

# **A genome-wide association study suggests new susceptibility loci for primary antiphospholipid syndrome**

Desiré Casares-Marfil<sup>1</sup>, Manuel Martínez-Bueno<sup>2</sup>, Maria Orietta Borghi<sup>3</sup>, Guillermo Pons-Estel<sup>4</sup>, PRECISEADS Clinical Consortium, Guillermo Reales<sup>5</sup>, Yu Zuo<sup>6</sup>, Gerard Espinosa<sup>4</sup>, Timothy Radstake<sup>7</sup>, Lucas L. van den Hoogen<sup>8</sup>, Chris Wallace<sup>9</sup>, Joel Guthridge<sup>10</sup>, Judith A James<sup>10</sup>, Ricard Cervera<sup>4</sup>, Pier Luigi Meroni<sup>11</sup>, Javier Martin<sup>12</sup>, Jason S. Knight<sup>6</sup>, Marta E. Alarcón-Riquelme<sup>2\*</sup>, Amr H. Sawalha<sup>13\*</sup>

<sup>1</sup> Division of Rheumatology, Department of Pediatrics, University of Pittsburgh, Pittsburgh, USA;

<sup>2</sup> Genetics of Complex Diseases Group, Department of Genomic Medicine, GENYO, Centre for Genomics and Oncological Research: Pfizer/University of Granada/Andalusian Regional Government, Granada, Spain;

<sup>3</sup> Division of Rheumatology, Department of Clinical Sciences and Community Health, University of Milan, Milan, Italy; Immunorheumatology research laboratory – IRCCS Istituto Auxologico Italiano, Milan, Italy;

<sup>4</sup> Department of Autoimmune Diseases, Reference Centre for Systemic Autoimmune Diseases (UEC/CSUR) of the Catalan and Spanish Health Systems-Member of ERN-ReCONNECT, Hospital Clínic, Universitat de Barcelona, Barcelona, Catalonia, Spain;

<sup>5</sup> Cambridge Institute of Therapeutic Immunology & Infectious Disease, Jeffrey Cheah Biomedical Centre, Cambridge Biomedical Campus, University of Cambridge, Puddicombe Way, Cambridge CB2 0AW, UK; Department of Medicine, University of Cambridge School of Clinical Medicine, Cambridge Biomedical Campus, Cambridge CB2 0QQ, UK;

<sup>6</sup> Division of Rheumatology, Department of Internal Medicine, University of Michigan, Ann Arbor, Michigan, USA;

<sup>7</sup> Laboratory for Translational Immunology and Department of Pediatric Rheumatology and Immunology, University Medical Center Utrecht, Utrecht, the Netherlands; Department of Rheumatology and Clinical Immunology, University Medical Center Utrecht, Utrecht, the Netherlands;

<sup>8</sup> Department of Rheumatology, Radboudumc and Sint Maartenskliniek, Nijmegen, The Netherlands;

<sup>9</sup> Cambridge Institute of Therapeutic Immunology & Infectious Disease, Jeffrey Cheah Biomedical Centre, Cambridge Biomedical Campus, University of Cambridge, Puddicombe Way, Cambridge CB2 0AW, UK; Department of Medicine, University of Cambridge School of Clinical Medicine, Cambridge Biomedical Campus, Cambridge CB2 0QQ, UK; MRC Biostatistics Unit, University of Cambridge, Cambridge Institute of Public Health, Forvie Site, Robinson Way, Cambridge Biomedical Campus, Cambridge CB2 0SR, UK;

<sup>10</sup> Arthritis and Clinical Immunology Program, Oklahoma Medical Research Foundation, Oklahoma City, OK, USA;

<sup>11</sup> Immunorheumatology research laboratory – IRCCS Istituto Auxologico Italiano, Milan, Italy;

<sup>12</sup> Institute of Parasitology and Biomedicine "López- Neyra", CSIC, Granada, Spain;

<sup>13</sup> Departments of Pediatrics, Medicine, and Immunology & Lupus Center of Excellence, University of Pittsburgh, Pittsburgh, USA.

\*These two authors equality contributed to this work.

Please address correspondence to Amr H. Sawalha, MD. Address: 7123 Rangos Research Center, 4401 Penn Avenue, Pittsburgh, PA 15224, USA. Phone: (412) 692-8140. Fax: (412) 412-692-5054. Email: [asawalha@pitt.edu](mailto:asawalha@pitt.edu)

**Keywords:** Primary antiphospholipid syndrome, genetics, autoimmune diseases.

**Running Title:** Genome-wide association study in primary antiphospholipid syndrome

**Conflict of interest:** None of the authors has any financial conflict of interest to disclose. Dr. Chris Wallace is partially funded by GSK and MSD, and holds a part time position at GSK. Neither funder had any involvement with or influence on this work.

## ABSTRACT

**Objectives:** Primary antiphospholipid syndrome (PAPS) is a rare autoimmune disease characterized by the presence of antiphospholipid antibodies and the occurrence of thrombotic events and pregnancy complications. Our study aimed to identify novel genetic susceptibility loci associated with PAPS.

**Methods:** We performed a genome-wide association study comprising 5,485 individuals (482 affected individuals) of European ancestry. Significant and suggestive independent variants from a meta-analysis of approximately 7 million variants were evaluated for functional and biological process enrichment. The genetic risk variability for PAPS in different populations was also assessed. Hierarchical clustering, Mahalanobis distance, and Dirichlet Process Mixtures with uncertainty clustering methods were used to assess genetic similarities between PAPS and other immune-mediated diseases.

**Results:** We revealed genetic associations with PAPS in a regulatory locus within the HLA class II region near *HLA-DRA* and in *STAT4* with a genome-wide level of significance. 34 additional suggestive genetic susceptibility loci for PAPS were also identified. The disease risk allele in the HLA class II locus is associated with overexpression of *HLA-DRB6*, *HLA-DRB9*, *HLA-DPB2*, *HLA-DQA2* and *HLA-DQB2*, and is independent of the association between PAPS and *HLA-DRB1\*1302*. Functional analyses highlighted immune and nervous system related pathways in PAPS-associated loci. The comparison with other immune-mediated diseases revealed a close genetic relatedness to neuromyelitis optica, systemic sclerosis, and Sjögren's syndrome, suggesting colocalized causal variations close to *STAT4*, *TNPO3*, and *BLK*.

**Conclusions:** This study represents a comprehensive large-scale genetic analysis for PAPS and provides new insights into the genetic basis and pathophysiology of this rare disease.

# 1. INTRODUCTION

Antiphospholipid syndrome (APS) is a rare autoimmune disease characterized by the occurrence of thrombotic events and obstetrical complications in the presence of antiphospholipid antibodies (aPL) [1]. The presence of aPL in conjunction with vascular damage can lead to other complications in APS patients, such as neurological, cardiac, and skin involvement [2]. APS is more frequent in women and is estimated to affect 50 per 100,000 population, with a mean diagnosis age of 50 years [3-5].

APS was originally described in patients with an underlying systemic autoimmune disease, commonly lupus (secondary APS) [6]. Nevertheless, in almost half of the cases, APS appears as an isolated disorder known as primary APS (PAPS) [7]. Beyond the presence of an underlying autoimmune disorder, no major clinical differences have been reported between the primary and secondary forms of APS [8].

Previous studies suggested the relevance of genetic susceptibility in primary and secondary APS [3, 8]. Several genetic loci within the human leukocyte antigen (HLA) class II region have been reported to be associated with PAPS, including *HLA-DRB1\*1302* [3, 9]. Outside the HLA region, several loci associated with PAPS have been identified in candidate gene studies, including *STAT4*, *IRF5*, *MTOR*, and *PTPN22*. Despite these findings, genetic associations with PAPS outside of the HLA region have been inconsistent [3]. Further, no genome-wide association study (GWAS) has been previously performed in PAPS.

Herein, we perform a large GWAS and meta-analysis in PAPS including patients and controls from five different populations of European ancestry, and provide important new insights into the genetic basis and potentially the pathogenesis of PAPS.

## 2. MATERIALS AND METHODS

### 2.1. Study population and genotyping

Five independent case-control cohorts of European ancestry, with a total of 482 patients with PAPS and 5,003 controls, were studied. These include cohorts from Spain (90 PAPS-affected individuals and 1,517 unaffected controls), Italy (133 PAPS-affected individuals and 1,271 unaffected controls), Northern European and European-American (69 PAPS-affected individuals and 599 unaffected controls), European-American (90 PAPS-affected individuals and 1,047 unaffected controls) and an additional European cohort (100 PAPS-affected individuals and 569 unaffected controls). Genotyping data for the 1,047 control individuals included in the European-American cohort were derived from the database of Genotypes and Phenotypes (dbGaP, study accession: phs000187.v1.p1) [10]. All PAPS patients in this study met the Sydney classification criteria for primary APS [11]. Genomic DNA was extracted from blood samples by standard methods and genotyping data were generated via Illumina platforms (Illumina, San Diego, CA, USA) according to the manufacturer's instructions (**Table S1**). The study was approved by the Ethics Committees at all participating institutions. Protocols followed the principles of the Declaration of Helsinki, and all individuals included in the study signed written informed consents.

### 2.2. Quality controls and imputation

Quality controls (QCs) of genotype data were performed for each cohort separately using Plink v.1.9 [12]. Individuals with a genotyping call rate <95% and those related or duplicated according to their identity by descent proportion (PI\_HAT >0.4) were excluded from further analyses. Single-nucleotide polymorphisms (SNPs) with a genotyping call rate <98%, minor allele frequency (MAF) <1%, or those deviated from Hardy-Weinberg equilibrium (HWE) in cases and controls ( $p$ -value <  $1 \times 10^{-3}$ ) were removed. In addition, A/T-C/G SNPs with a MAF

close to 50% were removed from genotype data to avoid errors during the imputation. Filtered genotype data were imputed using the TOPMed Imputation Server with Minimac4 [13] and the TOPMed version R2 reference panel [14]. In the case of the European-American cohort, imputation of cases and controls was performed separately due to the low number of overlapping SNPs among the genotyping arrays. After imputation, SNPs with MAF <1%, HWE  $p\text{-value} < 1 \times 10^{-3}$ , or imputation quality metric RSQ <0.9 were filtered out. A principal component analysis (PCA) was performed to control for population stratification using Eigensoft 6.1.4 [15] (**Figure S1**). This software calculates principal components in each population by using around 100,000 independent SNPs and estimating individuals' standard deviations from cluster centroids. Thus, those individuals with >6 standard deviations were considered outliers and removed from further analyses. The genomic inflation factor ( $\lambda$ ) was calculated for each cohort using the R package "gap", and quantile-quantile (Q-Q) plots were generated. All plots were produced with R version 4.2.0.

Imputation of HLA classical alleles within the expanded HLA region (chr6: 20 to 40 Mb) was performed using the Michigan Imputation Server with Minimac4 [13], following the same procedure as for the genome-wide imputation in all the cohorts. HLA class I and II classical alleles were imputed for each cohort at 2 fields of resolution. Only classical alleles with an imputation quality metric RSQ >0.9 were used for further analyses.

### 2.3. Data analysis

Logistic regressions were conducted in each cohort by adjusting for the first five principal components using Plink v.1.9 software [12]. The summary-level statistics of each cohort were meta-analyzed using the inverse variance method. Thus, variants with no evidence of heterogeneity (Cochran's Q test  $p\text{-value} > 0.1$  and heterogeneity index <50%) were considered under a fixed-effects model, while in variants that showed heterogeneity of effects between cohorts (Cochran's Q test  $p\text{-value} \leq 0.1$  and heterogeneity index  $\geq 50\%$ ) a random-effects model

was applied. The genome-wide significance level was established at a p-value  $<5 \times 10^{-8}$  and the suggestive association threshold was set at a p-value  $<1 \times 10^{-5}$ . To identify independent significant associations, we performed dependency analyses at the population level followed by inverse variance meta-analysis within and outside the HLA region. For this, the most associated variant identified in the meta-analysis for each significant locus at the genome-wide level was used as a covariate to perform the conditional logistic regression in each population. Then, the results from the conditional logistic regressions for each population were meta-analyzed, and the variant showing the most significant association was considered an independent association. Using a similar approach, a pairwise conditional analysis was performed within the HLA region to test the independence of the SNP identified in the meta-analysis and the classical allele associated with PAPS (p-value  $<1 \times 10^{-5}$ ).

## 2.4. Functional annotation and enrichment analysis

*In silico* approaches were used to evaluate potential biological causality of significant and suggestive independent SNPs (linkage disequilibrium, LD  $r^2 < 0.4$ ). To assess their genomic location and possible functional implication the Open Targets Genetic [16, 17], VannoPortal [18], HaploReg v4.2 [19], Genotype-Tissue Expression (GTEx) project [20], and RegulomeDB [21, 22] databases, and the WashU Epigenome browser [23] were used. In order to assess enrichment in histone marks, the software GARFIELD was used to explore regulatory features from the Roadmap Epigenomics project [24]. Briefly, this function assesses functional annotation from GWAS summary statistics using several p-value thresholds (p-value  $\leq 1 \times 10^{-5}$ , p-value  $\leq 1 \times 10^{-4}$ , p-value  $\leq 1 \times 10^{-3}$ , p-value  $\leq 1 \times 10^{-2}$ ) and considering LD, minor allele frequencies, and distance to the nearest gene of each variant.

Finally, key pathways were identified by a Gene Ontology (GO) enrichment analysis using the closest genes to significant and suggestive independent signals and using the

Metascope web-based portal [25]. Those Biological Process ontology terms with p-value  $<1 \times 10^{-3}$  and at least 3 genes in the GO term were reported.

## 2.5. Cumulative genetic risk score

To assess the variability of genetic risk to PAPS across populations, a cumulative genetic risk score (GRS) was calculated. For this analysis, a total of 2,504 individuals from the 1000 Genomes Project phase 3 major populations (African, n=661; Admixed American, n=347; East Asian, n=504; European, n=503; South Asian, n=489) were included [26]. Independent significant and suggestive variants at each locus reported in our meta-analysis were included in calculating the GRS, analyzing a total of 40 independent genetic variants (**Table S2**). For the variant located in the *ESR2* locus, a SNP in complete LD ( $r^2 = 1$ ) was used due to the lack of genetic information for the most significant SNP in this locus in the 1000 Genomes Project data. For GRS estimation, individuals' genotypes were coded as 0, 1 or 2 indicating the number of PAPS risk alleles for each SNP. For each variant, the OR obtained from the meta-analysis was used. Thus, cumulative GRS was obtained by multiplying the natural logarithm of the OR for each SNP by the number of risk alleles for its corresponding SNP ( $\sum_{k=1}^i \ln(OR_i) n_i$ ) [27]. To determine whether the cumulative GRS was significantly different across populations, a Welch's t-test was applied.

## 2.6. Relationship between PAPS genetic susceptibility with other immune-mediated diseases

We evaluated the relationship of PAPS with other immune-mediated diseases (IMDs) from a genetic perspective. A low-dimensional summary of the shared genetic component of selected IMDs, excluding the HLA region, was created using a PCA strategy to generate an object known as "basis" that comprises the multidimensional components of the shared genetic risk [28]. PAPS GWAS data were projected on this "basis" occupying a location with respect to



the shared genetic architecture of the included IMDs. Based on this information, we estimated the closeness of PAPS to the rest of the IMDs by using a complete linkage hierarchical clustering method. However, the clustering algorithm can be unstable when considering multiple dimensions, so we also compared PAPS to IMDs' projections by using the Mahalanobis distance. This distance was calculated as  $(P_j - P_{PAPS})' S^{-1} (P_j - P_{PAPS})$  where  $P_j$  and  $P_{PAPS}$  represent the projected vectors of trait  $j$  and PAPS, respectively, and  $S$  represents the correlation derived from the LD among the SNPs in the principal components. When projections from GWAS data are used in clustering, the uncertainty associated with each observation can vary systematically with sample size, especially in studies with small sample sizes. To address this, we complemented the above clustering methods with the Dirichlet Process Mixtures with uncertainty (DPMUnc) approach [29], a Bayesian clustering method that takes the uncertainty associated with each data point into consideration.

With the purpose of identifying the genetic variation behind the shared signals between PAPS and other IMDs, we selected the SNPs that influence the projections of each significant PC in PAPS, referred to as driver SNPs. The probability of being significant in both PAPS and another IMD was estimated for each driver SNP using an FDR-based approach, considering a pairwise FDR <0.05 to be significant:

$$P(H_1 \text{ true for PAPS \& IMD} \mid p_{PAPS} < \alpha, p_{IMD} < \alpha) = (1 - FDR_{PAPS})(1 - FDR_{IMD})$$

$$\begin{aligned} \text{Pairwise FDR} &= P(H_0 \text{ true for PAPS or IMD} \mid p_{PAPS} < \alpha, p_{IMD} < \alpha) \\ &= 1 - (1 - FDR_{PAPS})(1 - FDR_{IMD}) \end{aligned}$$

Driver SNPs represent a small fraction of the genome and they may be useful in the identification of potential true causal SNPs in LD with them for the diseases evaluated. In addition, true casual variants in LD with each driver SNP may differ between diseases. To address this, we used the colocalization method [30] to formally investigate whether a putatively

causal variant is shared between PAPS and other IMDs within a specific locus. This method was applied on 2 Mb regions centered on each driver SNP with no overlapping among regions. Thus, driver SNPs and the variants within these regions were considered for colocalization. SNPs with the highest posterior probability of being the shared causal variant in each region were reported.

### 3. RESULTS

#### 3.1. GWAS results and meta-analysis

Following genotype quality control measures and imputation, genetic data from a total of 5,222 individuals (410 patients with PAPS and 4,812 controls) were included in our genetic analysis (**Table S3**). A logistic regression adjusting for the first five PCs was performed on each independent cohort (**Figure S2**), followed by a meta-analysis of the summary statistics for all cohorts (**Figure 1**). After correction for population structure, deviation in genomic inflation factors was not detected in any cohort (**Figure S3 and Table S3**). The results of the meta-analysis revealed two associated loci (represented by 34 variants) at the genome-wide level of significance ( $p\text{-value} < 5 \times 10^{-8}$ ) and 349 suggestive associated variants ( $p\text{-value} < 1 \times 10^{-5}$ ) (**Table S4**). The significant associations were located within the *STAT4* and the HLA region (**Table 1**). The most significant variant in *STAT4* corresponds to an intronic SNP (rs11889341,  $p\text{-value} = 1.39 \times 10^{-9}$ , OR [95% CI]=0.61 [0.52-0.71]; **Figure 2A**), showing a consistent effect in all five cohorts analyzed. Regarding the association within the HLA region, the most significant SNP is located near the transcription start site of the *HLA-DRA* gene (rs9269041,  $p\text{-value} = 2.07 \times 10^{-8}$ , OR [95% CI]=0.63 [0.54-0.74]; **Figure 2B**).

To assess independent significant associations with PAPS ( $p\text{-value} < 5 \times 10^{-8}$ ) in the *STAT4* and the HLA regions, we conducted stepwise conditional analyses in these two loci. As a result, the most significant SNPs from each locus, i.e. rs11889341 for *STAT4* and rs9269041 for the *HLA-DRA* region, explained the entire genetic association in these two regions (**Figure S4**).

Next, we imputed HLA classical alleles to determine their association with PAPS by using a specific reference panel for this region. *HLA-DRB1\*1302* ( $p\text{-value} = 1.12 \times 10^{-6}$ , OR=2.06) was the classical allele showing the most significant association with PAPS. The results of the

logistic regressions and meta-analysis of HLA classical alleles after conditioning by *HLA-DRB1\*1302* did not reveal any other suggestive associations ( $p\text{-value} < 5 \times 10^{-5}$ ). To determine if the signal detected in the *HLA-DRB1\*1302* classical allele is independent of the PAPS-associated SNP close to *HLA-DRA*, i.e. rs9269041, we performed a pairwise conditional analysis. The results of this analysis showed a persistent, albeit slightly attenuated, genetic effect in each signal after adjusting for the other, suggesting an independent effect of these two markers in PAPS (**Table 2**).

In addition to the genetic variants showing association with a genome-wide level of significance, a total of 349 SNPs in the meta-analysis surpassed the suggestive significance threshold ( $p\text{-value} < 1 \times 10^{-5}$ ) (**Table S4**). Including these, a total of 40 SNPs correspond to independent variants ( $r^2 < 0.4$ ) located within 36 loci that are associated with PAPS with either a genome-wide level or suggestive level of significance (**Table S5**). Notably, some of these suggestive loci have been associated with clinical traits or molecular pathways that might be manifested in patients with PAPS. *ESR2* (estrogen receptor 2) (rs932657402,  $p\text{-value} = 1.13 \times 10^{-6}$ , OR [95% CI]=0.13 [0.06-0.30]) and *HGF* (hepatocyte growth factor) (rs75427037,  $p\text{-value} = 1.57 \times 10^{-6}$ , OR [95% CI]=0.23 [0.12-0.42]), have been associated with thrombotic events [31, 32], while *EPHA7* (rs117489460,  $p\text{-value} = 4.50 \times 10^{-6}$ , OR [95% CI]=0.19 [0.09-0.38]) is expressed in vascular endothelial cells and shows higher expression levels during inflammation [33]. Other suggestive loci identified in our meta-analysis have been previously associated with other immune-mediated diseases, such as *WNT11* (Wnt family member 11) and *DLG2* (discs large MAGUK scaffold protein 2) with inflammatory bowel disease, *ANKRD50* (ankyrin repeat domain containing 50) with systemic lupus erythematosus, and *PLCL1* (phospholipase C like 1) gene expression with rheumatoid arthritis [34-37].

Logistic regression for each cohort also revealed several loci with suggestive evidence of association with PAPS ( $p\text{-value} < 1 \times 10^{-5}$ ) (**Figure S2**).

### 3.2. Functional annotations and epigenetic enrichment analysis

In line with other autoimmune diseases, most PAPS-associated variants are in non-coding genetic regions and may influence the disease through regulatory elements. Therefore, we performed a functional annotation of significant and suggestive independent PAPS-associated variants (**Table S5**) by using *in silico* functional approaches.

First, we assessed the regulatory potential of significant and suggestive disease-associated independent variants identified in our meta-analysis. The functional annotation of these SNPs indicated their co-localization with epigenetic marks, suggesting a potential contribution to chromatin states across multiple tissues and cell types (**Table S6**). We found that 37 of these 40 disease-associated SNPs may alter regulatory motifs, which might affect transcription factor (TF) binding. DNase hypersensitivity sites were observed in eight of these variants, three of them also in protein binding sites. When all the significant and suggestive variants from the meta-analysis were evaluated for enrichment in histone marks, three histone modifications showed a significant enrichment. These marks corresponded to the H3K4me3 (p-value=3.62x10<sup>-2</sup>), H3K4me2 (p-value=1.12x10<sup>-2</sup>), and H4K20me1 (p-value=3.05x10<sup>-2</sup>) in blood cells, specifically in GM12878 lymphoblastoid cells.

Given the effect on gene expression that epigenetic marks can exert, we evaluated the relationship of significant and suggestive independent variants located at these epigenetic marks with changes in gene expression by examining eQTL. A total of eight PAPS-associated variants were linked to gene expression variation of at least one gene in multiple tissues (**Table S7**). The most significant eQTLs were identified for the variant associated with PAPS near *HLA-DRA*. The disease risk in this SNP (rs9269041) was associated with increased expression of several HLA class II genes, including *HLA-DRB6*, *HLA-DRB9*, *HLA-DPB2*, *HLA-DQA2* and *HLA-DQB2* in whole blood, vascular tissue, and nervous tissue (**Figure 3**, and **Figure S5**). Experimental ChIP-seq publicly available data show an epigenetic mark of active enhancer

(H3K27ac) near rs9269041 in CD14<sup>+</sup> monocytes, further supporting a regulatory potential for this PAPS-associated genetic susceptibility locus (**Figure S6**).

Furthermore, two PAPS-associated SNPs alter the gene expression of their nearest gene, producing expression changes in *SGCD* and *METTL13* in several tissues, including cardiovascular system tissues and the skin (**Table S7**).

Chromatin conformation might alter transcriptional regulation through physical interactions with other regions of the genome. To examine the presence of physical interactions within the regions containing PAPS-associated variants with gene bodies and/or gene promoters, the VannoPortal and the Open Targets Genetics public databases were used. From the list of significant and suggestive independent PAPS-associated variants, we identified 25 SNPs establishing physical interactions with gene bodies in the GM12878 lymphoblastoid cell line, and 11 promoter interactions in several other cell types (**Table S8**). The majority of these variants interact with multiple genes, some of them with genes whose expression levels were affected (**Table S9**). Some of these interactions are noteworthy given their functional relevance, as they are supported by eQTL evidence of increased gene expression in the presence of the risk alleles. These interactions correspond to the PAPS-associated variants rs35930163 and rs9411271 with the *DNM3* and *GPSM1* genes, respectively (**Figure 4**). These functional annotations suggest these loci as potential target genes for PAPS.

### 3.4. Pathway analysis

To identify possible biological processes enriched in the loci identified in the meta-analysis, we performed a GO term enrichment analysis including the closest genes to the significant ( $p\text{-value} < 5 \times 10^{-8}$ ) and suggestive ( $p\text{-value} < 1 \times 10^{-5}$ ) independent variants associated with PAPS. Six different clusters of GO terms were identified to be enriched in the reported loci: establishment or maintenance of cell polarity ( $p\text{-value} = 1.42 \times 10^{-7}$ ), cell junction assembly ( $p\text{-value} = 1.42 \times 10^{-7}$ ), cell junction assembly ( $p\text{-value} = 1.42 \times 10^{-7}$ ), cell junction assembly ( $p\text{-value} = 1.42 \times 10^{-7}$ ), cell junction assembly ( $p\text{-value} = 1.42 \times 10^{-7}$ ), cell junction assembly ( $p\text{-value} = 1.42 \times 10^{-7}$ ).

value= $8.73 \times 10^{-7}$ ), negative regulation of cell growth (p-value= $7.39 \times 10^{-5}$ ), establishment of cell polarity (p-value= $5.31 \times 10^{-4}$ ), regulation of hydrolase activity (p-value= $2.12 \times 10^{-3}$ ), and regulation of leukocyte mediated immunity (p-value= $3.20 \times 10^{-3}$ ). Interestingly, PAPS-associated genes showed enrichment in biological processes related to the central nervous and the immune systems (**Table S10**).

### 3.5. Cumulative GRS

Significant and suggestive independent variants identified in the meta-analysis were used to evaluate the genetic risk for PAPS in 2,504 individuals from the five major populations included in the 1000 Genomes Project. The highest genetic risk score (GRS) for PAPS was observed in the East Asian and African populations, while the lowest genetic risk was detected in the European and South Asian populations (**Figure 5**). When mean GRS values were compared, all populations showed significant differences in the pairwise comparisons except for the Admixed American versus South Asian populations (Welch's t-test p-value=0.25).

### 3.6. Relationship of PAPS with other immune-mediated diseases

The evaluation of the common genetic variation among traits or diseases can be useful in the identification of shared pathways that may be involved in the pathophysiology of a disease. To assess whether PAPS susceptibility variants might be shared with other traits or diseases, we conducted a search for significant and suggestive independent variants identified in our meta-analysis in public PheWAS data. From the 40 independent SNPs associated with PAPS in the meta-analysis, 14 showed a significant association (p-value<0.05) in PheWAS data with immune-related traits (**Table S11**). The most significant associations were observed for SNPs located in the *HLA-DRA* and *STAT4* loci, while the diseases showing a higher number of associations with PAPS susceptibility loci were systemic lupus erythematosus and rheumatoid arthritis.

Comparing the results of genetic association studies has proven to be a useful tool in the identification of shared genetic variation among different diseases. To assess genetic similarities among PAPS and other IMDs, we projected PAPS summary statistics into a low-dimensional representation of the shared risk of selected IMDs distributed in 13 components [28]. We found PAPS differed significantly ( $FDR < 5\%$ ) from controls on three of the 13 components, corresponding to the autoimmune (PC1), systemic lupus erythematosus (PC2), and the eosinophilic (PC13) signals (**Figure 6A**). The hierarchical clustering of the projections showed that PAPS is closer to IgG+ neuromyelitis optica (NMO) (**Figure 6A**), an inflammatory condition that affects the central nervous system [38]. This result was confirmed by the Mahalanobis and DPMUnc distance methods, which also suggested a close relationship of PAPS genetics with Sjögren's syndrome, systemic sclerosis (SSc), juvenile idiopathic arthritis (JIA), CREST syndrome, and primary biliary cholangitis (PBC) (**Figure 6B** and **Figure 6C**).

Next, we assessed the variation underlying this shared genetic component. For this, 229 driver SNPs that influence the projections for which PAPS was significant, i.e., PC1, PC2, and PC13 signals, were selected. The probability of driver SNPs to be significant in PAPS-IMD pairs was estimated, selecting for this analysis those IMDs resulting from distance methods: CREST syndrome, SSc, Sjögren's syndrome, IgG+ NMO, PBC, JIA, dermatomyositis, Felty syndrome, and localized scleroderma (morphea). We identified 9 significant PAPS-IMDs pairs comprising 4 driver SNPs that are significant in PAPS and other diseases from the pairs (**Table S12**). To assess whether these driver SNPs are acting as causal variants in both diseases of each PAPS-IMD pair or are close to causal variants, we applied a colocalization approach [30] by considering the genetic variation within a 1 Mb window for each driver SNP. Since driver SNPs on chromosome 7 (**Table S12**) are less than 1 Mb distant, only the one with the lowest pairwise FDR was considered for colocalization analysis. The colocalization analysis identified the most significant association in the *STAT4* locus from the meta-analysis as a causal variant



in this region that is shared with JIA, PBC, and Sjögren's syndrome (**Table S13**). Other SNPs in *TNPO3* and close to *BLK* were also identified as shared causal variants shared with SSc (**Table S13**).

## 4. DISCUSSION

We report the results of a large genetic association study in PAPS including affected and unaffected individuals of European ancestry. Two associations with a GWAS level of significance were identified near *HLA-DRA* and *STAT4*, contributing to PAPS genetic susceptibility. Further, 34 additional genetic susceptibility loci were identified with a suggestive level of association with the disease.

The most significant genetic association identified in our meta-analysis corresponds to a variant (rs11889341) located in *STAT4*, with a consistent effect in the five cohorts analyzed. *STAT4* encodes a signal transducer and activator of transcription and has been previously associated with PAPS in a candidate gene study [7]. In addition to PAPS, this variant has also been associated with other autoimmune conditions, including systemic lupus erythematosus and rheumatoid arthritis, and displaying a risk effect consistent with the results for PAPS [39]. The *STAT4* variant rs7574865, which is more extensively studied in systemic lupus erythematosus and rheumatoid arthritis, is in high LD ( $r^2=0.98$ ) with the PAPS-associated variant rs11889341. Indeed, rs7574865 also showed a significant risk association with PAPS in our study ( $p\text{-value}=1.11\times 10^{-8}$ , OR=1.70).

We identified a genetic association between PAPS and the HLA class II region, which we localized to a genetic region near *HLA-DRA*. Transcription factor binding and eQTL analyses suggest that this effect alters the binding of multiple transcription regulators and the expression of several genes within the HLA region. The PAPS-associated risk allele in rs9269041, which tags the effect in *HLA-DRA*, is associated with overexpression of *HLA-DRB6*, *HLA-DRB9*, *HLA-DPB2*, *HLA-DQA2*, and *HLA-DQB2* in several tissues including whole blood and tissues relevant to PAPS pathogenesis such as the vascular and nervous systems. In addition, we identified an association between *HLA-DRB1\*1302* and PAPS, which appears to be independent of the association in the *HLA-DRA* locus. The association between *HLA-*

*DRB1\*1302* and APS in cohorts combining both primary and secondary APS patients has been previously described [3]. Our study replicates and confirms this genetic association in the primary form of the disease.

Our study suggested a genetic susceptibility locus in *ESR2*, which encodes estrogen receptor 2, in patients with PAPS. Polymorphisms in this gene have been previously associated with the risk of deep vein thrombosis [31, 40]. Our data also suggested the genetic association between *HGF* and PAPS. *HGF* encodes hepatocyte growth factor and has been previously reported to be increased in the plasma in thrombosis-associated disorders in human and murine models [32, 41].

eQTL annotation revealed that the risk alleles of two PAPS-associated genetic variants, rs35930163 and rs9411271, increase the gene expression levels of *DNM3* and *GPSM1* genes. Indeed, chromatin interactions were revealed between these two disease-associated variants and gene body regions of *DNM3* and *GPSM1*. *DNM3* encodes Dynamin 3, which is a member of a family of binding proteins that have been previously associated with coagulation effects and venous thromboembolism in the FinnGen public data [42]. *GPSM1* encodes the G Protein Signaling Modulator 1 that has been found to be related to alterations in cardiovascular function and regulatory mechanisms of the immune function in mouse models [43, 44].

We identified several biological processes enriched in PAPS-associated genetic susceptibility loci. The most significant gene ontologies correspond to the maintenance of cell polarity, cell junction organization, and cell adhesion processes. Previous studies have shown increased production of cell adhesion molecules in endothelial cells stimulated with antiphospholipid antibodies, which has been related to increased inflammation and thrombosis due to leukocyte binding to vascular walls [45, 46]. Interestingly, our enrichment analysis also highlighted leukocyte-mediated immunity, confirming the relevance of PAPS genetics in the immune system pathways. It is noteworthy that biological processes related to the regulation of

the nervous system development were enriched in PAPS-associated loci. A high percentage of PAPS patients develop cognitive impairment, which in some cases precedes the diagnosis of the disease [47].

The cumulative GRS for PAPS suggested significant differences among populations. Individuals from East Asia, followed by African and Admixed American populations, showed the highest GRS values. The clinical variability in PAPS makes it challenging to accurately estimate the incidence and prevalence of this disease in different populations [4]. Despite this, population-based studies report a higher prevalence of PAPS, as well as a higher mortality, in Asian populations [48]. It is noteworthy that possible variability in the strength of genetic associations at the detected loci between populations might affect the accuracy of interpreting these results. Therefore, further validation studies in different populations and ethnic groups will be necessary.

Our genetic similarity analysis comparing PAPS to other immune-mediated diseases revealed a genetic relatedness for PAPS with NMO and Sjögren's syndrome, both characterized by the development of neurological manifestations [49, 50]. Although the neurological involvement of PAPS is poorly understood, it has been primarily attributed to the thrombotic aspect of the disease [51]. A role for anti-NMDA antibodies in cognitive impairment in PAPS has also been suggested [52]. Our findings suggest that neurological manifestations in PAPS may be beyond the thrombotic aspect, reinforcing the concept of immune-mediated etiology in PAPS manifestations. In addition, non-thrombotic neurological manifestations have been described in APS [47]. The genetic similarity analysis also showed a relatedness between PAPS with SSc, specifically with limited cutaneous SSc (CREST), where endothelial cell dysfunction and vascular involvement are prominent [53, 54]. Indeed, endothelial dysfunction appears to play a role in the pathogenesis of PAPS [46, 55].

In summary, we performed a multi-center genome-wide association study in PAPS, and characterized genetic susceptibility loci associated with this rare disease. Our findings revealed novel genetic insights and identified new candidate genes associated with the pathogenesis of PAPS. Genetic similarity analysis suggested relatedness between PAPS and other immune-mediated diseases characterized by neurological and vascular involvement. Additional studies with larger sample sizes and diverse populations are needed to confirm, replicate, and extend our results. Subsequent studies with adequate power to examine genetic susceptibility to specific clinical subphenotypes and serological specificities within PAPS are also warranted.

**Competing interests:** None of the authors has any financial conflict of interest to disclose. Dr. Chris Wallace is partially funded by GSK and MSD, and holds a part time position at GSK. Neither funder had any involvement with or influence on this work.

**Contributorship:** - All authors fulfilled the following criteria:

- Substantial contributions to the conception or design of the work, or the acquisition, analysis or interpretation of data.
- Drafting the work or revising it critically for important intellectual content.
- Final approval of the version published.

**Acknowledgements:** The authors would like to thank participating members of PRECISESADS for their contributions to this work (Supplemental information). This work was supported by the use of study data downloaded from the dbGaP website, under dbGaP: phs000187.v1.p1. The authors also acknowledge the participants and investigators of the FinnGen study.

**Funding information:** This work was supported by the National Institute of Allergy and Infectious Diseases (NIAID) of the National Institutes of Health (NIH) grant number R01 AI097134 and the National Institute of Arthritis and Musculoskeletal and Skin Diseases (NIAMS) of the NIH grant number R01AR070148. This work also received support from the EU/EFPIA Innovative Medicines Initiative Joint Undertaking (PRECISESADS, grant number 115565) including in-kind contributions from the EFPIA members involved.

C.W. and G.R. are funded by the Wellcome Trust (WT220788). C.W. is funded by the Medical Research Council (MRC; MC UU 00002/4) and supported by the NIHR Cambridge BRC (BRC-1215-20014). The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care.

**Ethical approval information:** The study was approved by the Institutional Review Board of the University of Pittsburgh, Pittsburgh, Pennsylvania, USA.

**Data sharing statement:** All data are included in the manuscript or supplementary material.

**Patient and public involvement:** It was not appropriate or possible to involve patients or the public in the design, or conduct, or reporting, or dissemination plans of our research.

## REFERENCES

1. Hughes GR. The antiphospholipid syndrome: ten years on. *Lancet*. 1993 Aug 7; 342(8867):341-344.
2. Knight JS, Branch DW, Ortel TL. Antiphospholipid syndrome: advances in diagnosis, pathogenesis, and management. *BMJ*. 2023 Feb 27; 380:e069717.
3. Ortiz-Fernandez L, Sawalha AH. Genetics of Antiphospholipid Syndrome. *Curr Rheumatol Rep*. 2019 Dec 5; 21(12):65.
4. Dabit JY, Valenzuela-Almada MO, Vallejo-Ramos S, Duarte-Garcia A. Epidemiology of Antiphospholipid Syndrome in the General Population. *Curr Rheumatol Rep*. 2022 Jan 5; 23(12):85.
5. Truglia S, Capozzi A, Mancuso S, Manganelli V, Rapino L, Riitano G, et al. Relationship Between Gender Differences and Clinical Outcome in Patients With the Antiphospholipid Syndrome. *Front Immunol*. 2022; 13:932181.
6. Shoenfeld Y, Meroni PL, Toubi E. Antiphospholipid syndrome and systemic lupus erythematosus: are they separate entities or just clinical presentations on the same scale? *Curr Opin Rheumatol*. 2009 Sep; 21(5):495-500.
7. Yin H, Borghi MO, Delgado-Vega AM, Tincani A, Meroni PL, Alarcon-Riquelme ME. Association of STAT4 and BLK, but not BANK1 or IRF5, with primary antiphospholipid syndrome. *Arthritis Rheum*. 2009 Aug; 60(8):2468-2471.
8. Durrani OM, Gordon C, Murray PI. Primary anti-phospholipid antibody syndrome (APS): current concepts. *Surv Ophthalmol*. 2002 May-Jun; 47(3):215-238.
9. Caliz R, Atsumi T, Kondeatis E, Amengual O, Khamashta MA, Vaughan RW, et al. HLA class II gene polymorphisms in antiphospholipid syndrome: haplotype analysis in 83 Caucasoid patients. *Rheumatology (Oxford)*. 2001 Jan; 40(1):31-36.
10. Tryka KA, Hao L, Sturcke A, Jin Y, Wang ZY, Ziyabari L, et al. NCBI's Database of Genotypes and Phenotypes: dbGaP. *Nucleic Acids Res*. 2014 Jan; 42(Database issue):D975-979.
11. Miyakis S, Lockshin MD, Atsumi T, Branch DW, Brey RL, Cervera R, et al. International consensus statement on an update of the classification criteria for definite antiphospholipid syndrome (APS). *J Thromb Haemost*. 2006 Feb; 4(2):295-306.
12. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience*. 2015; 4:7.
13. Das S, Forer L, Schonherr S, Sidore C, Locke AE, Kwong A, et al. Next-generation genotype imputation service and methods. *Nat Genet*. 2016 Oct; 48(10):1284-1287.
14. Taliun D, Harris DN, Kessler MD, Carlson J, Szpiech ZA, Torres R, et al. Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature*. 2021 Feb; 590(7845):290-299.
15. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet*. 2006 Aug; 38(8):904-909.
16. Ghoussaini M, Mountjoy E, Carmona M, Peat G, Schmidt EM, Hercules A, et al. Open Targets Genetics: systematic identification of trait-associated genes using large-scale genetics and functional genomics. *Nucleic Acids Res*. 2021 Jan 8; 49(D1):D1311-D1320.



17. Mountjoy E, Schmidt EM, Carmona M, Schwartzentruber J, Peat G, Miranda A, et al. An open approach to systematically prioritize causal variants and genes at all published human GWAS trait-associated loci. *Nat Genet.* 2021 Nov; 53(11):1527-1533.
18. Huang D, Zhou Y, Yi X, Fan X, Wang J, Yao H, et al. VannoPortal: multiscale functional annotation of human genetic variants for interrogating molecular mechanism of traits and diseases. *Nucleic Acids Res.* 2022 Jan 7; 50(D1):D1408-D1416.
19. Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res.* 2012 Jan; 40(Database issue):D930-934.
20. Consortium GT. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science.* 2015 May 8; 348(6235):648-660.
21. Boyle AP, Hong EL, Hariharan M, Cheng Y, Schaub MA, Kasowski M, et al. Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res.* 2012 Sep; 22(9):1790-1797.
22. Dong S, Zhao N, Spragins E, Kagda MS, Li M, Assis P, et al. Annotating and prioritizing human non-coding variants with RegulomeDB v.2. *Nat Genet.* 2023 May; 55(5):724-726.
23. Li D, Hsu S, Purushotham D, Sears RL, Wang T. WashU Epigenome Browser update 2019. *Nucleic Acids Res.* 2019 Jul 2; 47(W1):W158-W165.
24. Iotchkova V, Ritchie GRS, Geihs M, Morganella S, Min JL, Walter K, et al. GARFIELD classifies disease-relevant genomic features through integration of functional annotations with association signals. *Nat Genet.* 2019 Feb; 51(2):343-353.
25. Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun.* 2019 Apr 3; 10(1):1523.
26. Genomes Project C, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, et al. A global reference for human genetic variation. *Nature.* 2015 Oct 1; 526(7571):68-74.
27. Hughes T, Adler A, Merrill JT, Kelly JA, Kaufman KM, Williams A, et al. Analysis of autosomal genes reveals gene-sex interactions and higher total genetic risk in men with systemic lupus erythematosus. *Ann Rheum Dis.* 2012 May; 71(5):694-699.
28. Burren OS, Reales G, Wong L, Bowes J, Lee JC, Barton A, et al. Genetic feature engineering enables characterisation of shared risk factors in immune-mediated diseases. *Genome Med.* 2020 Nov 25; 12(1):106.
29. Bayesian clustering with uncertain data. *bioRxiv.* 2022 2022-01-01 00:00:00.
30. Giambartolomei C, Vukcevic D, Schadt EE, Franke L, Hingorani AD, Wallace C, et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet.* 2014 May; 10(5):e1004383.
31. Alessio AM, Hoehr NF, Siqueira LH, Ozelo MC, de Padua Mansur A, Annichino-Bizzacchi JM. Association between estrogen receptor alpha and beta gene polymorphisms and deep vein thrombosis. *Thromb Res.* 2007; 120(5):639-645.
32. Kinoshita M, Miyamoto T, Ohashi N, Sasayama S, Matsumori A. Thrombosis increases circulatory hepatocyte growth factor by degranulation of mast cells. *Circulation.* 2002 Dec 10; 106(24):3133-3138.
33. De T, Alarcon C, Hernandez W, Liko I, Cavallari LH, Duarte JD, et al. Association of Genetic Variants With Warfarin-Associated Bleeding Among Patients of African Descent. *JAMA.* 2018 Oct 23; 320(16):1670-1677.
34. Shi J, Chi S, Xue J, Yang J, Li F, Liu X. Emerging Role and Therapeutic Implication of Wnt Signaling Pathways in Autoimmune Diseases. *J Immunol Res.* 2016; 2016:9392132.
35. Keane S, Herring M, Rolny P, Wettergren Y, Ejekkar K. Inflammation suppresses DLG2 expression decreasing inflammasome formation. *J Cancer Res Clin Oncol.* 2022 Sep; 148(9):2295-2311.

36. Delgado-Vega AM, Martinez-Bueno M, Oparina NY, Lopez Herraez D, Kristjansdottir H, Steinsson K, et al. Whole Exome Sequencing of Patients from Multicase Families with Systemic Lupus Erythematosus Identifies Multiple Rare Variants. *Sci Rep*. 2018 Jun 8; 8(1):8775.
37. Luo S, Li XF, Yang YL, Song B, Wu S, Niu XN, et al. PLCL1 regulates fibroblast-like synoviocytes inflammation via NLRP3 inflammasomes in rheumatoid arthritis. *Adv Rheumatol*. 2022 Jul 11; 62(1):25.
38. Jarius S, Paul F, Weinshenker BG, Levy M, Kim HJ, Wildemann B. Neuromyelitis optica. *Nat Rev Dis Primers*. 2020 Oct 22; 6(1):85.
39. Sollis E, Mosaku A, Abid A, Buniello A, Cerezo M, Gil L, et al. The NHGRI-EBI GWAS Catalog: knowledgebase and deposition resource. *Nucleic Acids Res*. 2023 Jan 6; 51(D1):D977-D985.
40. Farmer-Boatwright MK, Roubey RA. Venous thrombosis in the antiphospholipid syndrome. *Arterioscler Thromb Vasc Biol*. 2009 Mar; 29(3):321-325.
41. Matsumori A, Takano H, Obata JE, Takeda S, Tsuyuguchi N, Ono K, et al. Circulating hepatocyte growth factor as a diagnostic marker of thrombus formation in patients with cerebral infarction. *Circ J*. 2002 Feb; 66(2):216-218.
42. Kurki MI, Karjalainen J, Palta P, Sipila TP, Kristiansson K, Donner KM, et al. FinnGen provides genetic insights from a well-phenotyped isolated population. *Nature*. 2023 Jan; 613(7944):508-518.
43. Blumer JB, Lord K, Saunders TL, Pacchioni A, Black C, Lazartigues E, et al. Activator of G protein signaling 3 null mice: I. Unexpected alterations in metabolic and cardiovascular function. *Endocrinology*. 2008 Aug; 149(8):3842-3849.
44. Branham-O'Connor M, Robichaux WG, 3rd, Zhang XK, Cho H, Kehrl JH, Lanier SM, et al. Defective chemokine signal integration in leukocytes lacking activator of G protein signaling 3 (AGS3). *J Biol Chem*. 2014 Apr 11; 289(15):10738-10747.
45. Rand JH. Molecular pathogenesis of the antiphospholipid syndrome. *Circ Res*. 2002 Jan 11; 90(1):29-37.
46. Cugno M, Borghi MO, Lonati LM, Ghiadoni L, Gerosa M, Grossi C, et al. Patients with antiphospholipid syndrome display endothelial perturbation. *J Autoimmun*. 2010 Mar; 34(2):105-110.
47. Fleetwood T, Cantello R, Comi C. Antiphospholipid Syndrome and the Neurologist: From Pathogenesis to Therapy. *Front Neurol*. 2018; 9:1001.
48. Niznik S, Rapoport MJ, Avnery O, Ellis MH, Hajyahia S, Agmon-Levin N. Ethnicity and Antiphospholipid Syndrome in Israel. *Arthritis Care Res (Hoboken)*. 2022 Nov; 74(11):1917-1923.
49. Carnero Contentti E, Correale J. Neuromyelitis optica spectrum disorders: from pathophysiology to therapeutic strategies. *J Neuroinflammation*. 2021 Sep 16; 18(1):208.
50. Perzynska-Mazan J, Maslinska M, Gasik R. Neurological manifestations of primary Sjogren's syndrome. *Reumatologia*. 2018; 56(2):99-105.
51. Leal Rato M, Bandeira M, Romao VC, Aguiar de Sousa D. Neurologic Manifestations of the Antiphospholipid Syndrome - an Update. *Curr Neurol Neurosci Rep*. 2021 Jun 14; 21(8):41.
52. Gerosa M, Poletti B, Pregolato F, Castellino G, Lafronza A, Silani V, et al. Antigliutamate Receptor Antibodies and Cognitive Impairment in Primary Antiphospholipid Syndrome and Systemic Lupus Erythematosus. *Front Immunol*. 2016; 7:5.
53. Pattanaik D, Brown M, Postlethwaite AE. Vascular involvement in systemic sclerosis (scleroderma). *J Inflamm Res*. 2011; 4:105-125.
54. Di Martino ML, Frau A, Losa F, Muggianu E, Mura MN, Rotta G, et al. Role of circulating endothelial cells in assessing the severity of systemic sclerosis and predicting its clinical worsening. *Sci Rep*. 2021 Jan 29; 11(1):2681.
55. Velasquez M, Rojas M, Abrahams VM, Escudero C, Cadavid AP. Mechanisms of Endothelial Dysfunction in Antiphospholipid Syndrome: Association With Clinical Manifestations. *Front Physiol*. 2018; 9:1840.



## FIGURE LEGENDS

**Figure 1.** A Manhattan plot depicting the results of the meta-analysis among the five PAPS cohorts included in this study. Y and X axes refer to the  $-\log_{10}$  p-values and chromosome positions, respectively. The red horizontal line represents the genome-wide association threshold (p-value  $<5 \times 10^{-8}$ ) and the blue line represents the suggestive threshold (p-value  $<1 \times 10^{-5}$ ).

**Figure 2.** Regional plots for the genome-wide associated signals in chromosome 2 (**A**) and chromosome 6 (**B**), respectively. The y-axis corresponds to the  $-\log_{10}$  p-values of evaluated variants while the x-axis indicates variants' positions. Purple diamonds correspond to the most associated variants in each region. The rest of the variants are colored according to their degree of linkage disequilibrium (LD) with each top signal based on pairwise  $r^2$  values from the European population from the 1000 Genomes Project. The grey dashed line corresponds to the genome-wide association threshold (p-value  $<5 \times 10^{-8}$ ) and the blue line to the estimated recombination rates. Genes within the selected regions are shown at the bottom of the plots.

**Figure 3.** Violin plots representing the differences in gene expression levels in whole blood (y-axis) of the HLA genes *HLA-DRB6*, *HLA-DRB9*, *HLA-DPB2*, *HLA-DQA2* and *HLA-DQB2* depending on the genotypes of the PAPS-associated variant within the HLA region, rs9269041 (x-axis). When the PAPS-risk allele is present (rs9269041-A), higher gene expression levels are observed.

**Figure 4.** Functional annotation results for the variants (**A**) rs35930163 and (**B**) rs9411271 located close to the genes *DNM3* and *GPSM1*, respectively. Left panels indicate epigenetic mark binding at each variant. Central panels represent the physical chromatin interactions established by the regions containing each variant in the lymphoblastoid cell line GM12878; the red lines represent all the interactions of each region while the yellow ones refer to the

interaction with the indicated genes. The right panels show violin plots representing the differences in expression levels (y-axis) of the interacting genes depending on the genotypes of each PAPS-associated variant (x-axis).

**Figure 5.** Dot plot of the cumulative genetic risk score (GRS) for PAPS susceptibility across the major populations of the 1000 Genomes project.

**Figure 6.** Results of etiologic similarity analyses of PAPS with other immune-mediated diseases (IMDs). **(A)** Hierarchical clustering of PAPS with other immune-mediated diseases GWAS data and self-reported disease traits in UK Biobank, FinnGen Project, and The International Multiple Sclerosis Genetics Consortium. PAPS appears in boldface. Heatmaps indicate delta values for each disease on each component PC1-PC13, with grey indicating 0 (no difference from control), and darker shades of blue or magenta showing departure from controls in one direction or the other. A filled circle indicates that the delta value was significantly non-zero with a significance of FDR <1%, while an unfilled circle indicates a significance of FDR <5%. **(B)** Top 20 closest Mahalanobis distances to PAPS of all tested traits. The Y axis displays the distance between the center of the distribution (PAPS) and different points (the IMDs). **(C)** Clustering analysis of PAPS with IMDs using the Dirichlet Process Mixtures with uncertainty method (DPMUnc). Features in which PAPS was significant at FDR 5% level are depicted, comparing PC1 versus PC13 (top) and PC1 versus PC2 (bottom).

**Table 1. Meta-analysis results showing the most significant genetic variant in each locus associated with PAPS at a genome-wide level of significance (p-value<5x10<sup>-8</sup>)**

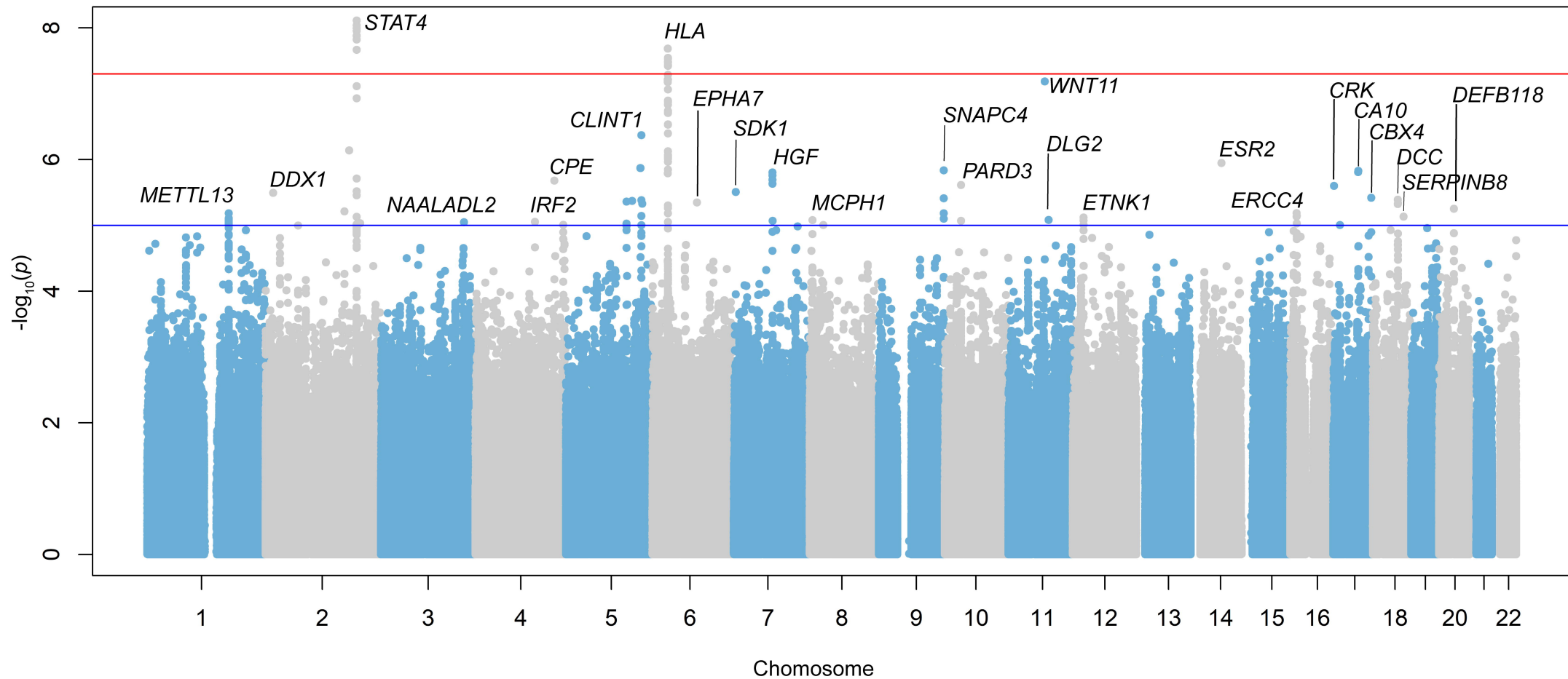
LOCUS	CHR	BP (HG38)	SNP	MINOR ALLELE	OR	95% CI	P-VALUE
<i>STAT4</i>	2	191079016	rs11889341	C	0.61	0.52-0.71	1.39x10 <sup>-9</sup>
<i>HLA-DRA</i>	6	32470466	rs9269041	G	0.63	0.54-0.74	2.07x10 <sup>-8</sup>

Abbreviations: Bp, base pair; chr chromosome; CI, confidence interval; OR, odds ratio; SNP, single-nucleotide polymorphism.

**Table 2. Pairwise conditional genetic analysis between rs9269041 and *HLA-DRB1\*13:02* in PAPS compared to healthy controls.**

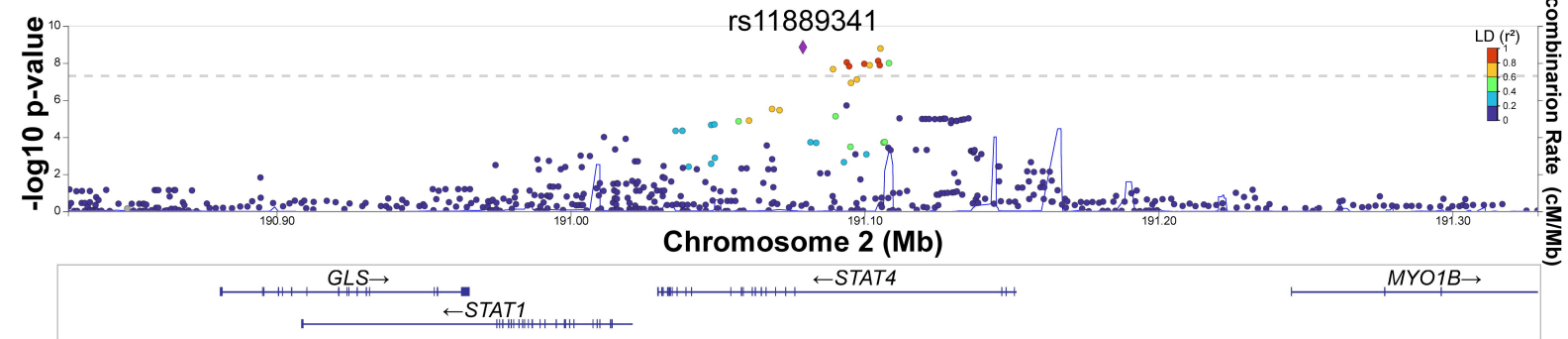
MARKER	COVARIATES	P-VALUE	OR
rs9269041	None	$2.07 \times 10^{-8}$	0.63
	<i>HLA-DRB1*13:02</i>	$9.10 \times 10^{-7}$	0.66
<i>HLA-DRB1*13:02</i>	None	$1.12 \times 10^{-6}$	2.06
	rs9269041	$1.76 \times 10^{-4}$	1.76

Abbreviations: OR, odds ratio.

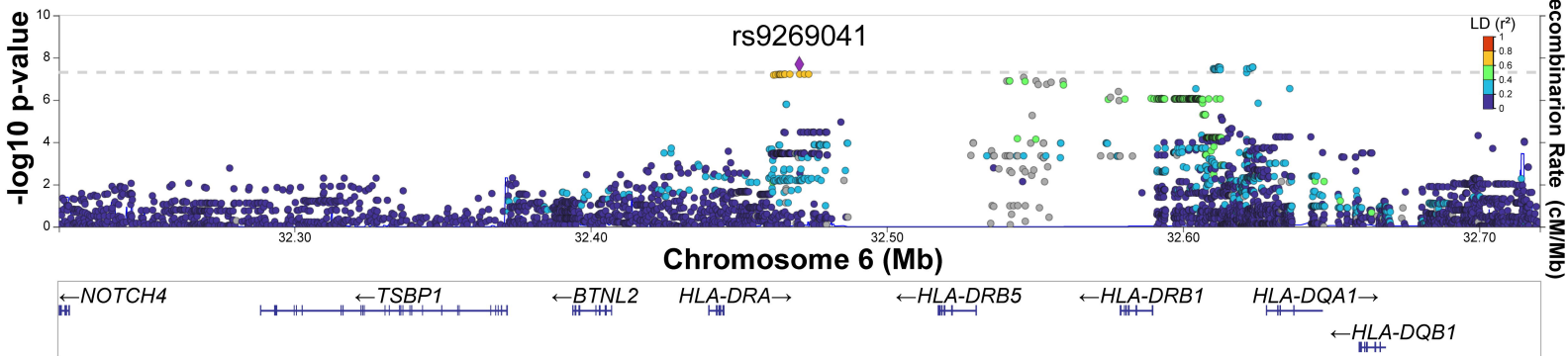




A

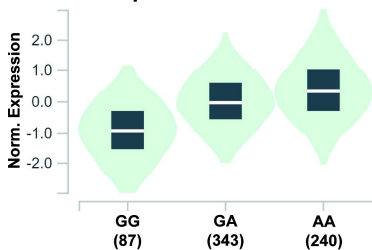


B



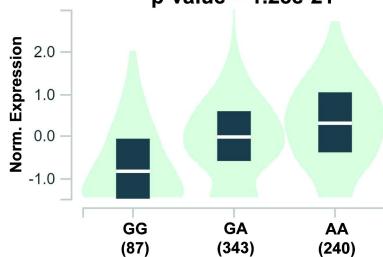
### ***HLA-DRB6***

p-value = 1.18e-42



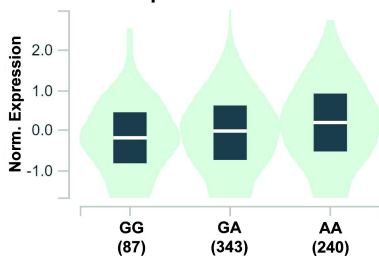
### ***HLA-DRB9***

p-value = 1.28e-21



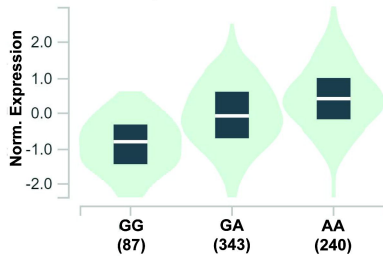
### ***HLA-DPB2***

p-value = 1.35e-4



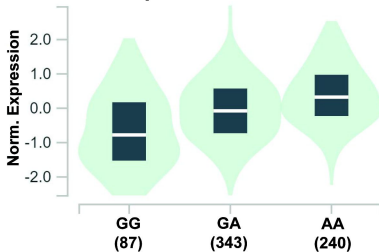
### ***HLA-DQA2***

p-value = 3.22e-36

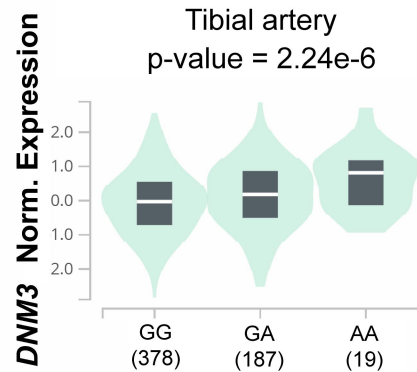
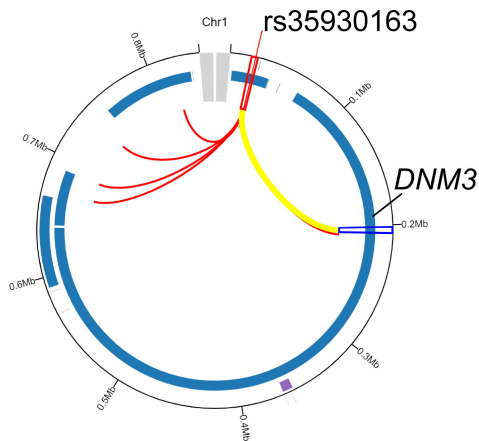
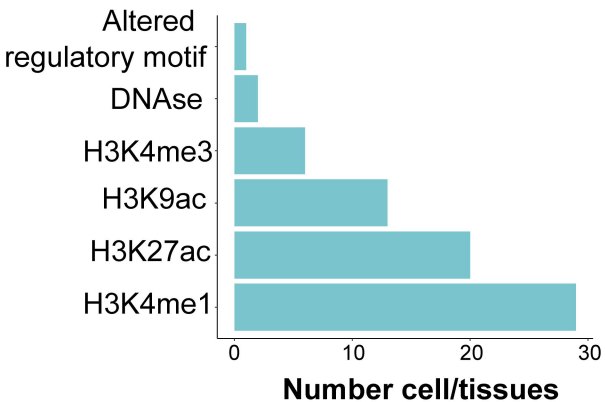


### ***HLA-DQB2***

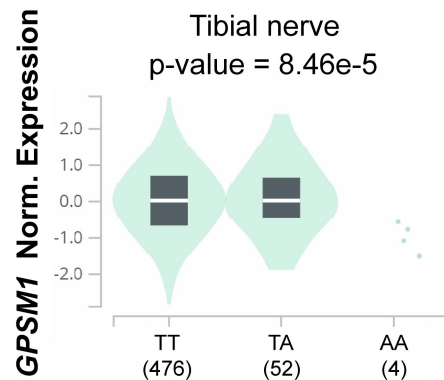
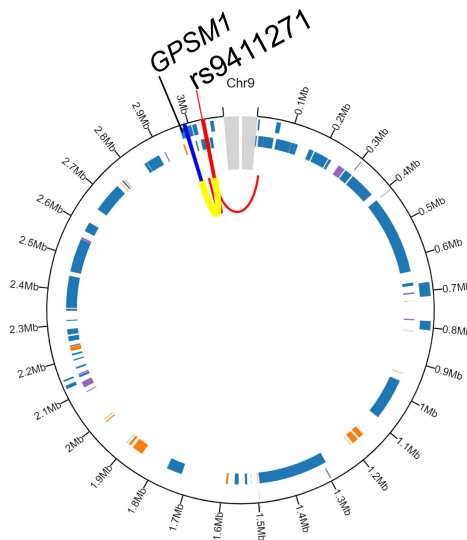
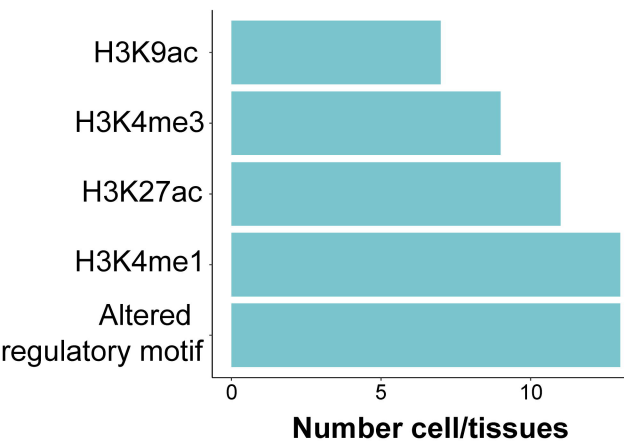
p-value = 5.97e-28

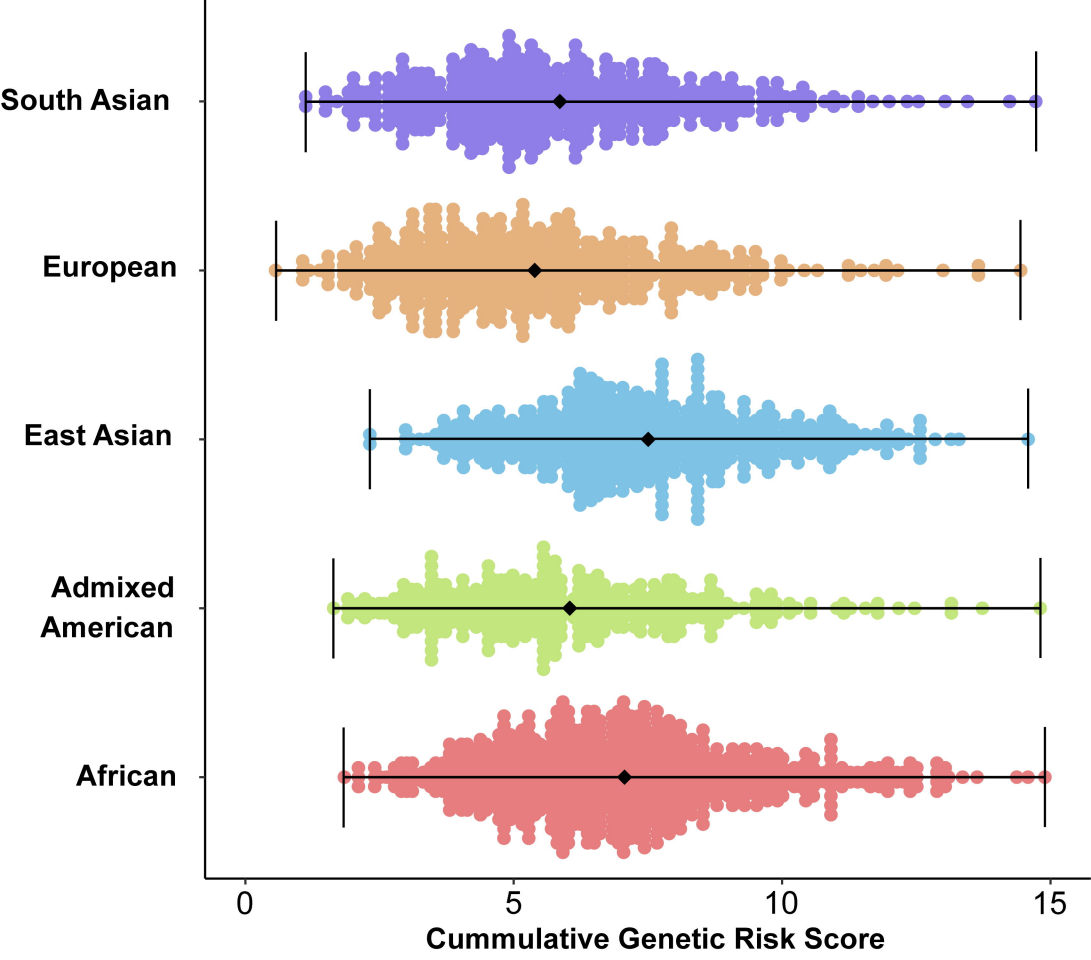


A



B





Heatmap A displays the correlation of 100 diseases with 13 principal components (PC1-PC13). The color scale ranges from -0.05 (blue) to 0.05 (red). Diseases are listed on the right, and PC labels are on the bottom. Dendrograms on the left and right show hierarchical clustering of diseases and components respectively.

Color scale: -0.05 (blue) to 0.05 (red).

Diseases (from top to bottom):

- Crohn's disease
- Inflammatory bowel disease
- Ulcerative colitis
- SLE
- Early-onset myasthenia gravis
- Rheumatoid arthritis
- Dermatopolymyositis
- Juvenile idiopathic arthritis
- Hyperthyroidism/Thyrotoxicosis
- Myasthenia gravis
- Autoimmune hyperthyroidism
- Autoimmune thyroid disease
- Hypothyroidism
- Felty syndrome
- IgG+ Neuromyelitis optica
- PAPS
- Nasal polyp
- Childhood-onset asthma
- Eosinophilic esophagitis
- ANCA negative EGPA
- MPO+ EGPA
- Palindromic rheumatism
- Type 1 diabetes
- Autoimmune Addison's disease
- Autoimmune thyroiditis
- Adrenocortical Insufficiency/Addison's disease
- Vitiligo
- Multiple sclerosis
- IgA deficiency
- Lichen planus
- Lichen sclerosus et atrophicus
- Graves disease
- Thyroiditis
- Certain disorders involving the immune mechanism
- Sarcoidosis
- Diabetic retinopathy
- Diabetic neuropathy
- Diabetic nephropathy
- Inflammation of eyelids
- Other and unspecified rosacea
- Rosacea
- Arthropathic psoriasis
- Psoriasis
- Peripheral atherosclerosis
- Type 2 diabetes
- Diseases of the blood and certain disorders involving the immune mechanism
- Gout
- Polymyalgia rheumatica
- Vasculitis
- Ankylosing spondylitis
- Atopic dermatitis
- Rhinitis-only vs. eczema-only
- Hay fever/Allergic rhinitis
- Adult-onset asthma
- Asthma
- Erythema nodosum
- Biological medication for rheuma
- Idiopathic thrombocytopenic purpura
- Unspecified disorders involving the immune mechanism not elsewhere classified
- Cholangitis (sclerosing)
- Glomerulonephritis
- Anaphylactic shock due to adverse effect of correcting drug/medicament properly administered
- Coeliac disease
- Juvenile rheuma
- Late-onset Myasthenia Gravis
- Localized scleroderma [morphea]
- Primary biliary cholangitis
- CREST syndrome
- Sjögren's syndrome
- Systemic sclerosis

Principal Components (PC1-PC13) are labeled on the bottom.

