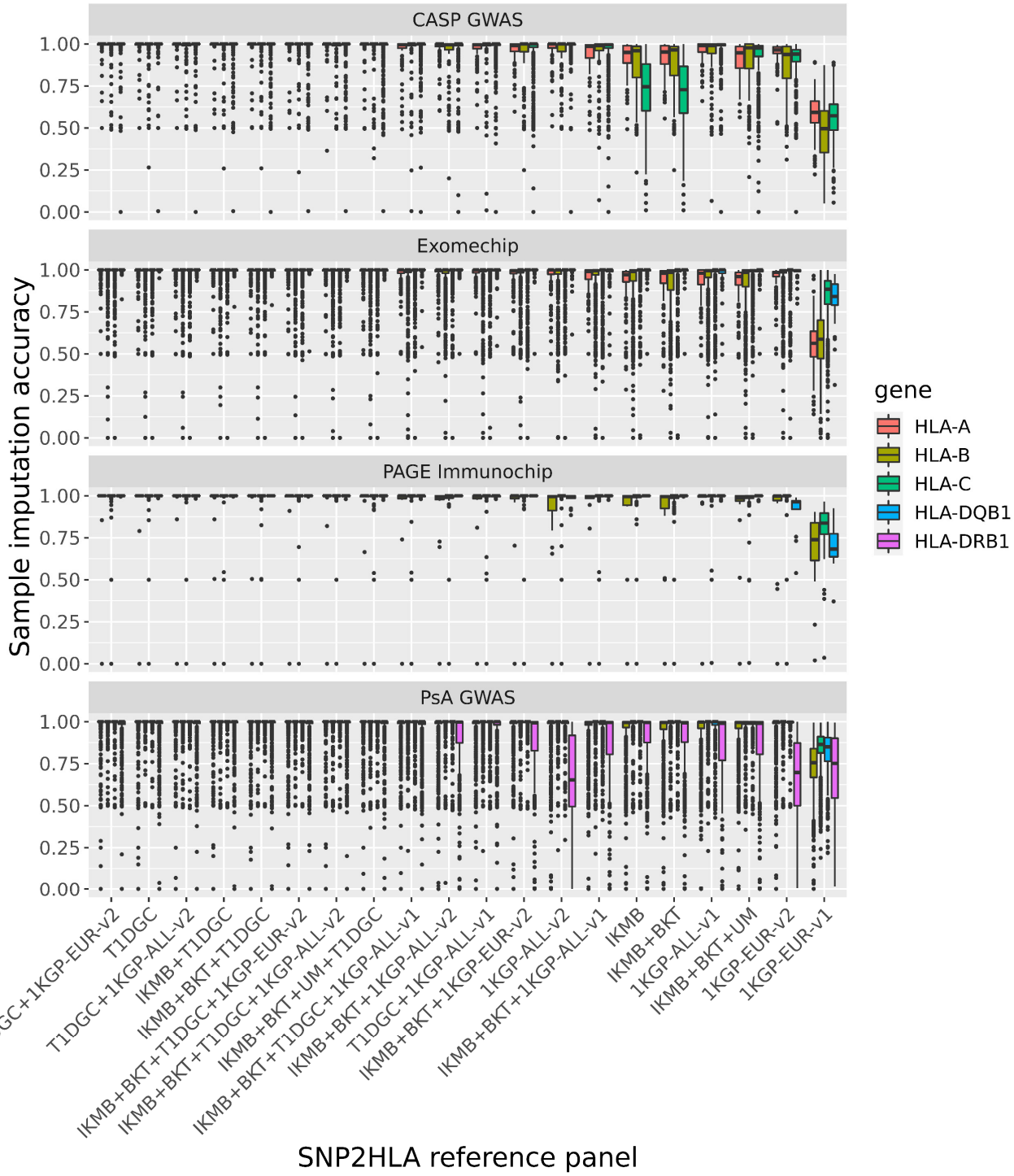

Figure S12. Boxplots comparing sample imputation accuracy of 2-field protein alleles for five HLA genes as a function of 20 different SNP2HLA reference panels for four validation sets of people of European ancestry. Panels are listed in decreasing order of the weighted mean across the four validation sets of their mean rank across 12 performance metrics based on the Wilcoxon signed rank test and the paired t-test. Sample imputation accuracy for a given HLA gene is measured for each individual in the validation set by subtracting, from 1, one-half of the sum for that individual of the positive differences in genotyped vs. imputed dosages of all 2-field alleles of that gene in the reference panel.

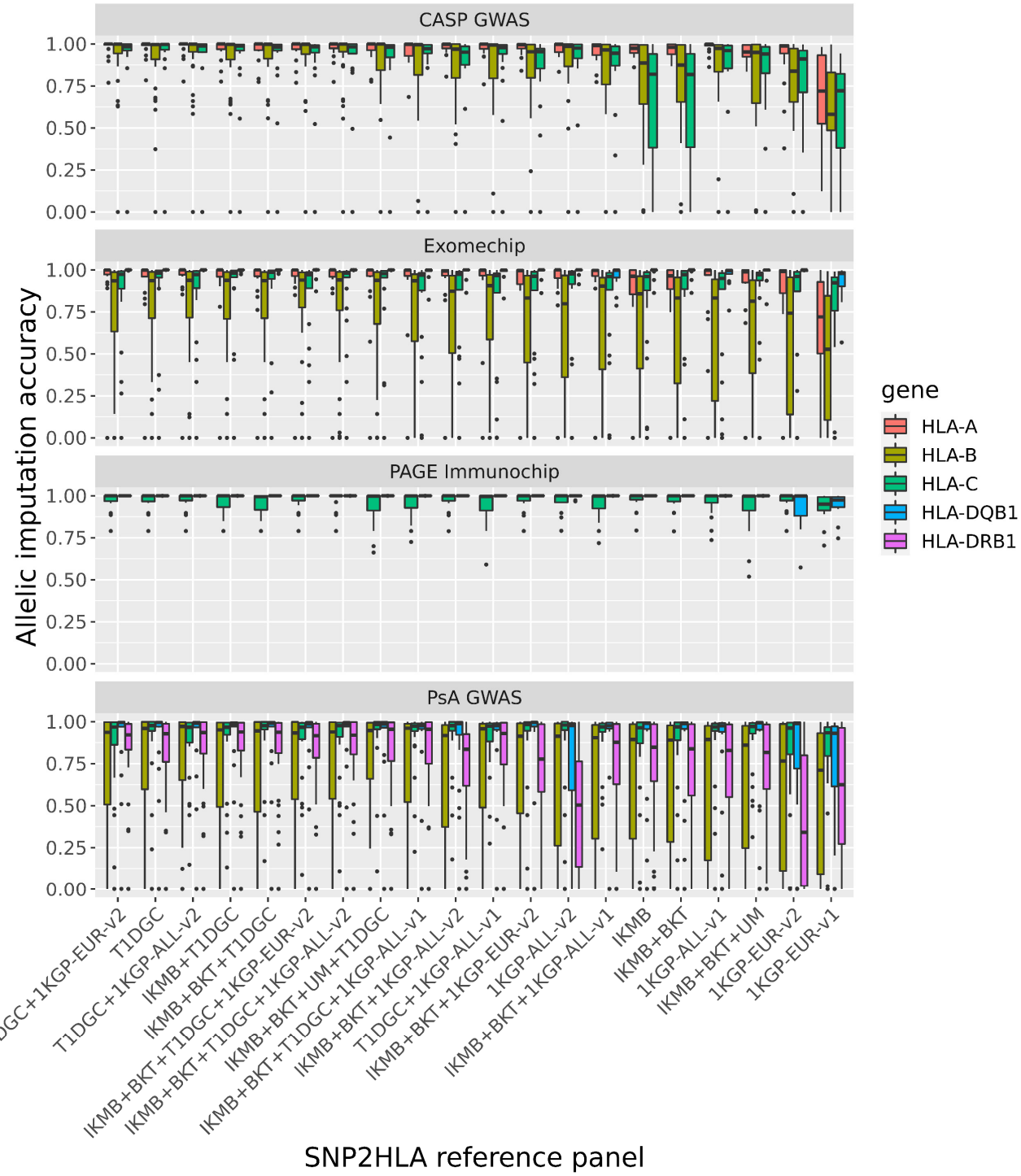

Figure S13. Boxplots comparing allelic imputation accuracy of 2-field protein alleles for five HLA genes as a function of 20 different SNP2HLA reference panels for four validation sets of people of European ancestry. Panels are listed in decreasing order of the weighted mean across the four validation sets of their mean rank across 12 performance metrics based on the Wilcoxon signed rank test and the paired t-test. Allelic imputation accuracy for a given HLA gene is measured for each 2-field allele of that gene in the reference panel by computing the squared Pearson correlation of vectors of genotyped and imputed dosages of that allele for all individuals in the validation set.

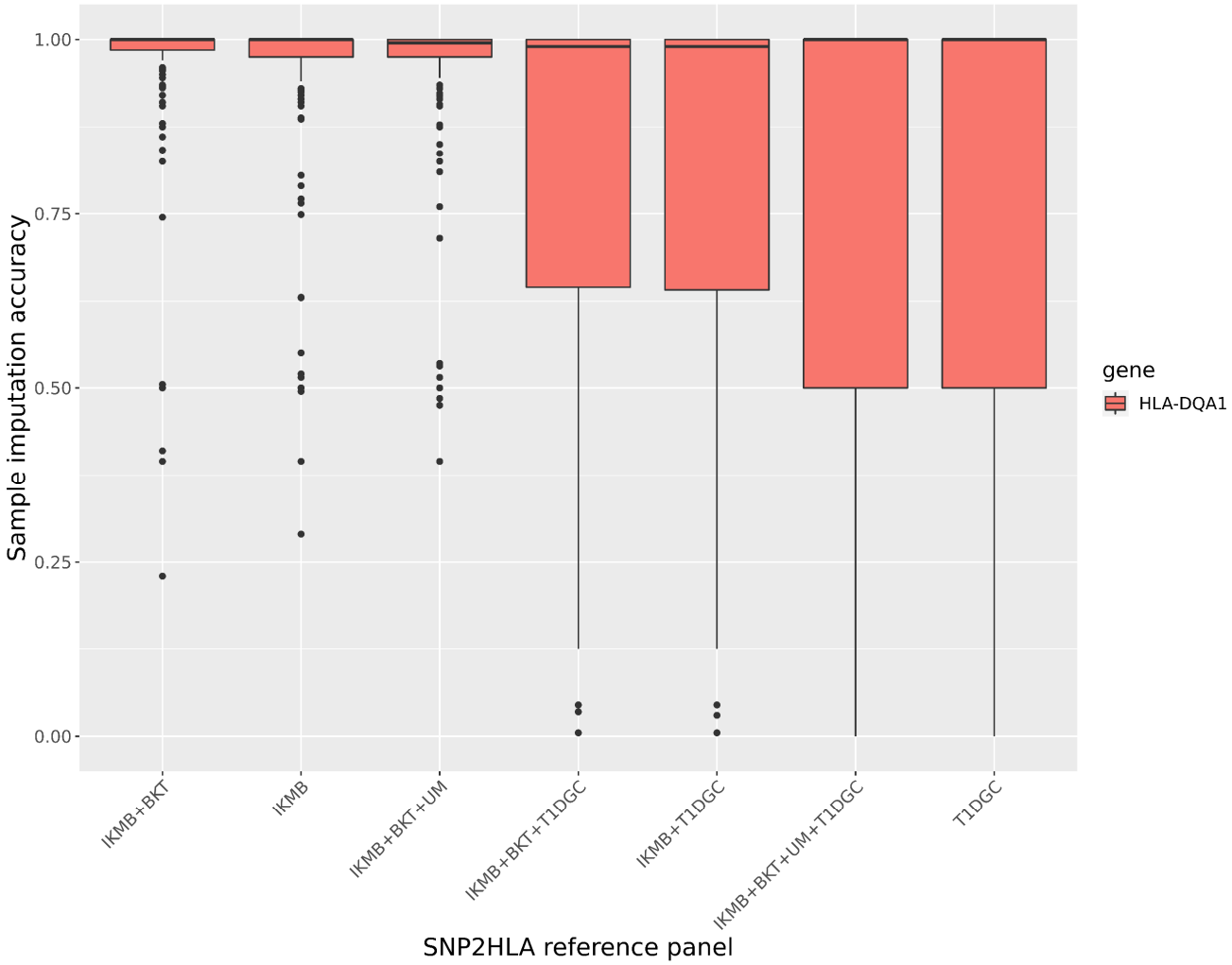

Figure S14. Boxplots comparing sample imputation accuracy of 2-field protein alleles for *HLA-DQA1* as a function of seven different SNP2HLA reference panels for a validation set of 174 people of European ancestry. Panels are listed in decreasing order of the weighted mean across the four validation sets of their mean rank across 12 performance metrics based on the Wilcoxon signed rank test and the paired t-test. Sample imputation accuracy is measured for each individual in the validation set by subtracting, from 1, one-half of the sum for that individual of the positive differences in genotyped vs. imputed dosages of all 2-field alleles of *HLA-DQA1* in the reference panel.

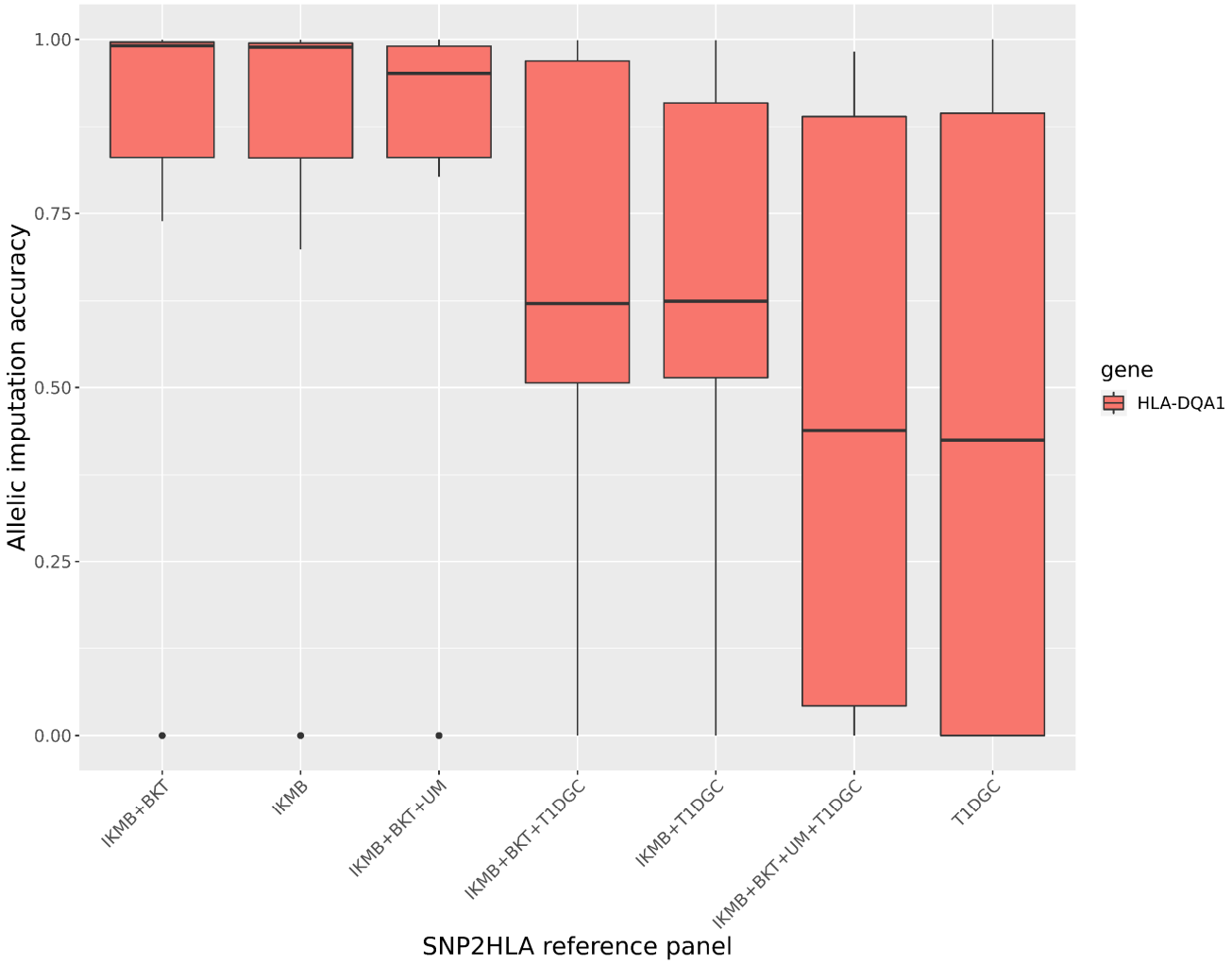

Figure S15. Boxplots comparing allelic imputation accuracy of 2-field protein alleles for *HLA-DQA1* as a function of seven different SNP2HLA reference panels for a validation set of 174 people of European ancestry. Panels are listed in decreasing order of the weighted mean across the four validation sets of their mean rank across 12 performance metrics based on the Wilcoxon signed rank test and the paired t-test. Allelic imputation accuracy is measured for each 2-field allele of *HLA-DQA1* in the reference panel by computing the squared Pearson correlation of vectors of genotyped and imputed dosages of that allele for all individuals in the validation set.

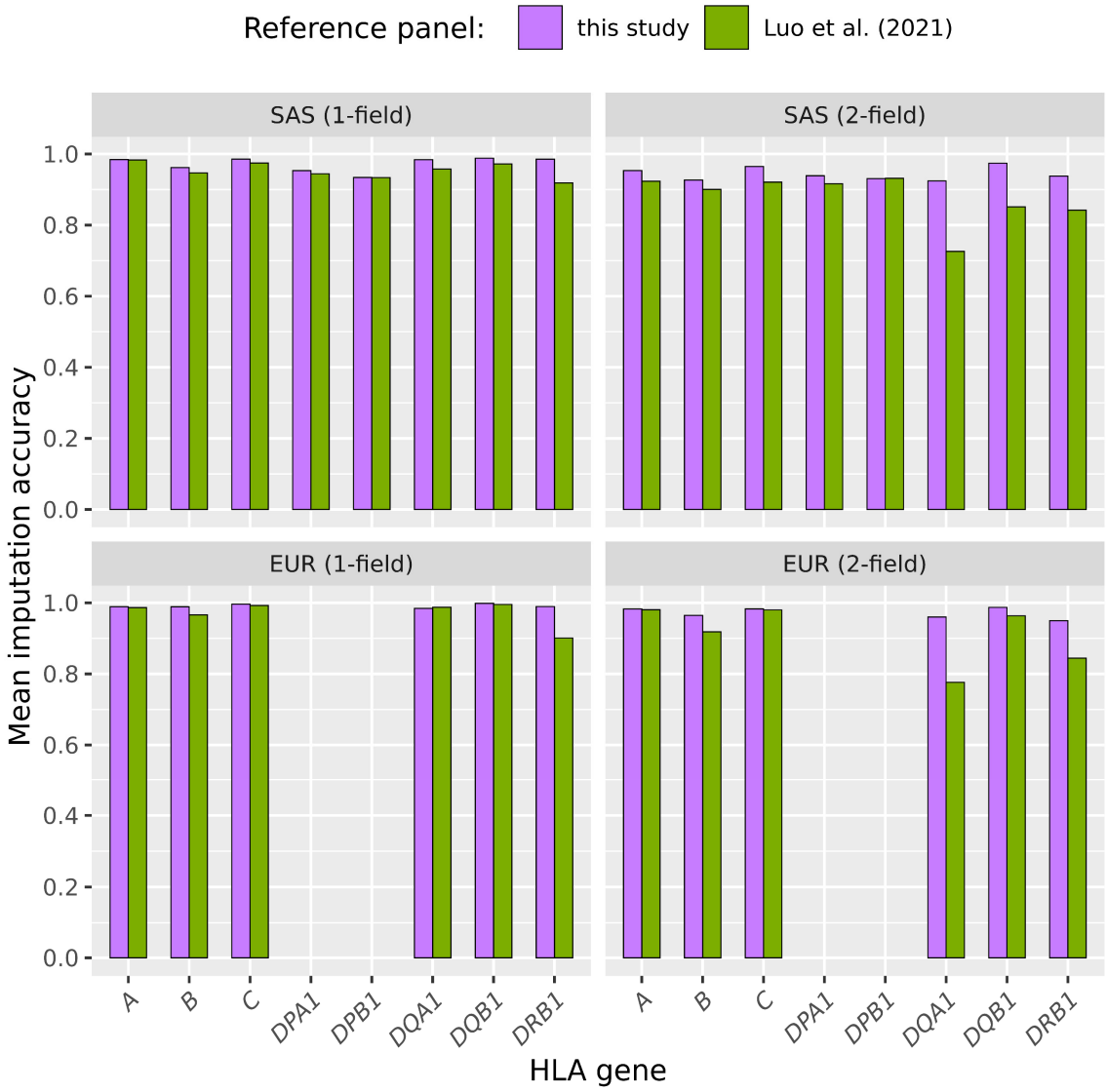

Figure S16. Comparison of HLA imputation accuracies achieved with the best-performing reference panels of this study vs. those obtained using a recently published multi-ancestry panel. Purple bars display accuracies achieved using the suite of best-performing HLA panels of this study for SAS targets (IKMB+BKT+UM+1KGP-ALL-v2 for *HLA-A,B,C,DQB1,DRB1*; IKMB+BKT+UM for *HLA-DPA1,DPB1*; IKMB+BKT+UM+T1DGC for *HLA-DQA1*) and for EUR targets (T1DGC+1KGP-EUR-v2 for *HLA-A,B,C,DQB1,DRB1*; IKMB+BKT for *HLA-DQA1*). Green bars display accuracy achieved for both SAS and EUR targets with a recently published multi-ancestry panel of 21,546 individuals[1] that is incorporated into the Michigan Imputation Server[2]. The target samples were the HLA-genotyped validation sets used by this study; namely 397 individuals of SAS ancestry and 174–3749 individuals of EUR ancestry. The top and bottom pairs of panels show results for 1- and 2-field allele resolution for SAS and EUR individuals, respectively.

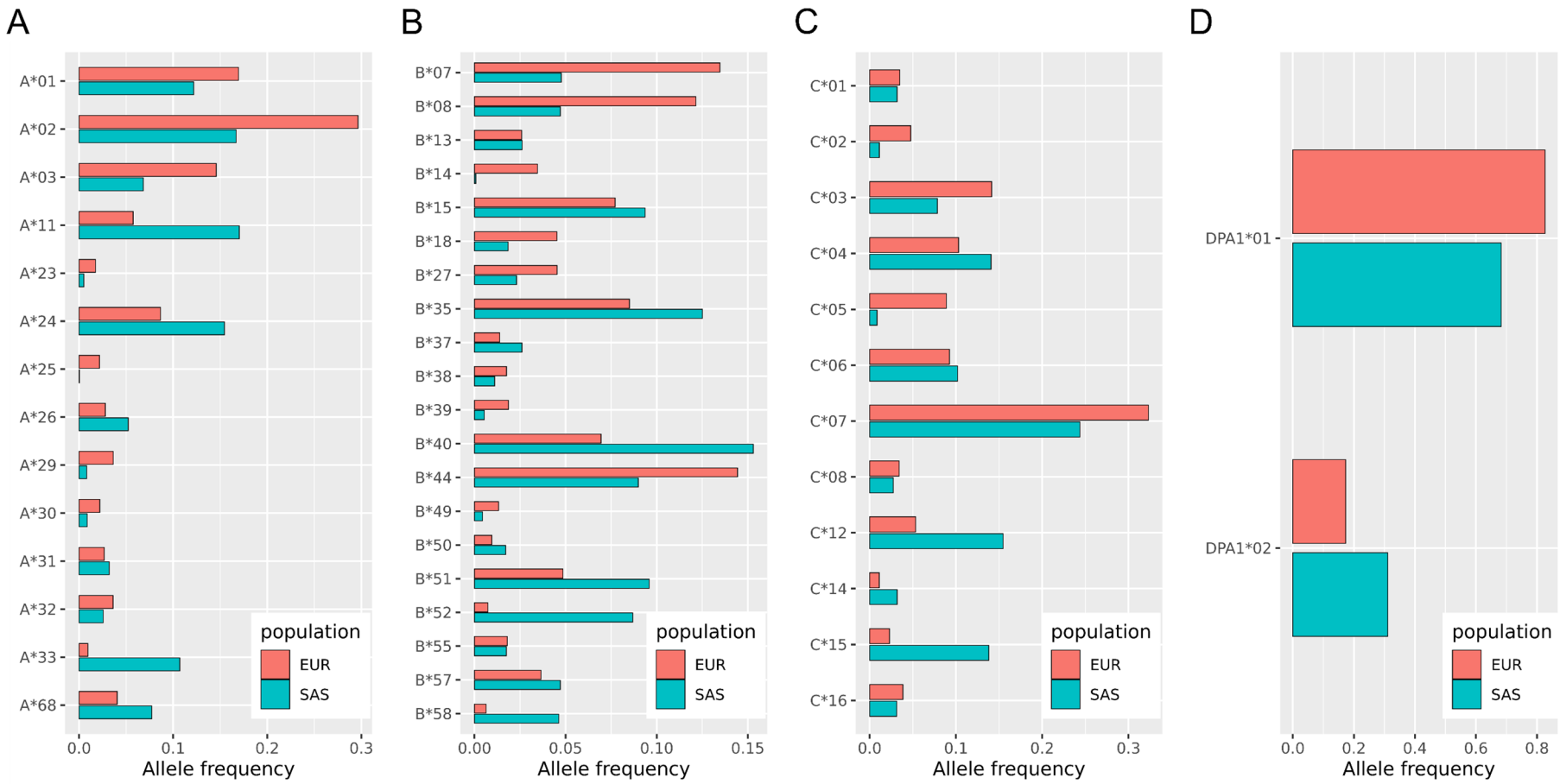

Figure S17. Imputed 1-field *HLA-A*, *HLA-B*, *HLA-C* and *HLA-DPA1* allele frequencies for the European (EUR) and South Asian (SAS) datasets of this study. Results for *HLA-A*, *HLA-B*, *HLA-C* and *HLA-DPA1* are shown in panels A, B, C and D, respectively. Frequencies for the EUR and SAS populations were estimated as a weighted average of frequencies in psoriasis cases and unaffected controls (weights = 0.015 and 0.985 for EUR and 0.003 and 0.997 for SAS). Alleles with frequencies < 0.01 in both populations are omitted.

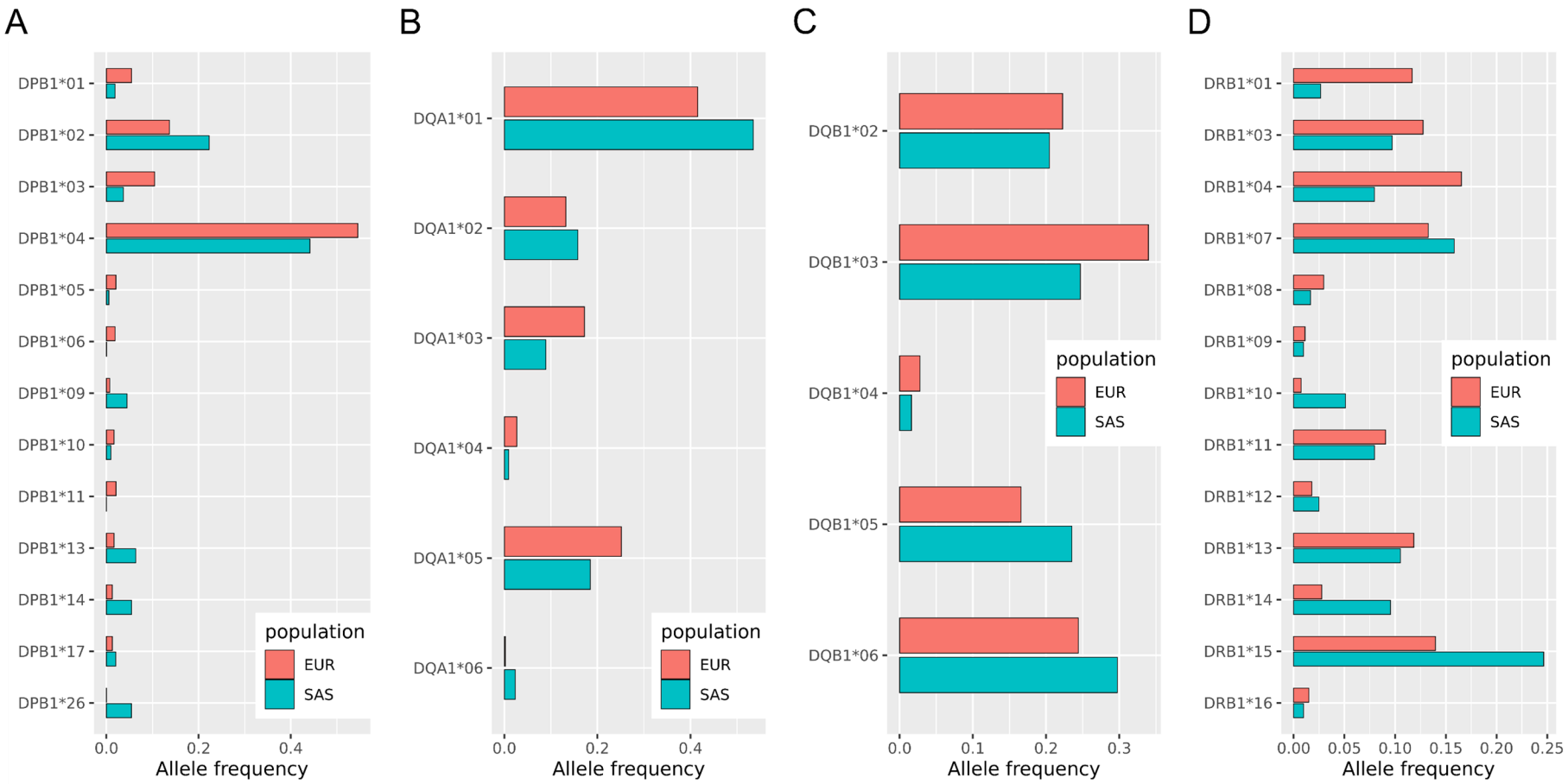

Figure S18. Imputed 1-field *HLA-DPB1*, *HLA-DQA1*, *HLA-DQB1* and *HLA-DRB1* allele frequencies for the European (EUR) and South Asian (SAS) datasets of this study. Results for *HLA-DPB1*, *HLA-DQA1*, *HLA-DQB1* and *HLA-DRB1* are shown in panels A, B, C and D, respectively. Frequencies for the EUR and SAS populations were estimated as a weighted average of frequencies in psoriasis cases and unaffected controls (weights = 0.015 and 0.985 for EUR and 0.003 and 0.997 for SAS). Alleles with frequencies < 0.01 in both populations are omitted.

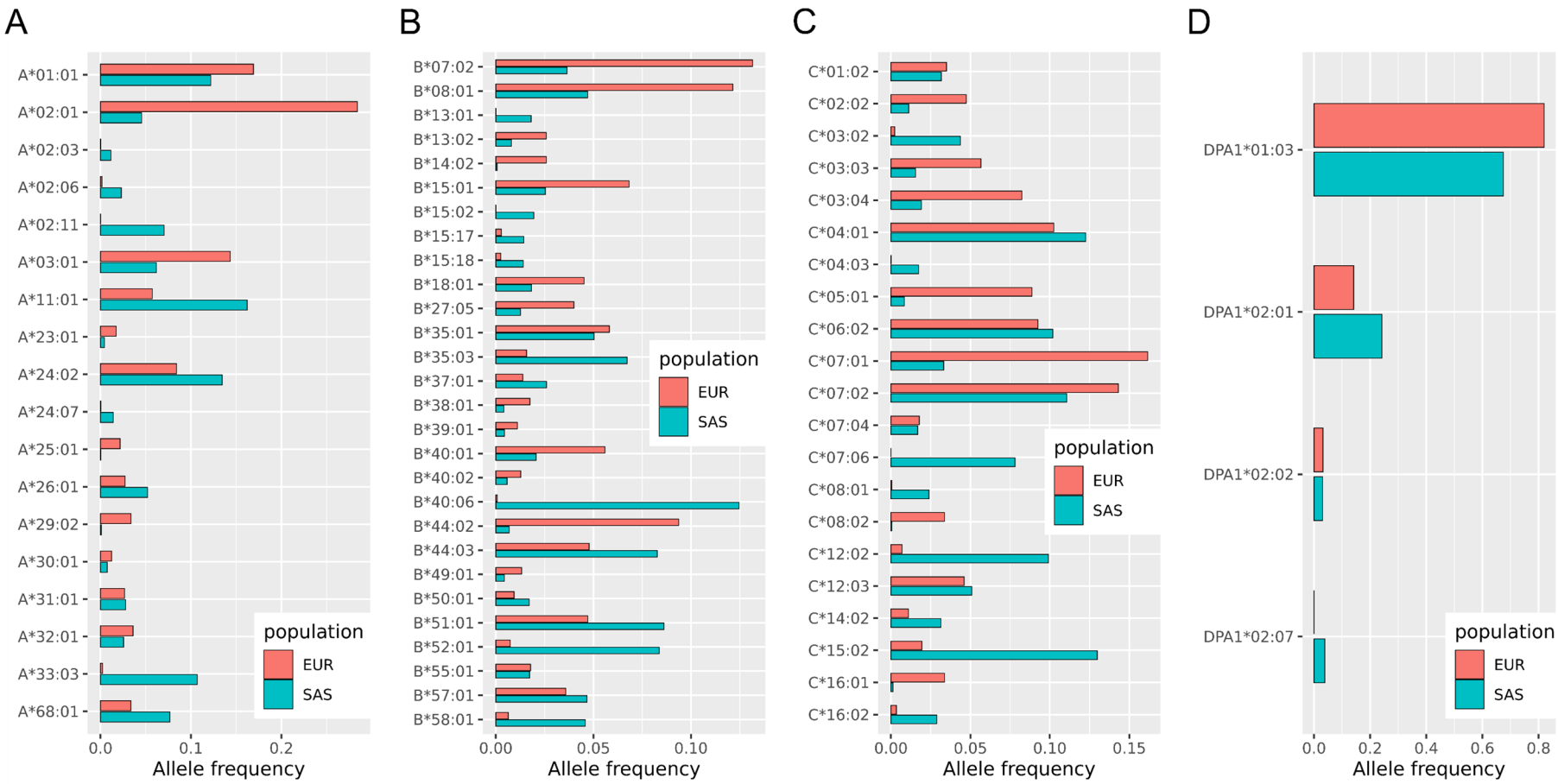

Figure S19. Imputed 2-field *HLA-A*, *HLA-B*, *HLA-C* and *HLA-DPA1* allele frequencies for the European (EUR) and South Asian (SAS) datasets of this study. Results for *HLA-A*, *HLA-B*, *HLA-C* and *HLA-DPA1* are shown in panels A, B, C and D, respectively. Frequencies for the EUR and SAS populations were estimated as a weighted average of frequencies in psoriasis cases and unaffected controls (weights = 0.0015 and 0.985 for EUR and 0.003 and 0.997 for SAS). Alleles with frequencies < 0.01 in both populations are omitted.

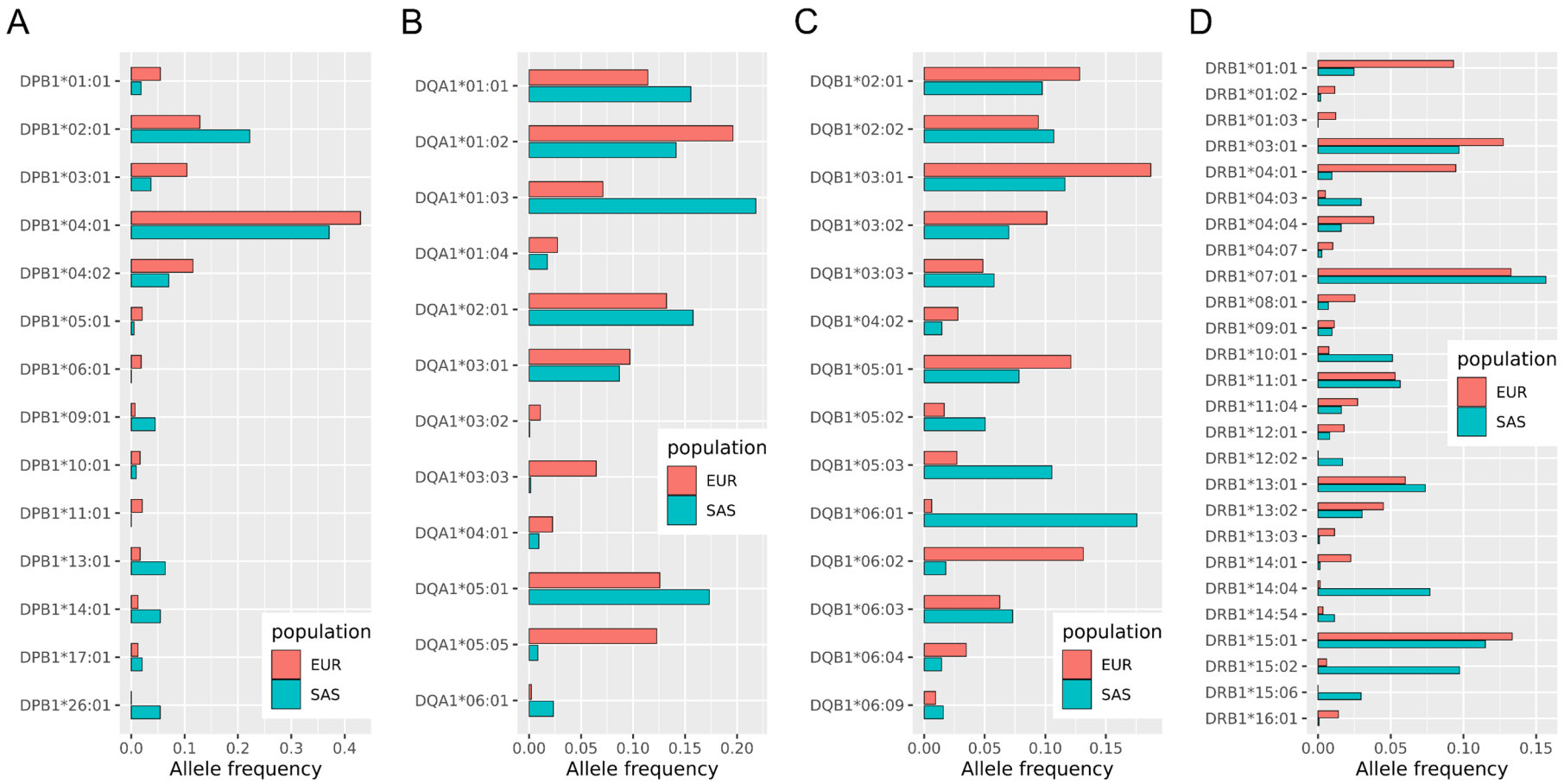

Figure S20. Imputed 2-field *HLA-DPB1*, *HLA-DQA1*, *HLA-DQB1* and *HLA-DRB1* allele frequencies for the European (EUR) and South Asian (SAS) datasets of this study. Results for *HLA-DPB1*, *HLA-DQA1*, *HLA-DQB1* and *HLA-DRB*1 are shown in panels A, B, C and D, respectively. Frequencies for the EUR and SAS populations were estimated as a weighted average of frequencies in psoriasis cases and unaffected controls (weights = 0.015 and 0.985 for EUR and 0.003 and 0.997 for SAS). Alleles with frequencies < 0.01 in both populations are omitted.

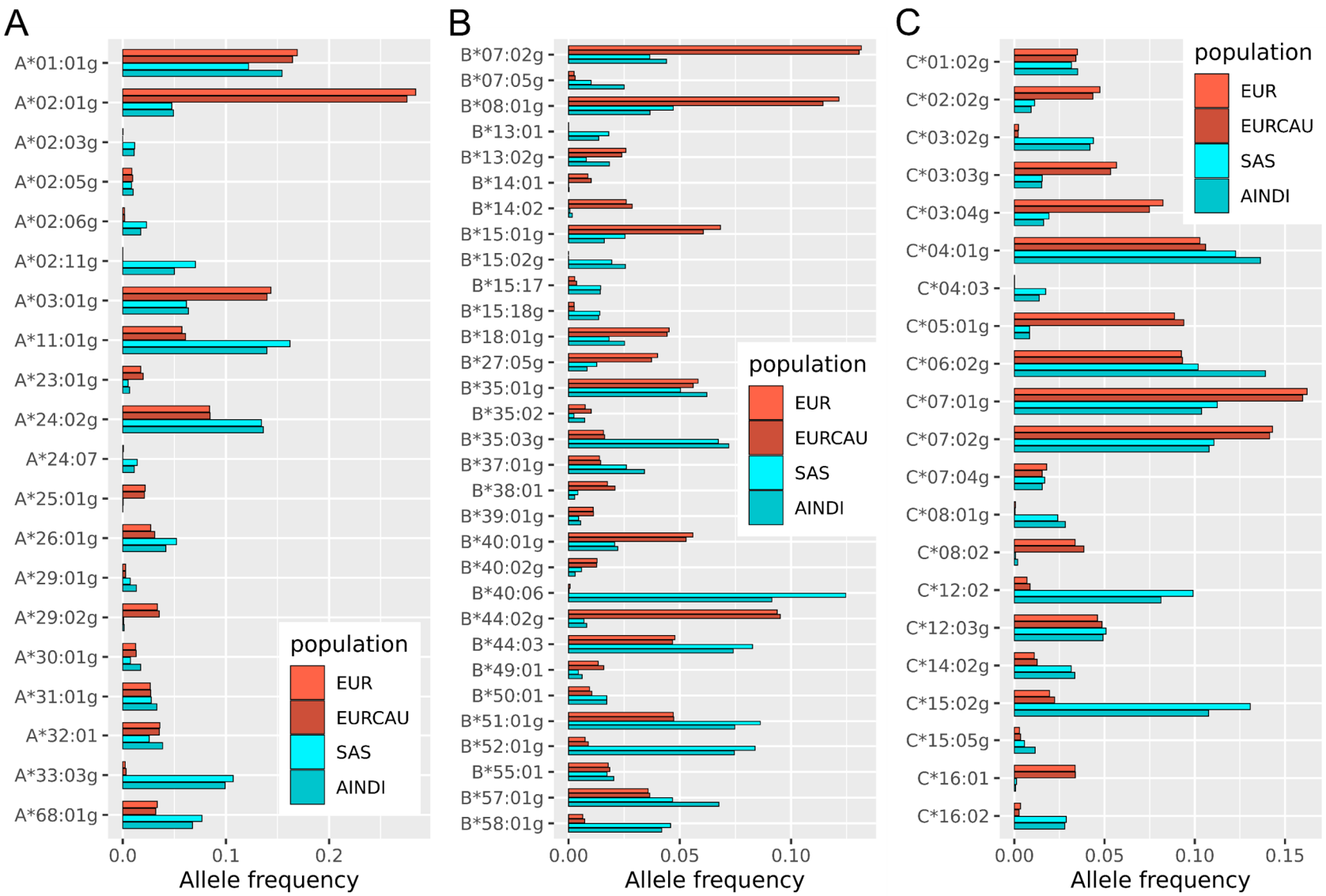

Figure S21. Comparison of imputed HLA class I frequencies for this study with genotyped HLA frequencies for corresponding populations in the National Marrow Donor Program (NMDP) database. HLA allele frequencies estimated for European (EUR) and South Asian (SAS) individuals of this study (bright red and bright cyan, respectively) are plotted next to allele frequencies for European (EURCAU) and South Asian (AINDI) individuals in the NMDP database (dark red and dark cyan, respectively).[3] Results for *HLA-A*, *HLA-B*, and *HLA-C* are shown in panels A, B and C, respectively. 2-field alleles for this study were aggregated, when necessary, to match the g-group designations used by NMDP, which distinguishes alleles that vary in amino acid sequence for the antigen recognition site coded for by exons 2 and 3 of class I and exon 2 of class II HLA genes. Alleles with frequencies < 0.01 in both populations are omitted.

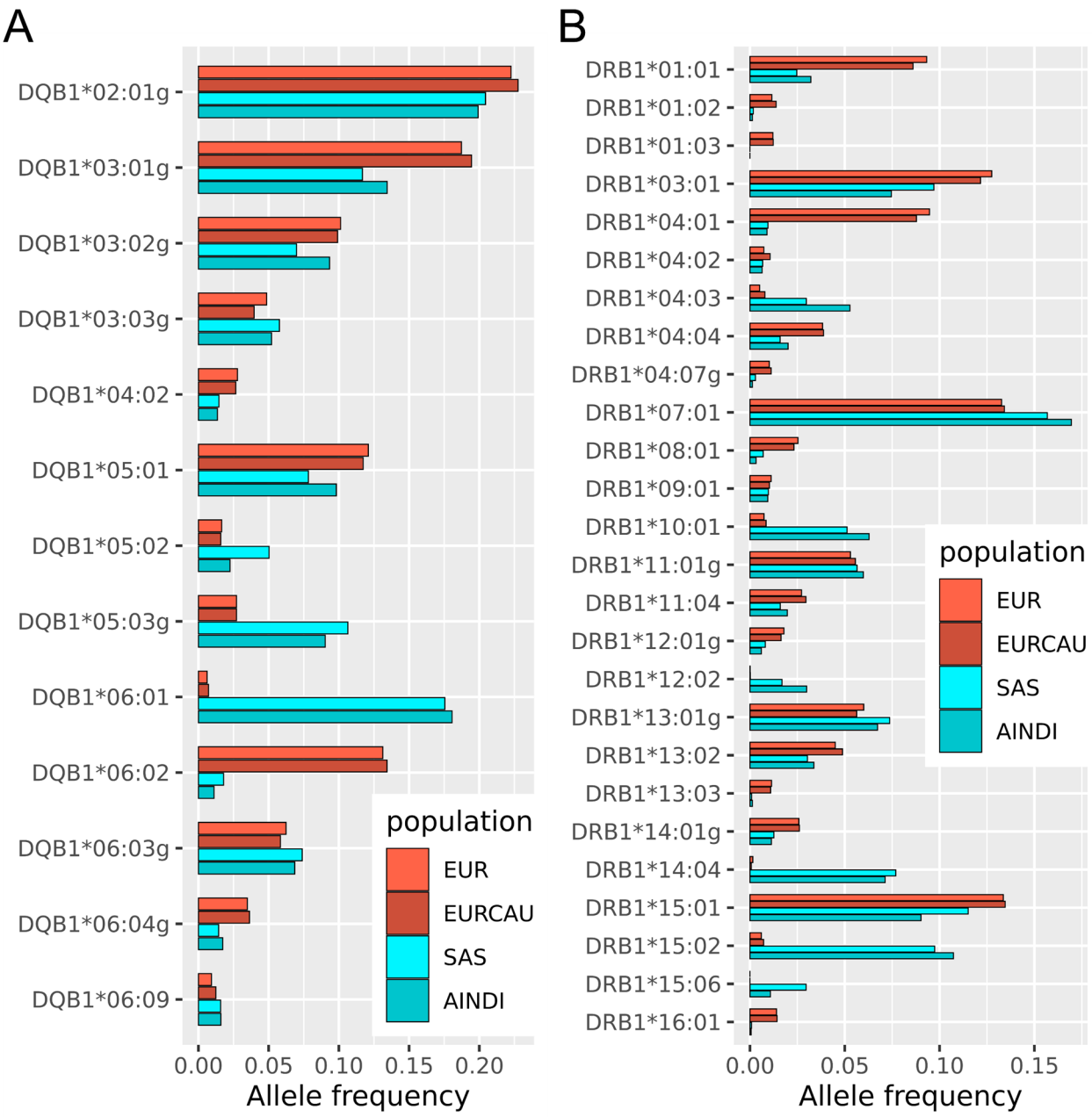

Figure S22. Comparison of imputed HLA class II frequencies for this study with genotyped HLA frequencies for corresponding populations in the National Marrow Donor Program (NMDP) database. HLA allele frequencies estimated for European (EUR) and South Asian (SAS) individuals of this study (bright red and bright cyan, respectively) are plotted next to allele frequencies for European (EURCAU) and South Asian (AINDI) individuals in the NMDP database (dark red and dark cyan, respectively).[3] Results for *HLA-DQB1* and *HLA-DRB1* are shown in panels A and B, respectively. 2-field alleles for this study were aggregated, when necessary, to match the g-group designations used by NMDP, which distinguishes alleles that vary in amino acid sequence for the antigen recognition site coded for by exons 2 and 3 of class I and exon 2 of class II HLA genes. Alleles with frequencies < 0.01 in both populations are omitted.

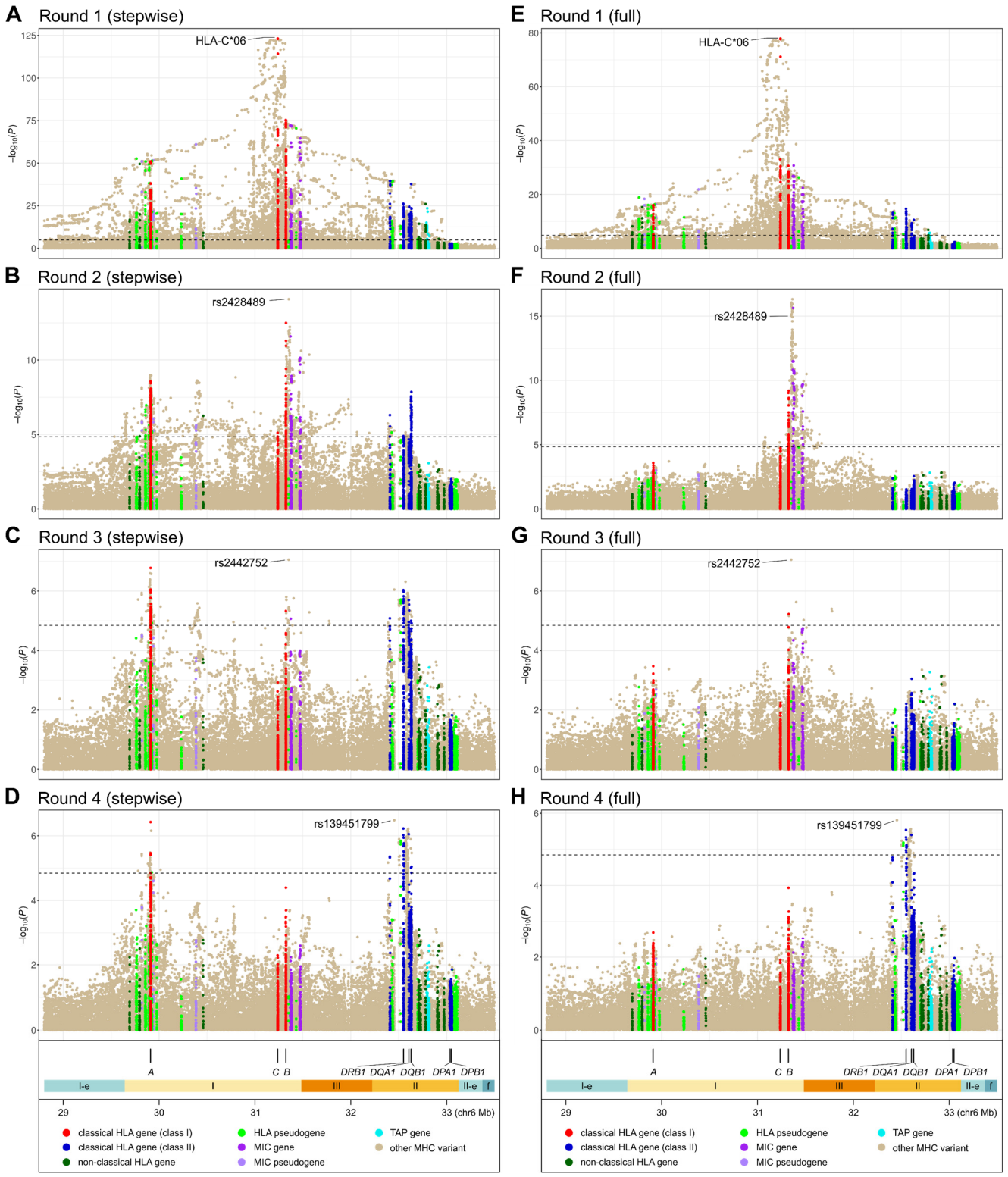

Figure S23. Plots of rounds 1–4 of stepwise analysis of psoriasis association in the extended MHC region in people of South Asian ancestry. The left 4 panels (A–D) show association results after each stepwise round; the right 4 panels (E–H) show association results for the final full regression model containing all variants identified by the stepwise analysis except the one selected at that specific round. Each circle represents the −log(p) of association of an imputed variant, color-coded based on its membership in various categories of MHC genes, as detailed in the key at the bottom. Dashed lines denote thresholds of Bonferroni-corrected significance of 0.05. The locations of the eight HLA genes for which amino acid and protein alleles were imputed are shown at the bottom, along with colored segments denoting the boundaries of the classical MHC region (class I, II and III), the extended MHC class I (I-e) and II (II-e) regions, and flanking MHC regions (f).

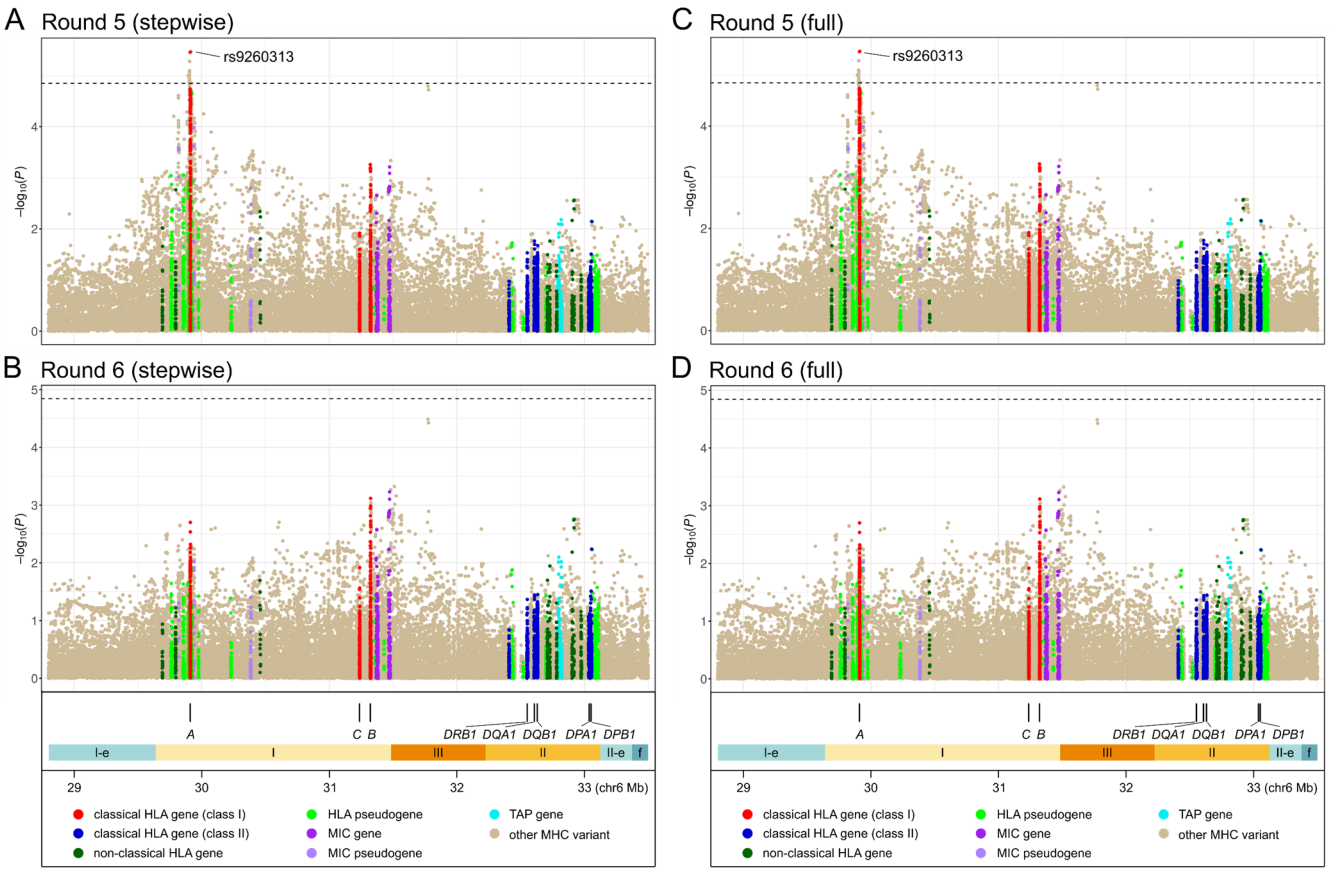

Figure S24. Plots of rounds 5–6 of stepwise analysis of psoriasis association in the extended MHC region in people of South Asian ancestry. The left 2 panels (A–B) show association results after each stepwise round; the right 2 panels (C–D) show association results for the final full regression model containing all variants identified by the stepwise analysis except the one selected at that specific round. Each circle represents the −log(p) of association of an imputed variant, color-coded based on its membership in various categories of MHC genes, as detailed in the key at the bottom. Dashed lines denote thresholds of Bonferroni-corrected significance of 0.05. The locations of the eight HLA genes for which amino acid and protein alleles were imputed are shown at the bottom, along with colored segments denoting the boundaries of the classical MHC region (class I, II and III), the extended MHC class I (I-e) and II (II-e) regions, and flanking MHC regions (f).

			South Asian model				
			HLA-A	HLA-C	—	MICA	HLA-DRB1
		Linkage disequilibrium (W_n^2) in South Asian dataset	rs9260313 (C,T)	HLA-C*06 (C*06,other)	rs2442752 (C,T)	rs2428489 (A,C,T)	rs139451799 (–,other)
South Asian model	HLA-A	rs9260313 (C,T)	–	0.06	0.03	0.04	0.00
	HLA-C	HLA-C*06 (C*06,other)	0.06	–	0.11	0.08	0.04
	—	rs2442752 (C,T)	0.03	0.11	–	0.17	0.00
	MICA	rs2428489 (A,C,T)	0.04	0.08	0.17	–	0.05
	HLA-DRB1	rs139451799 (–,other)	0.00	0.04	0.00	0.05	–

Figure S25. Matrix of pairwise linkage disequilibrium among variants of the South Asian association model for the MHC region. Variants are listed in the order of their position on chromosome 6, with the alleles used to compute LD in parentheses after each variant. LD values are for the multiallelic W_n^2 coefficient computed in the South Asian dataset, which reduces to the usual r^2 coefficient when both variants of a pair are biallelic. The magnitude of LD values is accentuated with red shading on a 0 (white) to 1 (dark red) scale. Variants are labeled with a gene name if the variant changes that gene's protein sequence or if it is in substantial LD ($W_n^2 \geq 0.4$) with a protein-changing variant for the gene.

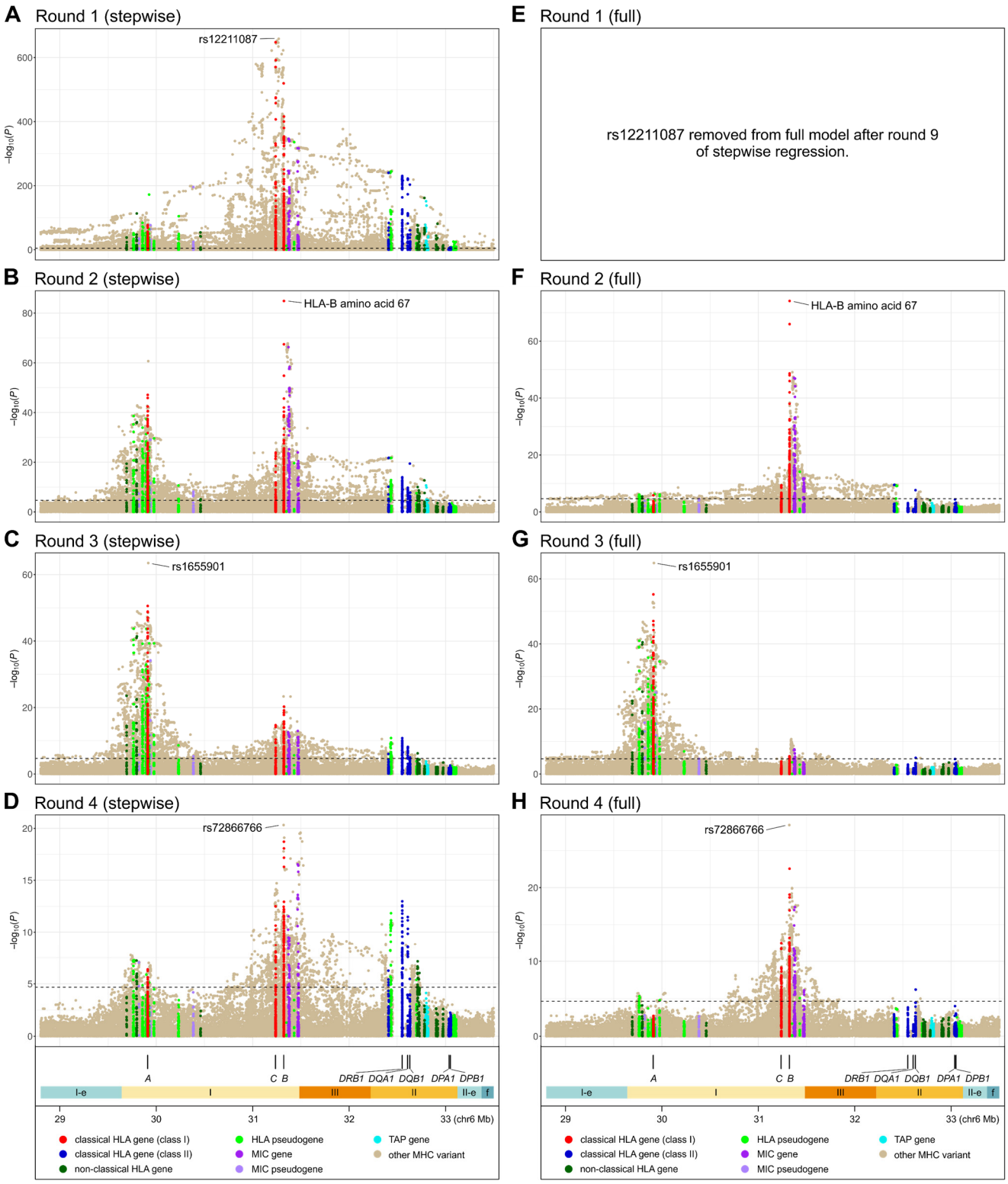

Figure S26. Plots of rounds 1–4 of stepwise analysis of psoriasis association in the extended MHC region in people of European ancestry. The left 4 panels (A–D) show association results after each stepwise round; the right 4 panels (E–H) show association results for the final full regression model containing all variants identified by the stepwise analysis except the one selected at that specific round. Each circle represents the −log(p) of association of an imputed variant, color-coded based on its membership in various categories of MHC genes, as detailed in the key at the bottom. Dashed lines denote thresholds of Bonferroni-corrected significance of 0.05. The locations of the eight HLA genes for which amino acid and protein alleles were imputed are shown at the bottom, along with colored segments denoting the boundaries of the classical MHC region (class I, II and III), the extended MHC class I (I-e) and II (II-e) regions, and flanking MHC regions (f).

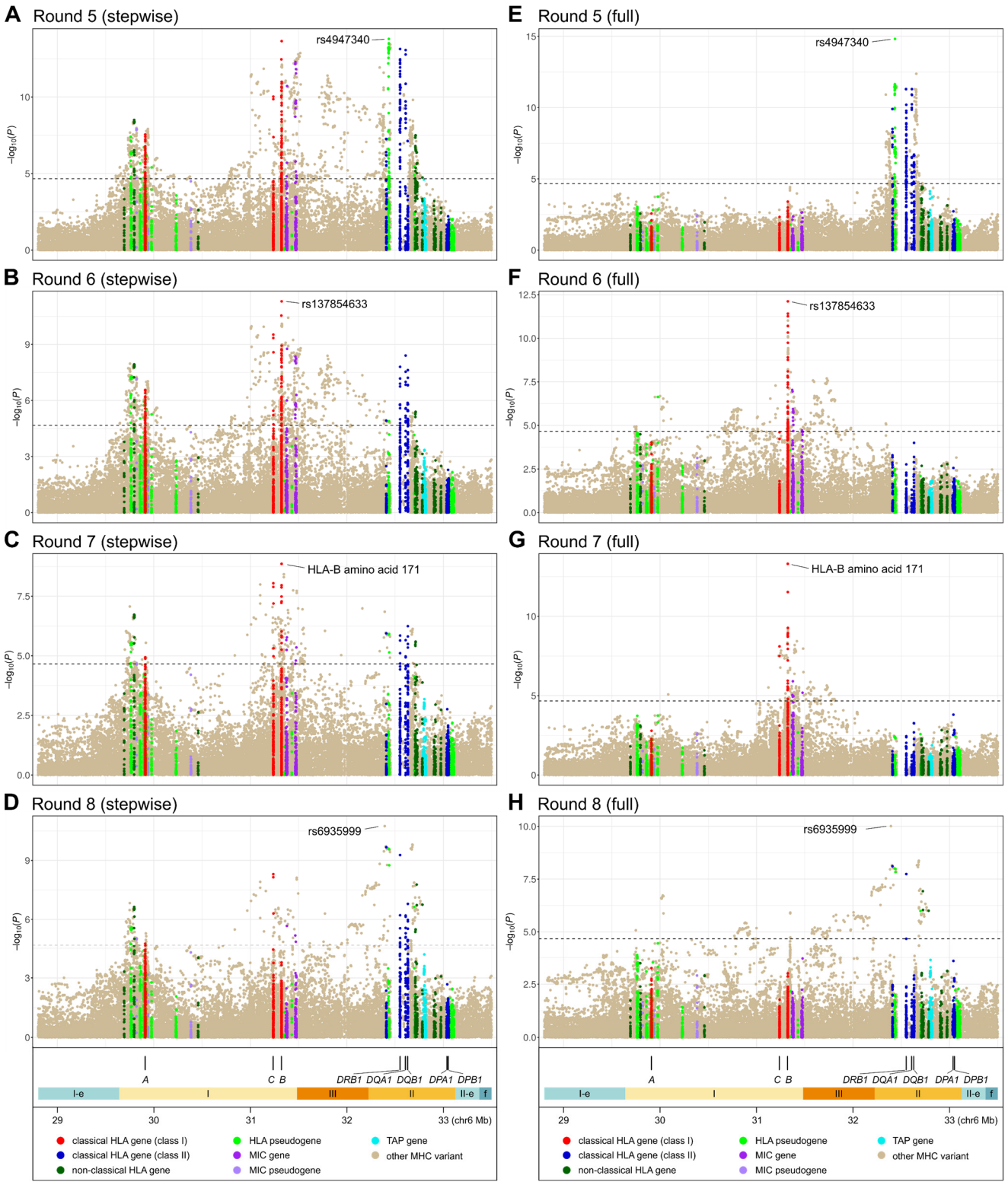

Figure S27. Plots of rounds 5–8 of stepwise analysis of psoriasis association in the extended MHC region in people of European ancestry. The left 4 panels (A–D) show association results after each stepwise round; the right 4 panels (E–H) show association results for the final full regression model containing all variants identified by the stepwise analysis except the one selected at that specific round. Each circle represents the −log(p) of association of an imputed variant, color-coded based on its membership in various categories of MHC genes, as detailed in the key at the bottom. Dashed lines denote thresholds of Bonferroni-corrected significance of 0.05. The locations of the eight HLA genes for which amino acid and protein alleles were imputed are shown at the bottom, along with colored segments denoting the boundaries of the classical MHC region (class I, II and III), the extended MHC class I (I-e) and II (II-e) regions, and flanking MHC regions (f).

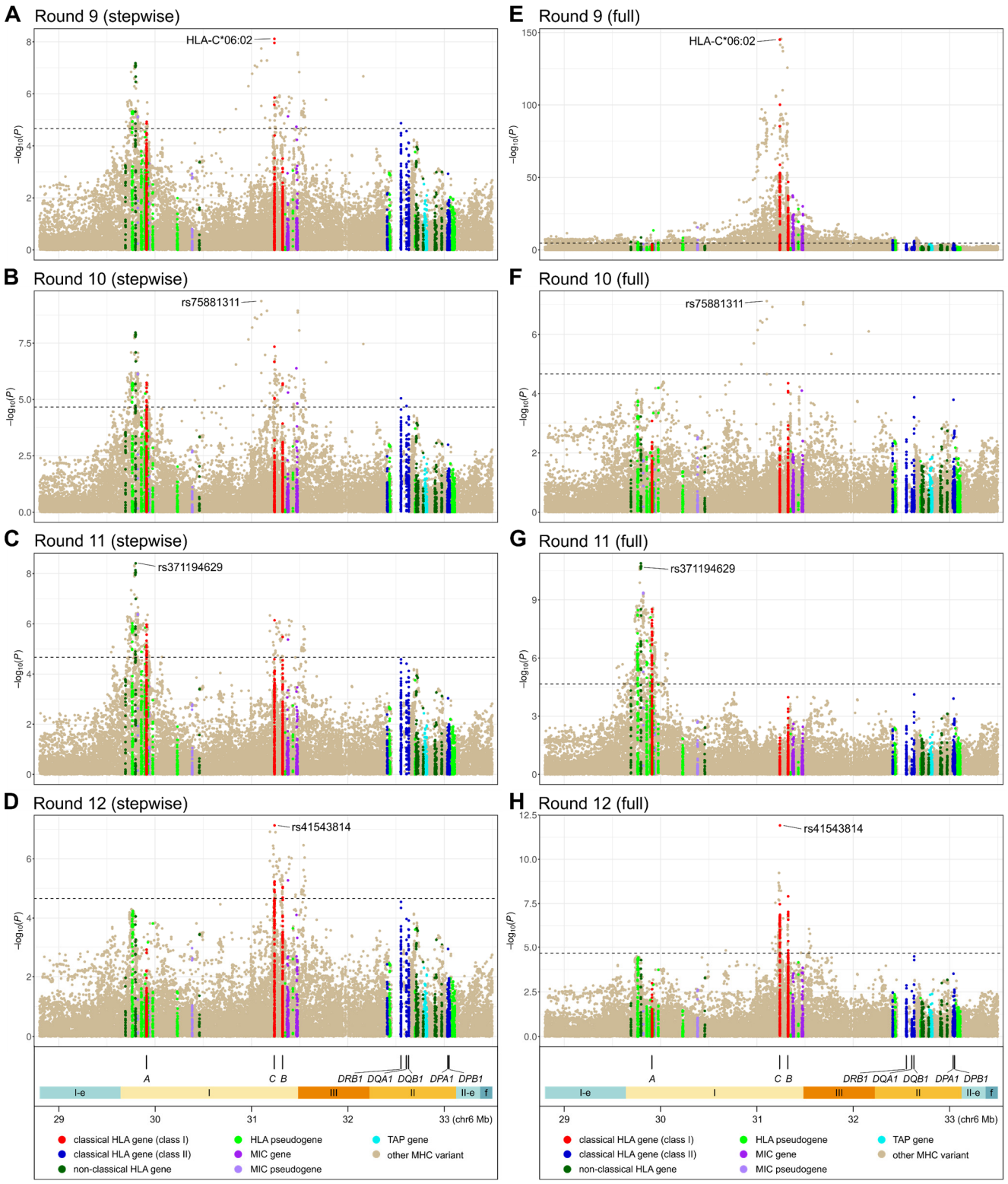

Figure S28. Plots of rounds 9–12 of stepwise analysis of psoriasis association in the extended MHC region in people of European ancestry. The left 4 panels (A–D) show association results after each stepwise round; the right 4 panels (E–H) show association results for the final full regression model containing all variants identified by the stepwise analysis except the one selected at that specific round. Each circle represents the −log(p) of association of an imputed variant, color-coded based on its membership in various categories of MHC genes, as detailed in the key at the bottom. Dashed lines denote thresholds of Bonferroni-corrected significance of 0.05. The locations of the eight HLA genes for which amino acid and protein alleles were imputed are shown at the bottom, along with colored segments denoting the boundaries of the classical MHC region (class I, II and III), the extended MHC class I (I-e) and II (II-e) regions, and flanking MHC regions (f).

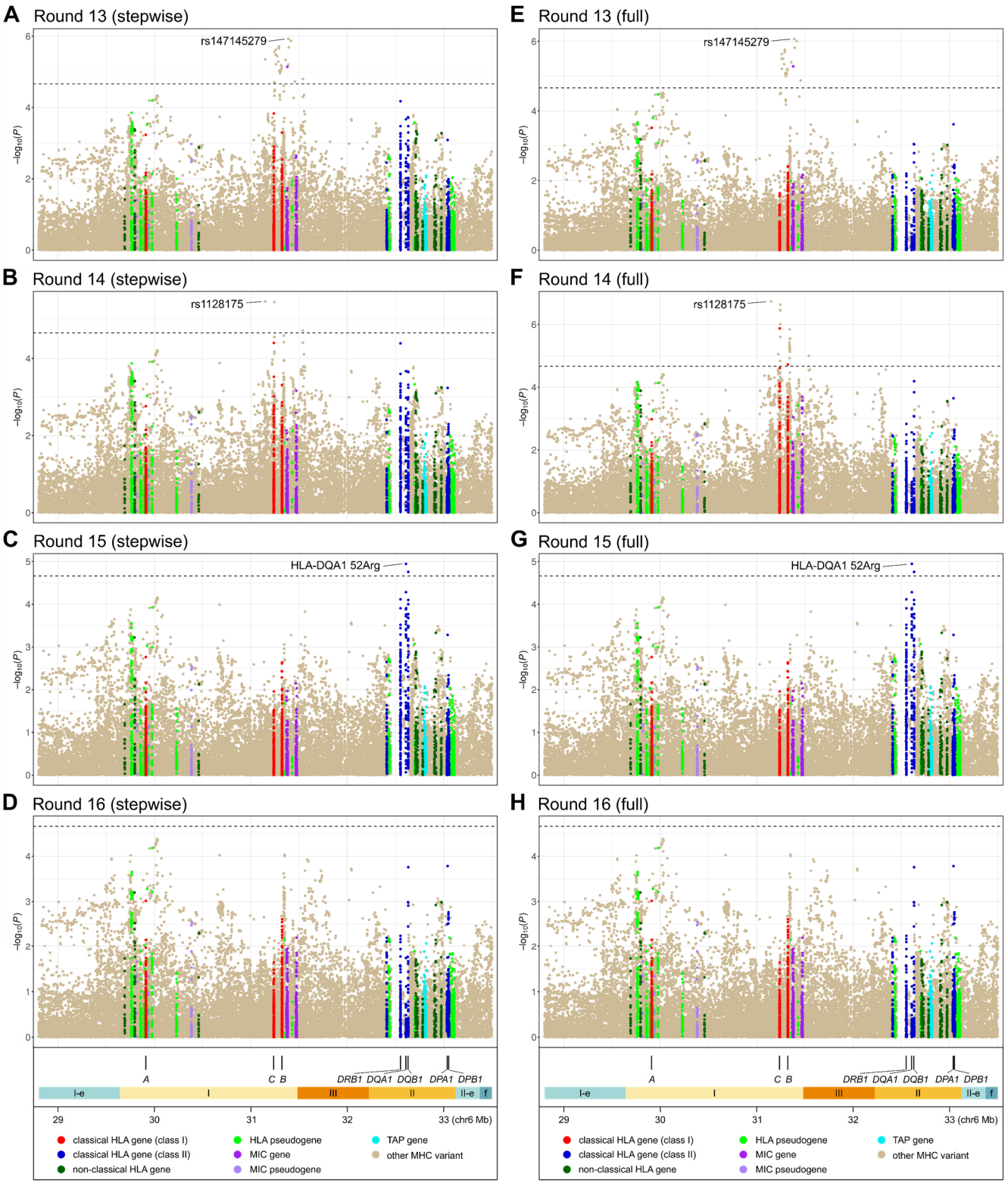

Figure S29. Plots of rounds 13–16 of stepwise analysis of psoriasis association in the extended MHC region in people of European ancestry. The left 4 panels (A–D) show association results after each stepwise round; the right 4 panels (E–H) show association results for the final full regression model containing all variants identified by the stepwise analysis except the one selected at that specific round. Each circle represents the −log(p) of association of an imputed variant, color-coded based on its membership in various categories of MHC genes, as detailed in the key at the bottom. Dashed lines denote thresholds of Bonferroni-corrected significance of 0.05. The locations of the eight HLA genes for which amino acid and protein alleles were imputed are shown at the bottom, along with colored segments denoting the boundaries of the classical MHC region (class I, II and III), the extended MHC class I (I-e) and II (II-e) regions, and flanking MHC regions (f).

			European model													
			HLA-A		—	HLA-C			HLA-B				MICA	HLA-DRB1		HLA-DQA1
		Linkage disequilibrium (W_n^2) in European dataset	rs371194629 (–,14mer)	rs1655901 (C,T)	rs75881311 (T,other)	rs1128175 (A,G)	HLA-C*06:02 (6*02,other)	rs41543814 (C,T)	rs72866766 (C,T)	HLA-B aa-171 (H,Y)	HLA-B aa-67 (C,F,M,S,Y)	rs137854633 (–,C,T)	rs147145279 (AA,other)	rs6935999 (A,G)	rs4947340 (C,T)	HLA-DQA1 aa-52 (R,other)
European model	HLA-A	rs371194629 (–,14mer)	–	0.00	0.00	0.03	0.01	0.03	0.00	0.00	0.07	0.01	0.00	0.00	0.00	0.01
		rs1655901 (C,T)	0.00	–	0.00	0.02	0.01	0.00	0.01	0.01	0.05	0.00	0.00	0.01	0.00	0.01
	—	rs75881311 (T,other)	0.00	0.00	–	0.02	0.00	0.00	0.04	0.00	0.03	0.00	0.00	0.00	0.00	0.00
	HLA-C	rs1128175 (A,G)	0.03	0.02	0.02	–	0.04	0.16	0.01	0.03	0.43	0.02	0.00	0.00	0.03	0.02
		HLA-C*06:02 (6*02,other)	0.01	0.01	0.00	0.04	–	0.11	0.01	0.02	0.37	0.26	0.00	0.00	0.07	0.04
		rs41543814 (C,T)	0.03	0.00	0.00	0.16	0.11	–	0.01	0.04	0.19	0.14	0.00	0.01	0.02	0.01
	HLA-B	rs72866766 (C,T)	0.00	0.01	0.04	0.01	0.01	0.01	–	0.00	0.17	0.02	0.00	0.00	0.00	0.00
		HLA-B aa-171 (H,Y)	0.00	0.01	0.00	0.03	0.02	0.04	0.00	–	0.06	0.05	0.00	0.05	0.00	0.00
		HLA-B aa-67 (C,F,M,S,Y)	0.07	0.05	0.03	0.43	0.37	0.19	0.17	0.06	–	0.18	0.01	0.04	0.12	0.08
		rs137854633 (–,C,T)	0.01	0.00	0.00	0.02	0.26	0.14	0.02	0.05	0.18	–	0.00	0.00	0.04	0.02
	MICA	rs147145279 (AA,other)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	–	0.00	0.00	0.00
	HLA-DRB1	rs6935999 (A,G)	0.00	0.01	0.00	0.00	0.00	0.01	0.00	0.05	0.04	0.00	0.00	–	0.01	0.01
		rs4947340 (C,T)	0.00	0.00	0.00	0.03	0.07	0.02	0.00	0.00	0.12	0.04	0.00	0.01	–	0.02
	HLA-DQA1	HLA-DQA1 aa-52 (R,other)	0.01	0.01	0.00	0.02	0.04	0.01	0.00	0.00	0.08	0.02	0.00	0.01	0.02	–

Figure S30. Pairwise linkage disequilibrium among variants of the European association model for the MHC region. Variants are listed in the order of their position on chromosome 6, with the alleles used to compute LD in parentheses after each variant. LD values are for the multiallelic W_n^2 coefficient computed in the European dataset, which reduces to the usual r^2 coefficient when both variants of a pair are biallelic. The magnitude of LD values is accentuated with red shading on a 0 (white) to 1 (dark red) scale. Variants are labeled with a gene name if the variant changes that gene's protein sequence or if it is in substantial LD ($W_n^2 \geq 0.4$) with a protein-changing variant for the gene. Variant ID rs147415279 is shorthand for a biallelic split of two biallelic indels (rs147415279 and rs559509014) that were genotyped as a single triallelic indel by the 1000 Genomes Project.

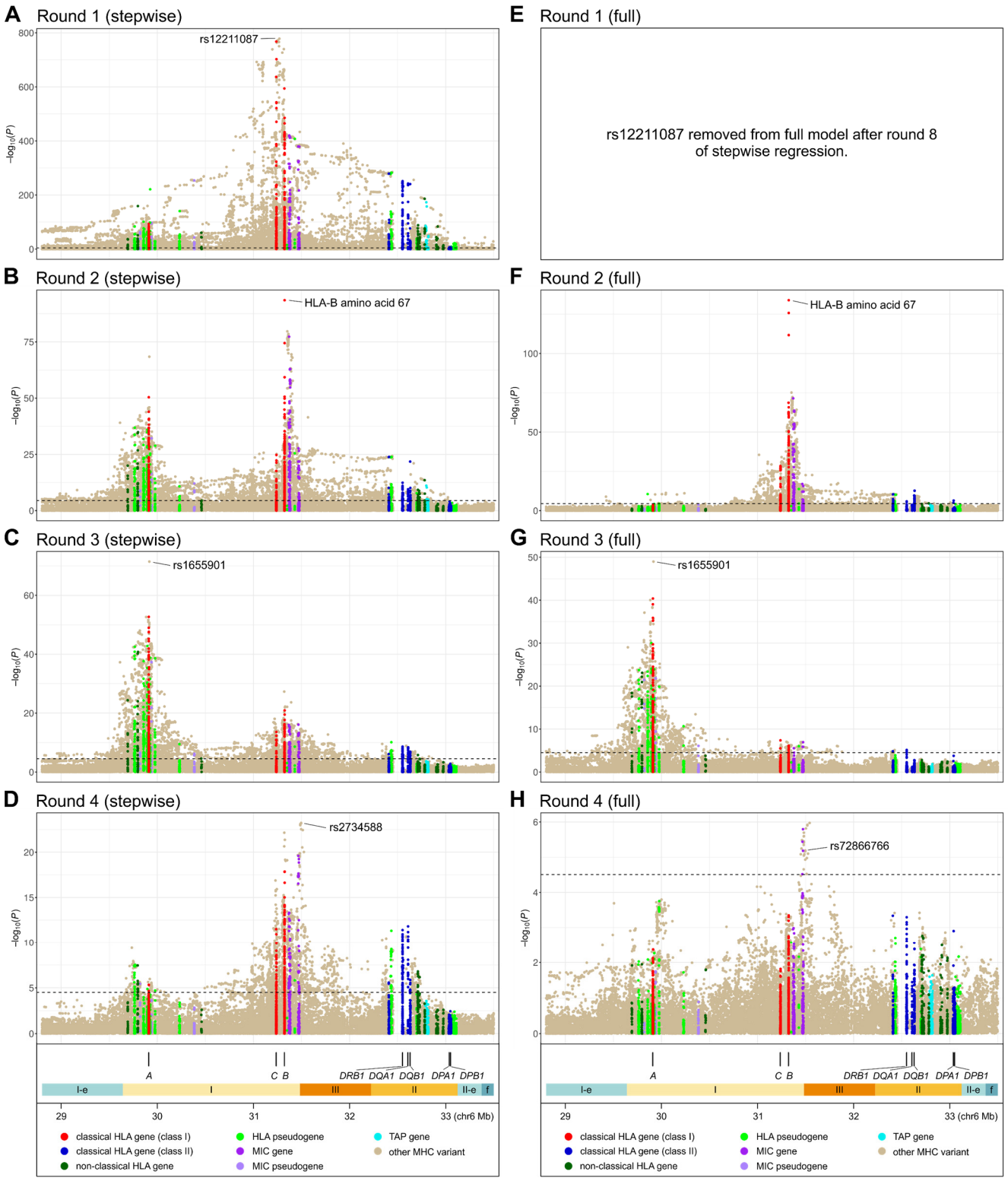

Figure S31. Plots of rounds 1–4 of stepwise analysis of psoriasis association in the extended MHC region in people of South Asian or European ancestry. The left 4 panels (A–D) show association results after each stepwise round; the right 4 panels (E–H) show association results for the final full regression model containing all variants identified by the stepwise analysis except the one selected at that specific round. Each circle represents the −log(p) of association of an imputed variant, color-coded based on its membership in various categories of MHC genes, as detailed in the key at the bottom. Dashed lines denote thresholds of Bonferroni-corrected significance of 0.05. The locations of the eight HLA genes for which amino acid and protein alleles were imputed are shown at the bottom, along with colored segments denoting the boundaries of the classical MHC region (class I, II and III), the extended MHC class I (I-e) and II (II-e) regions, and flanking MHC regions (f).

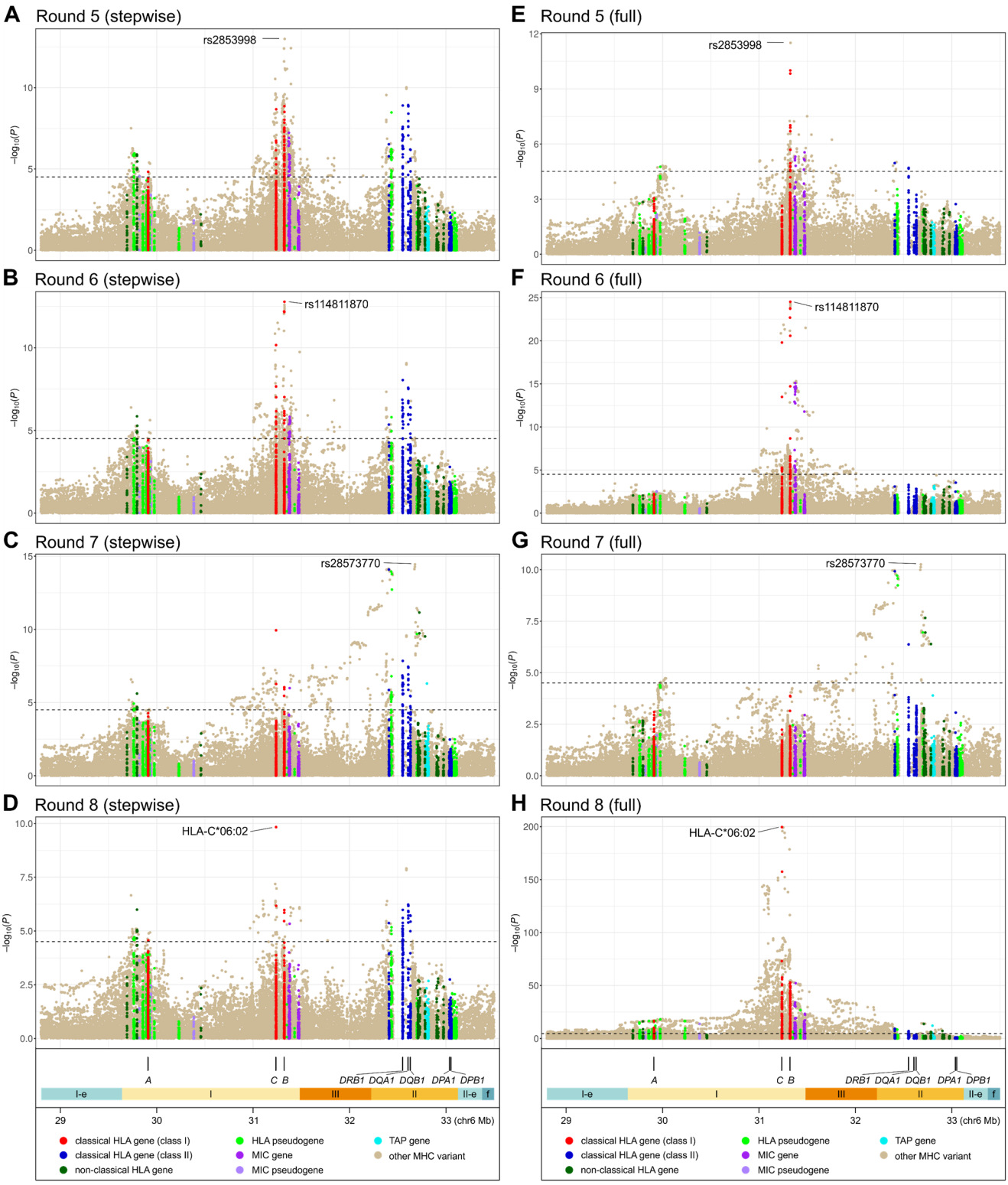

Figure S32. Plots of rounds 5–8 of stepwise analysis of psoriasis association in the extended MHC region in people of South Asian or European ancestry. The left 4 panels (A–D) show association results after each stepwise round; the right 4 panels (E–H) show association results for the final full regression model containing all variants identified by the stepwise analysis except the one selected at that specific round. Each circle represents the −log(p) of association of an imputed variant, color-coded based on its membership in various categories of MHC genes, as detailed in the key at the bottom. Dashed lines denote thresholds of Bonferroni-corrected significance of 0.05. The locations of the eight HLA genes for which amino acid and protein alleles were imputed are shown at the bottom, along with colored segments denoting the boundaries of the classical MHC region (class I, II and III), the extended MHC class I (I-e) and II (II-e) regions, and flanking MHC regions (f).

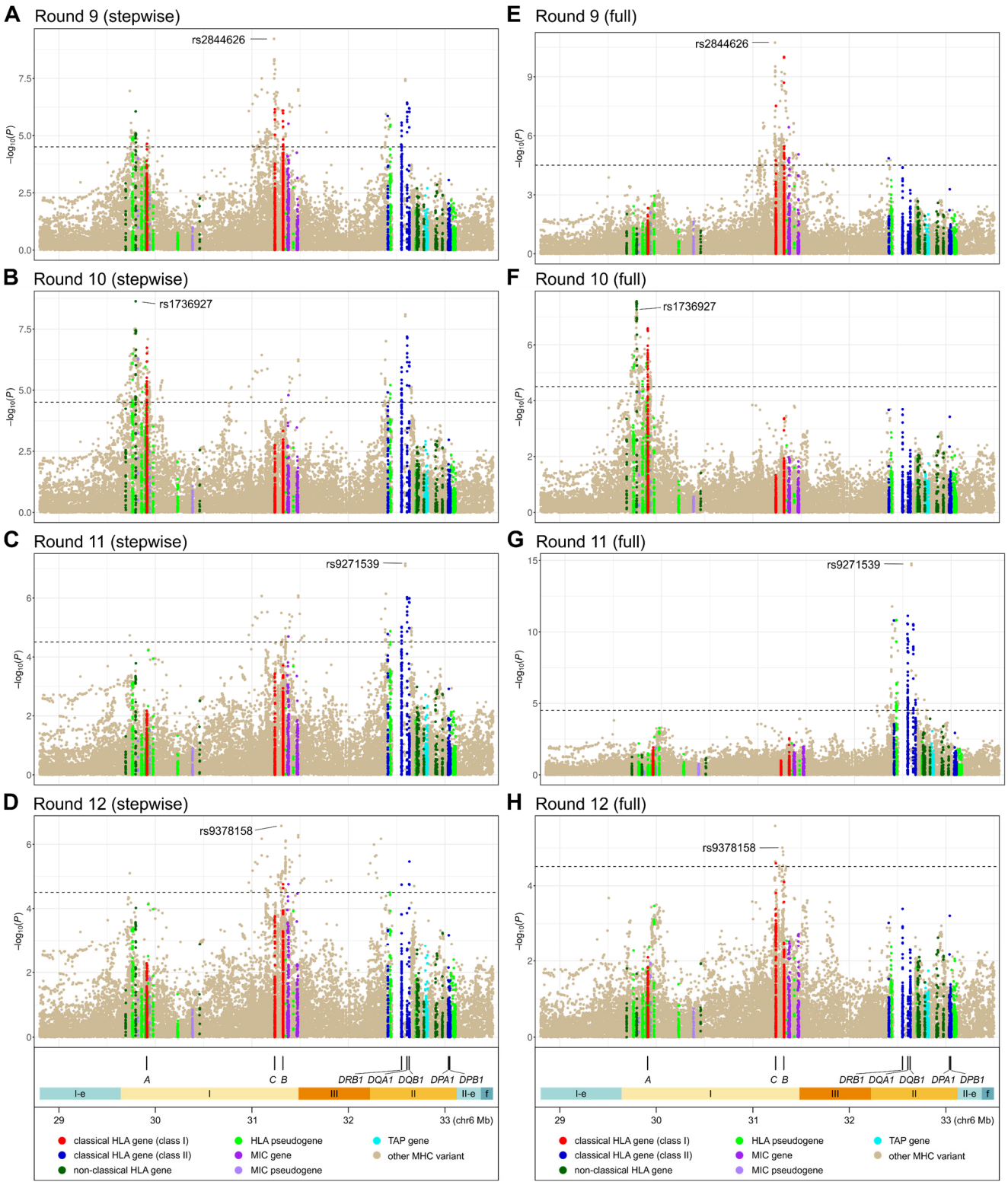

Figure S33. Plots of rounds 9–12 of stepwise analysis of psoriasis association in the extended MHC region in people of South Asian or European ancestry. The left 4 panels (A–D) show association results after each stepwise round; the right 4 panels (E–H) show association results for the final full regression model containing all variants identified by the stepwise analysis except the one selected at that specific round. Each circle represents the −log(p) of association of an imputed variant, color-coded based on its membership in various categories of MHC genes, as detailed in the key at the bottom. Dashed lines denote thresholds of Bonferroni-corrected significance of 0.05. The locations of the eight HLA genes for which amino acid and protein alleles were imputed are shown at the bottom, along with colored segments denoting the boundaries of the classical MHC region (class I, II and III), the extended MHC class I (I-e) and II (II-e) regions, and flanking MHC regions (f).

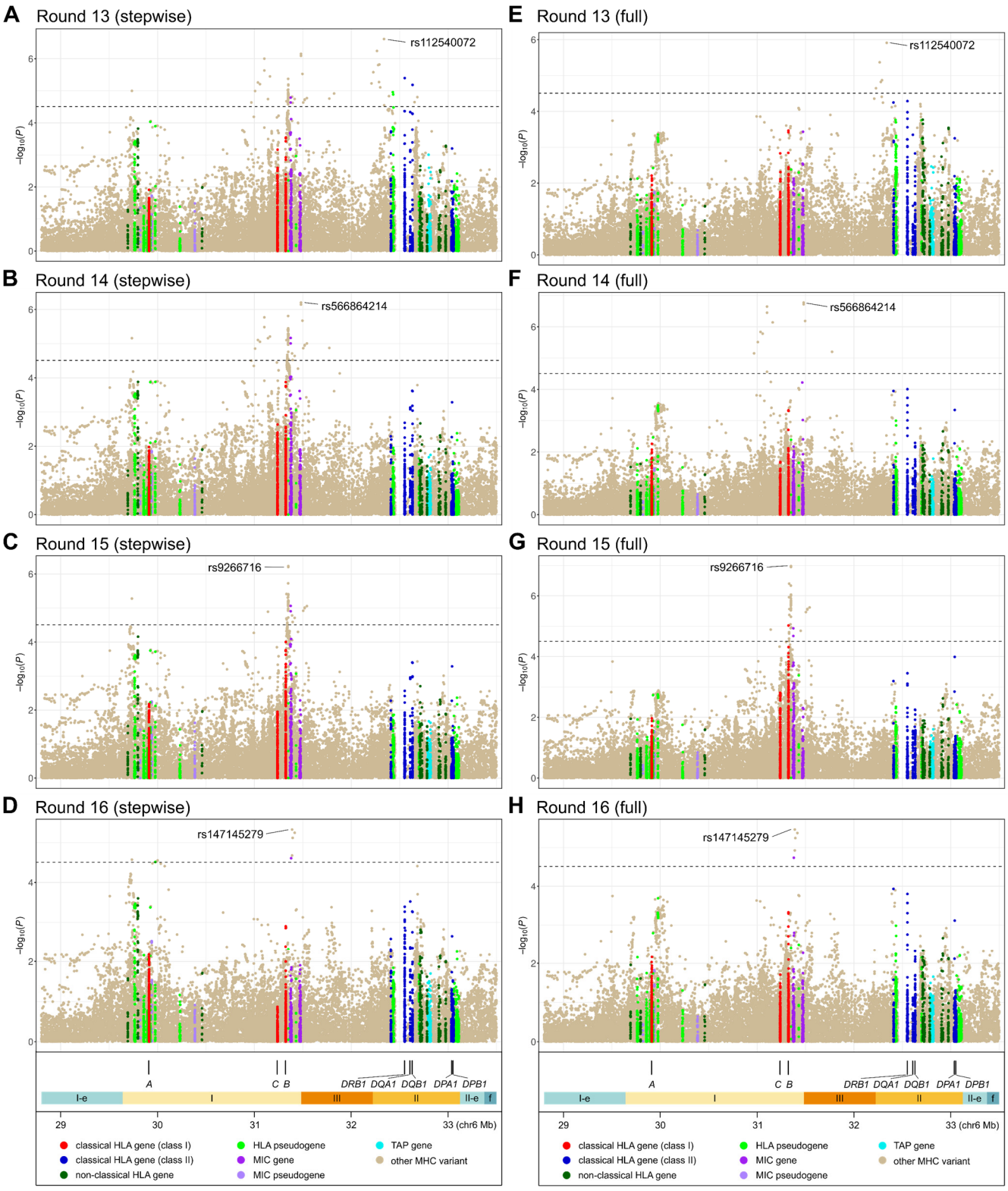

Figure S34. Plots of rounds 13–16 of stepwise analysis of psoriasis association in the extended MHC region in people of South Asian or European ancestry. The left 4 panels (A–D) show association results after each stepwise round; the right 4 panels (E–H) show association results for the final full regression model containing all variants identified by the stepwise analysis except the one selected at that specific round. Each circle represents the −log(p) of association of an imputed variant, color-coded based on its membership in various categories of MHC genes, as detailed in the key at the bottom. Dashed lines denote thresholds of Bonferroni-corrected significance of 0.05. The locations of the eight HLA genes for which amino acid and protein alleles were imputed are shown at the bottom, along with colored segments denoting the boundaries of the classical MHC region (class I, II and III), the extended MHC class I (I-e) and II (II-e) regions, and flanking MHC regions (f).

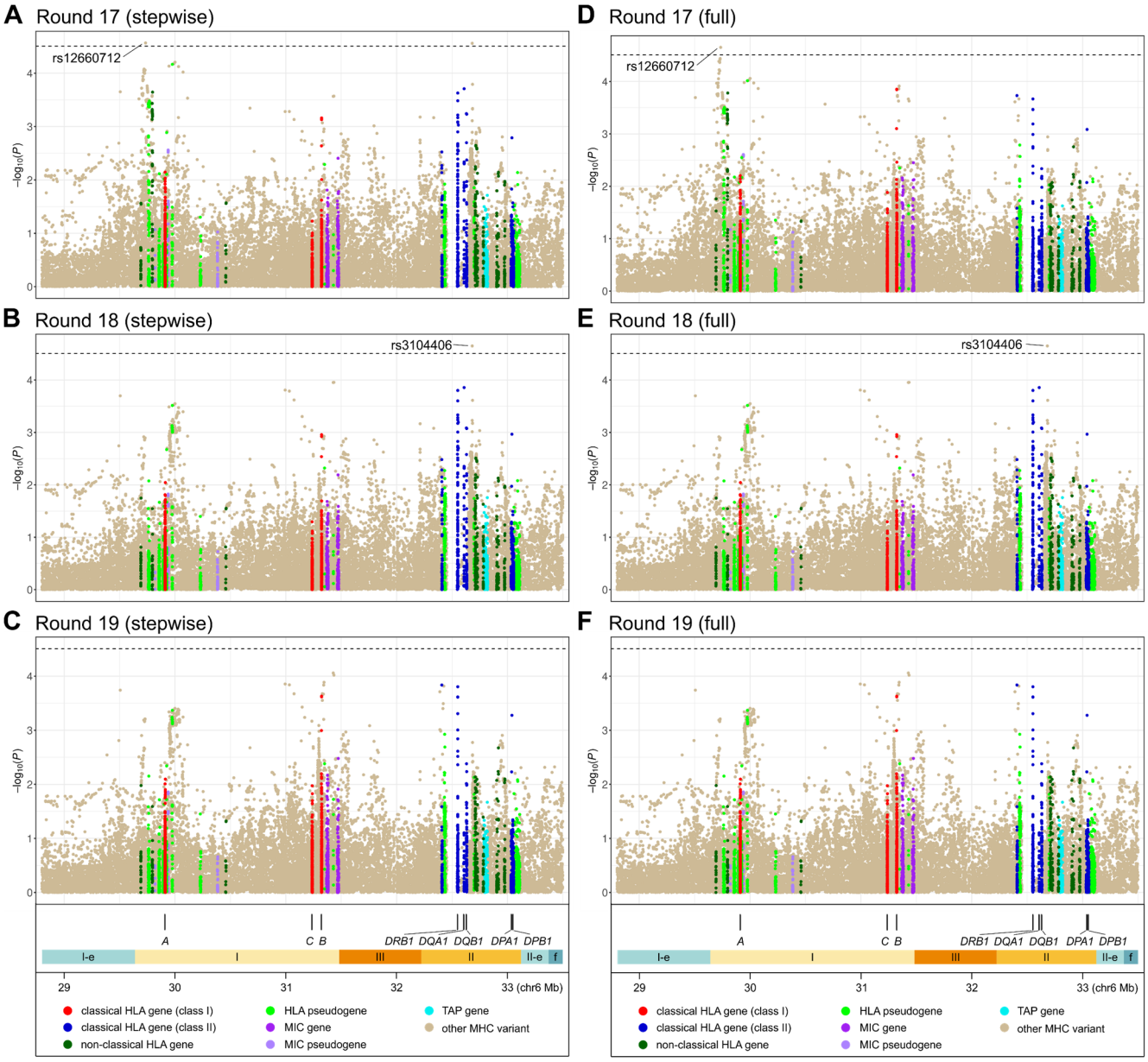

Figure S35. Plots of rounds 17–19 of stepwise analysis of psoriasis association in the extended MHC region in people of South Asian or European ancestry. The left 4 panels (A–D) show association results after each stepwise round; the right 4 panels (E–H) show association results for the final full regression model containing all variants identified by the stepwise analysis except the one selected at that specific round. Each circle represents the −log(p) of association of an imputed variant, color-coded based on its membership in various categories of MHC genes, as detailed in the key at the bottom. Dashed lines denote thresholds of Bonferroni-corrected significance of 0.05. The locations of the eight HLA genes for which amino acid and protein alleles were imputed are shown at the bottom, along with colored segments denoting the boundaries of the classical MHC region (class I, II and III), the extended MHC class I (I-e) and II (II-e) regions, and flanking MHC regions (f).

			South Asian + European model																	
			—	HLA-A		CCHCR1	HLA-C	—	HLA-B	—	HLA-B		MICA	—		HLA-DRB1		TAP2	—
		Linkage disequilibrium (W_n^2) in South Asian dataset (upper triangular); in European dataset lower triangular)	rs12660712 (A,C)	rs1736927 (A,C)	rs1655901 (C,T)	rs2844626 (A,T)	HLA-C*06:02 (6*02,other)	rs9378158 (A,G)	HLA-B aa-67 (C,F,M,S,Y)	rs114811870 (C,T)	rs2853998 (C,T)	rs9266716 (C,T)	rs147415279 (AA,other)	rs566864214 (A,G)	rs2734588 (C,other)	rs112540072 (A,G)	rs9271539 (A,G)	rs28573770 (A,T)	rs3104406 (A,G)
South Asian + European model	—	rs12660712 (A,C)	–	0.02	0.08	0.01	0.00	0.00	0.02	0.00	0.00	0.11	0.00	0.00	0.00	0.00	0.00	0.00	0.01
	HLA-A	rs1736927 (A,C)	0.03	–	0.02	0.01	0.02	0.00	0.01	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		rs1655901 (C,T)	0.03	0.02	–	0.03	0.05	0.00	0.04	0.00	0.00	0.05	0.00	0.00	0.02	0.00	0.01	0.00	0.00
	CCHCR1	rs2844626 (A,T)	0.00	0.07	0.01	–	0.44	0.07	0.14	0.00	0.03	0.01	0.00	0.00	0.02	0.00	0.01	0.00	0.01
	HLA-C	HLA-C*06:02 (6*02,other)	0.00	0.00	0.01	0.24	–	0.03	0.35	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.03	0.00	0.01
	—	rs9378158 (A,G)	0.00	0.00	0.00	0.04	0.01	–	0.06	0.00	0.01	0.01	0.00	0.00	0.02	0.00	0.00	0.00	0.01
	HLA-B	HLA-B aa-67 (C,F,M,S,Y)	0.00	0.05	0.05	0.36	0.37	0.03	–	0.01	0.28	0.19	0.00	0.01	0.07	0.02	0.07	0.01	0.02
	—	rs114811870 (C,T)	0.00	0.00	0.01	0.02	0.01	0.00	0.23	–	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.35	0.00
	HLA-B	rs2853998 (C,T)	0.01	0.01	0.01	0.13	0.01	0.01	0.33	0.02	–	0.04	0.00	0.00	0.01	0.00	0.01	0.00	0.01
		rs9266716 (C,T)	0.00	0.02	0.03	0.04	0.01	0.00	0.35	0.00	0.22	–	0.00	0.00	0.00	0.00	0.01	0.00	0.01
	MICA	rs147415279 (AA,other)	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.00	0.00	0.00	–	0.00	0.00	0.00	0.00	0.01	0.00
	—	rs566864214 (A,G)	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.00	0.00	–	0.00	0.00	0.00	0.00	0.00
		rs2734588 (C,other)	0.00	0.00	0.02	0.02	0.02	0.00	0.06	0.01	0.02	0.05	0.00	0.00	–	0.00	0.00	0.00	0.00
	HLA-DRB1	rs112540072 (A,G)	0.00	0.00	0.00	0.00	0.01	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.01	–	0.01	0.00	0.03
		rs9271539 (A,G)	0.00	0.01	0.01	0.01	0.03	0.01	0.09	0.00	0.01	0.06	0.00	0.00	0.01	0.03	–	0.00	0.00
	TAP2	rs28573770 (A,T)	0.00	0.00	0.01	0.00	0.00	0.00	0.05	0.21	0.00	0.00	0.00	0.00	0.00	0.00	0.01	–	0.00
	—	rs3104406 (A,G)	0.00	0.00	0.00	0.00	0.07	0.00	0.12	0.01	0.02	0.00	0.00	0.00	0.05	0.03	0.00	0.01	–

Figure S36. Pairwise linkage disequilibrium among variants of the transethnic association model for the MHC region. Variants are listed in the order of their position on chromosome 6, with the alleles used to compute LD in parentheses after each variant. LD values on the upper and lower triangulars are for the multiallelic W_n^2 coefficient computed in the South Asian and European datasets, respectively, which reduces to the usual r^2 coefficient when both variants of a pair are biallelic. The magnitude of LD values is accentuated with red shading on a 0 (white) to 1 (dark red) scale. Variants are labeled with a gene name if the variant changes that gene's protein sequence or if it is in substantial LD ($W_n^2 \geq 0.4$) with a protein-changing variant for the gene. Variant ID rs147415279 is shorthand for a biallelic split of two biallelic indels (rs147415279 and rs559509014) that were genotyped as a single triallelic indel by the 1000 Genomes Project.

			European model														
			HLA-A		—	HLA-C			HLA-B				MICA	HLA-DRB1		HLA-DQA1	
			rs371194629 (–,14mer)	rs1655901 (C,T)	rs75881311 (T,other)	rs1128175 (A,G)	HLA-C*06:02 (6*02,other)	rs41543814 (C,T)	rs72866766 (C,T)	HLA-B aa-171 (H,Y)	HLA-B aa-67 (C,F,M,S,Y)	rs137854633 (–,C,T)	rs147415279 (AA,other)	rs6935999 (A,G)	rs4947340 (C,T)	HLA-DQA1 aa52 (R,other)	
South Asian model	HLA-A	rs9260313 (C,T)	0.05	0.70	0.00	0.00	0.06	0.00	0.00	0.00	0.02	0.01	0.00	0.00	0.00	0.01	LD (W_n^2) in South Asians
	HLA-C	HLA-C*06 (C*06,other)	0.08	0.05	0.00	0.07	1.00	0.13	0.02	0.06	0.35	0.21	0.00	0.00	0.06	0.02	
	—	rs2442752 (C,T)	0.03	0.02	0.00	0.03	0.10	0.00	0.04	0.14	0.17	0.11	0.00	0.00	0.02	0.02	
	MICA	rs2428489 (A,C,T)	0.01	0.03	0.00	0.02	0.08	0.06	0.03	0.16	0.29	0.04	0.00	0.00	0.06	0.00	
	HLA-DRB1	rs139451799 (–,other)	0.01	0.00	0.00	0.01	0.04	0.01	0.00	0.01	0.05	0.01	0.00	0.00	0.00	0.10	
South Asian model	HLA-A	rs9260313 (C,T)	0.03	0.49	0.00	0.01	0.01	0.01	0.02	0.01	0.04	0.01	0.00	0.01	0.00	0.00	LD (W_n^2) in Europeans
	HLA-C	HLA-C*06 (C*06,other)	0.01	0.01	0.00	0.05	1.00	0.11	0.01	0.02	0.37	0.26	0.00	0.00	0.07	0.04	
	—	rs2442752 (C,T)	0.01	0.01	0.00	0.00	0.04	0.01	0.04	0.08	0.36	0.06	0.00	0.00	0.01	0.03	
	MICA	rs2428489 (A,C,T)	0.01	0.03	0.01	0.02	0.03	0.00	0.11	0.02	0.32	0.02	0.00	0.02	0.04	0.03	
	HLA-DRB1	rs139451799 (–,other)	0.01	0.00	0.00	0.08	0.01	0.01	0.00	0.00	0.15	0.01	0.00	0.00	0.09	0.11	

Figure S37. Pairwise linkage disequilibrium between variants of the South Asian and European association models for the MHC region. Variants are listed in the order of their position on chromosome 6, with the alleles used to compute LD in parentheses after each variant. LD values for the top five rows and bottom five rows are the multiallelic W_n^2 coefficient computed in the South Asian and European datasets, respectively, which reduces to the usual r^2 coefficient when both variants of a pair are biallelic. The magnitude of LD values is accentuated with red shading on a 0 (white) to 1 (dark red) scale. Variants are labeled with a gene name if the variant changes that gene's protein sequence or if it is in substantial LD ($W_n^2 \geq 0.4$) with a protein-changing variant for the gene. Variant ID rs147415279 is shorthand for a biallelic split of two biallelic indels (rs147415279 and rs559509014) that were genotyped as a single triallelic indel by the 1000 Genomes Project.

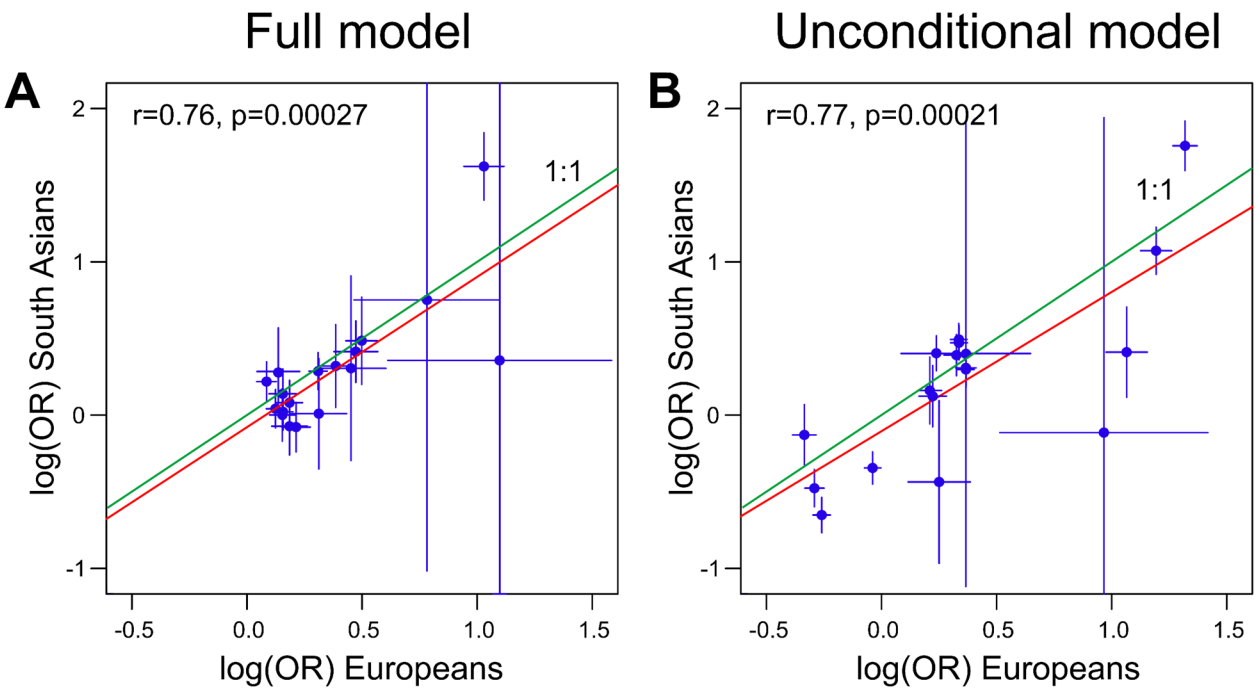

Figure S38. Plots comparing association effect sizes in South Asians vs. Europeans for all variants in the European regression model for the MHC region. The log(OR) of each of the 14 variants in the European model as estimated in the South Asian dataset is plotted against their estimates in the European dataset. The vertical and horizontal bars show the 95% confidence intervals for these estimates in each dataset. Multiallelic variants with *m* alleles are represented by the set of *m-1* decomposed biallelic variants used for the joint Wald test. Panel A shows OR estimates estimated in the full model containing all variants, and panel B shows unconditional OR estimates for each variant with no other variants in the regression model. Green and red lines depict a 1:1 correspondence and a linear fit, respectively. The Pearson correlation coefficient and its significance are shown in the upper left corner of each plot.

			South Asian + European model																	
			—	HLA-A		CCHCR1	HLA-C	—	HLA-B	—	HLA-B		MICA	—	—	HLA-DRB1		TAP2	—
		Linkage disequilibrium (W_n^2) in South Asian dataset	rs12660712 (A,C)	rs1736927 (A,C)	rs1655901 (C,T)	rs2844626 (A,T)	HLA-C*06:02 (6*02,other)	rs9378158 (A,G)	HLA-B aa-67 (C,F,M,S,Y)	rs114811870 (C,T)	rs2853998 (C,T)	rs9266716 (C,T)	rs147415279 (AA,other)	rs566864214 (A,G)	rs2734588 (C,other)	rs112540072 (A,G)	rs9271539 (A,G)	rs28573770 (A,T)	rs3104406 (A,G)
South Asian model	HLA-A	rs9260313 (C,T)	0.11	0.12	0.70	0.05	0.06	0.00	0.02	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.01	0.00	0.00
	HLA-C	HLA-C*06 (C*06,other)	0.00	0.02	0.05	0.44	1.00	0.03	0.35	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.03	0.00	0.01
	—	rs2442752 (C,T)	0.03	0.00	0.02	0.00	0.10	0.05	0.17	0.00	0.01	0.02	0.00	0.00	0.00	0.00	0.04	0.00	0.00
	MICA	rs2428489 (A,C,T)	0.22	0.01	0.03	0.23	0.08	0.01	0.29	0.00	0.15	0.51	0.00	0.00	0.02	0.01	0.01	0.00	0.03
	HLA-DRB1	rs139451799 (–,other)	0.00	0.01	0.00	0.01	0.04	0.00	0.05	0.00	0.02	0.01	0.00	0.00	0.01	0.01	0.09	0.00	0.11

Figure S39. Pairwise linkage disequilibrium between variants of the South Asian and transethnic association models for the MHC region. Variants are listed in the order of their position on chromosome 6, with the alleles used to compute LD in parentheses after each variant. LD values are for the multiallelic W_n^2 coefficient computed in the South Asian dataset, which reduces to the usual r^2 coefficient when both variants of a pair are biallelic. The magnitude of LD values is accentuated with red shading on a 0 (white) to 1 (dark red) scale. Variants are labeled with a gene name if the variant changes that gene's protein sequence or if it is in substantial LD ($W_n^2 \geq 0.4$) with a protein-changing variant for the gene. Variant ID rs147415279 is shorthand for a biallelic split of two biallelic indels (rs147415279 and rs559509014) that were genotyped as a single triallelic indel by the 1000 Genomes Project.

			South Asian + European model																
			—	HLA-A		CCHCR1	HLA-C	—	HLA-B	—	HLA-B		MICA	—	—	HLA-DRB1		TAP2	—
		Linkage disequilibrium (W_n^2) in European dataset	rs12660712 (A,C)	rs1736927 (A,C)	rs1655901 (C,T)	rs2844626 (A,T)	HLA-C*06:02 (6*02,other)	rs9378158 (A,G)	HLA-B aa-67 (C,F,M,S,Y)	rs114811870 (C,T)	rs2853998 (C,T)	rs9266716 (C,T)	rs147415279 (AA,other)	rs566864214 (A,G)	rs2734588 (C,other)	rs112540072 (A,G)	rs9271539 (A,G)	rs28573770 (A,T)	rs3104406 (A,G)
European model	HLA-A	rs371194629 (–,14mer)	0.03	0.73	0.00	0.07	0.01	0.00	0.07	0.00	0.01	0.03	0.00	0.00	0.00	0.00	0.01	0.00	0.00
		rs1655901 (C,T)	0.03	0.02	–	0.01	0.01	0.00	0.05	0.01	0.01	0.03	0.00	0.00	0.02	0.00	0.01	0.01	0.00
	—	rs75881311 (T,other)	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.55	0.00	0.00	0.00	0.00	0.00
	HLA-C	rs1128175 (A,G)	0.00	0.02	0.02	0.00	0.04	0.02	0.43	0.01	0.06	0.09	0.00	0.01	0.03	0.03	0.02	0.00	0.05
		HLA-C*06:02 (6*02,other)	0.00	0.00	0.01	0.24	–	0.01	0.37	0.01	0.01	0.01	0.00	0.00	0.02	0.01	0.03	0.00	0.07
		rs41543814 (C,T)	0.00	0.03	0.00	0.43	0.11	0.01	0.19	0.05	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.01	0.00
	HLA-B	rs72866766 (C,T)	0.00	0.01	0.01	0.01	0.01	0.01	0.17	0.00	0.01	0.04	0.00	0.03	0.10	0.00	0.00	0.00	0.00
		HLA-B aa-171 (H,Y)	0.00	0.00	0.01	0.03	0.02	0.02	0.06	0.24	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.05	0.01
		HLA-B aa-67 (C,F,M,S,Y)	0.00	0.05	0.05	0.36	0.37	0.03	–	0.23	0.33	0.35	0.01	0.03	0.06	0.02	0.09	0.05	0.12
		rs137854633 (–,C,T)	0.00	0.01	0.00	0.07	0.26	0.01	0.18	0.03	0.06	0.13	0.00	0.00	0.21	0.02	0.01	0.01	0.04
	MICA	rs147415279 (AA,other)	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.00	0.00	0.00	–	0.00	0.00	0.00	0.00	0.00	0.00
	HLA-DRB1	rs6935999 (A,G)	0.00	0.00	0.01	0.00	0.00	0.00	0.04	0.18	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.81	0.01
		rs4947340 (C,T)	0.00	0.00	0.00	0.00	0.07	0.00	0.12	0.01	0.02	0.01	0.00	0.00	0.02	0.02	0.24	0.02	0.12
	DQA1	HLA-DQA1 aa-52 (R,other)	0.00	0.00	0.01	0.00	0.04	0.00	0.08	0.01	0.00	0.06	0.00	0.00	0.02	0.02	0.13	0.01	0.01

Figure S40. Pairwise linkage disequilibrium between variants of the European and transethnic association models for the MHC region. Variants are listed in the order of their position on chromosome 6, with the alleles used to compute LD in parentheses after each variant. LD values are for the multiallelic W_n^2 coefficient computed in the European dataset, which reduces to the usual r^2 coefficient when both variants of a pair are biallelic. The magnitude of LD values is accentuated with red shading on a 0 (white) to 1 (dark red) scale. Variants are labeled with a gene name if the variant changes that gene's protein sequence or if it is in substantial LD ($W_n^2 \geq 0.4$) with a protein-changing variant for the gene. Variant ID rs147415279 is shorthand for a biallelic split of two biallelic indels (rs147415279 and rs559509014) that were genotyped as a single triallelic indel by the 1000 Genomes Project.

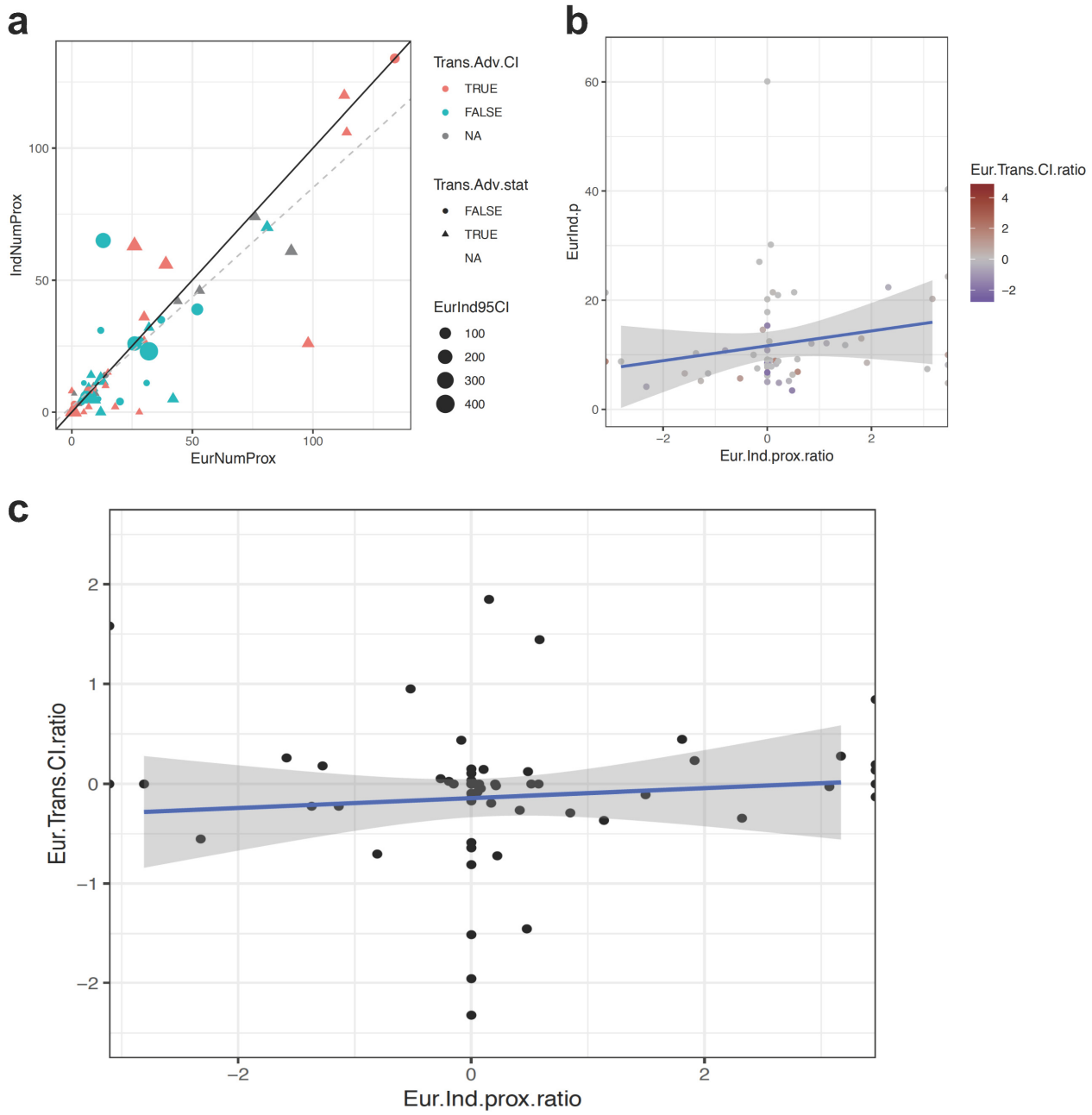

Figure S41. Comparisons between the number of population-specific LD proxies, significance of association, and size of credible interval sets for the psoriasis risk loci. a) The number of LD proxies in 1KGP phase 3 data for each non-MHC psoriasis-associated signal from the transethnic meta-analysis. X-axis shows the number of proxies from EUR, and the y-axis shows the number of proxies from SAS. Figure labels: Trans.Adv.CI: signal that has advantage (i.e. lower number of markers) in the 95% credible interval set in the transethnic meta-analysis; Trans.Adv.stat: signal that has advantage (more significant) in the transethnic meta-analysis; Trans.95CI: the number of markers in the 95% credible interval set in the transethnic meta-analysis. The gray dashed line represents a linear fit, and the black line represents a 1:1 relationship. **b)** The $\log_2$(ratio) of the number of LD proxies in EUR vs. SAS (x-axis) is plotted against the -log10(p-value) of each marker in the transethnic meta-analysis (y-axis). **c)** The $\log_2$(ratio) of the number of LD proxies in EUR vs. SAS (x-axis) is plotted against the ratio of the number of markers in the 95% credible interval in the EUR vs. transethnic meta-analysis (y-axis). In panels b and c, the blue line represents the linear fit, and the shaded area represents 95% confidence interval of the fitted linear model.

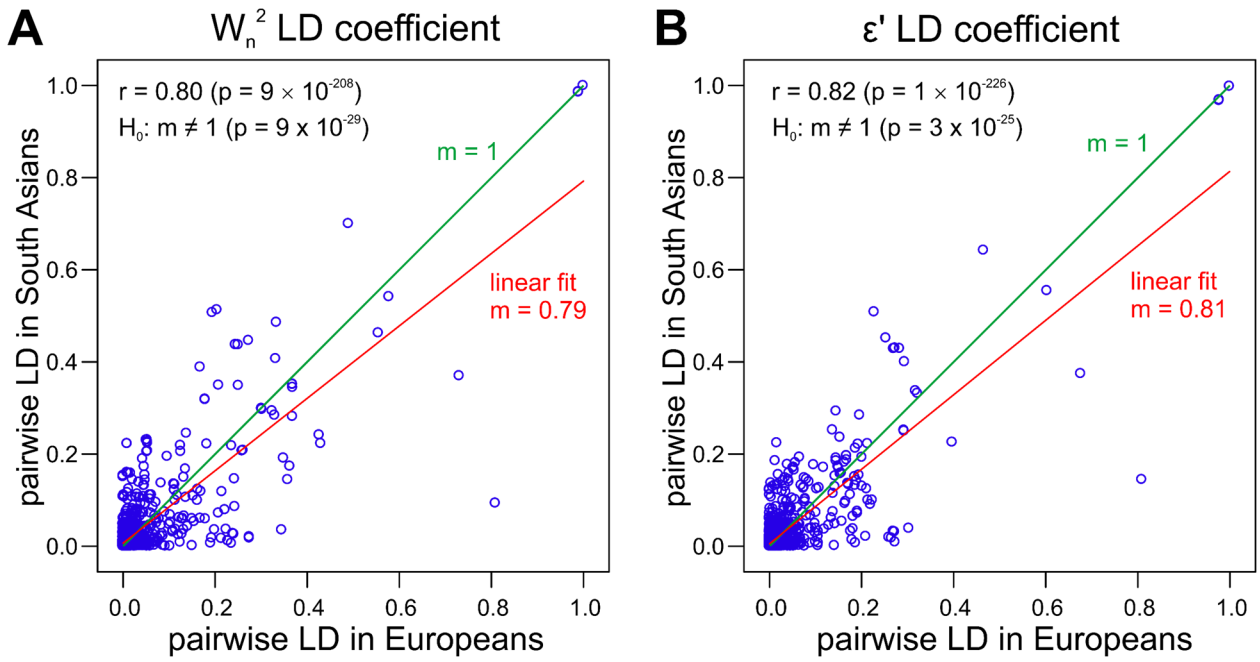

Figure S42. Scatterplots comparing strength of linkage disequilibrium between South Asians and Europeans for all MHC variants selected by stepwise association analysis for South Asians, Europeans, or South Asians and Europeans combined. Panels A and B show results for the W_n^2 and ε' multiallelic coefficients of LD, respectively. For the three multiallelic variants in the regression models, all biallelic split variants are analyzed in addition to the full multiallelic locus. The green and red lines depict a theoretical 1:1 linear correspondence and an empirical linear regression, respectively. The Pearson correlation coefficient and its significance is shown in the upper left corner of each plot, along with the p-value for testing the null hypothesis that the slope (m) of the linear regression fit is not equal to 1.

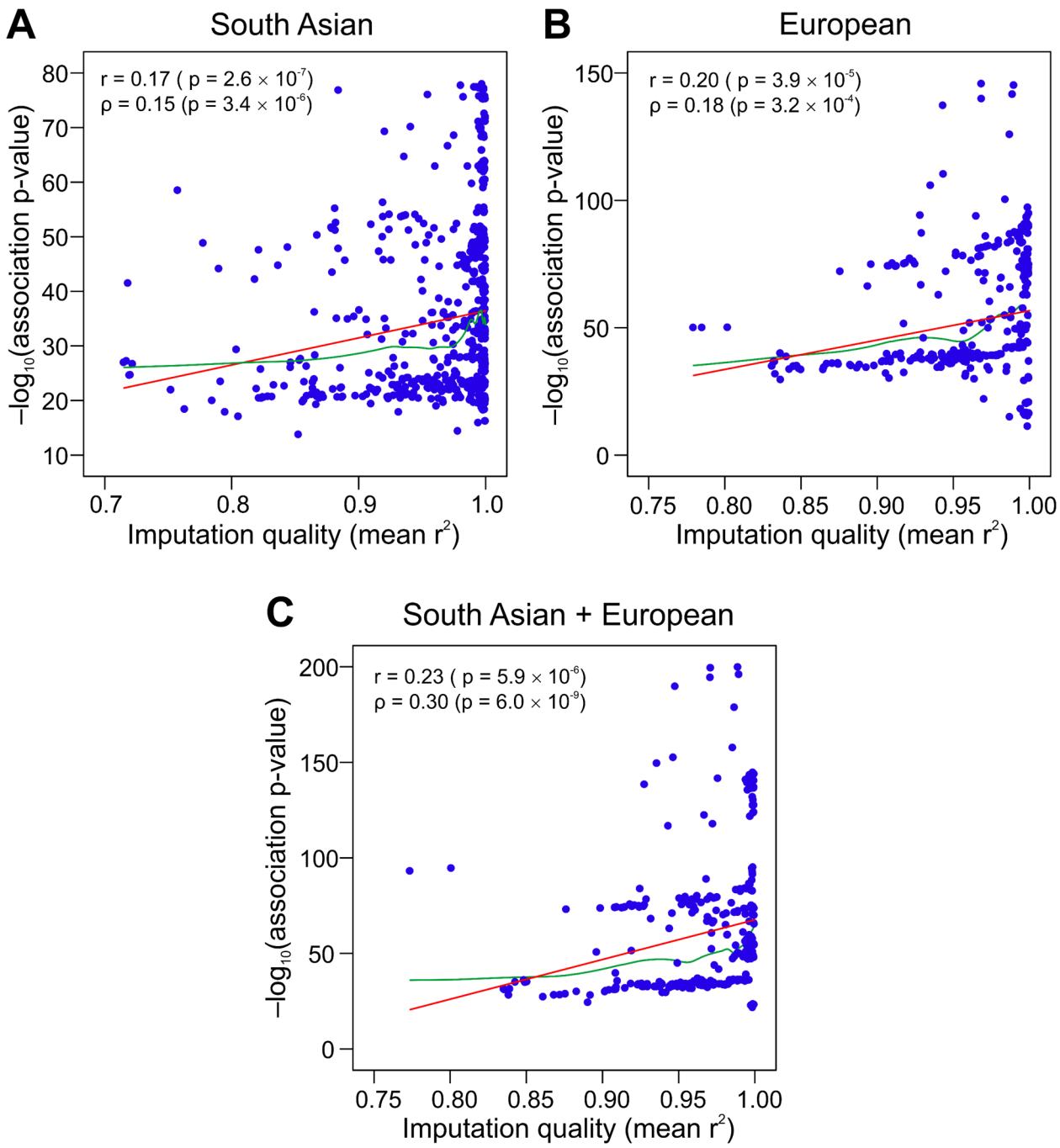

Figure S43. Scatterplots of the relationship of significance of MHC association with variant imputation quality. For variants in substantial linkage disequilibrium ($W_n^2 \geq 0.4$) with HLA-C*06, which is the most strongly associated locus in the full association model for all three ethnic datasets, the relationship of variant association ($-\log_{10}$ of full model p-value) with weighted mean variant imputation quality (mean Mach-r^2, effective sample size of each study in dataset as weights), is shown. Linear and loess fits are also plotted (red and green lines, respectively), and the Pearson *r* and Spearman *ρ* correlation coefficients and their significances are shown in the upper left corner of each plot. Panels A, B and C show results for the South Asian, European and transethnic datasets, respectively

Table S1. Characteristics of the 10 studies analyzed for psoriasis associations.

Study	Ancestry	No. individuals			N_{eff}[a]	No. Genotyped SNPs[b]			Type of Genotyping Microarray[d]
		PsV	Control	Total		classical MHC[c]	flanking MHC[c]	chr1–22	
South Asian, batch 1	Indian	952	855	1807	1802	6533	5201	909,101	exome + genome-wide
South Asian, batches 2+3	Indian, Pakistani	1638	865	2503	2264	3615	2851	508,389	exome + genome-wide
CASP GWAS	European	1336	1367	2703	2703	1235	1553	438,609	genome-wide
Exomechip	European	3845	4020	7865	7861	4547	2186	461,092	exome + genome-wide + custom PsV fine-mapping
GAPC Immunochip	European	2815	6730	9545	7939	1248	1603	489,501	autoimmune & inflammatory disease loci
Genizon GWAS	European	760	992	1752	1721	6032	2734	169,411	genome-wide
Kiel GWAS	European	464	1135	1599	1317	1344	1654	504,625	genome-wide
PAGE Immunochip	European	3167	7380	10,547	8864	5932	2665	160,228	autoimmune & inflammatory disease loci
PsA GWAS	European	1402	1398	2800	2800	10,929	16,961	972,453	genome-wide
WTCCC2 GWAS	European	2178	5172	7350	6130	1541	1737	515,579	genome-wide
All	—	18,557	29,914	48,471	43,401	—	—	—	—

Abbreviations: CASP, Collaborative Association Study of Psoriasis; GAPC, Genetic Analysis of Psoriasis Consortium; GWAS, genome-wide association study; N_{eff}, effective sample size; PAGE, Psoriasis Association of Genetics Extension; PsA, psoriatic arthritis; PsV, psoriasis vulgaris; SNP, single nucleotide polymorphism; WTCCC2, Wellcome Trust Case Control Consortium 2.

[a]Effective sample size, computed as $4/(1/N_{PsV} + 1/N_{control})$, is the size of a study with a balanced number of cases and controls having power to detect association equal to that of the actual study.

[b]After applying all quality control measures.

[c]MHC regions: classical (chr6:29.64–33.12 Mb), flanking (chr6:24–29.64 + 33.12–36 Mb); coordinates based on hg19 reference assembly.

[d]Microarrays for South Asian studies: batch1, Illumina OmniExpressExome-8v1-1_B; batch 2, Illumina HumanCoreExome-12v1-1_B; batch 3, Illumina HumanCoreExome-24v1-0_A.

Table S2. Validation sets used to assess accuracy of imputation for European ancestry individuals.

	No. samples in study with genotyped 1/2-field HLA alleles					
Psoriasis Study	**HLA-A**	**HLA-B**	**HLA-C**	**HLA-DQA1**	**HLA-DQB1**	**HLA-DRB1**
CASP GWAS	112/112	116/116	558/558	NA	NA	NA
Exomechip	1155/175	1159/904	1714/1714	174/174	91/89	91/0
PAGE Immunochip	96/0	117/16	121/86	NA	86/13	81/0
PsA GWAS	1369/0	1375/995	1392/1391	NA	1366/820	1373/494
Total	2732/287	2767/2031	3785/3749	174/174	1543/922	1545/494

Abbreviations: CASP, Collaborative Association Study of Psoriasis; GWAS, genome-wide association study; NA, not applicable; PAGE, Psoriasis Association of Genetics Extension; PsA, psoriatic arthritis.

Table S3. Control of population stratification for 10 studies analyzed for psoriasis associations in the MHC region.

Study	Ancestry	No. covariates			λ_{GC}[b]
		PC	Geographic[a]	Total	
South Asian, batch 1	Indian	5	0	5	1.020
South Asian, batches 2+3	Indian, Pakistani	9	0	9	1.024
CASP GWAS	European	5	2	7	1.008
Exomechip	European	7	4	11	1.015
GAPC Immunochip	European	16	5	21	1.048
Genizon GWAS	European	4	0	4	1.000
Kiel GWAS	European	3	0	3	1.036
PAGE Immunochip	European	6	2	8	1.013
PsA GWAS	European	4	2	6	1.008
WTCCC2 GWAS	European	7	0	7	1.037

Abbreviations: CASP, Collaborative Association Study of Psoriasis; GAPC, Genetic Analysis of Psoriasis Consortium; GWAS, genome-wide association study; PAGE, Psoriasis Association of Genetics Extension; PC, principal component; PsA, psoriatic arthritis; WTCCC2, Wellcome Trust Case Control Consortium 2.

[a]Number of geographic cohort indicator covariables after dropping one to avoid complete linear dependency.

[b]Robust estimate of the genomic control scaling parameter[4], a measure of residual population structure, based on association testing of microarray genotypes with genome-wide coverage.

Table S4. Heterogeneity of log(OR) effect sizes across studies in each meta-analysis.

	I^2 index (%)[b]			Cochran Q test p-value[c]		
Psoriasis risk variant[a]	SAS	EUR	SAS+EUR	SAS	EUR	SAS+EUR
1:8278116_G/A	0.0	6.1	0.0	4.42E-01	3.83E-01	5.29E-01
1:12054030_G/A	57.7	0.0	14.5	1.24E-01	5.19E-01	3.10E-01
1:24519437_C/A	46.7	0.7	13.7	1.71E-01	4.24E-01	3.17E-01
1:25297184_G/A	0.0	0.0	0.0	9.14E-01	9.53E-01	9.55E-01
1:67713346_T/C	0.0	0.0	0.0	3.83E-01	4.29E-01	5.27E-01
1:152593437_T/C	0.0	43	26.9	9.53E-01	9.18E-02	1.97E-01
1:168507463_C/T	0.0	24.2	15.4	8.19E-01	2.36E-01	3.01E-01
1:206648995_G/A	0.0	0.0	0.0	3.84E-01	5.94E-01	5.58E-01
2:61068822_C/CA	13.7	0.0	0.0	2.82E-01	9.48E-01	7.05E-01
2:62560332_A/G	0.0	0.0	0.0	5.74E-01	4.52E-01	5.52E-01
2:163110536_A/G	0.0	0.0	27.3	4.52E-01	8.42E-01	1.93E-01
3:16996035_G/A	0.0	0.0	0.0	8.34E-01	9.47E-01	9.87E-01
3:101616982_T/C	16.0	10.3	11.3	2.75E-01	3.50E-01	3.38E-01
5:96120198_TAAAC/T	0.0	41.8	28.0	5.17E-01	1.00E-01	1.86E-01
5:131996445_A/G	0.0	4.4	0.0	9.50E-01	3.96E-01	5.05E-01
5:150467189_G/C	0.0	0.0	0.0	8.27E-01	9.53E-01	9.08E-01
5:158829527_A/T	0.0	0.0	0.0	6.40E-01	8.54E-01	9.15E-01
6:577820_A/G	0.0	44.2	45.6	8.22E-01	8.38E-02	5.64E-02
6:20689945_G/A	11.3	57.0	53.0	2.88E-01	2.26E-02	2.38E-02
6:31269946_T/A	0.0	93.8	93.3	5.81E-01	2.85E-21	1.36E-24
6:111913262_C/T	0.0	69.5	60.8	9.73E-01	1.74E-03	6.26E-03
6:138197824_C/T	0.0	17.2	0.0	6.33E-01	2.94E-01	4.64E-01
6:159506600_C/T	0.0	0.0	0.0	5.35E-01	6.56E-01	7.01E-01
7:37385365_A/G	0.0	0.0	0.6	4.78E-01	4.82E-01	4.32E-01
9:32523737_T/C	0.0	0.0	0.0	5.87E-01	5.93E-01	6.83E-01
9:110814693_C/G	52.2	29.7	26.5	1.48E-01	1.91E-01	1.99E-01
10:75594050_G/T	0.0	12.1	10.7	3.67E-01	3.36E-01	3.44E-01
10:81043743_A/G	0.0	7.5	0.0	4.24E-01	3.72E-01	4.44E-01
10:102039458_G/A	0.0	22.5	0.0	6.87E-01	2.65E-01	4.61E-01
11:64123488_TG/T	53.4	0.0	0.0	1.43E-01	6.23E-01	5.80E-01
11:109973130_C/A	0.0	0.0	0.0	4.21E-01	9.44E-01	6.18E-01
11:128385169_G/A	11.3	0.0	0.0	2.88E-01	7.86E-01	6.30E-01
12:56750204_C/T	0.0	59.0	52.3	5.47E-01	1.69E-02	2.64E-02
12:112007756_C/T	0.0	42.0	35.8	9.06E-01	9.84E-02	1.21E-01
13:50794228_G/A	0.0	0.0	0.0	4.05E-01	9.18E-01	9.12E-01

Table S4. Continued.

Psoriasis risk variant[3]	I^2 index (%)[1]			Cochran Q test p-value[2]		
	SAS	EUR	SAS+EUR	SAS	EUR	SAS+EUR
13:99950260_G/A	0.0	14.9	0.0	9.82E-01	3.13E-01	4.85E-01
14:35839236_G/C	0.0	0.0	8.3	4.24E-01	5.72E-01	3.66E-01
14:98667928_T/C	0.0	25.7	8.8	9.64E-01	2.24E-01	3.61E-01
15:31637569_G/C	0.0	0.0	0.0	4.56E-01	6.60E-01	7.25E-01
16:11344903_C/T	0.0	0.0	0.0	7.32E-01	5.03E-01	5.78E-01
16:31057173_C/CAA	46.4	19.5	15.9	1.72E-01	2.75E-01	2.96E-01
17:26124908_G/A	0.0	0.0	0.0	4.33E-01	6.01E-01	5.56E-01
17:73851113_C/A	0.0	0.0	0.0	4.67E-01	7.98E-01	8.85E-01
18:12857002_G/T	0.0	0.0	0.0	9.33E-01	7.87E-01	9.13E-01
18:51791307_T/TTG	0.0	0.0	0.0	7.35E-01	4.49E-01	6.28E-01
19:10463118_G/C	0.0	0.0	0.0	3.55E-01	9.00E-01	6.11E-01
19:10819967_C/T	6.6	45.7	51.2	3.01E-01	7.48E-02	3.03E-02
19:49210869_A/G	0.0	0.0	6.7	6.28E-01	5.39E-01	3.80E-01
20:48561280_T/C	0.0	0.0	0.0	6.70E-01	6.39E-01	7.98E-01
22:21917479_C/T	53.3	0.0	0.0	1.44E-01	5.33E-01	5.15E-01

Abbreviations: EUR, European; SAS, South Asian.

[a]Psoriasis risk variants use a chrom:hg19_position_REF/ALT naming convention.

[b]I^2 index estimates the percentage of variability in the log(OR) effect sizes across studies that is due to real rather than chance differences.

[c]Cochran's Q test is an extension of the McNemar test that tests for heterogeneity of the log(OR) across studies in each meta-analysis. Nominal p-values are reported and can be multiplied by 50 to provide a Bonferroni correction for multiple testing.

Table S5. Credible interval analysis of the EUR and transethnic meta-analyses for established non-MHC psoriasis risk loci.

Established psoriasis locus					EUR 95% Bayesian CI				Transethnic 95% Bayesian CI			
					Best marker			No.	Best marker			No.
ID	Chr	Start pos	End pos	No. markers	Chr:pos	p-value	PP	markers	Chr:pos	p-value	PP	markers
chr1:8268095	1	8068095	8468095	1222	1:8282233	9.59E-11	32%	16	1:8278116	1.41E-11	30%	28
chr1:12053100	1	11853100	12253100	1475	1:12047028	1.29E-07	4%	83	1:12054030	1.22E-09	7%	23
chr1:24518643	1	24318643	24718643	1285	1:24513154	3.16E-18	30%	17	1:24519437	5.43E-21	22%	14
chr1:25293084	1	25093084	25493084	1267	1:25297184	2.14E-18	24%	43	1:25297184	1.17E-21	24%	43
chr1:67726104	1	67526104	67926104	1299	1:67707690	4.56E-24	21%	11	1:67713346	4.22E-25	21%	10
chr1:69788482	1	69588482	69988482	1065	1:69790226	3.87E-05	9%	192	1:69790226	6.81E-05	7%	281
chr1:78450517	1	78250517	78650517	1140	1:78444764	6.34E-07	79%	10	1:78444764	6.34E-07	80%	11
chr1:152590187	1	152390187	152790187	1238	1:152593549	1.44E-27	5%	25	1:152593437	6.30E-31	5%	25
chr1:168507463	1	168307463	168707463	1466	1:168507463	5.56E-07	61%	30	1:168507463	1.68E-09	91%	10
chr1:172675097	1	172475097	172875097	878	1:172675097	4.83E-07	8%	15	1:172674998	5.84E-07	8%	15
chr1:197671115	1	197471115	197871115	704	1:197814685	6.52E-06	11%	240	1:197704717	2.21E-06	6%	124
chr1:206655331	1	206455331	206855331	866	1:206648995	3.26E-09	28%	7	1:206648995	4.47E-09	28%	8
chr2:61083506	2	60883506	61283506	982	2:61072205	1.30E-23	6%	26	2:61068822	3.76E-23	12%	33
chr2:62551472	2	62351472	62751472	1284	2:62552321	4.21E-11	19%	9	2:62560332	8.59E-13	34%	11
chr2:163260691	2	163060691	163460691	575	2:163110536	1.76E-22	80%	2	2:163110536	1.14E-20	94%	2
chr3:16996035	3	16796035	17196035	1298	3:17011474	1.30E-12	14%	56	3:16996035	3.31E-13	7%	56
chr3:101663555	3	101463555	101863555	1320	3:101616982	1.08E-09	40%	53	3:101616982	1.16E-09	53%	56
chr3:189615475	3	189415475	189815475	1616	3:189662658	6.66E-07	75%	60	3:189662658	1.26E-07	77%	22
chr5:40370724	5	40170724	40570724	1535	5:40435058	8.55E-05	3%	167	5:40435058	0.00032	2%	456
chr5:96119273	5	95919273	96319273	1822	5:96120198	2.32E-20	19%	6	5:96120198	3.31E-22	19%	6
chr5:131996445	5	131796445	132196445	815	5:131996445	1.16E-17	40%	4	5:131996445	1.33E-18	52%	4
chr5:150467189	5	150267189	150667189	1700	5:150469973	1.72E-54	23%	8	5:150467189	7.25E-61	39%	8
chr5:158829527	5	158629527	159029527	1132	5:158829527	5.39E-80	100%	1	5:158829527	2.95E-81	99%	1
chr6:577820	6	377820	777820	1618	6:577820	6.44E-11	97%	1	6:577820	4.79E-10	90%	6
chr6:20678430	6	20478430	20878430	1530	6:20689945	6.34E-13	30%	7	6:20689945	7.22E-13	26%	9
chr6:111913262	6	111713262	112113262	1288	6:111929862	2.47E-45	71%	2	6:111913262	1.04E-48	90%	2
chr6:138197824	6	137997824	138397824	1226	6:138197824	1.15E-25	43%	4	6:138197824	8.11E-28	45%	4
chr6:159506600	6	159306600	159706600	1860	6:159506600	5.26E-09	41%	14	6:159506600	6.36E-09	30%	13
chr7:37386237	7	37186237	37586237	1446	7:37385365	1.19E-09	41%	5	7:37385365	3.00E-09	44%	25
chr9:32523737	9	32323737	32723737	1059	9:32523737	1.03E-10	53%	26	9:32523737	2.36E-12	96%	1
chr9:110817020	9	110617020	111017020	1470	9:110778738	8.47E-10	8%	80	9:110814693	6.38E-10	5%	80
chr9:117558703	9	117358703	117758703	1390	9:117611743	1.85E-05	14%	293	9:117611743	1.38E-05	28%	483
chr10:64369999	10	64169999	64569999	1416	10:64494348	1.04E-05	3%	127	10:64510934	9.61E-06	5%	143
chr10:75599127	10	75399127	75799127	632	10:75599127	1.49E-12	13%	13	10:75594050	1.64E-12	15%	14
chr10:81032532	10	80832532	81232532	1582	10:81043743	3.25E-11	47%	16	10:81043743	1.62E-11	28%	26

Table S5. Continued.

Established psoriasis locus					EUR 95% Bayesian CI				Transethnic 95% Bayesian CI			
					Best marker			No.	Best marker			No.
ID	Chr	Start pos	End pos	No. markers	Chr:pos	p-value	PP	markers	Chr:pos	p-value	PP	markers
chr10:89824771	10	89624771	90024771	1075	10:89637201	2.57E-07	68%	24	10:89637201	2.57E-07	59%	20
chr10:102038641	10	101838641	102238641	1157	10:102039458	5.87E-07	6%	57	10:102039458	4.11E-08	5%	58
chr11:64135298	11	63935298	64335298	846	11:64122279	1.62E-09	8%	55	11:64123488	9.57E-11	7%	53
chr11:65593444	11	65393444	65793444	1053	11:65656564	7.27E-07	17%	109	11:65656564	4.51E-07	14%	93
chr11:109962432	11	109762432	110162432	927	11:109959638	2.93E-17	13%	19	11:109973130	4.07E-16	8%	54
chr11:128406438	11	128206438	128606438	1174	11:128410264	3.99E-11	17%	15	11:128385169	1.02E-13	59%	11
chr12:10597207	12	10397207	10797207	1591	12:10594848	5.62E-06	2%	115	12:10573094	5.78E-06	4%	138
chr12:56750204	12	56550204	56950204	635	12:56750204	4.28E-21	3%	80	12:56750204	6.05E-21	3%	78
chr12:112059557	12	111859557	112259557	413	12:112007756	1.63E-07	25%	10	12:112007756	1.76E-08	25%	9
chr12:122668326	12	122468326	122868326	844	12:122661791	8.31E-06	12%	158	12:122661791	9.42E-06	8%	246
chr13:40333369	13	40133369	40533369	1289	13:40302608	1.39E-05	3%	169	13:40302608	6.27E-06	3%	149
chr13:45334194	13	45134194	45534194	968	13:45322344	1.24E-07	29%	6	13:45322344	2.47E-07	27%	7
chr13:50811220	13	50611220	51011220	1076	13:50811220	4.50E-09	20%	44	13:50794228	7.24E-09	19%	47
chr13:99950260	13	99750260	100150260	1438	13:99950260	5.34E-09	8%	28	13:99950260	9.52E-10	7%	42
chr14:35832666	14	35632666	36032666	1499	14:35832666	8.18E-20	86%	5	14:35839236	3.40E-22	100%	1
chr14:98668778:D	14	98468778	98868778	1957	14:98667928	3.08E-08	14%	55	14:98667928	3.18E-08	9%	54
chr15:31637666	15	31437666	31837666	1346	15:31637569	1.37E-09	50%	3	15:31637569	6.93E-10	52%	3
chr16:11365500	16	11165500	11565500	1981	16:11344903	1.20E-10	28%	9	16:11344903	1.05E-10	25%	5
chr16:31004812	16	30804812	31204812	533	16:31057173	2.23E-14	20%	76	16:31057173	2.29E-15	24%	56
chr17:26124908	17	25924908	26324908	1143	17:26124908	1.89E-21	100%	1	17:26124908	3.55E-22	100%	1
chr17:40561579	17	40361579	40761579	838	17:40536575	7.34E-09	17%	35	17:40536575	1.59E-07	12%	136
chr17:73890363	17	73690363	74090363	1011	17:73851113	1.71E-08	11%	43	17:73851113	7.27E-09	16%	47
chr17:78178893	17	77978893	78378893	1622	17:78178830	1.31E-06	13%	12	17:78178893	4.13E-07	16%	11
chr18:12857002	18	12657002	13057002	1381	18:12875316	6.92E-09	47%	3	18:12857002	1.57E-09	55%	3
chr18:51819750	18	51619750	52019750	1243	18:51791307	2.22E-08	2%	74	18:51791307	1.43E-09	3%	75
chr19:10463118	19	10263118	10663118	1200	19:10463118	2.02E-41	100%	1	19:10463118	4.78E-41	100%	1
chr19:10818092	19	10618092	11018092	1101	19:10819967	5.21E-12	14%	18	19:10819967	5.09E-11	14%	21
chr19:49206417	19	49006417	49406417	1456	19:49210869	2.46E-09	11%	33	19:49210869	1.42E-08	9%	34
chr20:48556229	20	48356229	48756229	1360	20:48590791	1.08E-20	9%	41	20:48561280	3.43E-22	5%	37
chr21:36470865	21	36270865	36670865	1287	21:36484440	2.95E-05	6%	150	21:36484440	1.56E-05	10%	131
chr22:21979289	22	21779289	22179289	802	22:21917450	1.69E-08	2%	86	22:21917479	2.62E-09	5%	73

Abbreviations: Chr, chromosome; CI, credible interval; EUR, European; pos, position; PP, posterior probability.

Table S6. The independent signals identified in the transethnic meta-analysis.

Established psoriasis locus						EUR 95% Bayesian CI of independent signal					EUR + SAS 95% Bayesian CI of independent signal				
						Best marker					Best marker				
ID	Rnd	Chr	Start pos	End Pos	No. Markers	Chr:pos	P-value	PP	No. markers	Length (bp)	Chr:pos	P-value	PP	No. markers	Length (bp)
chr1:67726104	1	1	67526104	67926104	1074	1:67623728	2.83E-13	8%	18	24349	1:67624304	9.68E-13	9%	17	24349
chr1:67726104	2	1	67526104	67926104	1060	1:67741314	8.35E-13	15%	13	18189	1:67705574	1.28E-12	53%	2	711
chr2:163260691	1	2	163060691	163460691	323	2:163128824	1.16E-17	100%	1	0	2:163128824	9.69E-17	100%	1	0
chr5:158829527	1	5	158629527	159029527	659	5:158777001	2.66E-52	13%	13	40751	5:158787385	3.04E-62	23%	11	40450
chr5:158829527	2	5	158629527	159029527	645	5:158811162	2.85E-20	15%	11	42850	5:158768365	4.54E-21	35%	10	42850
chr5:158829527	3	5	158629527	159029527	634	5:158687281	4.55E-07	8%	34	159860	5:158687281	8.15E-08	15%	27	170191
chr6:111913262	1	6	111713262	112113262	1006	6:111839019	5.09E-08	51%	9	135649	6:111839019	3.32E-08	58%	9	135649
chr9:110817020	1	9	110617020	111017020	347	9:110845060	7.63E-08	15%	19	16709	9:110845060	2.26E-09	12%	18	21353
chr14:35832666	1	14	35632666	36032666	1115	14:35828741	1.57E-11	81%	4	39773	14:35828741	9.27E-11	69%	4	39773
chr17:26124908	1	17	25924908	26324908	922	17:26103703	1.59E-10	33%	19	42680	17:26103703	8.32E-13	61%	14	33399
chr17:40561579	1	17	40361579	40761579	683	17:40519890	3.54E-09	8%	29	44035	17:40536575	1.46E-08	28%	48	238188
chr20:48556229	1	20	48356229	48756229	885	20:48642702	3.57E-08	54%	13	49855	20:48642702	6.86E-08	50%	14	49855

Note: The best marker(s) of the European + South Asian transethnic meta-analysis from the previous round(s) of the conditional analysis were used.
Abbreviations: Chr, chromosome; CI, credible interval; EUR, European; pos, position; PP, posterior probability; Rnd, round; SAS, South Asian.

Table S7. Accuracy of 1-field and 2-field HLA alleles imputed by a SNP2HLA reference panel of 397 individuals of South Asian ancestry.

HLA allele resolution[a]	Performance Metric	HLA-A	HLA-B	HLA-C	HLA-DPA1	HLA-DPB1	HLA-DQA1	HLA-DQB1	HLA-DRB1
1-field	sample accuracy (mean)[b]	0.9244	0.8548	0.9368	0.9400	0.8552	0.9693	0.9400	0.9102
	sample accuracy (median)[b]	0.9550	0.8960	0.9600	0.9900	0.9302	0.9850	0.9552	0.9250
	allelic accuracy (mean)[c]	0.8059	0.8451	0.9477	0.8408	0.5579	0.9352	0.9341	0.9417
	allelic accuracy (median)[c]	0.9421	0.8963	0.9727	0.8408	0.7020	0.9427	0.9498	0.9554
2-field	sample accuracy (mean)[b]	0.8166	0.7696	0.8944	0.9298	0.8490	0.8338	0.8932	0.7950
	sample accuracy (median)[b]	0.8719	0.8325	0.9403	0.9851	0.9250	0.8844	0.9261	0.8400
	allelic accuracy (mean)[c]	0.4743	0.6222	0.6212	0.4431	0.5778	0.7224	0.8482	0.5659
	allelic accuracy (median)[c]	0.6070	0.7755	0.8726	0.4651	0.7607	0.8504	0.9396	0.6796

Imputation accuracy of the 397-individual South Asian SNP2HLA panel (UM+BKT+IKMB-SAS) was assessed by leave-one-out cross-validation.

[a]Nomenclature of HLA alleles follows the most recent system,[5] where 1-field alleles describe the allele family (often corresponding to serological antigens), and 2-field alleles describe variations in the amino acid sequence.

[b]Sample accuracy is measured for each individual in the panel by first normalizing the imputed dosages so they sum to 2.0 for each individual across all gene alleles, and then subtracting one-half of the sum for that individual of the positive differences in genotyped vs. imputed dosages of all 1-field or 2-field HLA protein alleles in the reference panel from 1.

[c]Allelic accuracy is measured for each HLA 1-field or 2-field allele in the reference panel by first normalizing the imputed dosages so they sum to 2.0 for each individual across all gene alleles, and then computing the squared Pearson correlation of vectors of genotyped and imputed dosages of that allele for all individuals in the panel.

Table S8. Data sources for SNP2HLA reference panels constructed and tested by this study.

Source dataset	No. Individuals							HLA-DQA1 quality[b]	Reference
	South Asian	European	East Asian	African[a]	Admixed American	Iranian	Total		
UM	233	0	0	0	0	0	233	good	this study
BKT[c]	23	0	0	0	0	0	23	good	Liu (2015);[6] Degenhardt (2019)[7]
IKMB	143	322	451	312	0	132	1360	good	Liu (2015);[6] Degenhardt (2019)[7]
T1DGC	0	5225	0	0	0	0	5225	poor	Mychaleckyj (2010);[8] Jia (2013);[9] Onengut-Gumuscu (2015)[10]
Pan-Asian	120	0	410	0	0	0	530	poor	Pillai (2014);[11] Okada (2014)[12]
1KGP-ALL-v1[d]	541	526	528	700	371	0	2666	NA	The 1000 Genomes Project Consortium (2015);[13] Abi-Rached (2018)[14]
1KGP-ALL-v2[e]	489	503	504	661	347	0	2504	NA	The 1000 Genomes Project Consortium (2015);[13] Abi-Rached (2018)[14]

Abbreviations: 1KGP-ALL, 1000 Genomes Project, all populations; BKT, B.K. Thelma, IKMB, Institute of Clinical Molecular Biology (Kiel, Germany); NA, not applicable; SAS, South Asian; T1DGC, Type 1 Diabetes Genetics Consortium; UM, University of Michigan.

[a]Includes admixed Africans (all 312 of IKMB, 164 of 1KGP-ALL-v1 and 157 of 1KGP-ALL-v2 datasets).

[b]Older data sources (Pan-Asian and T1DGC) used a method for HLA-DQA1 genotyping that misclassified several alleles that commonly occur in many world populations, including HLA-DQA1*01:04–01:06, HLA-DQA1*03:02–03:03, and HLA-DQA1*05:02–05:09.

[c]BKT is a small subset of the Immunochip-typed IBD case-control Indian samples of B.K. Thelma described in Liu et al.[6] that were HLA-genotyped by Degenhardt et al.[7] but were not included in the final IKMB reference panel (Frauke Degenhardt, 2018, personal communication).

[d]Phase 3 1000 Genome Project samples for which both Affymetrix 6.0 microarray and HLA genotypes were available.

[e]Phase 3 1000 Genome Project samples for which both phased chromosomal haplotypes of sequence-based variant calls and HLA genotypes were available.

Table S9. Composition of 19 SNP2HLA reference panels constructed and tested for imputation of MHC variants in South Asians.

SNP2HLA reference panel	No. individuals			No. variants							No. HLA genes[a]
	SAS	EUR	Total	1-field HLA	2-field HLA	HLA AA	HLA SNP	HLA indel	non-HLA SNP	Total	
1KGP-ALL-v1	541	526	2666	87	411	749	378	3	2106	3734	5
1KGP-ALL-v2	489	503	2504	87	398	746	376	3	63,106	64,716	5
1KGP-SAS-v1	541	0	541	70	191	622	335	2	2093	3313	5
1KGP-SAS-v2	489	0	489	69	179	621	334	2	57,360	58,565	5
IKMB	143	322	1360	140	369	822	447	5	8803	10,586	8
IKMB+BKT	166	322	1383	140	369	822	447	5	8781	10,564	8
IKMB+BKT+UM	397	322	1614	141	388	828	450	5	2221	4033	8
IKMB+BKT+UM+1KGP-ALL-v2	886	825	4118	87	453	764	385	5	2165	3859	5
IKMB+BKT+UM+T1DGC	397	5547	6839	146	439	854	455	5	1840	3739	8
IKMB+BKT+UM+T1DGC+1KGP-ALL-v2	886	6050	9343	87	472	787	388	5	1817	3556	5
Pan-Asian	120	0	530	95	178	606	1415	1	6169	8464	8
Pan-Asian+T1DGC	120	5225	5755	129	319	753	1244	3	2641	5089	8
T1DGC	0	5225	5225	126	298	715	1615	3	5868	8625	8
UM+BKT+IKMB-SAS	397	0	397	99	226	728	434	2	2284	3773	8
UM+BKT+IKMB-SAS+1KGP-ALL-v2	886	503	2901	87	417	754	379	3	2235	3875	5
UM+BKT+IKMB-SAS+Pan-Asian	517	0	927	108	263	747	444	2	1591	3155	8
UM+BKT+IKMB-SAS+Pan-Asian+T1DGC	517	5225	6152	133	366	816	456	3	1357	3131	8
UM+BKT+IKMB-SAS+T1DGC	397	5225	5622	130	347	813	455	3	1882	3630	8
UM+BKT+IKMB-SAS+T1DGC+1KGP-ALL-v2	886	5728	8126	87	440	776	382	3	1851	3539	5

Abbreviations: 1KGP, 1000 Genomes Project; ALL, all populations (1000 Genomes); AA, amino acid; BKT, B.K. Thelma, EUR, European; IKMB, Institute of Clinical Molecular Biology (Kiel, Germany); indel, insertion-deletion; SAS, South Asian; SNP, single nucleotide polymorphism; T1DGC, Type 1 Diabetes Genetics Consortium; UM, University of Michigan.

[a]Number of HLA genes with genotypes in panel; HLA-A, -B, -C, -DQB1, -DRB1 for panels with 5 HLA genes and HLA-A, -B, -C, -DPA1, -DPB1, -DQA1, -DQB1, -DRB1 for panels with 8 HLA genes.

Table S10. Accuracy of 2-field HLA protein alleles imputed by the best-performing SNP2HLA panels for individuals of South Asian ancestry.

	Genes Imputed by SNP2HLA Reference Panel							
	IKMB+BKT+UM+1KGP-ALL-v2					IKMB+BKT+UM		IKMB+BKT+UM+ T1DGC
Performance Metric	**HLA-A**	**HLA-B**	**HLA-C**	**HLA-DQB1**	**HLA-DRB1**	**HLA-DPA1**	**HLA-DPB1**	**HLA-DQA1**
sample accuracy (mean)[a]	0.9535	0.9266	0.9648	0.9741	0.9379	0.9389	0.9308	0.9242
sample accuracy (median)[a]	1.0000	0.9901	1.0000	1.0000	1.0000	1.0000	0.9850	0.9950
allelic accuracy (mean)[b]	0.7915	0.8184	0.7659	0.8784	0.8059	0.5525	0.7760	0.7794
allelic accuracy (median)[b]	0.9634	0.9217	0.9625	0.9647	0.9341	0.8186	0.9545	0.9024
mean performance rank[c]	2.1	5.2	2.8	1.4	1.3	3.5	1.9	2.0

Imputation performance for each of the three panels was assessed by leave-one-out cross-validation for a set of 397 HLA-genotyped individuals of South Asian ancestry (IKMB-SAS+BKT+UM) that are also part of each reference panel.

Abbreviations: 1KGP, 1000 Genomes Project; ALL, all populations (1000 Genomes); BKT, B.K. Thelma; IKMB, Institute of Clinical Molecular Biology (Kiel, Germany); SAS, South Asian; T1DGC, Type 1 Diabetes Genetics Consortium; UM, University of Michigan.

[a]Sample accuracy is measured for each individual in the validation set by first normalizing the imputed dosages so they sum to 2.0 for each individual across all gene alleles, and then subtracting one-half of the sum for that individual of the positive differences in genotyped vs. imputed dosages of all 1-field or 2-field HLA alleles in the validation set from 1.

[b]Allelic accuracy is measured for each HLA 1-field or 2-field allele in the validation set by first normalizing the imputed dosages so they sum to 2.0 for each individual across all gene alleles, and then computing the squared Pearson correlation of vectors of genotyped and imputed dosages of that allele for all individuals in the validation set.

[c]Mean performance rank is computed as the mean of the rank of the best panel for 12 different metrics comparing it with all others being assessed (total of 19 panels for HLA-A, -B, -C, -DQB1 and -DRB1 and 11 panels for HLA-DPA1, -DPB1 and -DQA1). These 12 metrics consist of 3 paired-sample comparison measures based on the Wilcoxon signed rank test (mean biserial correlation, no. comparisons where rank sum of panel i > rank sum panel j, no. comparisons where rank sum panel i is significantly > rank sum panel j) and 3 paired-sample measures based on the paired t-test (mean paired difference, no. pairs where mean difference of panel i minus panel j is > 0, no. pairs where mean difference of panel i – panel j is significantly > 0 based on bootstrapping); each of these 6 paired-sample measures is applied to both the set of 2-field sample accuracies for all individuals in the validation set and the set of allelic accuracies for all 2-field alleles in the validation set.

Table S11. Accuracy of HLA alleles imputed by the best-performing SAS panel for populations in the phase 3 1000 Genomes dataset.

		mean sample accuracy for 1KGP super population[a]					mean allelic accuracy for 1KGP super population[b]				
Resolution	**Gene**	**AFR**	**AMR**	**EAS**	**EUR**	**SAS**	**AFR**	**AMR**	**EAS**	**EUR**	**SAS**
1-field	HLA-A	0.9901	0.9911	0.9913	0.9910	0.9879	0.9598	0.9686	0.9223	0.9763	0.9777
	HLA-B	0.9547	0.9566	0.9655	0.9829	0.9743	0.9079	0.9387	0.9231	0.9305	0.8884
	HLA-C	0.9971	0.9967	0.9986	0.9974	0.9930	0.9955	0.9919	0.9987	0.9995	0.9873
	HLA-DQB1	0.9889	0.9895	0.9922	0.9959	0.9908	0.9868	0.9849	0.9928	0.9942	0.9852
	HLA-DRB1	0.9627	0.9806	0.9889	0.9862	0.9864	0.9667	0.9810	0.9923	0.9829	0.9831
2-field	HLA-A	0.9634	0.9122	0.9309	0.9702	0.9513	0.7452	0.7105	0.7291	0.8855	0.6626
	HLA-B	0.9236	0.8311	0.9311	0.9531	0.9352	0.7436	0.5990	0.8339	0.8339	0.7881
	HLA-C	0.9717	0.9612	0.9722	0.9809	0.9767	0.7516	0.8298	0.8561	0.8444	0.7199
	HLA-DQB1	0.9658	0.9707	0.9668	0.9818	0.9675	0.8240	0.7789	0.7388	0.9423	0.7501
	HLA-DRB1	0.9360	0.8647	0.9421	0.9289	0.9306	0.7463	0.7622	0.7951	0.6895	0.6765

Imputation performance of SAS panel IKMB+BKT+UM+1KGP-ALL-v2 was assessed by leave-one-out cross-validation for a set of 661, 347, 504, 503 and 489 HLA-genotyped individuals of the five super populations of phase 3 of the 1000 Genomes project (1000 Genomes), which are also part of this reference panel.

Abbreviations: 1KGP, 1000 Genomes Project, AFR (African 1KGP super population), AMR (mixed American 1KGP super population), EAS (East Asian 1KGP super population), EUR (European 1KGP super population), SAS (South Asian 1KGP super population).[13]

[a]Sample accuracy is measured for each individual in the validation set by first normalizing the imputed dosages so they sum to 2.0 for each individual across all gene alleles, and then subtracting one-half of the sum for that individual of the positive differences in genotyped vs. imputed dosages of all 1-field or 2-field HLA alleles in the validation set from 1.

[b]Allelic accuracy is measured for each HLA 1-field or 2-field allele in the validation set by panel by first normalizing the imputed dosages so they sum to 2.0 for each individual across all gene alleles, and then computing the squared Pearson correlation of vectors of genotyped and imputed dosages of that allele for all individuals in the validation set.

Table S12. Composition of 20 SNP2HLA reference panels constructed and tested for imputation of MHC variants in Europeans.

SNP2HLA reference panel	No. individuals			No. variants							No. HLA genes[a]
	SAS	EUR	Total	1-field HLA	2-field HLA	HLA AA	HLA SNP	HLA indel	non-HLA SNP	Total	
1KGP-ALL-v1	541	526	2666	87	411	749	378	3	2106	3734	5
1KGP-ALL-v2	489	503	2504	87	398	746	376	3	63,106	64,716	5
1KGP-EUR-v1	0	526	526	75	165	512	597	2	2108	3459	5
1KGP-EUR-v2	0	503	503	75	165	512	597	2	56,218	57,569	5
IKMB	143	322	1360	140	369	822	447	5	8803	10586	8
IKMB+BKT	166	322	1383	140	369	822	447	5	8781	10564	8
IKMB+BKT+1KGP-ALL-v1	707	848	4049	87	456	764	385	5	1179	2876	5
IKMB+BKT+1KGP-ALL-v2	655	825	3887	87	443	761	383	5	6722	8401	5
IKMB+BKT+1KGP-EUR-v2	655	825	1886	87	306	670	361	5	6497	7926	5
IKMB+BKT+T1DGC	166	5547	6608	145	423	848	452	5	5562	7435	8
IKMB+BKT+T1DGC+1KGP-ALL-v1	707	6073	9274	87	474	785	387	5	932	2670	5
IKMB+BKT+T1DGC+1KGP-ALL-v2	655	6050	9112	87	463	784	386	5	5426	7151	5
IKMB+BKT+T1DGC+1KGP-EUR-v2	166	6050	7111	87	351	697	366	5	5407	6913	5
IKMB+BKT+UM	397	322	1614	141	388	828	450	5	2221	4033	8
IKMB+BKT+UM+T1DGC	397	5547	6839	146	439	854	455	5	1840	3739	8
IKMB+T1DGC	143	5547	6585	145	423	848	452	5	5575	7448	8
T1DGC	0	5225	5225	126	298	715	1615	3	5868	8625	8
T1DGC+1KGP-ALL-v1	541	5751	7891	87	433	769	380	3	956	2628	5
T1DGC+1KGP-ALL-v2	489	5728	7729	87	422	768	379	3	5660	7319	5
T1DGC+1KGP-EUR-v2	0	5728	5728	85	267	617	436	3	5639	7047	5

Abbreviations: 1KGP, 1000 Genomes Project; ALL, all populations (1000 Genomes); AA, amino acid; BKT, B.K. Thelma, EUR, European; IKMB, Institute of Clinical Molecular Biology (Kiel, Germany); indel, insertion-deletion; SAS, South Asian; SNP, single nucleotide polymorphism; T1DGC, Type 1 Diabetes Genetics Consortium; UM, University of Michigan.

[a]Number of HLA genes with genotypes in panel; HLA-A, -B, -C, -DQB1, -DRB1 for panels with 5 HLA genes and HLA-A, -B, -C, -DPA1, -DPB1, -DQA1, -DQB1, -DRB1 for panels with 8 HLA genes.

Table S13. Accuracy of 2-field HLA protein alleles imputed by the best-performing SNP2HLA panels for individuals of European ancestry.

	Genes Imputed by SNP2HLA Reference Panel					
	T1DGC+1KGP-EUR-v2					IKMB+BKT
Performance Metric	***HLA-A***	***HLA-B***	***HLA-C***	***HLA-DQB1***	***HLA-DRB1***	***HLA-DQA1***
sample accuracy (mean)[a]	0.9827	0.9650	0.9830	0.9871	0.9498	0.9601
sample accuracy (median)[a]	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
allelic accuracy (mean)[b]	0.9284	0.7429	0.8552	0.9010	0.8226	0.8704
allelic accuracy (median)[b]	0.9995	0.9396	0.9735	0.9956	0.9218	0.9908
mean performance rank[c]	2.7	4.5	8.3	3.8	2.6	1.1

Imputation performance for each of the two panels was assessed by independent validation for a set of up to 3749 HLA-genotyped individuals of European ancestry from four psoriasis case-control studies (CASP-GWAS, Exomechip, PAGE Immunochip, PsA GWAS) that are independent of all assessed reference panels.

Abbreviations: 1KGP, 1000 Genomes Project; BKT, B.K. Thelma; IKMB, Institute of Clinical Molecular Biology (Kiel, Germany); EUR, European; T1DGC, Type 1 Diabetes Genetics Consortium; UM.

[a]Sample accuracy is measured for each individual in the validation set by first normalizing the imputed dosages so they sum to 2.0 for each individual across all gene alleles, and then subtracting one-half of the sum for that individual of the positive differences in genotyped vs. imputed dosages of all 1-field or 2-field HLA alleles in the validation set from 1.

[b]Allelic accuracy is measured for each HLA 1-field or 2-field allele in the validation set by first normalizing the imputed dosages so they sum to 2.0 for each individual across all gene alleles, and then computing the squared Pearson correlation of vectors of genotyped and imputed dosages of that allele for all individuals in the validation set.

[c]Mean performance rank is computed as the mean of the rank of the best panel for 12 different metrics comparing it with all others being assessed (total of 20 panels for *HLA-A*, *-B*, *-C*, *-DQB1* and *-DRB1* and 7 panels for *HLA -DQA1*). These 12 metrics consist of 3 paired-sample comparison measures based on the Wilcoxon signed rank test (mean biserial correlation, no. comparisons where rank sum of panel i > rank sum panel j, no. comparisons where rank sum panel i is significantly > rank sum panel j) and 3 paired-sample measures based on the paired t-test (mean paired difference, no. pairs where mean difference of panel i minus panel j is > 0, no. pairs where mean difference of panel i – panel j is significantly > 0 based on bootstrapping); each of these 6 paired-sample measures is applied to both the set of 2-field sample accuracies for all individuals in the validation set and the set of allelic accuracies for all 2-field alleles in the validation set.

Table S14. Counts and densities of coding, non-coding and immune-related genes in the extended MHC region.

	chr6 position (Mb hg19)			No. genes[b]				Density of genes (per Mb)			
Genomic Region[a]	Region start	Region end	Length (Mb)	Coding	Non-coding	Total	Immune	Coding	Non-coding	Total	Immune
Flanking MHC, telomeric	24.00	25.73	1.73	18	18	36	0	10.4	10.4	20.8	0.0
Horton Extended MHC, telomeric	25.73	29.64	3.91	109	57	166	2	27.9	14.6	42.5	0.5
Classical MHC	29.64	33.12	3.48	137	76	213	37	39.4	21.8	61.2	10.6
Horton Extended MHC, centromeric	33.12	33.37	0.25	16	4	20	3	64.0	16.0	80.0	12.0
Flanking MHC, centromeric	33.37	36.00	2.63	41	25	66	3	15.6	9.5	25.1	1.1
Extended MHC	24.00	36.00	12.00	321	180	501	45	26.8	15.0	41.8	3.7
chr 1-22	NA	NA	2881.03	20,486	24,234	44,720	1752	7.1	8.4	15.5	0.6

Abbreviations: NA, not applicable.

[a]The extended MHC region of this study (chr6:24-36 Mb) was broken into five segments based on the bounds of the classical MHC and smaller extended MHC regions defined by Horton et al.[15]

[b]The numbers of coding and non-coding genes were determined by tallying unique gene symbols in the basic gene annotation set of GENCODE version 33lift37. The numbers of immune-related genes were based on a comprehensive list of immunologically important genes downloaded in July, 2020 from the ImmPort Shared Data repository (see Web Resources).

Table S15. Frequency distribution of variants in the imputed genotype datasets for the MHC region, cross-classified by reference panel source, MHC region and ancestry.

Reference panel source[a]	Classical MHC[b]			Flanking MHC[b]			Extended MHC[b]		
	SAS	EUR	SAS+EUR	SAS	EUR	SAS+EUR	SAS	EUR	SAS+EUR
1KGP	67,516	62,830	63,932	95,527	85,383	87,451	163,043	148,213	151,383
HRC	39,452	43,896	43,077	83,250	93,473	91,438	122,672	137,369	134,515
SNP2HLA (5 gene)	2074	1687	1635	0	0	0	2074	1687	1635
SNP2HLA (2 gene)	217	398	177	0	0	0	217	398	177
SNP2HLA (1 gene)	217	214	215	0	0	0	217	214	215
All	109,476	109,025	109,036	178,747	178,856	178,889	288,223	287,881	287,925

Abbreviations: 1KGP, 1000 Genomes Project; HRC, Haplotype Reference Consortium; EUR, European; SAS, South Asian.

For multiallelic variants, all counts include both the number of biallelic splits and the number of sets of biallelic splits because these splits are intended to be tested both individually and jointly for association.

[a]For the 1KGP and HRC reference panel sources, frequency counts are means because the number of imputed variants from these two sources in the final merged panel varies among the individual case-control studies. Three SNP2HLA reference panels were used for imputing HLA gene variants in the South Asian and European datasets; 5 gene, 2 gene and 1 gene refer, respectively, to the panels used to impute HLA-A/B/C/DQB1/DRB1, HLA-DPA1/DPB1, or HLA-DQA1 variants.

[b]MHC regions: classical (chr6:29.64-33.12 Mb), flanking (chr6:24-29.64 + 33.12-36 Mb), extended (chr6:24-36 Mb); coordinates based on hg19 reference assembly.

Table S16. Frequency distribution of variants in the imputed genotype datasets for the MHC region, cross-classified by minor allele frequency, imputation quality, MHC region and ancestry.

MHC region[b]	Ancestry	MAF[a]	No. variants with mean imputation quality[a]							
			0.00–0.30	0.30–0.50	0.50–0.70	0.70–0.80	0.80–0.90	0.90–0.95	0.95–1.00	Total
Classical	SAS	≥ 0.05	5022	1495	2321	1705	3477	2555	29,221	45,796
		0.01–0.05	1123	344	399	372	875	917	8159	12,189
		< 0.01	31,701	4389	4197	1507	1815	1177	5273	50,059
		Total	37,846	6228	6917	3584	6167	4649	42,653	108,044
	EUR	≥ 0.05	6138	1901	1732	1301	3188	3103	27,859	45,222
		0.01–0.05	1273	324	419	330	840	849	7002	11,037
		< 0.01	25,148	5939	5181	2826	3122	1943	7157	51,136
		Total	32,559	8164	7332	4457	7150	5895	42,018	107,575
	SAS+EUR	≥ 0.05	5967	2006	1842	1326	3365	3257	27,689	45,452
		0.01–0.05	1206	387	442	346	1017	1008	7655	12,061
		< 0.01	25,376	6198	5466	3038	3331	2389	4269	50,067
		Total	32,549	859	7750	4710	7713	6654	39,613	107,580
Flanking	SAS	≥ 0.05	17	214	597	661	1719	2208	17,679	23,095
		0.01–0.05	184	613	1069	972	1809	1351	6218	12,216
		< 0.01	102,052	14,543	9799	4218	4412	2525	4575	142,124
		Total	102,253	15,370	11,465	5851	7940	6084	28,472	177,435
	EUR	≥ 0.05	40	398	2935	2249	2869	2836	12,179	23,506
		0.01–0.05	143	568	2122	1527	1685	970	3802	10,817
		< 0.01	80,800	21,619	16,883	6497	6577	3920	6806	143,102
		Total	80,983	22,585	21,940	10,273	11,131	7226	22,787	177,425
	SAS+EUR	≥ 0.05	29	381	2415	2576	3102	3162	11,996	23,661
		0.01–0.05	122	563	2208	1608	1726	1097	3563	10,887
		< 0.01	82,119	22,701	17,002	6315	7267	3290	4183	142,877
		Total	82,270	23,645	21,625	10,499	12,095	7549	19,742	177,425

Table S16. (Continued)

MHC region[b]	Ancestry	MAF[a]	No. variants with mean imputation quality[a]							
			0.00–0.30	0.30–0.50	0.50–0.70	0.70–0.80	0.80–0.90	0.90–0.95	0.95–1.00	Total
Extended	SAS	≥ 0.05	5039	1709	2918	2366	5196	4763	46,900	68,891
		0.01–0.05	1307	957	1468	1344	2684	2268	14,377	24,405
		< 0.01	133,753	18,932	13,996	5725	6227	3702	9848	192,183
		Total	140,099	21,598	18,382	9435	14,107	10,733	71,125	285,479
	EUR	≥ 0.05	6178	2299	4667	3550	6057	5939	40,038	68,728
		0.01–0.05	1416	892	2541	1857	2525	1819	10,804	21,854
		< 0.01	105,948	27,558	22,064	9323	9699	5863	13,963	194,418
		Total	113,542	30,749	29,272	14,730	18,281	13,621	64,805	285,000
	SAS+EUR	≥ 0.05	5996	2387	4257	3902	6467	6419	39,685	69,113
		0.01–0.05	1328	950	2650	1954	2743	2105	11,218	22,948
		< 0.01	107,495	28,899	22,468	9353	10,598	5679	8452	192,944
		Total	114,819	32,236	29,375	15,209	19,808	14,203	59,355	285,005

Frequency counts are restricted to biallelic variants, including splits of multiallelic variants.

Abbreviations: EUR, European; MAF, minor allele frequency; SAS, South Asian.

[a]Membership in MAF and imputation quality bins based on the weighted mean of MAF and empirical imputation quality (R^2) values, respectively, across studies in the ancestry group, with the effective sample size of each study as weights.

[b]MHC regions: classical (chr6:29.64–33.12 Mb), flanking (chr6:24-29.64 + 33.12-36 Mb), extended (chr6:24–36 Mb); coordinates based on hg19 reference assembly.

Table S17. Frequency distribution of variants in the imputed genotype datasets for the MHC region, cross-classified by variant type, MHC region and ancestry.

	Classical MHC[a]			Flanking MHC[a]			Extended MHC[a]		
Variant Type	**SAS**	**EUR**	**SAS+EUR**	**SAS**	**EUR**	**SAS+EUR**	**SAS**	**EUR**	**SAS+EUR**
SNP	100,186	100,235	100,227	166,306	166,386	166,414	266,492	266,621	266,641
Indel	7146	7146	7153	12,330	12,359	12,364	19,476	19,505	19,517
Structural (≥ 50 bp)	96	96	96	111	111	111	207	207	207
Multiallelic, split[b]	5646	5159	5079	4814	4814	4814	10,460	9973	9893
Multiallelic, full[c]	1432	1450	1456	1312	1431	1464	2744	2881	2920
HLA 1-field allele	141	127	122	0	0	0	141	127	122
HLA 2-field allele	530	328	304	0	0	0	530	328	304
HLA amino acid	1377	1093	1134	0	0	0	1377	1093	1134
HLA SNP	13,164	13,173	13,150	0	0	0	13,164	13,173	13,150
HLA indel	714	707	711	0	0	0	714	707	711
HLA protein-changing	3013	2477	2489	0	0	0	3013	2477	2489
non-HLA protein-changing	2096	2094	2095	2195	2189	2189	4291	4283	4284
All[d]	109,476	109,025	109,036	178,747	178,856	178,889	288,223	287,881	287,925

Abbreviations: EUR, European; Indel, insertion-deletion polymorphism; SAS, South Asian; SNP, single nucleotide polymorphism.

[a]MHC regions: classical (chr6:29.64-33.12 Mb), flanking (chr6:24-29.64 + 33.12-36 Mb), extended (chr6:24-36 Mb); coordinates based on hg19 reference assembly.

[b]Refers to the individual biallelic variants in the sets that result from splitting of multiallelic variants.

[c]Refers to the full sets of multiallelic variants.

[d]Total number of variants, which is less than the sum of all variant type counts because of overlap among some of the variant type categories.

Table S18. Proportion of MHC reference panel variants analyzed for association with psoriasis, cross-classified by minor allele frequency, MHC region and ancestry.

	Proportion of variants analyzed for association[a]								
	Classical MHC[b]			Flanking MHC[b]			Extended MHC[b]		
MAF	SAS	EUR	SAS+EUR	SAS	EUR	SAS+EUR	SAS	EUR	SAS+EUR
≥ 0.05	0.797	0.693	0.682	0.953	0.712	0.679	0.849	0.699	0.681
0.01–0.05	0.836	0.759	0.715	0.815	0.547	0.454	0.826	0.655	0.591
< 0.01	0.179	0.173	0.092	0.092	0.079	0.038	0.115	0.104	0.052
Total	0.515	0.454	0.415	0.254	0.191	0.150	0.353	0.290	0.250

Abbreviations: EUR, European; MAF, minor allele frequency; SAS, South Asian.

[a]To qualify for association testing, the minimum imputation quality (R^2) for a variant among all case-control studies in the analyzed dataset must be ≥ 0.70.

[b]MHC regions: classical (chr6:29.64-33.12 Mb), flanking (chr6:24-29.64 + 33.12-36 Mb), extended (chr6:24-36 Mb); coordinates based on hg19 reference assembly.

Table S19. Frequency distribution of imputed MHC variants analyzed for association with psoriasis, cross-classified by reference panel source, MHC region and ancestry.

Reference panel source[a]	Classical MHC[b]			Flanking MHC[b]			Extended MHC[b]		
	SAS	EUR	SAS+EUR	SAS	EUR	SAS+EUR	SAS	EUR	SAS+EUR
1KGP	32,242	22,535	21,804	18,771	7809	6901	51,013	30,344	28,705
HRC	22,744	25,275	22,005	26,483	26,136	19,782	49,227	51,411	41,787
SNP2HLA (5 gene)	1163	1149	1106	0	0	0	1163	1149	1106
SNP2HLA (2 gene)	102	312	103	0	0	0	102	312	103
SNP2HLA (1 gene)	135	136	131	0	0	0	135	136	131
All	56,386	49,407	45,149	45,254	33,945	26,683	101,640	83,352	71,832

Abbreviations: 1KGP, 1000 Genomes Project; HRC, Haplotype Reference Consortium; EUR, European; SAS, South Asian.
For multiallelic variants, all counts include both the number of biallelic splits and the number of sets of biallelic splits because these splits are tested both individually and jointly for association.
[a]For the 1KG and HRC reference panel sources frequency counts are means because the number of imputed variants tested for association from these two sources varies across the individual case-control studies. Three SNP2HLA reference panels were used for imputing HLA gene variants in the South Asian and European datasets; 5 gene, 2 gene and 1 gene refer, respectively, to the panels used to impute HLA-A/B/C/DQB1/DRB1, HLA-DPA1/DPB1, or HLA-DQA1 variants.
[b]MHC regions: classical (chr6:29.64-33.12 Mb), flanking (chr6:24-29.64 + 33.12-36 Mb), extended (chr6:24-36 Mb); coordinates based on hg19 reference assembly.

Table S20. Frequency distribution of imputed MHC variants analyzed for association with psoriasis, cross-classified by minor allele frequency, imputation quality, MHC region and ancestry.

MHC region[b]	Ancestry	MAF[a]	No. variants with mean imputation quality[a]							
			0.70–0.75	0.75–0.80	0.80–0.85	0.85–0.90	0.90–0.95	0.95–0.99	0.99–1.00	Total
Classical	SAS	≥ 0.05	494	781	1257	2194	2556	5848	23,374	36,504
		0.01–0.05	71	208	356	504	917	2164	5995	10,215
		< 0.01	198	530	797	960	1177	1943	3330	8935
		Total	763	1519	2410	3658	4650	9955	32,699	55,654
	EUR	≥ 0.05	11	107	367	961	2232	5725	22,041	31,444
		0.01–0.05	7	55	177	344	758	2033	4969	8343
		< 0.01	4	59	273	589	1197	3081	3834	9037
		Total	22	221	817	1894	4187	10,839	30,844	48,824
	SAS+EUR	≥ 0.05	8	84	391	1027	2143	5832	21,594	31,079
		0.01–0.05	3	35	111	276	658	2131	5436	8650
		< 0.01	0	17	72	180	592	1958	2062	4881
		Total	11	136	574	1483	3393	9921	29,092	44,610
Flanking	SAS	≥ 0.05	115	315	585	1105	2208	7249	10,430	22,007
		0.01–0.05	131	459	737	1043	1351	2896	3322	9939
		< 0.01	470	470	1985	2248	2525	2901	1674	12,273
		Total	716	1244	3307	4396	6084	13,046	15,426	44,219
	EUR	≥ 0.05	15	87	226	1391	2827	4209	7970	16,725
		0.01–0.05	6	101	330	664	964	1939	1863	5867
		< 0.01	6	169	692	1769	2399	3922	2299	11,256
		Total	27	357	1248	3824	6190	10,070	12,132	33,848
	SAS+EUR	≥ 0.05	6	55	164	1143	2782	4496	7374	16,020
		0.01–0.05	1	11	105	410	780	1888	1590	4785
		< 0.01	0	13	182	684	1054	2306	1565	5804
		Total	7	79	451	2237	4616	8690	10,529	26,609

Table S20. (Continued).

MHC region[b]	Ancestry	MAF[a]	No. variants with mean imputation quality[a]							
			0.70–0.75	0.75–0.80	0.80–0.85	0.85–0.90	0.90–0.95	0.95–0.99	0.99–1.00	Total
Extended	SAS	≥ 0.05	609	1096	1842	3299	4764	13,097	33,804	58,511
		0.01–0.05	202	667	1093	1547	2268	5060	9317	20,154
		< 0.01	668	1838	2782	3208	3702	4844	5004	22,046
		Total	1479	3601	5717	8054	10,734	23,001	48,125	100,711
	EUR	≥ 0.05	26	194	593	2352	5059	9934	30,011	48,169
		0.01–0.05	13	156	507	1008	1722	3972	6832	14,210
		< 0.01	10	228	965	2358	3596	7003	6133	20,293
		Total	49	578	2065	5718	10,377	20,909	42,976	82,672
	SAS+EUR	≥ 0.05	14	139	555	2170	4925	10,328	28,968	47,099
		0.01–0.05	4	46	216	686	1438	4019	7026	13,435
		< 0.01	0	30	254	864	1646	4264	3627	10,685
		Total	18	215	1025	3720	8009	18,611	39,621	71,219

Frequency counts are restricted to biallelic variants, including splits of multiallelic variants.

Abbreviations: EUR, European; MAF, minor allele frequency; SAS, South Asian.

[a]Membership in MAF and imputation quality bins based on the weighted mean of MAF and empirical imputation quality (R^2) values, respectively, across studies in the ancestry group, with the effective sample size of each study as weights.

[b]MHC regions: classical (chr6:29.64–33.12 Mb), flanking (chr6:24-29.64 + 33.12-36 Mb), extended (chr6:24–36 Mb); coordinates based on hg19 reference assembly.

Table S21. Frequency distribution of imputed MHC variants analyzed for association with psoriasis, cross-classified by variant type, MHC region and ethnic population.

	Classical MHC[a]			Flanking MHC[a]			Extended MHC[a]		
Variant Type	**SAS**	**EUR**	**SAS+EUR**	**SAS**	**EUR**	**SAS+EUR**	**SAS**	**EUR**	**SAS+EUR**
SNP	50,513	44,207	40,183	40,383	31,178	24,292	90,896	75,385	64,475
Indel	4828	4280	4029	4844	2755	2381	9672	7035	6410
Structural (≥ 50 bp)	39	31	29	27	12	10	66	43	39
Multiallelic, split[b]	3550	2994	2809	1583	916	793	5133	3910	3602
Multiallelic, full[c]	732	583	539	197	97	94	929	680	613
HLA 1-field allele	80	78	72	0	0	0	80	78	72
HLA 2-field allele	119	107	83	0	0	0	119	107	83
HLA amino acid	807	704	753	0	0	0	807	704	753
HLA SNP	7330	5548	5193	0	0	0	7330	5548	5193
HLA indel	511	380	373	0	0	0	511	380	373
HLA protein-changing	1659	1497	1483	0	0	0	1659	1497	1483
non-HLA protein-changing	716	628	522	531	345	284	1247	973	806
All[d]	56,386	49,407	45,149	45,254	33,945	26,683	101,640	83,352	71,832

Abbreviations: EUR, European; Indel, insertion-deletion polymorphism; SAS, South Asian; SNP, single nucleotide polymorphism.

[a]MHC regions: classical (chr6:29.64-33.12 Mb), flanking (chr6:24-29.64 + 33.12-36 Mb), extended (chr6:24-36 Mb); coordinates based on hg19 reference assembly.

[b]Refers to the individual biallelic variants in the sets that result from splitting of multiallelic variants.

[c]Refers to the full sets of multiallelic variants.

[d]Total number of variants; less than the sum of all variant type counts because of overlap among some of the variant type categories.

Table S22. Annotations and protein-changing surrogates for associated variants in the extended MHC region for people of South Asian ancestry.

				Best protein-changing surrogate based on			
Step[a]	**Variant[b]**	**hg19 position[c]**	**nearest gene (consequence)**	**LD[d] (W_n^2, rank)**	**LD[d] (ε', rank)**	**p-value[e] (p-value, rank)**	**Bayesian PP[e] (PP, rank)**
1	*HLA-C*06*	31238192	*HLA-C* (protein)	*HLA-C*06* (NA, 1)	*HLA-C*06* (NA, 1)	*HLA-C*06* (1.3×10^{-78}, 2)	*HLA-C*06* (0.255, 2)
2	rs2428489	31352972	*AL645933.2* (9.1 kb downstream)	rs1051786 (*MICA*) (0.8349, 336)	rs1051786 (*MICA*) (0.6271, 183)	rs1051786 (*MICA*) (1.7×10^{-11}, 146)	rs1051786 (*MICA*) (0.000, 146)
3	rs2442752	31351764	*AL671883.3* (7.9 kb downstream)	HLA-B aa-45 (0.3794, 75)	HLA-B Lys45 (0.2733, 103)	HLA-B aa-45 (6.0×10^{-6}, 6)	HLA-B aa-45 (0.005, 8)
4	rs139451799	32454479	*HLA-DRB5* (31 kb downstream)	HLA-DRB1 aa-13 (0.9705, 2)	HLA-DRB1 Arg13 (0.9355, 2)	HLA-DRB1 aa-142 (2.9×10^{-6}, 4)	HLA-DRB1 aa-142 (0.008, 4)
5	rs9260313	29916885	*HLA-A* (3.2 kb downstream)	HLA-A aa- 97 (0.8210, 2)	HLA-A Met97 (0.7023, 2)	HLA-A Arg62 (1.8×10^{-5}, 31)	HLA-A Arg62 (0.005, 29)

Abbreviations: aa, amino acid; chr6, chromosome 6; LD, linkage disequilibrium; NA, not applicable, PP, posterior probability.

[a]Round of stepwise regression analysis.

[b]Build 151 dbSNP rsID when applicable,

[c]Base pair position in hg19 human reference; for classical HLA proteins the position of the center of the coding unit is given; for indels (all of which are insertions into the reference sequence), the position immediately before the insertion point is given.

[d]W_n^2 and ε' coefficients of linkage disequilibrium can accommodate multiallelic variants; see Subjects and Methods for more details.

[e]P-values and Bayesian posterior probabilities based on the final full regression model for associated variants.

Table S23. Parameters of 95% Bayesian credible sets for the five MHC psoriasis association signals in the final full regression model for people of South Asian ancestry.

	Stepwise-selected variant			Top variant in 95% BCS			Size of 95% BCS			
Step[a]	ID[b]	chr6 position[c]	PP	ID[b]	chr6 position[c]	PP	No. variants	Left bound[d]	Right bound[d]	Length[d]
1	*HLA-C*06*	31238192	0.255	rs12199223	31242731	0.266	12	31155785	31269946	114,162
2	rs2428489	31352972	0.003	rs72657660	31363697	0.075	31	31352972	31371101	18,130
3	rs2442752	31351764	0.583	rs2442752	31351764	0.583	2868	30859268	31851354	992,087
4	rs139451799	32454479	0.013	rs139451799	32454479	0.013	571	32386554	32636433	249,880
5	rs9260313	29916885	0.022	rs9260313	29916885	0.022	2587	29425584	30416864	991,281

Abbreviations: BCS, Bayesian credible set; chr6, chromosome 6; PP, posterior probability.

[a]Round of stepwise regression analysis.

[b]ID is build 151 dbSNP rsID when applicable.

[c]Base pair position in hg19 human reference; for classical HLA proteins the position of the center of the coding unit is given; for indels (all of which are insertions into the reference sequence), the position immediately before the insertion point is given.

[d]All values are bp on chromosome 6 of the hg19 human reference assembly.

Table S24. Psoriasis associations from stepwise analysis of the extended MHC region for eight studies of European ancestry.

			alleles		risk allele frequency		association at entry into model		association in final full model		
Step[a]	Variant[b]	chr6 position[c]	risk[d]	nonrisk	cases	controls	OR (95% CI)	p-value	OR (95% CI)	p-value	Vg
1	rs12211087	31269946	A	T	0.2470	0.0915	3.93 (3.75-4.13)	2.7×10^{-659}	NA[e]	NA[e]	NA[e]
2	HLA-B amino acid 67	31324536	C,F,M,Y	S	NA	NA	NA	1.3×10^{-85}	NA	8.1×10^{-75}	0.00913
			C	other	0.1476	0.1185	1.53 (1.45-1.61)	5.1×10^{-57}	1.65 (1.55-1.75)	1.4×10^{-55}	NA
			F	other	0.2049	0.2586	1.07 (1.02-1.12)	0.0025	1.17 (1.11-1.23)	1.9×10^{-8}	NA
			M	other	0.1277	0.0449	1.57 (1.46-1.69)	1.1×10^{-34}	1.61 (1.47-1.75)	3.3×10^{-27}	NA
			Y	other	0.1202	0.1601	1.00 (0.95-1.05)	0.97	1.15 (1.05-1.25)	0.0014	NA
			other	S	0.6004	0.5821	ref	ref	ref	ref	NA
3	rs1655901	29916804	C	T	0.5915	0.5187	1.35 (1.30-1.39)	3.3×10^{-64}	1.36 (1.32-1.41)	1.4×10^{-65}	0.00739
4	rs72866766	31321184	C	T	0.1093	0.0858	1.34 (1.26-1.43)	4.8×10^{-21}	1.47 (1.37-1.57)	3.7×10^{-29}	0.00382
5	rs4947340	32435338	T	C	0.4586	0.3756	1.15 (1.11-1.19)	1.6×10^{-14}	1.17 (1.13-1.22)	1.5×10^{-15}	0.00182
6	rs137854633	31324851	C,T	–	NA	NA	NA	5.0×10^{-12}	NA	7.7×10^{-13}	0.00226
			C	other	0.3941	0.4545	1.09 (1.04-1.14)	1.1×10^{-4}	1.17 (1.11-1.23)	1.3×10^{-9}	NA
			T	other	0.0702	0.0253	1.45 (1.30-1.61)	1.1×10^{-11}	1.37 (1.23-1.52)	1.7×10^{-8}	NA
			other	–	0.4642	0.4799	ref	ref	ref	ref	NA
7	HLA-B amino acid 171	31323979	Y	H	0.8938	0.8639	1.18 (1.12-1.24)	1.4×10^{-9}	1.24 (1.17-1.31)	4.9×10^{-14}	0.00164
8	rs6935999	32392757	A	G	0.0195	0.0138	1.59 (1.39-1.82)	1.8×10^{-11}	1.57 (1.37-1.80)	9.7×10^{-11}	0.00093
9	*HLA-C*06:02*	31238192	*C*06:02*	other	0.2478	0.0902	2.16 (1.66-2.80)	7.7×10^{-9}	2.80 (2.59-3.03)	6.4×10^{-146}	0.03006
10	rs75881311	31102273	T	other	0.9961	0.9944	2.39 (1.82-3.14)	4.3×10^{-10}	2.19 (1.64-2.91)	7.5×10^{-8}	0.00096
11	rs371194629	29798581	–	14mer	0.5803	0.5839	1.11 (1.07-1.15)	3.9×10^{-9}	1.13 (1.09-1.17)	2.0×10^{-11}	0.00117
12	rs41543814	31239430	C	T	0.6616	0.5793	1.13 (1.08-1.18)	7.4×10^{-8}	1.20 (1.14-1.27)	1.2×10^{-12}	0.00259
13	rs147145279	31388958	AA	other	0.0023	0.0009	2.95 (1.91-4.57)	1.2×10^{-6}	3.00 (1.94-4.64)	8.5×10^{-7}	0.00042
14	rs1128175	31150435	G	A	0.8272	0.7702	1.18 (1.10-1.26)	3.3×10^{-6}	1.20 (1.12-1.29)	1.9×10^{-7}	0.00187
15	HLA-DQA1 Arg52	32609228	other	R	0.6040	0.5466	1.09 (1.05-1.13)	1.1×10^{-5}	1.09 (1.05-1.13)	1.1×10^{-5}	0.00057

Abbreviations: chr6, chromosome 6; CI, confidence interval; NA, not applicable; OR, odds ratio; ref, reference; V_g, variance in liability explained by the genetic variant.[16]

[a]Round of stepwise regression analysis.

[b]Variant notes: variant ID is build 151 dbSNP rsID when applicable; *HLA-C*06:02* is one biallelic split from a decomposed set of 41 classical 2-field HLA-C alleles; the stepwise-selected variant for triallelic SNP rs75881311 is one of its biallelic splits with G/A+T alleles; rs147145279 is shorthand for two biallelic

indels (rs559509014 and rs147145279) that were genotyped as a single triallelic indel by phase 3 1000 Genomes with –, A and AA alleles, and the stepwise-selected variant is one of the biallelic splits of this triallelic indel; HLA-DQA1 Arg52 is a biallelic split (R/H+S alleles) of a triallelic amino acid variant.

[c]Base pair position in hg19 human reference; for amino acids and classical HLA proteins the position of the center of the coding unit is given; for indels (all of which are insertions into the reference sequence), the position immediately before the insertion point is given.

[d]Risk allele is based on final full regression model.

[e]Variant removed after addition of *HLA-C*06:02* in ninth round reduced its significance below the backward elimination inclusion threshold.

Table S25. Annotations and protein-changing surrogates for associated variants in the extended MHC region for people of European ancestry.

				Best protein-changing surrogate based on			
Step[a]	Variant[b]	hg19 position[c]	nearest gene (consequence)	LD[d] (W_n^2, rank)	LD[d] (ε', rank)	p-value[e] (p-value, rank)	Bayesian PP[e] (PP, rank)
1	rs12211087	31269946	*LINC02571* (527 bp upstream)	*HLA-C*06* (0.9897, 5)	*HLA-C*06* (0.9773, 5)	NA	NA
2	HLA-B aa-67	31324536	*HLA-B* (protein)	HLA-B aa-67 (NA, 1)	HLA-B aa-67 (NA, 1)	HLA-B aa-67 (8.1×10^{-75}, 1)	HLA-B aa-67 (1.000, 1)
3	rs1655901	29916804	*HLA-A* (3.1 kb downstream)	HLA-A aa-62 (0.5049, 48)	HLA-A Val99 (0.3705, 99)	HLA-A aa-95 (9.3×10^{-48}, 7)	HLA-A aa-95 (0.000, 7)
4	rs72866766	31321184	*HLA-B* (468 bp downstream)	HLA-B aa-158 (0.4424, 14)	HLA-B aa-158 (0.4601, 12)	HLA-B aa- 97 (2.8×10^{-23}, 2)	HLA-B aa-97 (0.000, 2)
5	rs4947340	32435338	*HLA-DRA* (23 kb downstream)	HLA-DRB1 aa-13 (0.8585, 2)	HLA-DQA1 Glu175 (0.7484, 2)	HLA-DRB1 aa-104 (5.0×10^{-12}, 47)	HLA-DRB1 aa-98 (0.000, 47)
6	rs137854633	31324851	*HLA-B* (intron)	HLA-B*13 (0.9976, 15)	HLA-B aa-10 (0.5509, 4)	HLA-B aa-10 (1.9×10^{-11}, 5)	HLA-B aa-10 (0.033, 5)
7	HLA-B aa-171	31323979	*HLA-B* (protein)	HLA-B aa-171 (NA, 1)	HLA-B aa-171 (NA, 1)	HLA-B aa-171 (4.9×10^{-14}, 1)	HLA-B aa-171 (0.996, 1)
8	rs6935999	32392757	*HLA-DRA* (15 kb upstream)	HLA-DRB1*01:02 (0.8376, 13)	HLA-DRB1*01:02 (0.8365, 10)	HLA-DRB1*01:02 (1.8×10^{-8}, 20)	HLA-DRB1*01:02 (0.005, 20)
9	HLA-C*06:02	31238192	*HLA-C* (protein)	*HLA-C*06:02* (NA, 1)	*HLA-C*06:02* (NA, 1)	*HLA-C*06:02* (6.4×10^{-146}, 2)	*HLA-C*06:02* (0.288, 2)
10	rs75881311	31102273	*PSORS1C1* (intron)	rs115749638 (*MUCL3*) (0.2173, 12)	rs115749638 (*MUCL3*) (0.2812, 12)	HLA-B aa-97 (4.4×10^{-5}, 22)	*HLA-B*39:01* (0.000, 17)
11	rs371194629	29798581	*HLA-G* (3'-UTR)	rs9260156 (*HLA-A*) (0.6027, 420)	HLA-A Arg163 (0.5091, 387)	HLA-A Val95 (3.0×10^{-9}, 93)	HLA-A Val95 (0.000, 93)
12	rs41543814	31239430	*HLA-C* (protein)	rs41543814 (*HLA-C*) (NA, 1)	rs41543814 (*HLA-C*) (NA, 1)	rs41543814 (*HLA-C*) (1.2×10^{-12}, 1)	rs41543814 (*HLA-C*) (0.992, 1)
13	rs147145279[f]	31388958	*HCP5* (intron)	rs41556715 (*MICA*) (0.9505, 4)	rs41556715 (*MICA*) (0.9474, 4)	rs41556715 (*MICA*) (5.3×10^{-6}, 14)	rs41556715 (*MICA*) (0.025, 11)
14	rs1128175	31150435	*POU5F1* (1.9 kb upstream)	HLA-C aa-99 (0.8454, 3)	rs3130981 (*CDSN*) (0.6574, 14)	HLA-C aa-99 (1.3×10^{-6}, 6)	HLA-C Cys99 (0.011, 11)
15	HLA-DQA1 Arg52	32609228	*HLA-DQA1* (protein)	HLA-DQA1 Arg52 (NA, 1)	HLA-DQA1 Arg52 (NA, 1)	HLA-DQA1 Arg52 (1.1×10^{-5}, 1)	HLA-DQA1 Arg52 (0.048,1)

Abbreviations: aa, amino acid; chr6, chromosome 6; LD, linkage disequilibrium; NA, not applicable, PP, posterior probability.

[a]Round of stepwise regression analysis.

[b]Build 151 dbSNP rsID when applicable

[c]Base pair position in hg19 human reference; for classical HLA proteins the position of the center of the coding unit is given; for indels (all of which are insertions into the reference sequence), the position immediately before the insertion point is given.
[d]W_n^2 and ε' coefficients of linkage disequilibrium can accommodate multiallelic variants; see Subjects and Methods for more details.
[e]P-values and Bayesian posterior probabilities based on the final full regression model for associated variants.
[f]rs147145279 is shorthand for two biallelic indels (rs559509014 and rs147145279) that were genotyped as a single triallelic indel by phase 3 1000 Genomes Project with –, A and AA alleles, and the stepwise-selected variant is one of the biallelic splits of this triallelic indel.

Table S26. Parameters of 95% Bayesian credible sets for the 14 MHC psoriasis association signals in the final full regression model for people of European ancestry.

	Stepwise-selected variant			Top variant in 95% BCS			Size of 95% BCS			
Step[a]	ID[b]	chr6 position[c]	PP	ID[b]	chr6 position[c]	PP	No. variants	Left bound[d]	Right bound[d]	Length[d]
2	HLA-B aa-67	31324536	1.000	HLA-B aa-67	31324536	1.000	1	31324535	31324537	3
3	rs1655901	29916804	1.000	rs1655901	29916804	1.000	1	29916804	29916804	1
4	rs72866766	31321184	1.000	rs72866766	31321184	1.000	1	31321184	31321184	1
5	rs4947340	32435338	0.965	rs4947340	32435338	0.965	1	32435338	32435338	1
6	rs137854633	31324851	0.543	rs137854633	31324851	0.543	7	31324851	31325932	1082
7	HLA-B aa-171	31323979	0.996	HLA-B aa-171	31323979	0.996	1	31323978	31323980	3
8	rs6935999	32392757	0.752	rs6935999	32392757	0.752	18	32384283	32681927	297,645
9	*HLA-C*06:02*	31238192	0.288	rs12189871	31251924	0.712	2	31236526	31251924	15,399
10	rs75881311	31102273	0.217	rs75881311	31102273	0.217	23	30968398	31495960	527,563
11	rs371194629	29798581	0.022	rs1736927	29796115	0.033	47	29784514	29820218	35,705
12	rs41543814	31239430	0.992	rs41543814	31239430	0.992	1	31239430	31239430	1
13	rs147145279[e]	31388958	0.124	rs147145279[e]	31388959	0.124	1852	30890201	31884909	994,709
14	rs1128175	31150435	0.180	rs1128175	31150435	0.180	662	30666027	31646746	980,720
15	HLA-DQA1 Arg52	32609228	0.048	HLA-DQA1 Arg52	32609228	0.048	6766	32109547	33108809	999,263

Abbreviations: aa, amino acid; BCS, Bayesian credible set; chr6, chromosome 6; PP, posterior probability.

[a]Round of stepwise regression analysis.

[b]ID is build 151 dbSNP rsID when applicable.

[c]Base pair position in hg19 human reference; for amino acids and classical HLA proteins the position of the center of the coding unit is given; for indels (all of which are insertions into the reference sequence), the position immediately before the insertion point is given.

[d]All values are bp on chromosome 6 of the hg19 human reference assembly.

[e]rs147145279 is shorthand for two biallelic indels (rs559509014 and rs147145279) that were genotyped as a single triallelic indel by phase 3 1000 Genomes Project with –, A and AA alleles, and the stepwise-selected variant is one of the biallelic splits of this triallelic indel.

Table S27. Psoriasis associations from stepwise analysis of the extended MHC region for ten studies of South Asian or European ancestry.

			alleles		risk allele frequency		association at entry into model		association in final full model	
Step[a]	Variant[b]	chr6 position[c]	risk[d]	nonrisk	cases	controls	OR (95% CI)	p-value	OR (95% CI)	p-value
1	rs12211087	31269946	A	T	0.2593	0.0921	4.09 (3.90-4.28)	5.3×10^{-779}	NA[e]	NA[e]
2	HLA-B amino acid 67	31324536	C,M,Y	F,S	NA	NA	NA	3.0×10^{-94}	NA	1.3×10^{-134}
			C	other	0.1363	0.1148	1.55 (1.47-1.63)	5.3×10^{-64}	1.98 (1.86-2.10)	4.0×10^{-105}
			other	F	0.7957	0.7404	0.93 (0.89-0.97)	6.2×10^{-4}	1.04 (0.98-1.10)	0.22
			M	other	0.1445	0.0485	1.55 (1.45-1.66)	3.3×10^{-39}	1.30 (1.20-1.40)	1.3×10^{-11}
			Y	other	0.1125	0.1550	1.03 (0.98-1.08)	0.31	1.25 (1.18-1.33)	1.3×10^{-13}
			other	S	0.5974	0.5779	ref	ref	ref	ref
3	rs1655901	29916804	C	T	0.5882	0.5151	1.35 (1.31-1.39)	3.5×10^{-72}	1.30 (1.25-1.34)	1.0×10^{-49}
4	rs2734588	31496467	other	C	0.2036	0.1659	1.24 (1.19-1.30)	5.9×10^{-24}	1.11 (1.06-1.17)	6.5×10^{-6}
5	rs2853998	31327164	T	C	0.7081	0.6544	1.17 (1.12-1.22)	1.0×10^{-13}	1.17 (1.12-1.23)	3.1×10^{-12}
6	rs114811870	31324938	C	T	0.9738	0.9669	1.50 (1.35-1.67)	1.7×10^{-13}	1.91 (1.69-2.15)	3.0×10^{-25}
7	rs28573770	32679275	A	T	0.0146	0.0108	1.90 (1.62-2.24)	3.6×10^{-15}	1.72 (1.46-2.02)	5.5×10^{-11}
8	*HLA-C*06:02*	31238192	*C*06:02*	other	0.2601	0.0908	2.25 (1.76-2.88)	1.5×10^{-10}	3.18 (2.95-3.43)	3.2×10^{-200}
9	rs2844626	31229552	T	A	0.4995	0.3626	1.18 (1.12-1.24)	5.9×10^{-10}	1.20 (1.14-1.26)	1.8×10^{-11}
10	rs1736927	29796115	A	C	0.4718	0.4863	1.11 (1.07-1.15)	2.3×10^{-9}	1.10 (1.06-1.14)	5.5×10^{-8}
11	rs9271539	32590028	G	A	0.7165	0.6610	1.11 (1.07-1.15)	6.8×10^{-8}	1.17 (1.12-1.21)	1.7×10^{-15}
12	rs9378158	31305761	G	A	0.0647	0.0676	1.19 (1.11-1.27)	2.7×10^{-7}	1.16 (1.09-1.24)	9.9×10^{-6}
13	rs112540072	32338289	G	A	0.9481	0.9266	1.20 (1.12-1.28)	2.4×10^{-7}	1.19 (1.11-1.27)	1.2×10^{-6}
14	rs566864214	31480770	A	G	0.9964	0.9945	1.87 (1.46-2.40)	6.3×10^{-7}	1.94 (1.51-2.48)	1.7×10^{-7}
15	rs9266716	31349727	T	C	0.7259	0.7149	1.13 (1.08-1.18)	5.8×10^{-7}	1.14 (1.09-1.20)	1.0×10^{-7}
16	rs147145279	31388958	AA	other	0.0021	0.0009	2.74 (1.78-4.21)	4.7×10^{-6}	2.78 (1.80-4.28)	3.5×10^{-6}
17	rs12660712	29734634	C	A	0.9756	0.9679	1.24 (1.12-1.36)	2.7×10^{-5}	1.24 (1.12-1.37)	2.2×10^{-5}
18	rs3104406	32682443	A	G	0.4881	0.3971	1.08 (1.04-1.12)	2.2×10^{-5}	1.08 (1.04-1.12)	2.2×10^{-5}

Abbreviations: chr6, chromosome 6; CI, confidence interval; NA, not applicable; OR, odds ratio; PP, posterior probability (Bayesian); ref, reference.

[a]Round of stepwise regression analysis.

[b]Variant notes: variant ID is build 151 dbSNP rsID when applicable; the stepwise-selected variant for triallelic SNP rs2734588 is one of its biallelic splits with C/A+G alleles; HLA-C*06:02 is one biallelic split from decomposed sets of 74 and 41 classical 2-field HLA-C alleles for the South Asian and European studies, respectively; rs147145279 is shorthand for two biallelic indels (rs559509014 and rs147145279) that were genotyped as a single triallelic indel by phase 3 1000 Genomes Project with –, A and AA alleles, and the stepwise-selected variant is one of the biallelic splits of this triallelic indel.

[c]Base pair position in hg19 human reference; for amino acids and classical HLA proteins the position of the center of the coding unit is given; for indels (all of which are insertions into the reference sequence), the position immediately before the insertion point is given.
[d]Risk allele is based on final full regression model.
[e]Variant removed after addition of HLA-C*06:02 in eighth round reduced its significance below the backward elimination inclusion threshold.

Table S28. Annotations and protein-changing surrogates for associated variants in the extended MHC region for people of South Asian or European ancestry.

				Best protein-changing surrogate based on					
Step[a]	**Variant[b]**	**hg19 position[c]**	**nearest gene (consequence)**	**LD in EUR[d] (W_n^2, rank)**	**LD in EUR[d] (ε', rank)**	**LD in SAS[d] (W_n^2, rank)**	**LD in SAS[d] (ε', rank)**	**p-value[e] (p-value, rank)**	**Bayesian PP[e] (PP, rank)**
1	rs12211087	31269946	*LINC02571* (527 bp upstream)	*HLA-C*06* (0.9897, 5)	*HLA-C*06* (0.9773, 5)	*HLA-C*06* (0.9861, 5)	*HLA-C*06* (0.9686, 5)	NA	NA
2	HLA-B aa-67	31324536	*HLA-B* (protein)	HLA-B aa-67 (NA, 1)	HLA-B aa-67 (NA, 1)	HLA-B aa-67 (NA, 1)	HLA-B aa-67 (NA, 1)	HLA-B aa-67 (1.3×10^{-134}, 1)	HLA-B aa-67 (1.000, 1)
3	rs1655901	29916804	HLA-A (3.1 kb downstream)	HLA-A aa-62 (0.5049, 48)	HLA-A Val95 (0.3705, 100)	HLA-A aa-62 (0.5563, 12)	HLA-A Ala152 (0.5702, 39)	HLA-A Glu62 (9.0×10^{-40}, 4)	HLA-A Glu62 (0.000, 4)
4	rs2734588	31496467	*MCCD1* (272 bp upstream)	HLA-C aa-156 (0.2110, 154)	HLA-B*13:02 (0.1933, 173)	rs28399976 (*MSH5*) (0.1715, 161)	rs28399976 (*MSH5*) (0.1875, 116)	rs2229094 (*LTA*) (8.7×10^{-5}, 42)	rs144223778 (*HSPA1B*) (0.001, 39)
5	rs2853998	31327164	*HLA-B* (2.2 kb upstream)	HLA-B aa-9 (0.5429, 2)	HLA-B His9 (0.5019, 2)	HLA-B aa-9 (0.7884, 2)	HLA-B His9 (0.7329, 2)	HLA-B Asp9 (9.9×10^{-11}, 2)	HLA-B aa-9 (0.034, 2)
6	rs114811870	31324938	*HLA-B* (5'-UTR)	HLA-C*08 (0.9419, 10)	HLA-C*08 (0.9195, 10)	rs41560824 (MICA) (0.7998, 21)	rs41560824 (MICA) (0.8322, 21)	rs41556417 (HLA-B) (2.6×10^{-21}, 13)	rs41556417 (HLA-B) (0.000, 13)
7	rs28573770	32679275	*AL662789*.1 (6.5 kb upstream)	rs140654840 (*TAP2*) (0.6124, 47)	rs140654840 (*TAP2*) (0.6411, 34)	rs140654840 (*TAP2*) (0.7140, 23)	rs140654840 (*TAP2*) (0.7627, 22)	rs16870005 (*TSBP1*) (6.5×10^{-9}, 28)	rs16870005 (*TSBP1*) (0.001, 28)
8	*HLA-C*06:02*	31238192	*HLA-C* (protein)	*HLA-C*06:02* (NA, 1)	*HLA-C*06:02* (NA, 1)	*HLA-C*06:02* (NA, 1)	*HLA-C*06:02* (NA, 1)	*HLA-C*06:02* (3.2×10^{-200}, 2)	*HLA-C*06:02* (0.664, 1)
9	rs2844626	31229552	*HLA-C* (7.0 kb downstream)	rs1576 (*CCHCR1*) (0.5764, 49)	rs1576 (*CCHCR1*) (0.4794, 51)	rs1576 (*CCHCR1*) (0.5713, 50)	rs2308592 (*HLA-C*) (0.4905, 72)	*HLA-B*51* (9.8×10^{-11}, 2)	*HLA-B*51* (0.106, 2)
10	rs1736927	29796115	*HLA-G* (intron)	rs9260156 (*HLA-A*) (0.6659, 97)	HLA-A Leu156 (0.5826, 108)	HLA-A aa-95 (0.5352, 184)	HLA-A Val95 (0.5083, 176)	rs543623321 (*HLA-A*) (2.6×10^{-7}, 86)	rs543623321 (*HLA-A*) (0.003, 84)
11	rs9271539	32590028	*HLA-DQA1* (15.1 kb upstream)	HLA-DRB1 aa-233 (0.8311, 3)	HLA-DRB1 aa-233 (0.7739, 3)	HLA-DRB1 aa-13 (0.8799, 3)	HLA-DRB1 Ser12 (e=0.80368, 3)	HLA-DRB1 aa-13 (7.5×10^{-12}, 4)	HLA-DRB1 aa-13 (0.000, 4)
12	rs9378158	31305761	*HLA-B* (15.9 kb downstream)	HLA-C aa-156 (0.2674, 173)	*HLA-C*07:04* (0.2961, 240)	*HLA-B*15* (0.3347, 141)	*HLA-C*07:01* (0.3244, 146)	*HLA-C*07:01* (2.5×10^{-5}, 6)	*HLA-C*07:01* (0.016, 6)

Table S28. (Continued).

Step[a]	Variant[b]	hg19 position[c]	nearest gene (consequence)	Best protein-changing surrogate based on					
				LD in EUR[d] (W_n^2, rank)	LD in EUR[d] (ε', rank)	LD in SAS[d] (W_n^2, rank)	LD in SAS[d] (ε', rank)	p-value[e] (p-value, rank)	Bayesian PP[e] (PP, rank)
13	rs112540072	32338289	*TSBP1* (intron)	*HLA-DRB1*01:01* (0.4935, 31)	*HLA-DRB1*01:01* (0.4653, 45)	HLA-DRB1 aa-30 (0.6365, 17)	*HLA-DRB1*01:01* (0.6330, 17)	*HLA-DRB1*01:01* (5.2×10^{-5}, 8)	*HLA-DRB1*01:01* (0.009, 8)
14	rs566864214	31480770	*MICB* (1.9 kb downstream)	HLA-B aa-158 (0.0737, 23)	HLA-B aa-158 (0.1029, 23)	HLA-B aa-158 (0.0207, 60)	HLA-B aa-158 (e=0.0334, 60)	rs16822820 (*HLA-DRB1*) (9.8×10^{-5}, 16)	*HLA-B*51:01* (0.000, 24)
15	rs9266716	31349727	*AL671883.3* (5.9 kb downstream)	HLA-B aa-163 (0.4128, 168)	*HLA-B*07* (0.3491, 170)	HLA-B aa-116 (0.4515, 211)	HLA-B*44:03 (0.3672, 232)	HLA-B Thr163 (9.5×10^{-6}, 68)	HLA-B Thr163 (0.002, 68)
16	rs147145279[f]	31388958	*HCP5* (intron)	rs41556715 (*MICA*) (0.9505, 4)	rs41556715 (*MICA*) (0.9474, 4)	rs41556715 (*MICA*) (0.7999, 5)	rs41556715 (*MICA*) (0.8368, 5)	rs41556715 (*MICA*) (1.8×10^{-5}, 5)	rs41556715 (*MICA*) (0.043, 5)
17	rs12660712	29734634	*AL645939.5* (3.6 kb downstream)	rs41551813 (*HLA-G*) (0.7624, 136)	rs41551813 (*HLA-G*) (0.7547, 136)	rs61730331 (*OR2I1P*) (0.4951, 49)	rs61730331 (*OR2I1P*) (0.4887, 49)	*HLA-B*51:01* (1.4×10^{-4}, 24)	rs41551813 (*HLA-G*) (0.000, 275)
18	rs3104406	32682443	*AL662789.1* (3.3 kb upstream)	HLA-DQA1 aa-25 (0.3647, 145)	HLA-DQA1 aa-25 (0.3334, 144)	HLA-DRB1 aa-67 (0.5896, 138)	HLA-DRB1 Leu67 (0.4870, 139)	HLA-DQA1 aa-25 (1.4×10^{-4}, 4)	HLA-DQA1 aa-25 (0.013, 2)

Abbreviations: aa, amino acid; chr6, chromosome 6; EUR, people of European ancestry; LD, linkage disequilibrium; NA, not applicable, PP, posterior probability; SAS, people of South Asian ancestry.

[a]Round of stepwise regression analysis.

[b]Build 151 dbSNP rsID when applicable

[c]Base pair position in hg19 human reference; for classical HLA proteins the position of the center of the coding unit is given; for indels (all of which are insertions into the reference sequence), the position immediately before the insertion point is given.

[d]W_n^2 and ε' coefficients of linkage disequilibrium can accommodate multiallelic variants; see Subjects and Methods for more details.

[e]P-values and Bayesian posterior probabilities based on the final full regression model for associated variants.

[f]rs147145279 is shorthand for two biallelic indels (rs559509014 and rs147145279) that were genotyped as a single triallelic indel by phase 3 1000 Genomes Project with –, A and AA alleles, and the stepwise-selected variant is one of the biallelic splits of this triallelic indel.

Table S29. Parameters of 95% Bayesian credible sets for the 17 MHC psoriasis association signals in the final full regression model for people of South Asian or European ancestry.

	Stepwise-selected variant			Top variant in 95% BCS			Size of 95% BCS			
Step[a]	ID[b]	chr6 position[c]	PP	ID[b]	chr6 position[c]	PP	No. variants	Left bound[d]	Right bound[d]	Length[d]
2	HLA-B aa-67	31324536	1.000	HLA-B aa-67	31324536	1.000	1	31324535	31324537	3
3	rs1655901	29916804	1.000	rs1655901	29916804	1.000	1	29916804	29916804	1
4	rs2734588	31496467	0.014	rs3093547	31541848	0.081	459	31002742	31973511	970,770
5	rs2853998	31327164	0.942	rs2853998	31327164	0.942	2	31324709	31327164	2456
6	rs114811870	31324938	0.349	rs114811870	31324938	0.349	6	31323480	31326410	2931
7	rs28573770	32679275	0.133	rs28573770	32679275	0.133	17	32340792	32679275	338,484
8	*HLA-C*06:02*	31238192	0.664	*HLA-C*06:02*	31238192	0.664	2	31236946	31251924	14,979
9	rs2844626	31229552	0.541	rs2844626	31229552	0.541	13	31229552	31324935	95,384
10	rs1736927	29796115	0.012	rs3215482	29796126	0.023	92	29784514	29913232	128,719
11	rs9271539	32590028	0.562	rs9271539	32590028	0.562	2	32589791	32590028	238
12	rs9378158	31305761	0.040	rs9501557	31231306	0.137	5376	30805799	31804729	998,931
13	rs112540072	32338289	0.275	rs112540072	32338289	0.275	3057	31840408	32836111	995,704
14	rs566864214	31480770	0.253	rs566864214	31480770	0.253	10	31007124	31773521	766,398
15	rs9266716	31349727	0.119	rs9266716	31349727	0.119	106	31140068	31541848	401,781
16	rs147145279[e]	31388958	0.184	rs147145279[e]	31388959	0.184	8935	30889260	31888869	999,610
17	rs12660712	29734634	0.032	rs12660712	29734634	0.032	6416	29234827	30233996	999,170
18	rs3104406	32682443	0.061	rs3104406	32682443	0.061	8965	32183589	33181884	998,296

Abbreviations: aa, amino acid; BCS, Bayesian credible set; chr6, chromosome 6; PP, posterior probability.

[a]Round of stepwise regression analysis.

[b]ID is build 151 dbSNP rsID when applicable.

[c]Base pair position in hg19 human reference; for amino acids and classical HLA proteins the position of the center of the coding unit is given; for indels (all of which are insertions into the reference sequence), the position immediately before the insertion point is given.

[d]All values are bp on chromosome 6 of the hg19 human reference assembly.

[e]rs147145279 is shorthand for two biallelic indels (rs559509014 and rs147145279) that were genotyped as a single triallelic indel by phase 3 1000 Genomes Project with –, A and AA alleles, and the stepwise-selected variant is one of the biallelic splits of this triallelic indel.

Table S30. Enrichment of variant types in full MHC regression models compared to the set of analyzed MHC variants.

Population	Variant type	Number of variants			Fold-enrichment		Enrichment p-value	
		Regression model	Classical MHC[a]	Extended MHC[a]	Classical MHC[a]	Extended MHC[a]	Classical MHC[a]	Extended MHC[a]
South Asian	all	5	56,390	101,640	NA	NA	NA	NA
	HLA protein-changing	1	1659	1659	6.80	12.25	0.14	0.079
	non-HLA protein-changing	0	716	1247	0.00	0.00	1.00	1.00
	indel	1	4828	9672	2.34	2.10	0.36	0.39
	structural (≥ 50 bp)	0	39	66	0.00	0.00	1.00	1.00
	full multiallelic	1	732	929	15.41	21.88	0.063	0.045
	split multiallelic	2	3550	5133	6.35	7.92	0.035	0.023
	multiallelic (all)	3	4282	6062	7.90	10.06	0.0039	0.0019
European	all	14	49,407	83,352	NA	NA	NA	NA
	HLA protein-changing	5	1497	1497	11.79	19.89	4.0×10^{-5}	3.2×10^{-6}
	non-HLA protein-changing	0	628	973	0.00	0.00	1.00	1.00
	indel	3	4280	7035	2.47	2.54	0.12	0.11
	structural (≥ 50 bp)	0	31	43	0.00	0.00	1.00	1.00
	full multiallelic	2	583	680	12.11	17.51	0.012	0.0057
	split multiallelic	4	2994	3910	4.71	6.09	0.0082	0.0033
	multiallelic (all)	6	3577	4590	5.92	7.78	2.6×10^{-4}	5.7×10^{-5}
South Asian + European	all	17	45,159	71,832	NA	NA	NA	NA
	HLA protein-changing	2	1483	1483	3.58	5.70	0.11	0.047
	non-HLA protein-changing	0	522	806	0.00	0.00	1.00	1.00
	indel	1	4029	6410	0.66	0.66	0.80	0.80
	structural (≥ 50 bp)	0	29	39	0.00	0.00	1.00	1.00
	full multiallelic	1	539	613	4.93	6.89	0.18	0.14
	split multiallelic	3	2809	3602	2.84	3.52	0.085	0.051
	multiallelic (all)	4	3348	4215	3.17	4.01	0.033	0.015

Abbreviations: NA, not applicable.

[a]MHC regions: classical (chr6:29.64-33.12 Mb), extended (chr6:24-36 Mb); coordinates based on hg19 reference assembly.

Table S31. Enrichment of variant types in the MHC region compared to the whole genome.

Population[a]	Variant type[b]	Number of variants			Fold-enrichment		Enrichment p-value	
		Classical MHC[c]	Extended MHC[d]	Whole genome	Classical MHC[c]	Extended MHC[d]	Classical MHC[c]	Extended MHC[d]
All	all	58,996	102,123	13,661,501	NA	NA	NA	NA
	protein-changing	938	1,241	40,966	5.30	4.05	1.5×10^{-820}	7.0×10^{-822}
	indel	4,231	9,436	1,478,852	0.66	0.85	1.00	1.00
	structural (≥ 50 bp)	65	100	9,753	1.54	1.37	6.2×10^{-4}	0.0014
	multiallelic	1,065	1,841	236,146	1.04	1.04	0.079	0.036
South Asian	all	54,257	88,781	10,215,054	NA	NA	NA	NA
	protein-changing	857	1,126	31,922	5.05	4.06	8.0×10^{-718}	3.2×10^{-749}
	indel	3,899	8,241	1,171,219	0.63	0.81	1.00	1.00
	structural (≥ 50 bp)	61	90	7,331	1.57	1.41	6.2×10^{-4}	0.0010
	multiallelic	1,021	1,689	182,099	1.06	1.07	0.042	0.0038
European	all	53,326	87,148	9,763,639	NA	NA	NA	NA
	protein-changing	824	1,089	30,831	4.89	3.96	1.0×10^{-668}	8.5×10^{-704}
	indel	3,814	8,088	1,129,099	0.62	0.80	1.00	1.00
	structural (≥ 50 bp)	58	83	7,036	1.51	1.32	0.0019	0.0081
	multiallelic	1,004	1,665	174,417	1.05	1.07	0.049	0.0031

Abbreviations: NA, not applicable.

[a]Populations as defined by phase 3 1000 Genomes: All includes 2504 individuals from 26 different populations; South Asian includes 489 individuals from five South Asian populations, and European includes 503 individuals from five European populations.

[b]Variant types based on autosomal variants in release 5 of phase 3 of the 1000 Genomes Project with minor allele frequency ≥ 0.01 in the designated population.

[c]Classical MHC: chr6:29.64–33.12 Mbp in hg19 reference assembly.

[d]Extended MHC: chr6:24–36 Mbp in hg19 reference assembly.

Table S32. Complete results for stepwise conditional analysis of South Asian, European and transethnic psoriasis associations in the extended MHC region.

See Excel file mmc2.xlsx in Supplemental information. The COL_KEY and Variant_KEY worksheets of this file describe the column headers, and how to interpret the nature of the various types of variants analyzed for association.

Table S33. Comparison of total and decomposed goodness of fit of within-population vs. cross-population association models for the MHC region.

model 1 (within-pop.)[a]	model 2 (cross-pop.)[b]	Model Subset	df		AIC				Tjur's R^{2}[e]		
			model 1	model 2	model 1	model 2	Δ[c]	Evidence Ratio[d]	model 1	model 2	Δ
SAS in SAS	EUR-top5 in SAS	full	21	23	-1228.35	-1169.07	-59.28	7.4×10^{12}	0.268	0.257	0.011
		covariates	15	15	-345.73	-347.29	1.56	4.6×10^{-1}	0.079	0.080	-0.001
		*HLA-C*06*	1	1	-582.18	-594.15	11.97	2.5×10^{-3}	0.123	0.127	-0.004
		other variants	5	7	-300.44	-227.64	-72.80	6.4×10^{15}	0.066	0.051	0.015
SAS in SAS	EUR-full in SAS	full	21	33	-1228.35	-1171.72	-56.63	2.0×10^{12}	0.268	0.262	0.006
		covariates	15	15	-345.73	-335.64	-10.09	1.6×10^{2}	0.079	0.077	0.002
		*HLA-C*06*	1	1	-582.18	-521.35	-60.84	1.6×10^{13}	0.123	0.111	0.012
		other variants	5	17	-300.44	-314.74	14.30	7.9×10^{-4}	0.066	0.074	-0.008
EUR-top5 in EUR	SAS in EUR	full	82	80	-14805.54	-14378.27	-427.27	6.0×10^{92}	0.304	0.296	0.009
		covariates	74	74	-9991.09	-10049.17	58.09	2.4×10^{-13}	0.198	0.200	-0.002
		*HLA-C*06*	1	1	-2878.65	-3259.53	380.88	2.0×10^{-83}	0.064	0.072	-0.008
		other variants	7	5	-1935.80	-1069.57	-866.24	1.3×10^{188}	0.042	0.023	0.019
EUR-full in EUR	SAS in EUR	full	92	80	-15086.99	-14378.27	-708.72	7.9×10^{153}	0.310	0.296	0.014
		covariates	74	74	-9853.78	-10049.17	195.40	3.7×10^{-43}	0.195	0.200	-0.005
		*HLA-C*06*	1	1	-2292.53	-3259.53	967.00	1.0×10^{-210}	0.051	0.072	-0.021
		other variants	17	5	-2940.68	-1069.57	-1871.12	2.0×10^{406}	0.064	0.023	0.041

Abbreviations: AIC, Akaike Information Criterion; df, degrees of freedom; EUR, European; SAS, South Asian.

[a]Within-population models are built by stepwise analysis within a single population, and the model parameters (coefficients and standard errors) are estimated for that same population (e.g., "EUR-full in EUR" refers to the full regression model that was selected and estimated in the same European dataset).

[b]Cross-population models are built by stepwise analysis in one population, but the model parameters are estimated in another population (e.g., "EUR-top5 in SAS" refers to the five most significant variants of the full regression model that was selected in the European dataset but with parameters re-estimated in the South Asian dataset).

[c]The difference in AIC of models 1 and 2. Negative values indicate that model 1 has more support; positive values that model 2 has more support. Absolute differences of 10 or more indicate that the poorer model has essentially no support of being the best approximating model of the two being compared.

[d]Evidence ratio = relative likelihood of model 1 versus model 2 = $1/\exp((AIC_1\text{-}AIC_2)/2)$.

[e]Also known as Tjur's D (coefficient of discrimination), equal to the difference in the average model-predicted probability of having psoriasis between psoriasis cases and unaffected controls.

Table S34. Mean imputation quality of variants in the imputed genotype datasets for the classical MHC region, cross-classified by minor allele frequency, ancestry, and reference panel.

		Mean imputation quality (R^2) for variants imputed using reference panel[a]				
MAF[b]	Ancestry	1KGP+HRC	SNP2HLA (5 genes)[c]	SNP2HLA (2 genes)[c]	SNP2HLA (1 gene)[c]	All
≥ 0.05	SAS	0.8265	0.9762	0.9724	0.9917	0.8298
	EUR	0.8051	0.9887	0.9877	0.9625	0.8098
	SAS+EUR	0.8066	0.9871	0.9750	0.9632	0.8105
0.01–0.05	SAS	0.8544	0.8848	0.8518	0.7274	0.8545
	EUR	0.8280	0.9490	0.9199	0.8914	0.8304
	SAS+EUR	0.8402	0.9285	0.8756	0.8772	0.8418
< 0.01	SAS	0.2974	0.2161	0.1922	0.2917	0.2959
	EUR	0.4176	0.3627	0.4414	0.1516	0.4170
	SAS+EUR	0.3968	0.3471	0.3756	0.1602	0.3962
Total	SAS	0.5841	0.6448	0.5361	0.7599	0.5852
	EUR	0.6212	0.7975	0.8399	0.7554	0.6246
	SAS+EUR	0.6183	0.8124	0.7229	0.7536	0.6212

Only biallelic variants are considered, including splits of multiallelic variants.

Abbreviations: 1KGP, 1000 Genomes Project; EUR, European; HRC, Haplotype Reference Consortium; MAF, minor allele frequency; NA, not applicable; SAS, South Asian.

[a]Average imputation quality computed as a weighted mean of empirical values, weighted by the effective sample size of each study in the ancestry dataset.

[b]Membership in MAF bins is based on a weighted mean of MAF values for each study in the ancestry group, with the effective study sample size as weights.

[c]Three SNP2HLA reference panels were used for imputing HLA gene variants in the South Asian and European datasets; 5 gene, 2 gene and 1 gene refer, respectively, to the panels used to impute HLA-A/B/C/DQB1/DRB1, HLA-DPA1/DPB1, or HLA-DQA1 variants.

Table S35. Comparison of 95% Bayesian credible sets for four MHC association signals occurring in both the transethnic association model and in at least one of the two monoethnic association models.

Stepwise-selected variant	95% Bayesian credible set											
	Top variant			PP of top variant			No. variants			Interval length (kb)		
	SAS	EUR	SAS+EUR	SAS	EUR	SAS+EUR	SAS	EUR	SAS+EUR	SAS	EUR	SAS+EUR
*HLA-C*06:02*	rs12199223	rs12189871	*HLA-C*06:02*	0.266	0.712	0.664	11	2	2	108.7	15.0	15.0
HLA-B aa-67	NA	HLA-B aa-67	HLA-B aa-67	NA	1.000	1.000	NA	1	1	NA	0.003	0.003
rs1655901	NA	rs1655901	rs1655901	NA	1.000	1.000	NA	1	1	NA	0.001	0.001
rs147145279[a]	NA	rs147145279[a]	rs147145279[a]	NA	0.229	0.184	NA	3910	8935	NA	997.6	999.6

To ensure an unbiased comparison, credible sets were recomputed for all three ancestry association models after restricting the variant sets to markers passing the imputation quality threshold of $r^2 \geq 0.70$ for all eight European and both South Asian studies.

Abbreviations: aa, amino acid; EUR, European; NA, not applicable; PP, posterior probability; SAS, South Asian.

[a]Variant rs147145279 is shorthand for two biallelic indels (rs559509014 and rs147145279) that were genotyped as a single triallelic indel by phase 3 1000 Genomes Project with –, A and AA alleles, and the stepwise-selected variant is one of the biallelic splits of this triallelic indel.

Table S36. Comparison of total and decomposed goodness of fit of transethnic vs. monoethnic association models for the MHC region.

model 1 (transethnic)[a]	model 2 (monethnic)[b]	Model Subset	df		AIC				Tjur's $R^{2(e)}$		
			model 1	model 2	model 1	model 2	Δ[c]	Evidence Ratio[d]	model 1	model 2	Δ
SAS+EUR in SAS	SAS in SAS	full	35	21	-1213.32	-1228.35	15.04	5.4×10^{-4}	0.271	0.268	0.003
		covariates	20	15	-416.52	-345.73	-70.79	2.4×10^{15}	0.094	0.079	0.014
		*HLA-C*06*	1	1	-442.02	-582.18	140.17	3.7×10^{-31}	0.095	0.123	-0.028
		other variants	19	5	-354.78	-300.44	-54.34	6.3×10^{11}	0.083	0.066	0.017
SAS+EUR in SAS	EUR in SAS	full	35	33	-1213.32	-1171.72	-41.59	1.1×10^{9}	0.271	0.262	0.009
		covariates	20	15	-416.52	-335.64	-80.88	3.7×10^{17}	0.094	0.077	0.017
		*HLA-C*06*	1	1	-442.02	-521.35	79.33	5.9×10^{-18}	0.095	0.111	-0.017
		other variants	19	17	-354.78	-314.74	-40.04	5.0×10^{8}	0.083	0.074	0.009
SAS+EUR in EUR	SAS in EUR	full	94	80	-15117.71	-14378.27	-739.44	3.7×10^{160}	0.311	0.296	0.015
		covariates	74	74	-9860.92	-10049.17	188.25	1.3×10^{-41}	0.195	0.200	-0.005
		*HLA-C*06*	1	1	-2355.60	-3259.53	903.93	5.2×10^{-197}	0.053	0.072	-0.019
		other variants	19	5	-2901.19	-1069.57	-1831.62	2.7×10^{451}	0.063	0.023	0.040
SAS+EUR in EUR	EUR in EUR	full	94	92	-15117.71	-15086.99	-30.72	4.7×10^{6}	0.311	0.310	0.001
		covariates	74	74	-9860.92	-9853.78	-7.15	3.6×10^{1}	0.195	0.195	0.000
		*HLA-C*06*	1	1	-2355.60	-2292.53	-63.07	5.0×10^{13}	0.053	0.051	0.002
		other variants	19	17	-2901.19	-2940.68	39.50	2.7×10^{-9}	0.063	0.064	-0.001
SAS+EUR in SAS+EUR	SAS in SAS+EUR	full	110	96	-17237.47	-16461.37	-776.10	3.4×10^{168}	0.320	0.306	0.014
		covariates	90	90	-10944.17	-11163.09	218.91	2.9×10^{-48}	0.195	0.201	-0.006
		*HLA-C*06*	1	1	-2931.78	-4015.65	1083.87	4.4×10^{-236}	0.059	0.080	-0.021
		other variants	19	5	-3361.52	-1282.63	-2078.89	5.4×10^{397}	0.066	0.025	0.041
SAS+EUR in SAS+EUR	EUR in SAS+EUR	full	110	108	-17237.47	-17123.59	-113.88	5.4×10^{24}	0.320	0.318	0.002
		covariates	90	90	-10944.17	-10884.18	-59.99	1.1×10^{13}	0.195	0.194	0.001
		*HLA-C*06*	1	1	-2931.78	-2915.15	-16.63	4.1×10^{3}	0.059	0.058	0.000
		other variants	19	17	-3361.52	-3324.26	-37.26	1.2×10^{8}	0.066	0.065	0.001

Abbreviations: AIC, Akaike Information Criterion; df, degrees of freedom; EUR, European; SAS, South Asian.

[a]Transethnic models are built by stepwise analysis in a dataset combining two ethnic groups, and the parameters (coefficients and standard errors) are estimated in either of the individual ethnic groups or their combination (e.g., "SAS+EUR in EUR" refers to the regression model that was selected in the transethnic South Asian + European dataset but with parameters re-estimated in the European dataset)

[b]Monoethnic models are built by stepwise analysis in a dataset for a single ethnic group, and the parameters are estimated in either of the individual ethnic groups or their combination (e.g., "SAS in SAS+EUR" refers to the regression model that was selected in the South Asian dataset but with parameters re-estimated in the combined South Asian + European dataset)

[c]The difference in AIC of models 1 and 2. Negative values indicate that model 1 has more support; positive values that model 2 has more support. Absolute differences of 10 or more indicate that the poorer model has essentially no support of being the best approximating model of the two being compared.

[d]Evidence ratio = relative likelihood of model 1 versus model 2 = $1/\exp((AIC_1-AIC_2)/2)$.

[e]Also known as Tjur's D (coefficient of discrimination), equal to the difference in the average model-predicted probability of having psoriasis between psoriasis cases and unaffected controls.

Table S37. Most significant cis-eQTL effects in relevant tissues for noncoding psoriasis-associated MHC variants with a Bayesian posterior probability > 0.50.

			most significant cis-eQTL result (target gene, sign risk allele effect, p-value) for variant								
Tissue[a]	Study[b]	N	rs1655901	rs2844626	rs72866766	rs137854633	rs2853998	rs2442752	rs6935999	rs4947340	rs9271529
Skin (normal, PsV lesional)	PsV RNAseq	171	—	*PSORS1C1* (+, 1e-04)	*MICA* (−, 1e-03)	—	—	*C4A* (+, 3e-06)	—	*HLA-DRB1* (−, 1e-13)	*C4A* (+, 2e-05)
Skin (not sun-exposed)	GTEx	517	*ZFP57* (+, 1e-09)	*C4A* (−, 7e-15)	*MICA* (−, 5e-10)	—	*MIR6891* (+, 4e-08)	*PSORS1C3* (+, 2e-11)	NS	—	*HLA-DRB5* (+, 2e-41)
Skin (sun exposed)	GTEx	605	*HCP5B* (+, 4e-15)	*HLA-C* (+, 7e-15)	*CCHCR1* (−, 7e-14)	—	*MIR6891* (+, 3e-09)	*C4A* (+, 5e-08)	*HLA-DRB5* (−, 1e-04)	—	*HLA-DRB5* (+, 3e-45)
Skin	TwinsUK	370	*ZFP57* (+, 1e-10)	*CCHCR1* (+, 2e-10)	*PSORS1C3* (+, 2e-07)	—	*HLA-B* (−, 2e-23)	*PSORS1C3* (+, 1e-13)	—	*HLA-DQA2* (+, 1e-44)	—
Fibroblasts (cultured)	GTEx	483	*HCG4P7* (+, 9e-26)	*AL645933.2* (−, 9e-11)	*MICA* (−, 8e-15)	—	*AL645933.2* (−, 5e-06)	*C4A* (+, 1e-05)	*LY6G5C* (−, 4e-04)	—	*C4B* (−, 2e-08)
Fibroblasts	GENCORD	186	*GABBR1* (+, 1e-02)	*AL645933.2* (−, 2e-06)	*MICA* (−, 7e-11)	*HLA-B* (−, 6e-14)	*AL645933.2* (−, 5e-08)	*AL645933.2* (+, 6e-03)	*CSNK2B* (−, 3e-03)	*ZBTB22* (+, 3e-03)	*HLA-DRB5* (+, 3e-03)
Whole blood, PBMCs	eQTLGen Consortium	32K	—	*C4B* (+, 8e-93)	*MICA* (−, 4e-103)	—	—	*HLA-B* (+, 8e-192)	*HLA-DRB6* (−, 1e-10)	*HLA-DQA2* (+, 3e-310)	*HLA-DRB6* (−, 3e-10)
Whole blood	GTEx	670	*HCP5B* (+, 2e-24)	*HLA-C* (+, 1e-39)	*MICA* (−, 3e-13)	—	*MIR6891* (+, 6e-15)	*NOTCH4* (−, 2e-14)	NS	—	*HLA-DRB5* (+, 1e-39)
Whole blood	Lepik (2018)	471	*AL645939.5* (+, 4e-12)	*HLA-C* (+, 3e-08)	*MICA* (−, 1e-17)	—	*HLA-B* (−, 1e-11)	*IER3* (−, 3e-08)	—	—	*HLA-DRB5* (+, 3e-39)
Whole blood	TwinsUK	195	*HCP5B* (+, 9e-10)	*CCHCR1* (+, 3e-08)	*PSORS1C3* (+, 7e-05)	—	*HLA-B* (−, 7e-08)	*LINC00243* (−, 2e-07)	—	*HLA-DQA2* (+, 9e-36)	—
B cells ($CD19^+$)	CEDAR	262	*HLA-A* (+, 7e-13)	*TNXB* (+, 2e-03)	*HLA-C* (+, 1e-05)	*HLA-C* (−, 7e-04)	*HLA-C* (−, 1e-02)	*HLA-C* (+, 3e-07)	*NCR3* (−, 9e-07)	*HLA-DRB5* (−, 2e-07)	*HLA-DRB1* (+, 2e-16)
B cells ($CD19^+$)	Fairfax (2012)	282	*HLA-A* (+, 7e-13)	*HCG27* (+, 2e-05)	*HLA-C* (+, 2e-04)	*HLA-C* (−, 2e-09)	*ATP6V1G2* (+, 1e-05)	*HCG22* (+, 2e-05)	*NCR3* (−, 2e-06)	*HLA-DRB1* (−, 1e-11)	*HLA-DRB1* (+, 1e-17)
B cells (naïve)	Schmiedel (2018)	91	*DHX16* (+, 4e-03)	*MICB* (−, 1e-03)	*POU5F1* (−, 4e-03)	*HLA-B* (−, 1e-03)	*HCG22* (−, 5e-03)	*HLA-B* (+, 2e-02)	*AGPAT1* (−, 3e-03)	*HLA-DRB1* (−, 9e-11)	*HLA-DRB5* (+, 3e-05)

Table S37. (Continued).

Tissue[a]	Study[b]	N	most significant cis-eQTL result (target gene, sign risk allele effect, p-value) for variant								
			rs1655901	rs2844626	rs72866766	rs137854633	rs2853998	rs2442752	rs6935999	rs4947340	rs9271529
LCLs	GTEx	147	*HLA-F-AS1* (+, 7e-06)	HLA-C (+, 4e-07)	NS	—	*HCG22* (−, 7e-08)	*HLA-C* (−, 2e-08)	NS	—	*HLA-DRB9* (−, 1e-14)
LCLs	GENCORD	190	*ZFP57* (+, 4e-06)	*HCG22* (−, 7e-06)	*MICA* (−, 2e-06)	*HLA-B* (−, 1e-12)	*HLA-B* (−, 3e-08)	*SKIV2L* (−, 4e-04)	*NOTCH4* (+, 3e-03)	*HLA-DRB1* (−, 1e-17)	*HLA-DRB5* (+, 3e-13)
LCLs	GEUVADIS	445	*AL645939.5* (+, 8e-14)	*AL645933.2* (−, 1e-12)	*MICA* (−, 4e-10)	*HLA-B* (−, 2e-16)	*AL645933.2* (−, 1e-12)	*HLA-C* (−, 6e-09)	*HSPA1A* (−, 8e-04)	*HLA-DQA1* (−, 5e-47)	*HLA-DRB5* (+, 8e-19)
LCLs	TwinsUK	418	*HLA-F-AS1* (+, 4e-26)	*RNF5* (−, 6e-18)	*MICA* (−, 2e-09)	—	*HLA-B* (−, 5e-25)	*HCG22* (+, 1e-16)	—	*HLA-DQA2* (+, 1e-71)	—
NK cells ($CD56^{dim}$ $CD16^{+}$)	Schmiedel (2018)	90	*HLA-F-AS1* (+, 6e-06)	*CCHCR1* (+, 9e-05)	*SKIV2L* (−, 8e-03)	*DDAH2* (−, 7e-03)	*MDC1-AS1* (−, 9e-03)	*PPP1R18* (+, 5e-03)	*HSPA1L* (−, 2e-02)	*HLA-DQA2* (+, 1e-06)	*HLA-DRB5* (+, 7e-05)
Primary T cells (PHA-stim)	GENCORD	184	*AL645939.5* (+, 9e-06)	*AL645933.2* (−, 8e-05)	*MICA* (−, 6e-07)	*HLA-B* (−, 2e-16)	*HLA-B* (−, 3e-06)	*ABCF1* (−, 1e-03)	*PSMB8* (+, 3e-03)	*HLA-DQA2* (+, 1e-21)	*HLA-DRB5* (+, 1e-15)
$CD4^{+}$ T cells (naïve)	BLUEPRINT	167	—	*CCHCR1* (+, 1e-12)	*AL662844.4* (−, 1e-05)	—	—	*HLA-B* (+, 8e-09)	—	—	—
$CD4^{+}$ T cells (naïve)	CEDAR	290	*HLA-A* (+, 2e-05)	*ABCF1* (+, 6e-03)	*HLA-C* (+, 2e-06)	*HLA-C* (−, 1e-05)	*HLA-C* (−, 1e-04)	*HLA-C* (+, 1e-09)	*HLA-DQB1* (+, 2e-03)	*HLA-DRB1* (−, 3e-08)	*HLA-DRB1* (+, 4e-15)
$CD4^{+}$ T cells (naïve)	Kasela (2017)	282	*VARS2* (−, 3e-03)	*HCG27* (+, 6e-06)	*AGPAT1* (−, 2e-03)	*VARS2* (+, 2e-07)	*HCG27* (+, 5e-05)	*HLA-C* (+, 3e-05)	—	*HLA-DRB1* (−, 3e-12)	*HLA-DRB1* (+, 3e-12)
$CD4^{+}$ T cells (naïve)	Schmiedel (2018)	88	*HLA-F-AS1* (+, 5e-05)	*CCHCR1* (+, 6e-04)	*CCHRC1* (−, 4e-03)	*HLA-B* (−, 5e-07)	*MRPS18B* (+, 1e-03)	*MPIG6B* (+, 7e-03)	*HSPA1L* (−, 1e-03)	*HLA-DQA1* (−, 3e-08)	*HLA-DRB5* (+, 8e-05)
$CD4^{+}$ T cells (anti-CD3-CD28)	Schmiedel (2018)	88	*PPP1R11* (−, 5e-03)	*HLA-C* (+, 1e-03)	*VWA7* (+, 2e-02)	*AL645933.2* (+, 8e-03)	*MDC1* (−, 6e-04)	*ABHD16A* (−, 2e-02)	*LTB* (+, 2e-02)	*HLA-DQB2* (+, 2e-08)	*HLA-DRB5* (+, 6e-06)

Table S37. (Continued).

Tissue[a]	Study[b]	N	most significant cis-eQTL result (target gene, sign risk allele effect, p-value) for variant								
			rs1655901	rs2844626	rs72866766	rs137854633	rs2853998	rs2442752	rs6935999	rs4947340	rs9271529
$CD8^+$ T cells (naive)	CEDAR	277	*HLA-A* (+, 5e-08)	*HLA-E* (+, 3e-04)	*HLA-C* (−, 2e-05)	*HLA-C* (−, 4e-05)	*HLA-C* (−, 9e-08)	*HLA-C* (+, 9e-09)	NS	*HLA-DRB5* (−, 2e-07)	*HLA-DRB1* (+, 1e-15)
$CD8^+$ T cells (naive)	Kasela (2017)	271	*HLA-A* (+, 2e-04)	*ATP6V1G2* (+, 1e-03)	*ATAT1* (+, 5e-03)	*HLA-C* (−, 2e-06)	*HCG27* (+, 9e-05)	*HLA-C* (+, 6e-05)	—	*HLA-DRB1* (−, 2e-09)	*HLA-DRB1* (+, 2e-13)
$CD8^+$ T cells (naïve)	Schmiedel (2018)	89	*HLA-F-AS1* (+, 1e-05)	*CCHCR1* (+, 8e-04)	*NFKBIL1* (+, 2e-02)	*HLA-B* (−, 3e-05)	*MDC1* (−, 2e-03)	*NELFE* (−, 2e-03)	*LY6G5C* (−, 3e-03)	*HLA-DQA2* (+, 1e-05)	*HLA-DRB5* (+, 7e-07)
$CD8^+$ T cells (anti-CD3-CD28)	Schmiedel (2018)	88	*GNL1* (−, 8e-05)	*HCG27* (+, 1e-03)	*CCHCR1* (−, 3e-03)	*PRRC2A* (−, 4e-03)	*MDC1* (−, 4e-03)	*PPP1R10* (−, 2e-04)	*RNF5* (−, 8e-03)	*HLA-DQA1* (−, 2e-08)	*HLA-DRB5* (+, 2e-04)
T_{REG} cells (naïve)	Schmiedel (2018)	89	*HLA-F-AS1* (+, 2e-05)	*CCHCR1* (+, 7e-04)	*ATAT1* (−, 3e-03)	*HLA-B* (−, 2e-07)	*DHX16* (+, 1e-02)	*PRRT1* (−, 2e-02)	*HLA-DQB2* (−, 1e-02)	*HLA-DQB2* (+, 2e-06)	*HLA-DRB5* (+, 3e-05)
T_{REG} cells (memory)	Schmiedel (2018)	89	*HLA-F-AS1* (+, 6e-11)	*NRM* (−, 1e-02)	*PPT2* (−, 3e-04)	*MICB* (+, 2e-03)	*NCR3* (−, 3e-03)	*FLOT1* (−, 3e-03)	PSMB9 (−, 5e-04)	*HLA-DQA2* (+, 5e-06)	*HLA-DRB5* (+, 9e-06)
T_H1 cells	Schmiedel (2018)	82	*HLA-F-AS1* (+, 6e-07)	*CCHCR1* (+, 2e-04)	*PRRC2A* (+, 4e-04)	*HLA-B* (−, 2e-05)	*DDX39B* (−, 2e-02)	*VWA7* (+, 3e-03)	HSPA1L (−, 6e-03)	*HLA-DQB2* (+, 4e-05)	*RXRB* (+, 8e-04)
$T_H1/17$ cells	Schmiedel (2018)	88	*HLA-F-AS1* (+, 5e-09)	*CCHCR1* (+, 9e-03)	*DDX39B* (+, 1e-03)	*HLA-B* (−, 1e-06)	*HLA-B* (−, 2e-02)	*VWA7* (+, 2e-02)	*AIF1* (−, 4e-03)	*HLA-DQB2* (+, 3e-06)	*LY6G5C* (−, 8e-03)
T_H17 cells	Schmiedel (2018)	89	*HLA-F-AS1* (+, 5e-05)	*CCHCR1* (+, 3e-05)	*DDX39B* (+, 2e-03)	*HLA-B* (−, 2e-05)	*FLOT1* (−, 3e-03)	*TNF* (+, 2e-02)	*HLA-DQB2* (−, 1e-03)	*HLA-DQB2* (+, 2e-07)	*HLA-DRB5* (+, 1e-04)
T_H2 cells	Schmiedel (2018)	89	*HLA-F-AS1* (+, 3e-04)	*CCHCR1* (+, 5e-05)	*DDX39B* (+, 1e-03)	*HLA-B* (−, 7e-05)	*HLA-B* (−, 2e-02)	*TNF* (+, 2e-03)	*C6orf47* (−, 5e-04)	*HLA-DQA2* (+, 9e-07)	*HLA-DRB5* (+, 1e-05)
T_{FH} cells	Schmiedel (2018)	89	*HLA-F-AS1* (+, 8e-08)	*CCHCR1* (+, 9e-05)	*DDX39B* (+, 2e-03)	*HLA-B* (−, 2e-05)	*SNHG32* (+, 1e-02)	*HLA-E* (+, 1e-03)	*PSMB8* (−, 3e-03)	*HLA-DQB2* (+, 8e-05)	*HLA-DRB5* (+, 5e-05)

Table S37. (Continued).

Tissue[a]	Study[b]	N	most significant cis-eQTL result (target gene, sign risk allele effect, p-value) for variant								
			rs1655901	rs2844626	rs72866766	rs137854633	rs2853998	rs2442752	rs6935999	rs4947340	rs9271529
Monocytes	BLUEPRINT	191	—	*CCHCR1* (+, 2e-13)	*MICA* (−, 3e-11)	—	—	*HLA-B* (+, 9e-07)	—	—	—
Monocytes	CEDAR	286	*HLA-A* (+, 4e-08)	*HLA-E* (+, 2e-03)	*HLA-C* (+, 6e-05)	*HLA-C* (−, 2e-08)	*HLA-C* (−, 7e-05)	*HLA-C* (+, 3e-11)	*CSNK2B* (−, 2e-02)	*HLA-DRB5* (−, 3e-08)	*HLA-DRB1* (+, 1e-18)
Monocytes (naïve)	Quach (2016)	200	*HLA-F* (−, 1e-03)	*HLA-C* (+, 1e-20)	*MICA* (−, 7e-10)	*HLA-B* (−, 3e-05)	*AL645933.2* (−, 2e-11)	*HLA-C* (−, 6e-10)	*AGER* (+, 2e-05)	*HLA-DQA2* (+, 7e-18)	*HLA-DRB5* (+, 7e-20)
Monocytes (naïve)	Fairfax (2014)	421	*HLA-A* (+, 4e-20)	*HCG27* (+, 1e-08)	*HLA-C* (+, 4e-04)	*HLA-C* (−, 2e-14)	*PBX2* (−, 1e-06)	*HLA-C* (+, 4e-05)	*HLA-DRB1* (+, 2e-04)	*HLA-DRB1* (−, 4e-12)	*HLA-DRB1* (+, 2e-26)
Monocytes ($CD14^{high}$ $CD16^{-}$)	Schmiedel (2018)	79	*HLA-F* (−, 6e-04)	*CCHCR1* (+, 5e-05)	*HSPA1B* (−, 4e-03)	*CSNK2B* (−, 1e-02)	*EHMT2* (−, 3e-02)	*GPANK1* (−, 4e-03)	*HLA-DQA1* (+, 2e-03)	*HLA-DQA2* (+, 3e-08)	*HLA-DRB5* (+, 2e-03)
Monocytes ($CD14^{-}$ $CD16^{+}$)	Schmiedel (2018)	90	*HLA-F-AS1* (+, 1e-03)	*PSORS1C3* (−, 2e-02)	*DDX39B* (+, 6e-03)	*HLA-B* (−, 2e-04)	*CSNK2B* (+, 8e-03)	*FLOT1* (−, 3e-03)	*HSPA1L* (−, 1e-03)	*HLA-DQA2* (+, 2e-08)	*HLA-DQA1* (+, 9e-05)
Monocytes (LPS stim.)	Quach (2016)	184	*HLA-F* (−, 3e-02)	*HLA-C* (+, 7e-15)	*MICA* (−, 2e-07)	*HLA-B* (−, 2e-05)	*AL645933.2* (−, 5e-08)	*HLA-C* (−, 1e-06)	*CSNK2B* (−, 4e-04)	*HLA-DQA1* (−, 3e-13)	*HLA-DRB5* (+, 2e-17)
Monocytes (Pam_3CSK_4 stim.)	Quach (2016)	196	*HLA-F* (−, 9e-03)	*HLA-C* (+, 3e-18)	*MICA* (−, 3e-07)	*HLA-B* (−, 3e-03)	*AL645933.2* (−, 3e-09)	*HLA-C* (−, 1e-06)	*AGER* (+, 5e-03)	*HLA-DQA1* (−, 2e-16)	*HLA-DRB5* (+, 3e-18)
Monocytes (R848 stim.)	Quach (2016)	191	*HLA-F* (−, 6e-04)	*HLA-C* (+, 7e-09)	*MICA* (−, 9e-08)	*HLA-B* (−, 2e-09)	*AL645933.2* (−, 2e-09)	*HLA-C* (−, 7e-06)	*AGER* (+, 5e-03)	*HLA-DQA2* (+, 8e-14)	*HLA-DRB5* (+, 7e-17)
Monocytes (IAV stim.)	Quach (2016)	198	*HLA-F* (−, 1e-03)	*HLA-C* (+, 1e-11)	*MICA* (−, 4e-09)	*AL671883.3* (+, 4e-09)	*AL645933.2* (−, 8e-09)	*HLA-C* (−, 1e-05)	*HLA-DOB* (−, 6e-04)	*HLA-DQA2* (+, 1e-22)	*HLA-DRB5* (+, 2e-19)
Monocytes (IFN24 stim.)	Fairfax (2014)	370	*HLA-A* (+, 1e-08)	*LST1* (+, 6e-05)	*HLA-C* (+, 4e-03)	*HLA-C* (−, 2e-10)	*HLA-C* (−, 1e-06)	*HLA-C* (+, 4e-05)	*STK19* (+, 3e-04)	*HLA-DRB5* (−, 1e-11)	*HLA-DRB1* (+, 1e-28)
Monocytes (LPS2 stim.)	Fairfax (2014)	256	*HLA-A* (+, 2e-22)	*VARS2* (−, 3e-04)	*ATP6V1G2* (−, 3e-03)	*HLA-C* (−, 2e-08)	*HLA-C* (−, 4e-07)	*HLA-C* (+, 9e-03)	*DDX39B* (−, 3e-03)	*HLA-DRB5* (−, 4e-08)	*HLA-DRB1* (+, 6e-22)
Monocytes (LPS24 stim.)	Fairfax (2014)	325	*HLA-A* (+, 8e-18)	*C2* (+, 1e-04)	*SNHG32* (−, 3e-03)	*VARS2* (+, 8e-09)	*HLA-C* (−, 5e-07)	*HLA-C* (+, 1e-04)	*HLA-DRB1* (+, 3e-03)	*HLA-DRB1* (−, 3e-11)	*HLA-DRB1* (+, 5e-20)

Table S37. (Continued).

Tissue[a]	Study[b]	N	most significant cis-eQTL result (target gene, sign risk allele effect, p-value) for variant								
			rs1655901	rs2844626	rs72866766	rs137854633	rs2853998	rs2442752	rs6935999	rs4947340	rs9271529
Macrophages (naïve)	Alasoo (2018)	84	*TRIM26* (+, 8e-03)	*RNF5* (−, 1e-03)	*MICA* (−, 1e-05)	*HLA-C* (−, 2e-04)	*RNF5* (−, 8e-05)	*RNF5* (+, 7e-03)	—	*HLA-DRB1* (−, 3e-05)	*SAPCD1-AS1* (−, 5e-03)
Macrophages (naïve)	Nedelec (2016)	163	*HLA-G* (−, 5e-04)	*HLA-C* (+, 4e-07)	*MICA* (−, 4e-05)	*MICB* (+, 3e-03)	*AL645933.2* (−, 3e-05)	*PSORS1C3* (+, 3e-04)	*RING1* (+, 8e-03)	*HLA-DQA2* (+, 1e-17)	*HLA-DRB5* (+, 2e-10)
Macrophages (IFNγ)	Alasoo (2018)	84	*AL645939.5* (+, 2e-03)	*AL645933.2* (−, 4e-03)	*PRR3* (+, 1e-04)	*HLA-B* (−, 6e-10)	*CCHCR1* (−, 3e-03)	*LTA* (−, 2e-04)	—	*HLA-DQA2* (+, 7e-11)	*PSMB9* (−, 6e-06)
Macrophages (Salmonella)	Alasoo (2018)	84	*RPP21* (+, 4e-03)	*PPP1R10* (−, 2e-03)	*MICA* (−, 2e-05)	*HLA-B* (−, 6e-07)	*C2* (+, 3e-02)	*IER3* (−, 5e-03)	—	*HLA-DRB1* (−, 3e-06)	*HLA-DQA1* (−, 8e-10)
Macrophages (IFNg+Salmonella)	Alasoo (2018)	84	*TRIM26* (+, 6e-04)	*CSNK2B* (+, 5e-03)	*HSPA1B* (−, 2e-03)	*HLA-B* (−, 5e-08)	*HSPA1L* (−, 2e-02)	*RNF5* (+, 3e-03)	—	*HLA-DRB1* (−, 2e-08)	*PSMB9* (−, 6e-04)
Macrophages (Listeria)	Nedelec (2016)	163	*PPP1R11* (−, 3e-04)	*HLA-C* (+, 9e-06)	*MICA* (−, 6e-05)	*TNF* (−, 2e-03)	*AL645933.2* (−, 2e-06)	*DDX39B* (+, 5e-04)	*BAG6* (+, 7e-03)	*HLA-DQA2* (+, 1e-15)	*HLA-DRB5* (+, 3e-12)
Macrophages (Salmonella)	Nedelec (2016)	167	*HLA-G* (−, 2e-04)	*HLA-C* (+, 7e-08)	*MICA* (−, 3e-04)	*MICB* (+, 1e-03)	*AL645933.2* (−, 1e-05)	*HLA-C* (−, 3e-03)	*TAP2* (−, 1e-02)	*HLA-DQA2* (+, 4e-12)	*HLA-DRB5* (+, 1e-11)
Neutrophils ($CD66b^+$ $CD16^+$)	BLUEPRINT	196	—	*AL645933.2* (−, 2e-08)	*AL645933.2* (−, 5e-09)	—	—	*HLA-B* (+, 2e-12)	—	—	—
Neutrophils ($CD15^+$)	CEDAR	280	*HLA-A* (+, 2e-05)	*HCG27* (+, 4e-04)	*HLA-C* (+, 2e-06)	*LST1* (−, 1e-04)	*HCG27* (+, 3e-05)	*HLA-C* (+, 3e-08)	*AGPAT1* (+, 2e-03)	*HLA-DRB1* (−, 4e-05)	*HLA-DMA* (−, 1e-07)
Neutrophils ($CD16^+$)	Naranbhai (2015)	93	*HCG9* (+, 3e-10)	*PSORS1C3* (−, 7e-04)	*HLA-C* (+, 1e-02)	*GPANK1* (+, 8e-03)	*HSPA1A* (+, 4e-03)	*PSORS1C3* (+, 1e-03)	*HLA-DRB1* (+, 3e-03)	*HLA-DRB1* (−, 2e-03)	*HLA-DRB1* (+, 1e-04)

Abbreviations: IAV, influenza A virus; IFNγ, interferon gamma; LPS, lipopolysaccharide; LPS2, 2 h LPS stimulation; LPS24, 24 h LPS stimulation; N, number of samples; NS (FDR > 0.05 for GTEx; p > 0.05 for all other studies); Pam_3CSK_4, Pam3CysSerLys4; PHA, phytohemagglutinin; PsV, psoriasis vulgaris; R848, resiquimod.

[a]Tissues with relevance to psoriasis (skin, fibroblasts, whole blood, B cells, LCLs, NK cells, T cells, monocytes, macrophages, neutrophils) were selected.

[b]Results compiled from 16 different RNA expression studies, including PsV RNAseq,[17] GTEx,[18] the eQTLGen consortium,[19] as well as public datasets whose eQTLs were recomputed by the eQTL Catalogue,[20] including Alasoo *et al.*,[21] BLUEPRINT,[22] CEDAR,[23] GENCORD,[24] Fairfax *et al.*,[25,26] GEUVADIS,[27] Kasela *et al.*,[28] Lepik *et al.*,[29] Naranbhai *et al.*,[30] Nedelec *et al.*,[31] Quach *et al.*,[32] and TwinsUK.[33]

Table S38. Functional annotation of noncoding psoriasis-associated MHC variants with a Bayesian posterior probability exceeding 0.50.

				Conservation[a]				TF[b]		Chromatin state[c]					
Variant	Ancestry	PP	Nearest Gene (position)	PhastCons	PhyloP	Gerp++	SiPhy-π lod	No. bound TF proteins	No. altered TFBS motifs	Promoter histone marks	Enhancer histone marks	DNase HS sites	RegulomeDB rank[d]	RegulomeDB score[e]	CADD Phred[f]
rs1655901	EUR	1.000	*HLA-A* (3.1 kb downstream)	0.03	0.38	0.36	0	1	2	0	2	2	3a	0.55	12.3
rs2844626	SAS+EUR	0.541	*HLA-C* (7.0 kb downstream)	0.00	-0.66	0.20	0	0	6	0	0	0	1f	0.84	7.3
rs72866766	EUR	1.000	*HLA-B* (468 bp downstream)	0.00	-0.61	1.37	0	2	0	1	6	4	4	0.61	9.7
rs137854633	EUR	0.543	*HLA-B* (intron 1)	0.00	-0.03	0.62	6.6	10	0	24	1	16	3a	0.79	11.2
rs2853998	SAS+EUR	0.942	*HLA-B* (2.2 kb upstream)	0.00	-0.82	0.12	0	0	0	0	3	0	4	0.61	5.5
rs2442752	SAS	0.583	*AL671883.3* (7.9 kb downstream)	0.01	0.15	0.00	0	0	0	0	1	0	1f	0.22	8.5
rs6935999	EUR	0.752	*HLA-DRA* (15 kb upstream)	0.02	0.52	0.37	0	0	4	0	2	2	7	0.18	1.6
rs4947340	EUR	0.965	*HLA-DRA* (23 kb upstream)	0.54	0.12	0.00	0	0	1	0	0	0	5	0.59	11.0
rs9271539	SAS+EUR	0.562	*HLA-DQA1* (15.1 kb upstream)	0.00	-3.17	-5.27	0	0	2	1	5	3	2b	0.69	6.2

Abbreviations: HS, hypersensitivity; PP, posterior probability (Bayesian); TF, transcription factor; TFBS, transcription factor binding site

[a]PhastCons, PhyloP and Gerp++ evolutionary conservation scores based on tracks phastCons100way, phyloP100wayAll and allHg19RS_BW, respectively, from the UCSC table browser.[34] SiPhy-π lod scores were downloaded from supplemental information for Lindblad-Toh *et al.*[35] SiPhy-π lod scores with an FDR > 10% were considered non-significant and are represented with a 0 score.

[b]Number of bound transcription factors and number of altered transcription factor binding site motifs; based on v 4.1 of HaploReg.[36]

[c]Number of cell types with promoter histone marks, enhancer histone marks, or DNase hypersensitivity sites; based on v 4.1 of HaploReg.[36]

[d]Rank from v2.0 of RegulomeDB:[37] 1f = likely to affect binding and linked to expression of a gene target (eQTL + TF binding/DNase peak), 2b = likely to affect binding (TF binding + any motif + DNase footprint + DNase peak), 3a = less likely to affect binding (TF binding + any motif + DNase peak), 4 = minimal binding evidence (TF binding + Dnase peak), 5 = minimal binding evidence (TF binding or DNase peak), 6 = minimal binding evidence (Motif hit), 7 = uncurated region.

[e]Probability score from v 2.0 of Regulome DB,[37] ranging from 0–1, of the likelihood of being a regulatory variant.

[f]Weighted mean Phred score from v 1.6 of CADD,[38] an indicator of variant deleteriousness.

Supplemental methods

Additional methods for construction of SNP2HLA reference panels

We constructed an additional 18 SNP2HLA reference panels for imputing our South Asian GWAS samples by rebuilding existing HLA reference panels and by forming various combinations of the UM, IKMB-SAS and BKT components of the 397-person SAS panel with four other datasets—the non-SAS subset of the multiethnic IKMB reference panel, the European ancestry Type 1 Diabetes Genetics Consortium (T1DGC) SNP2HLA panel,[8-10] the pan-Asian SNP2HLA panel,[11,12] and data from phase 3 of the 1000 Genomes Project (1KGP).[13,14] HLA genotypes for 1KGP were combined separately with both microarray-based (v1) and sequence-based (v2) MHC genotypes for both the full 1KGP dataset (1KGP-ALL) and its SAS subset (1KGP-SAS), resulting in four versions of 1KGP data used for construction of South Asian panels: 1KGP-ALL-v1, 1KGP-SAS-v1, 1KGP-ALL-v2, 1KGP-SAS-v2 with 2666, 541, 2504 and 489 individuals, respectively. After removal of the 141 Indian samples used for our original South Asian panel, the IKMB panel dataset consisted of genotypes for 8,803 MHC SNPs and two-field genotypes of eight HLA genes for 1,217 people of European, East Asian, African-American and Iranian ancestry along with two additional Indian individuals. For the T1DGC panel we procured the datasets of genotypes for eight HLA genes and 5,868 MHC SNPs for 5,225 people that were used to construct the original panel;[9] for the pan-Asian panel we extracted genotypes for eight HLA genes and 6,173 MHC SNPs for 530 individuals of Chinese, Japanese, Tamil Indian or Malaysian ancestry from the published panel;[11,12] see Web Resources. Recently-published 2-field genotypes of five HLA genes (*HLA-A*, *-B*, *-C*, *-DQB1*, *-DRB1*) for 2,693 samples of phase 3 of the 1000 Genomes Project;[14] see Web Resources) were processed by downcoding ambiguous 2-field allele designations to their 1-field equivalent and setting genotypes with ambiguous 1-field

and 2-field designations to missing. Genotypes of SNPs in the MHC region were drawn from two sources—the reduced set of sequence-based integrated variant calls for version 5 of the phase 3 release and genotype data typed on the Affymetrix 6.0 microarray (Web Resources). Genotypes for all 2,124 variants in the chr6:29-34 Mb interval were extracted from the Affymetrix 6.0 microarray data, as were genotypes for all 63,106 SNPs with MAF ≥ 0.01 in this same interval that were present in the sequence-based integrated variant call set. We decided to use the much sparser microarray data in addition to the sequence-based calls because of high error rates for sequence-based SNP genotype calls in HLA genes that has been reported for phase 1 1000 Genomes data, which was caused by a strong mapping bias that overestimates reference allele frequencies.[39]

Three published HLA panels (IKMB, T1DGC, pan-Asian) were rebuilt from their component HLA and MHC genotypes with our updated MakeReference script and updated HLA SNP and amino acid sequence dictionaries. SNP2HLA panels for the four phase 3 1000 Genome datasets we created (1KGP-ALL-v1, 1KGP-SAS-v1, 1KGP-ALL-v2, 1KGP-SAS-v2) were also constructed with our updated method. Finally, we merged MHC genotypes for eleven combinations of the seven datasets previously described by identifying for each pair of datasets in a combination the set of shared MHC SNPs where we could confidently match strand orientation and where allele frequencies differed by no more than 0.15 or 0.30 between datasets of similar or dissimilar ancestries, respectively. 2-field HLA genotypes from datasets were combined by simple concatenation, with missing values for genes not typed in a particular dataset. SNP2HLA panels were then built for each combination by processing the merged MHC and HLA genotypes with the updated MakeReference method. By default, the MakeReference

script removes SNPs that violate Hardy-Weinberg equilibrium ($p < 1 \times 10^{-6}$); this quality control filter was imposed only for panels where all individuals have the same continental ancestry.

We also built 20 SNP2HLA reference panels for imputation of HLA variants in people of European ancestry by the same methods just described for South Asians. Datasets used for these panels consisted of many of those used for South Asians (T1DGC, UM, BKT, IKMB, 1KGP-ALL-v1, 1KGP-ALL-v2) as well as the EUR subset of phase 3 1KGP with HLA and either microarray or sequence-based MHC data (1KGP-EUR-v1 and 1KGP-EUR-v2 with 526 and 503 people, respectively).

For all SNP2HLA reference panels created for this study, we requested 10 burn-in iterations of the Beagle 4.0 phasing algorithm followed by 25 iterations of the more accurate phasing algorithm of Beagle 4.1.

Additional methods for validation of SNP2HLA reference panels

The relative performance of SNP2HLA reference panels for imputing into datasets of South Asian and European ancestry was assessed as follows. For each ancestry, panels were compared as two different groups—the full set of 19 South Asian or 20 European panels with genotypes for five HLA genes (*HLA-A*, *-B*, *-C*, *-DQB1*, *-DRB1*), and the subset of 11 South Asian or 7 European panels with genotypes for eight HLA genes (including *HLA-DQA1*, *HLA-DPA1* and *HLA-DQB1*). The 11 South Asian panels with genotyping for all eight HLA genes were assessed separately for their accuracy of *HLA-DQA1* imputation vs. their accuracy of HLA-DPA1 and HLA-DPB1 imputation because the method of *HLA-DQA1* genotyping for two of the component datasets of many of these panels (T1DGC and pan-Asian) could not discriminate several commonly occurring 2-field alleles from each other. Similar separate assessment of *HLA-DQA1* imputation accuracy was carried out for the 7 European panels with genotyping for all eight

genes, but because we had no independent *HLA-DPA1*/*DPB1* genotypes for Europeans we instead used the mean imputation performance for *HLA-A*, *-B*, *-C*, *-DQB1* and *-DRB1* as a proxy.

The relative performance of each panel within a set of panels being compared was assessed as its mean rank across 12 metrics consisting of three paired-sample comparison measures based on the Wilcoxon signed rank test (mean rank-biserial correlation, no. paired comparisons where rank sum of panel i > rank sum of panel j, no. paired comparisons where rank sum of panel i is significantly > rank sum of panel j) and three paired-sample comparison measures based on the paired t-test (mean paired difference, no. pairs where the mean difference of panel i minus panel j is > 0, no. pairs where mean difference of panel i – panel j is significantly > 0 based on bootstrapping); each of these six paired-sample measures was applied to both per-individual and per-allele imputation accuracies. For the SAS panels, which were evaluated with a single validation set, mean ranks for the 12 metrics were determined for 2-field imputed alleles of each of the eight HLA genes and then separately averaged across all genes within each of the three gene sets for which panels were tested *(HLA-A*,*B*,*C*,*DQB1*,*DRB1*; *HLA-DPA1*,*DPB1*; *HLA-DQA1*). For the EUR panels, mean ranks for the 12 metrics were first determined for each combination of the eight HLA genes and four validation sets, a validation set size weighted mean was then determined for each gene across the validation sets, and then unweighted mean ranks were computed across all genes in each of the three gene sets.

Additional methods for association analysis of MHC variants

Association testing was restricted to variants with a predicted r^2 imputation quality of at least 0.7 for all K case-control studies in an analysis. We identified candidate variants for the multiple independent psoriasis association signals in the MHC region using a stepwise regression approach that combines iterative forward selection of the best variant meeting a

significance threshold with backward elimination at each step of the worst included variant if it is no longer significant. We used the same cutoff probability for adding and removing variables, with the proviso that for inclusion the p-value of the variant must be ≤ cutoff probability and for elimination the p-value must be > cutoff probability. The specific cutoff probability for each stepwise analysis was based on the number of effectively independent variants tested, which was in turn estimated using the LD-pruning function of Plink 1.9 with an r^2 threshold of 0.3, as recommended by Sobota *et al*.[40] We estimated the number of effectively independent tested variants as 3499, 2305, and 1609 for the South Asian, European, and transethnic analyses, respectively, resulting in Bonferroni-adjusted cutoff probabilities of 1.4×10^{-5}, 2.2×10^{-5}, and 3.1×10^{-5}.

Most downstream analyses of the regression models fitted by this procedure made use of full model association statistics. For candidate variants in the final regression model these are simply the magnitude and variance of their β coefficients. Full model association statistics for all other variants constituting the broader association peak surrounding a candidate variant were derived by refitting the regression model after dropping that candidate variant but leaving all other candidate variants and covariates intact.

Selection and processing of imputed multiallelic variants

Before selecting multiallelic MHC variants, which are represented in the 1KGP and HRC panels by sets of biallelic splits for each of the alternate alleles, we first augmented them by computing and adding for each variant dosages of a biallelic split that treats the reference allele as the alternate allele. This augmentation served several purposes—all m biallelic splits of a variant with m alleles were needed to compute LD between it and other variants, testing all m biallelic splits of the multiallelic G1K and HRC panel variants individually in addition to jointly

testing their $m-1$ set provided a more comprehensive assessment of their association, and representing 1KGP and HRC multiallelic variants by sets of m rather than $m-1$ biallelic splits harmonized their treatment with that of multiallelic HLA variants in the SNP2HLA reference panels. The imputed dosage of the biallelic split for the reference allele was computed as $\max(0, 2-\sum_{a=1}^{m-1} x_{ia})$, where x_{ia} is the imputed dosage for biallelic split a of individual i and $m-1$ is the number of alternate alleles for the variant. Study-specific imputation quality (r^2) for each augmented biallelic split was estimated as the ratio of the observed sample variance of the computed dosages for that study to their expected variance based on observed allele frequencies and an assumption of Hardy-Weinberg equilibrium. For 1KGP multiallelic variants, the summation term for the dosage computation rarely (0.0065% of the time) produced negative dosages, and the mean and minimum of these negative dosages (−0.0018 and −0.081, respectively) were small enough to be explained by roundoff error. For HRC multiallelic variants, however, the summation for the computed dosages was frequently (5.92%) negative, with a very large mean and minimum for negative dosages (−0.594 and −3.963, respectively), indicating a serious issue with many of these variants in the HRC panel. In fact, we identified 249 problematic HRC multiallelic SNPs in the extended MHC region based on the criterion of its augmented biallelic split having at least one computed dosage < −0.10; these SNPs were removed from further consideration. After augmentation, from the extracted HRC-imputed datasets we selected only those multiallelic variants unique to the HRC panel; dosages for all other multiallelic variants (those unique to 1KGP and those shared by both panels) were selected from the 1KGP-imputed datasets.

Principal components analysis of South Asians

We performed principal components analysis (PCA) for 6,420 individuals of various South Asian ancestries drawn from four different sources: (i) 1,807 people of Indian ancestry collected in New Delhi with 902,747 genotyped autosomal variants from batch 1 of our psoriasis GWAS, (ii) 2,503 people of Pakistani and Indian ancestry with 501,166 genotyped autosomal variants from the combined batches 2 and 3 of the psoriasis GWAS, (iii) 1,621 Immunochip-typed North Indians with 187,115 genotyped autosomal variants that were part of a large meta-analysis of inflammatory bowel disease,[6] and (iv) 489 South Asians with 47,109,439 sequence-based autosomal variants from a reduced set (no monomorphic or singleton sites) of version 5 of the phase 3 release of the 1000 Genomes Project[13] (see Web Resources). Individuals and variants in the psoriasis GWAS and IBD Immunochip datasets had first passed all standard sample and variant quality control filters, including removal of population outliers. Datasets were merged successively, in the order given (i.e., the first two sets were merged, then this combined set with the third, etc.). At each step, genotypes for variants common to both datasets were combined provided they were biallelic autosomal SNPs with a difference in MAF of ≤ 0.15 (after strand flipping if necessary), with an additional stipulation that the mean MAF of A/T and C/G SNPs was no more than 0.40. 14,499 genotyped SNPs common to all 6,420 individuals remained after merging, which were further processed by removing variants in regions of known psoriasis susceptibility, with a MAF < 0.005, or within large chromosomal inversions, and then LD pruning to quasi-independence with the *-indep-pairwise* command of Plink 1.9 (parameters set to window size = 1500, step size = 150, and r^2 threshold = 0.20). PCA was then performed with Plink 1.9.

Phenotypic variance explained

For biallelic MHC variants detected by this study, we used a liability threshold model to compute separately for each variant the percent variance of liability for disease explained.[16] Relative risk was approximated by the log-additive OR from the final full regression model, frequency of the risk allele in the underlying population was estimated as a weighted average of its frequency in psoriasis cases and unaffected controls (weights = 0.015 and 0.985 for Europeans and 0.003 and 0.997 for South Asians, respectively), and the population prevalence of psoriasis in Europeans and South Asians was assumed to be 0.015 and 0.003 based on estimates from the Global Psoriasis Atlas;[41] see Web Resources. For multiallelic variants we applied a multiallelic extension of this biallelic method.[42]

MHC variant annotation

Basic functional annotation of variants in the MHC region was performed using the UCSC Variant Annotation Integrator tool[43] with the basic gene annotation set from GENCODE Version 31[44] that has been lifted over to hg19 reference assembly coordinates. The UCSC Table Browser[34] was used to extract PhastCons, PhyloP, and Gerp++ conservation scores, from hg19 reference assembly tracks phastCons100way, phyloP100wayAll, allHg19RS_BW, and cpgIslandExt, respectively. SiPhy-π lod scores, a measure of evolutionary constraint, were downloaded from the supplemental data of Linblad-Toh *et al.*[35] Version 4.1 of HaploReg[36] was used to determine how many cell types among the ENCODE and Roadmap reference epigenomes have promoter histone marks, enhancer histone marks, or DNase I hypersensitive sites overlapping a variant, as well as ChIP-seq evidence of transcription factors that bind to the interval containing the variant, and which transcription factor binding site motifs are changed by a variant. Version 2.0 of RegulomeDB[37] was used to obtain scores of how likely it is that a variant has regulatory function, and version 1.6 of CADD[38] provided scores of variant

deleteriousness. The overlap of associated variants with chromatin states of 33 cell types relevant to psoriasis, which were derived from a 15-state segmentation model based on five chromatin marks, was based on results of analyses generated by the Roadmap Epigenomics Project.[45] Results for cis-eQTLs were compiled from 16 RNA expression studies for 58 tissues and cells types with relevance to psoriasis, including our RNA-seq study of psoriatic and normal skin,[17] GTEx,[18] the eQTLGen Consortium,[19] as well as multiple public datasets whose eQTLs were recomputed by the eQTL Catalogue,[20] including BLUEPRINT,[22] CEDAR,[23] GENCORD,[24] TwinsUK[33] and several others.[21,25-32]

Enrichment analysis

Two MHC variant enrichment analyses were performed. The first compared counts of different types of MHC variants in the final full regression model for each ethnic dataset with the corresponding variant counts for the full set of MHC variants tested for association. The second compared counts of variant types within the MHC region to their counts in the whole genome based on all autosomal variants in release 5 of phase 3 of the 1000 Genomes Project;[13] this was done separately for all 2504 individuals in 1000 Genomes, the 489 South Asian ancestry individuals, and the 503 European ancestry individuals. For association-tested MHC variants, variant types were determined based on the nature of their alleles, and functional annotations were determined with the UCSC Variant Annotation Integrator. For phase 3 1000 Genomes variants, variant types were extracted from the VCF files for the full integrated variant call set, and functional annotations were determined from filtered annotation files produced using the Ensembl Variant Effect Predictor (see Web Resources).

Fold-enrichment for each variant type was computed as the ratio of the proportion of that type in the target set (final regression model or MHC region) to the proportion of that type

in the background set (association-tested variants or whole genome). The significance of enrichment was determined using an upper-tailed hypergeometric test.

Supplemental web resources

1000 Genomes, phase 3, Affymetrix 6.0 microarray data, http://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502/supporting/hd_genotype_chip/

1000 Genomes, phase 3, functional annotations, http://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502/supporting/functional_annotation/filtered/

1000 Genomes, phase 3, HLA genotypes, http://ftp.1000genomes.ebi.ac.uk/vol1/ftp/data_collections/HLA_types/

1000 Genomes, v5 phase3, reduced integrated variant call set (no monomorphic or singleton sites), http://csg.sph.umich.edu/abecasis/MACH/download/1000G.Phase3.v5.html

Beagle 4.1, https://faculty.washington.edu/browning/beagle/b4_1.html

CADD v1.6, https://cadd.gs.washington.edu/

eQTL Catalogue, https://www.ebi.ac.uk/eqtl/

eQTLGen Consortium, https://www.eqtlgen.org/

Global Psoriasis Atlas, https://globalpsoriasisatlas.org/statistics/prevalence?

GTEx Portal, https://www.gtexportal.org/home/

HaploReg v4.1, https://pubs.broadinstitute.org/mammals/haploreg/haploreg.php

ImmPort list of immune-related genes, https://www.immport.org/home

Pan-Asian SNP2HLA reference panel, http://software.broadinstitute.org/mpg/snp2hla/

Plink 1.9, https://www.cog-genomics.org/plink/

RegulomeDB 2.0, https://regulomedb.org/regulome-search/

Roadmap Epigenomics Project, core 15-state model, https://egg2.wustl.edu/roadmap/web_portal/chr_state_learning.html#core_15state

SiPhy-π scores, https://www.broadinstitute.org/mammals-models/29-mammals-project-supplementary-info

T1DGC SNP2HLA reference panel, https://repository.niddk.nih.gov/studies/t1dgc-special/

UCSC Table Browser, https://genome.ucsc.edu/cgi-bin/hgTables

Supplemental References

1. Luo, M., Kanai, M., Chow, W., Li, X., Yamamoto, K., Ogawa, K., Gutierrez-Arcelus, M., Gregersen, P.K., Stuart, P.E., Elder, J.T., et al. (2021). A high-resolution HLA reference panel capturing global population diversity enables multi-ethnic fine-mapping in HIV host response. Nature Genetics in press.
2. Das, S., Forer, L., Schonherr, S., Sidore, C., Locke, A.E., Kwong, A., Vrieze, S.I., Chew, E.Y., Levy, S., McGue, M., et al. (2016). Next-generation genotype imputation service and methods. Nat Genet 48, 1284-1287.
3. Gragert, L., Madbouly, A., Freeman, J., and Maiers, M. (2013). Six-locus high resolution HLA haplotype frequencies derived from mixed-resolution DNA typing for the entire US donor registry. Hum Immunol 74, 1313-1320.
4. Devlin, B., Roeder, K., and Wasserman, L. (2001). Genomic control, a new approach to genetic-based association studies. Theor Popul Biol 60, 155-166.
5. Marsh, S.G., Albert, E.D., Bodmer, W.F., Bontrop, R.E., Dupont, B., Erlich, H.A., Fernandez-Vina, M., Geraghty, D.E., Holdsworth, R., Hurley, C.K., et al. (2010). An update to HLA nomenclature, 2010. Bone Marrow Transplant 45, 846-848.
6. Liu, J.Z., van Sommeren, S., Huang, H., Ng, S.C., Alberts, R., Takahashi, A., Ripke, S., Lee, J.C., Jostins, L., Shah, T., et al. (2015). Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. Nat Genet 47, 979-986.
7. Degenhardt, F., Wendorff, M., Wittig, M., Ellinghaus, E., Datta, L.W., Schembri, J., Ng, S.C., Rosati, E., Hubenthal, M., Ellinghaus, D., et al. (2019). Construction and benchmarking of a multi-ethnic reference panel for the imputation of HLA class I and II alleles. Hum Mol Genet 28, 2078-2092.
8. Mychaleckyj, J.C., Noble, J.A., Moonsamy, P.V., Carlson, J.A., Varney, M.D., Post, J., Helmberg, W., Pierce, J.J., Bonella, P., Fear, A.L., et al. (2010). HLA genotyping in the international Type 1 Diabetes Genetics Consortium. Clin Trials 7, S75-87.
9. Jia, X., Han, B., Onengut-Gumuscu, S., Chen, W.M., Concannon, P.J., Rich, S.S., Raychaudhuri, S., and de Bakker, P.I. (2013). Imputing amino acid polymorphisms in human leukocyte antigens. PLoS One 8, e64683.
10. Onengut-Gumuscu, S., Chen, W.M., Burren, O., Cooper, N.J., Quinlan, A.R., Mychaleckyj, J.C., Farber, E., Bonnie, J.K., Szpak, M., Schofield, E., et al. (2015). Fine mapping of type 1 diabetes susceptibility loci and evidence for colocalization of causal variants with lymphoid gene enhancers. Nat Genet 47, 381-386.
11. Pillai, N.E., Okada, Y., Saw, W.Y., Ong, R.T., Wang, X., Tantoso, E., Xu, W., Peterson, T.A., Bielawny, T., Ali, M., et al. (2014). Predicting HLA alleles from high-resolution SNP data in three Southeast Asian populations. Hum Mol Genet 23, 4443-4451.
12. Okada, Y., Kim, K., Han, B., Pillai, N.E., Ong, R.T., Saw, W.Y., Luo, M., Jiang, L., Yin, J., Bang, S.Y., et al. (2014). Risk for ACPA-positive rheumatoid arthritis is driven by shared HLA amino acid polymorphisms in Asian and European populations. Hum Mol Genet 23, 6916-6926.
13. Genomes Project, C., Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A., et al. (2015). A global reference for human genetic variation. Nature 526, 68-74.
14. Abi-Rached, L., Gouret, P., Yeh, J.H., Di Cristofaro, J., Pontarotti, P., Picard, C., and Paganini, J. (2018). Immune diversity sheds light on missing variation in worldwide genetic diversity panels. PLoS One 13, e0206512.
15. Horton, R., Wilming, L., Rand, V., Lovering, R.C., Bruford, E.A., Khodiyar, V.K., Lush, M.J., Povey, S., Talbot, C.C., Jr., Wright, M.W., et al. (2004). Gene map of the extended human MHC. Nat Rev Genet 5, 889-899.
16. So, H.C., Gui, A.H., Cherny, S.S., and Sham, P.C. (2011). Evaluating the heritability explained by known susceptibility variants: a survey of ten complex diseases. Genet Epidemiol 35, 310-317.

17. Tsoi, L.C., Iyer, M.K., Stuart, P.E., Swindell, W.R., Gudjonsson, J.E., Tejasvi, T., Sarkar, M.K., Li, B., Ding, J., Voorhees, J.J., et al. (2015). Analysis of long non-coding RNAs highlights tissue-specific expression patterns and epigenetic profiles in normal and psoriatic skin. Genome Biol 16, 24.
18. Consortium, G.T., Laboratory, D.A., Coordinating Center -Analysis Working, G., Statistical Methods groups-Analysis Working, G., Enhancing, G.g., Fund, N.I.H.C., Nih/Nci, Nih/Nhgri, Nih/Nimh, Nih/Nida, et al. (2017). Genetic effects on gene expression across human tissues. Nature 550, 204-213.
19. Vosa, U., Claringbould, A., Westra, H.-J., Bonder, M.J., Deelen, P., Zeng, B., Kirsten, H., others], Visscher, P.M., Scholz, M., et al. (2018). Unraveling the polygenic architecture of complex traits using blood eQTL meta-analysis. bioRxiv, 1-57.
20. Kerimov, N., Hayhurst, j.D., Manning, J.R., Walter, P., Kolber, L., Peikova, K., Samovica, M., Burdett, T., Jupp, S., Parkinson, H., et al. (2020). eQTL Catalogue: a compendium of uniformly processed human gene expression and splicing eQTLs. bioRxiv https://doi.org/10.1101/2020.01.29.924266.
21. Alasoo, K., Rodrigues, J., Mukhopadhyay, S., Knights, A.J., Mann, A.L., Kundu, K., Consortium, H., Hale, C., Dougan, G., and Gaffney, D.J. (2018). Shared genetic effects on chromatin and gene expression indicate a role for enhancer priming in immune response. Nat Genet 50, 424-431.
22. Chen, L., Ge, B., Casale, F.P., Vasquez, L., Kwan, T., Garrido-Martin, D., Watt, S., Yan, Y., Kundu, K., Ecker, S., et al. (2016). Genetic Drivers of Epigenetic and Transcriptional Variation in Human Immune Cells. Cell 167, 1398-1414 e1324.
23. Momozawa, Y., Dmitrieva, J., Theatre, E., Deffontaine, V., Rahmouni, S., Charloteaux, B., Crins, F., Docampo, E., Elansary, M., Gori, A.S., et al. (2018). IBD risk loci are enriched in multigenic regulatory modules encompassing putative causative genes. Nat Commun 9, 2427.
24. Gutierrez-Arcelus, M., Lappalainen, T., Montgomery, S.B., Buil, A., Ongen, H., Yurovsky, A., Bryois, J., Giger, T., Romano, L., Planchon, A., et al. (2013). Passive and active DNA methylation and the interplay with genetic variation in gene regulation. Elife 2, e00523.
25. Fairfax, B.P., Makino, S., Radhakrishnan, J., Plant, K., Leslie, S., Dilthey, A., Ellis, P., Langford, C., Vannberg, F.O., and Knight, J.C. (2012). Genetics of gene expression in primary immune cells identifies cell type-specific master regulators and roles of HLA alleles. Nat Genet 44, 502-510.
26. Fairfax, B.P., Humburg, P., Makino, S., Naranbhai, V., Wong, D., Lau, E., Jostins, L., Plant, K., Andrews, R., McGee, C., et al. (2014). Innate immune activity conditions the effect of regulatory variants upon monocyte gene expression. Science 343, 1246949.
27. Lappalainen, T., Sammeth, M., Friedlander, M.R., t Hoen, P.A., Monlong, J., Rivas, M.A., Gonzalez-Porta, M., Kurbatova, N., Griebel, T., Ferreira, P.G., et al. (2013). Transcriptome and genome sequencing uncovers functional variation in humans. Nature 501, 506-511.
28. Kasela, S., Kisand, K., Tserel, L., Kaleviste, E., Remm, A., Fischer, K., Esko, T., Westra, H.J., Fairfax, B.P., Makino, S., et al. (2017). Pathogenic implications for autoimmune mechanisms derived by comparative eQTL analysis of CD4+ versus CD8+ T cells. PLoS Genet 13, e1006643.
29. Lepik, K., Annilo, T., Kukuskina, V., e, Q.C., Kisand, K., Kutalik, Z., Peterson, P., and Peterson, H. (2017). C-reactive protein upregulates the whole blood expression of CD59 - an integrative analysis. PLoS Comput Biol 13, e1005766.
30. Naranbhai, V., Fairfax, B.P., Makino, S., Humburg, P., Wong, D., Ng, E., Hill, A.V., and Knight, J.C. (2015). Genomic modulators of gene expression in human neutrophils. Nat Commun 6, 7545.
31. Nedelec, Y., Sanz, J., Baharian, G., Szpiech, Z.A., Pacis, A., Dumaine, A., Grenier, J.C., Freiman, A., Sams, A.J., Hebert, S., et al. (2016). Genetic Ancestry and Natural Selection Drive Population Differences in Immune Responses to Pathogens. Cell 167, 657-669 e621.
32. Quach, H., Rotival, M., Pothlichet, J., Loh, Y.E., Dannemann, M., Zidane, N., Laval, G., Patin, E., Harmant, C., Lopez, M., et al. (2016). Genetic Adaptation and Neandertal Admixture Shaped the Immune System of Human Populations. Cell 167, 643-656 e617.

33. Buil, A., Brown, A.A., Lappalainen, T., Vinuela, A., Davies, M.N., Zheng, H.F., Richards, J.B., Glass, D., Small, K.S., Durbin, R., et al. (2015). Gene-gene and gene-environment interactions detected by transcriptome sequence analysis in twins. Nat Genet 47, 88-91.
34. Karolchik, D., Hinrichs, A.S., Furey, T.S., Roskin, K.M., Sugnet, C.W., Haussler, D., and Kent, W.J. (2004). The UCSC Table Browser data retrieval tool. Nucleic Acids Res 32, D493-496.
35. Lindblad-Toh, K., Garber, M., Zuk, O., Lin, M.F., Parker, B.J., Washietl, S., Kheradpour, P., Ernst, J., Jordan, G., Mauceli, E., et al. (2011). A high-resolution map of human evolutionary constraint using 29 mammals. Nature 478, 476-482.
36. Ward, L.D., and Kellis, M. (2012). HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. Nucleic Acids Res 40, D930-934.
37. Boyle, A.P., Hong, E.L., Hariharan, M., Cheng, Y., Schaub, M.A., Kasowski, M., Karczewski, K.J., Park, J., Hitz, B.C., Weng, S., et al. (2012). Annotation of functional variation in personal genomes using RegulomeDB. Genome Res 22, 1790-1797.
38. Rentzsch, P., Witten, D., Cooper, G.M., Shendure, J., and Kircher, M. (2019). CADD: predicting the deleteriousness of variants throughout the human genome. Nucleic Acids Res 47, D886-D894.
39. Brandt, D.Y., Aguiar, V.R., Bitarello, B.D., Nunes, K., Goudet, J., and Meyer, D. (2015). Mapping Bias Overestimates Reference Allele Frequencies at the HLA Genes in the 1000 Genomes Project Phase I Data. G3 (Bethesda) 5, 931-941.
40. Sobota, R.S., Shriner, D., Kodaman, N., Goodloe, R., Zheng, W., Gao, Y.T., Edwards, T.L., Amos, C.I., and Williams, S.M. (2015). Addressing population-specific multiple testing burdens in genetic association studies. Ann Hum Genet 79, 136-147.
41. Griffiths, C.E.M., van der Walt, J.M., Ashcroft, D.M., Flohr, C., Naldi, L., Nijsten, T., and Augustin, M. (2017). The global state of psoriasis disease epidemiology: a workshop report. Br J Dermatol 177, e4-e7.
42. Okada, Y., Han, B., Tsoi, L.C., Stuart, P.E., Ellinghaus, E., Tejasvi, T., Chandran, V., Pellett, F., Pollock, R., Bowcock, A.M., et al. (2014). Fine mapping major histocompatibility complex associations in psoriasis and its clinical subtypes. Am J Hum Genet 95, 162-172.
43. Hinrichs, A.S., Raney, B.J., Speir, M.L., Rhead, B., Casper, J., Karolchik, D., Kuhn, R.M., Rosenbloom, K.R., Zweig, A.S., Haussler, D., et al. (2016). UCSC Data Integrator and Variant Annotation Integrator. Bioinformatics 32, 1430-1432.
44. Frankish, A., Diekhans, M., Ferreira, A.M., Johnson, R., Jungreis, I., Loveland, J., Mudge, J.M., Sisu, C., Wright, J., Armstrong, J., et al. (2019). GENCODE reference annotation for the human and mouse genomes. Nucleic Acids Res 47, D766-D773.
45. Roadmap Epigenomics, C., Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., et al. (2015). Integrative analysis of 111 reference human epigenomes. Nature 518, 317-330.