

## **Cross-Phenotype GWAS Supports Shared Genetic Susceptibility to Systemic Sclerosis and Primary Biliary Cholangitis**

Yiming Luo<sup>1</sup>, Atlas Khan<sup>2</sup>, Lili Liu<sup>2</sup>, Cue Hyunkyu Lee<sup>3</sup>, Gabriel J Perreault<sup>4</sup>, Sydney F Pomenti<sup>4</sup>, Pravitt Gourh<sup>5</sup>, Krzysztof Kiryluk\*<sup>2</sup> and Elana J Bernstein\*<sup>1</sup>

<sup>1</sup>Division of Rheumatology, Department of Medicine, Columbia University Irving Medical Center, New York, NY

<sup>2</sup>Division of Nephrology, Department of Medicine, Columbia University Irving Medical Center, New York, NY

<sup>3</sup>Department of Biostatistics, Mailman School of Public Health, Columbia University Irving Medical Center, New York, NY

<sup>4</sup>Division of Digestive and Liver Diseases, Department of Medicine, Columbia University Irving Medical Center, New York, NY

<sup>5</sup>Scleroderma Genomics and Health Disparities Unit, National Institute of Arthritis and Musculoskeletal and Skin Diseases, National Institutes of Health, Bethesda, MD

\*These authors jointly supervised this work: Krzysztof Kiryluk and Elana Bernstein

Correspondence to Dr. Elana J Bernstein, Division of Rheumatology, Department of Medicine, Columbia University Irving Medical Center, New York, NY 10032, USA; [ejb2153@cumc.columbia.edu](mailto:ejb2153@cumc.columbia.edu)

## Key messages:

### What is already known on this topic

- Systemic sclerosis (SSc) and primary biliary cholangitis (PBC) are autoimmune disorders that exhibit overlapping clinical and histological features.
- The prevalence of PBC is higher in patients with SSc compared to the general population.
- Multiple susceptibility genomic loci have been identified for SSc and PBC through genome-wide association studies (GWAS).

### What this study adds

- There is a strong genetic correlation between SSc and PBC, comparable in magnitude to the genetic correlation between SSc and systemic lupus erythematosus (SLE). This shared genetic susceptibility aligns with the observed increased relative risk of developing PBC and SLE in individuals with SSc.
- Using cross-phenotype GWAS and colocalization analysis, we have discovered nine genomic loci that account for the shared genetic etiology. Five of the nine loci were novel.
- Using an integrative approach, we have prioritized five novel candidate causal genes: *CD40*, *ERAP1*, *PLD4*, *SPPL3* and *CCDC113*.
- The *CD40* risk allele for SSc and PBC is paradoxically associated with reduced CD40 levels. Causal inference analyses indicate that this reduction in CD40 levels, due to *CD40* locus polymorphism, leads to an increase in various plasma proteins involved in B cell activation, including the CD40 ligand.

### How this study might affect research, practice or policy

- Mechanistic studies are needed to confirm the candidate causal genes prioritized by our *in silico* analyses.
- Our study advocates for heightened awareness among rheumatologists regarding the possibility of concurrent PBC in patients with SSc.

## Abstract

### Objective

An increased risk of primary biliary cholangitis (PBC) has been reported in patients with systemic sclerosis (SSc). Our study aims to investigate the shared genetic susceptibility between the two disorders and to define candidate causal genes using cross-phenotype GWAS meta-analysis.

### Methods

We performed cross-phenotype GWAS meta-analysis and colocalization analysis for SSc and PBC. We performed both genome-wide and locus-based analysis, including tissue and pathway enrichment analyses, fine-mapping, colocalization analyses with expression quantitative trait loci (eQTL) and protein quantitative trait loci (pQTL) datasets, and phenome-wide association studies (PheWAS). Finally, we used an integrative approach to prioritize candidate causal genes from the novel loci.

### Results

We detected a strong genetic correlation between SSc and PBC ( $r_g = 0.84$ ,  $p = 1.7 \times 10^{-6}$ ). In the cross-phenotype GWAS meta-analysis, we identified 44 non-HLA loci that reached genome-wide significance ( $p < 5 \times 10^{-8}$ ). Evidence of shared causal variants between SSc and PBC was found for nine loci, five of which were novel. Integrating multiple sources of evidence, we prioritized *CD40*, *ERAP1*, *PLD4*, *SPPL3*, and *CCDC113* as novel candidate causal genes. The *CD40* risk locus colocalized with trans-pQTLs of multiple plasma proteins involved in B cell function.

### Conclusion

Our study supports a strong shared genetic susceptibility between SSc and PBC. Through cross-phenotype analyses, we have prioritized several novel candidate causal genes and pathways for these disorders.



## Introduction

Systemic sclerosis (SSc) is a multi-system autoimmune disease characterized by a complex interplay of fibrosis, vasculopathy, and inflammation. Unlike in other systemic autoimmune rheumatic diseases, the therapeutic response to immunosuppressive medications in SSc is organ-dependent. Certain organ involvement, including gastrointestinal tract fibrosis, has not been found to be responsive to immunosuppressive therapy.<sup>1</sup>

Primary biliary cholangitis (PBC) is an autoimmune liver disease characterized by inflammation of the intrahepatic bile ducts leading to liver fibrosis. Similar to the gastrointestinal involvement of SSc, the efficacy of immunosuppressive therapies in PBC has not yet been established.<sup>2</sup> The prevalence of PBC in patients with SSc is 2-2.5%, substantially higher than its prevalence of 0.4% in the general population.<sup>3,4</sup> Thus, there is likely an overlap of etiopathogenesis between SSc and PBC.

Genome-wide association studies (GWAS) have been conducted in both SSc and PBC, identifying numerous genomic loci associated with these two disorders.<sup>5,6</sup> Cross-phenotype GWAS analytic approaches that leverage existing GWAS summary statistics have emerged as a powerful new strategy for identifying shared mechanisms and novel risk loci.<sup>7</sup> We used this approach to systematically assess overlapping susceptibility and identify novel candidate causal genes that contribute to the common etiopathogenesis of the two disorders.

## Methods

### Study Design

An overview of the study design is shown in **Figure 1**.

### GWAS Summary Statistics

We obtained summary statistics of SSc and PBC from their recent GWAS meta-analyses. The GWAS for SSc was comprised of 26,679 individuals (9,095 cases and 17,584 controls),<sup>5</sup> while the GWAS for PBC was comprised of 24,510 individuals (8,021 cases and 16,489 controls).<sup>6</sup> For comparison, we also obtained GWAS summary statistics for rheumatoid arthritis (RA)<sup>8</sup> and systemic lupus erythematosus (SLE)<sup>8,9</sup> since both are prevalent in patients with SSc. We also obtained GWAS summary statistics from expression quantitative trait locus (eQTL) datasets, including the eQTLGen, Genotype-Tissue Expression (GTEx, for skin, liver, and lung) and Correlated Expression and Disease Association Research (CEDAR), as well as plasma proteomics quantitative trait locus (pQTL) datasets from the UK Biobank.<sup>10-13</sup> The GWAS summary statistics were harmonized using reference data from the 1000 Genome Project (phase 3) and underwent quality control with MungeSumstats.<sup>14,15</sup> The included GWAS datasets are summarized in **Table 1**.

### **Genetic Correlation Analysis**

To quantify the degree of shared genetic susceptibility, we used linkage disequilibrium score regression (LDSC) to estimate the global genetic correlation ( $r_g$ ), excluding the human leukocyte antigens (HLA) region, between each phenotype pair (SSc, PBC, RA, and SLE).<sup>16,17</sup> A Bonferroni-corrected p-value of  $8.3 \times 10^{-3}$  was used as the significance threshold.

### **Cross-Phenotype GWAS Meta-Analysis**

We performed a cross-phenotype GWAS meta-analysis to identify pleiotropic loci shared between SSc and PBC. We combined the summary statistics of SSc and PBC using the fixed-effect model with effect size estimates and standard errors using METAL.<sup>18</sup> Genomic control correction was applied to the summary statistics of each phenotype before the meta-analysis.<sup>19</sup> After the meta-analysis, we excluded SNPs in the HLA region or with evidence of heterogeneity (heterogeneity  $p < 0.05$ ). We used wANNOVAR to annotate the lead SNPs of the significant loci.<sup>20</sup> We defined novel loci as those that were significant in

the cross-phenotype meta-analysis but not significant in either SSc or PBC input GWAS. The loci were named based on the annotation of the lead SNPs from WANNVAR, which relies on the distance to nearby genes.<sup>20</sup>

The fixed-effect model is limited in examining SNPs with heterogeneity of effects. Therefore, we performed a sensitivity analysis using the PLEIO (Pleiotropic Locus Exploration and Interpretation using Optimal test) method.<sup>21</sup> PLEIO is designed for cross-phenotype meta-analysis and can account for heritability, genetic correlation, and sample overlap. There is no established method for extrapolating Z-scores from PLEIO statistics. Therefore, we used the fixed-effect model statistics for subsequent analyses.

### **Fine-Mapping and Credible Set Analyses**

We prioritized the most likely causal SNPs from the cross-phenotype meta-analysis statistics by calculating the 99% credible sets using CARMA (CAusal Robust Mapping method with Annotations).<sup>22</sup> These sets represent the smallest sets of SNPs with the probability of including the causal variant exceeding 99%. CARMA is a novel Bayesian model for fine-mapping that can better account for the uneven measurement of SNPs in each GWAS study of a meta-analysis, as well as the discrepancies between summary statistics and LD from external reference panels. We incorporated functional annotation into CARMA using the prior causal probabilities based on the meta-analysis of 15 UK Biobank traits from PolyFun (POLYgenic FUNctionally-informed fine-mapping).<sup>23</sup>

### **Tissue and Pathway Enrichment Analyses**

We used two methods to prioritize the pathways and tissues that contribute to the pleiotropy of SSc and PBC: MAGMA (Multi-marker Analysis of GenoMic Annotation) and DEPICT (Data-driven Expression Prioritized Integration for Complex Traits).<sup>24-26</sup> MAGMA and DEPICT perform enrichment analyses at pathway and tissue levels but use different approaches to associate loci with genes. MAGMA annotates SNPs based on their locations relative to genic regions (transcription start and stop sites +/- 10 kb



window). DEPICT prioritizes genes in a locus if genes in different loci have similar predicted functions. We reported the Bonferroni-corrected p values.

### **Colocalization between SSc and PBC**

We performed colocalization analyses between SSc and PBC in loci that are significant in the cross-phenotype meta-analysis with the fixed-effect model. Colocalization analyses infer the probability that a single genetic variant is causal to both traits of interest – SSc and PBC. We performed colocalization using the Wakefield's method from the R package “coloc”.<sup>27</sup> For loci with independent signals in the conditional analysis ( $p < 5 \times 10^{-8}$ ), we performed additional colocalization analyses using statistics from the conditional analyses. Evidence of colocalization was defined as at least one signal with a colocalization probability (PP4) above 70%.

### **Colocalization between Meta-Analysis Statistics and eQTL/pQTL Datasets**

We prioritized loci that met the following criteria for additional colocalization analyses: (1) lead SNPs without evidence of heterogeneity ( $P_{\text{het}} \geq 0.05$ ); (2) significant in the fixed-effect cross-phenotype meta-analysis; and (3) colocalized between SSc and PBC. We performed colocalization analyses between the SSc-PBC meta-analysis statistics of these loci and cis-eQTL and cis-pQTL datasets to predict transcripts or proteins associated with the genomic signal. For the eQTL colocalization analyses, we included relevant tissues and cells, including blood, skin, lung, liver, and immune cells. We first screened genes using SNPs in that locus's 99% credible set to query the eQTL databases. Next, we selected genes with significant eQTL signals for the colocalization analyses. For loci that colocalized with a cis-pQTL signal, we performed additional colocalization analyses between the meta-analysis statistics and trans-pQTLs measured at that locus to investigate the downstream effects of the candidate causal gene. A colocalization probability (PP4) above 70% was used as the significance threshold, and 50% as the suggestive threshold.

## **Phenome-wide association studies (PheWAS)**

We performed PheWAS for the lead SNP for each novel locus that colocalized between SSc and PBC. The PheWAS was performed within three biobanks: the Electronic Medical Records and Genomics III (eMERGE-III), All of Us, and the UK Biobank. Meta-PheWAS statistics were then calculated by fixed-effect meta-analysis of PheWAS results across the three biobanks.<sup>28-30</sup> The details of the genotyping methods, imputation, quality control, ancestry inference, covariate adjustment, and phenotype identification were described in our previous studies.<sup>31-33</sup>

To further explore pleiotropic associations, we performed meta-PheWAS analyses on the polygenic risk scores (PRS) of the cross-phenotype meta-analysis (PRS-meta-PheWAS). We used PRS-CS (continuous shrinkage), a method based on high-dimensional Bayesian regression, to generate the weights for PRS.<sup>34</sup> The HLA region was excluded. We set the Bonferroni-corrected statistical significance threshold for phenome-wide significance at  $2.75 \times 10^{-5}$  (0.05/1,817 phecodes tested). Lastly, we manually queried the top SNPs using PheWAS results from the Open Targets Genetics webpage<sup>35,36</sup>, which includes data from the GWAS Catalogue, UK Biobank, and FinnGen. We designated the effect allele as the GWAS risk allele in SSc-PBC cross-phenotype analysis.

## **Integrative prioritization of novel candidate causal genes**

For each novel locus, we prioritized candidate causal genes using an integrative approach. Each candidate causal gene was scored with the following criteria and we then calculated the number of the satisfied criteria (“priority score”): (1) genes most proximal to the lead SNP at the locus; (2) genes colocalizing with the locus in the examined cis-eQTL datasets; (3) genes colocalizing with the locus in the examined cis-pQTL datasets; (4) genes with a nonsynonymous coding variant in the credible set; (5) Genes prioritized by MAGMA (false discovery rate [FDR]  $q$ -value  $< 0.05$ )<sup>24</sup>; (6) genes prioritized by DEPICT (FDR  $q$ -value  $< 0.05$ )<sup>26</sup>; (7) genes receiving the top score from the Variant-to-Gene (V2G) pipeline

on the Open Targets Genetics webpage<sup>35,36</sup>, using the fine-mapped SNPs (PIP > 10%); (8) genes whose predicted regulatory elements from the ENCODE-rE2G model<sup>37</sup> intersected with the fine-mapped SNPs (PIP > 10%); (9) genes prioritized by the large language model (LLM) GPT-4, which was recently suggested as a systematic way to mine literature for candidate causal gene prioritization (input described in the **Supplemental Note**).<sup>38</sup> We prioritized the genes with the highest priority scores within each locus. Patients or the public were not involved in the design, or conduct, or reporting, or dissemination plans of our research.

## Results

### Global Genetic Correlation

There was a strong global genetic correlation between SSc and PBC ( $r_g = 0.84$ ,  $p = 1.7 \times 10^{-5}$ ), in which the effect estimate is comparable to the genetic correlation between SSc and SLE ( $r_g = 0.84$ ,  $p = 1.6 \times 10^{-15}$ ). The pairwise comparison of global genetic correlation in SSc, PBC, RA and SLE is shown in **Figure 2**.

### Cross-Phenotype GWAS Meta-Analysis

We performed a cross-phenotype meta-analysis for SSc and PBC using the fixed-effect model. The Manhattan plot is shown in **Figure 3**. There were 44 non-HLA loci that reached genome-wide significance ( $p < 5 \times 10^{-8}$ , **Supplemental Table 1**). The genomic inflation factor ( $\lambda$ ) was 1.065 and the LDSC intercept was 1.010 (standard error 0.012). In 16 out of the 44 significant loci (36%), there was evidence of heterogeneity in the lead SNPs ( $P_{het} < 0.05$ ). However, these loci remained significant after removing SNPs with heterogeneity.

Given the high proportion of loci with evidence of heterogeneity, we performed another cross-phenotype meta-analysis between SSc and PBC using PELIO as a sensitivity analysis. There were 58 non-HLA loci that reached genome-wide significance ( $p < 5 \times 10^{-8}$ , **Supplemental Table 2**). The genomic

inflation factor ( $\lambda$ ) was 1.091. Forty-one of the 44 loci (93%) that were significant in the fixed-effect model were also significant in PLEIO. Regarding the novel loci, five out of the seven (71%) identified in the fixed-effect model were also significant in PLEIO. The two novel loci that were only significant in the fixed-effect model, *CD40* and *AHNAK2*, had p-values of  $7.37 \times 10^{-8}$  and  $3.37 \times 10^{-7}$  in PLEIO. Two novel loci, *NDFIP1* and *PPHLN1*, were significant only in PLEIO but not in the fixed-effect model. However, the evidence for the association of these loci with SSc was insufficient (lead SNP p values  $3.93 \times 10^{-2}$  and  $1.33 \times 10^{-2}$ , respectively). Thus, no further analyses were performed for these loci.

### Tissue and Pathway Enrichment Analyses

We used two data-driven genome-wide methods, MAGMA and DEPICT, to explore pathway and tissue enrichment across SSc and PBC. Tissue enrichment analysis using MAGMA prioritized spleen ( $p = 1.21 \times 10^{-6}$ ), whole blood ( $p = 1.59 \times 10^{-6}$ ), Epstein-Barr Virus (EBV)-transformed lymphocytes ( $p = 4.16 \times 10^{-4}$ ), lung ( $p = 2.81 \times 10^{-3}$ ), and terminal ileum ( $p = 4.63 \times 10^{-3}$ ) (**Figure 4a, Supplemental Table 3**).

Concurrently, the DEPICT method significantly enriched 20 tissues and cells, most related to the immune system or respiratory tract. The top-ranked tissues/cells were mononuclear leukocytes ( $p = 2.95 \times 10^{-5}$ ), oropharynx ( $p = 3.80 \times 10^{-5}$ ), palatine tonsil ( $p = 3.80 \times 10^{-5}$ ), and synovial fluid ( $p = 2.63 \times 10^{-4}$ ) (**Figure 4b, Supplemental Table 4**).

Pathway enrichment analysis using MAGMA revealed significant enrichment in 73 gene sets (**Figure 4c, Supplemental Table 6**). The top-ranked gene sets were “Th1/Th2 pathway” ( $p = 4.92 \times 10^{-12}$ ), “positive regulation of immune system process” ( $p = 3.67 \times 10^{-9}$ ), “IL-27 pathway” ( $p = 5.60 \times 10^{-9}$ ) and “IL-35 pathway” ( $p = 7.63 \times 10^{-9}$ ). A separate pathway enrichment analysis using DEPICT revealed significant enrichment in 263 gene sets, which similarly involved multiple aspects of the human immune system (**Supplemental Data 1**).

### Fine-Mapping and Credible Set Analyses

To infer the causal SNPs underlying the association signals, we performed fine-mapping and generated 99% credible sets. There were 53 predicted causal signals across the 44 significant loci. Eleven (21%) of the credible sets contained only one SNP. Twenty-six (49%) credible sets contained fewer than five SNPs. Thirty-nine (74%) credible sets contained fewer than ten SNPs. In 30 (68%) loci, the maximum posterior inclusion probability (PIP) in their credible sets contained the lead SNPs of that loci. There were no nonsynonymous coding variants in any of the credible sets. The credible SNP sets in each locus are summarized in **Supplemental Table 5**.

### **Colocalization between SSc and PBC**

Due to LD, significant SNPs within genomic loci may not necessarily be causal for the associated trait. Therefore, we conducted colocalization analyses to determine whether there was at least one shared causal variant between SSc and PBC in loci significant in the cross-phenotype meta-analysis. We identified 9 loci that colocalized between SSc and PBC ( $PP_4 > 70\%$ ) and did not have evidence of heterogeneity ( $P_{\text{het}} \geq 0.05$ ), as detailed in the **Table 2** and **Supplemental Figure 1**. Notably, five were novel: *CSNK2A2/CCDC113*, *SPPL3*, *CAST/ERAP1*, *AHNAK2*, and *CD40*.

### **Colocalization with Tissue- and Immune Cell-Specific cis-eQTL**

In the nine loci that were significant in cross-phenotype meta-analysis without evidence of heterogeneity and colocalized with SSc and PBC, we further performed colocalization between SSc-PBC meta-analysis statistics and eQTL statistics in blood, skin, lung, liver and immune cells. In seven of the nine loci, the SSc-PBC meta-analysis statistics colocalized with at least one transcript in the examined eQTL statistics ( $PP_4 > 50\%$ ), prioritizing candidate causal genes at these loci (**Figure 5**). The eQTL colocalized transcripts were *IRF5*, *TNPO3*, *ANP32B*, *IL12RB1*, *ERAP1*, *ERAP2*, *SPPL3*, *AKT1*, *PLD4*, *LINC00638* and *CD40*.

### **Colocalization with plasma cis- and trans-pQTL**

Across the nine loci mentioned above, only the *IL12RB1* and *CD40* loci encode secreted proteins previously assessed in blood pQTL datasets. These plasma proteins include interleukin 12 receptor subunit beta 1 (IL12RB1) and CD40, respectively. The SSc-PBC meta-analysis statistics colocalized with cis-pQTL for CD40 protein levels (PP4 = 98%) but not with IL12RB1 protein levels (PP4 = 0.4%). The SSc-PBC risk allele was associated with lower plasma CD40 levels. Moreover, the SSc-PBC risk alleles at the CD40 locus colocalized in trans to reduced *BAFF* levels (PP4 = 99%) and increased levels of *CD40L* (PP4 = 99%), *FCER2* (PP4 = 99%), *CD22* (PP4 = 99%), *TRAF2* (PP4 = 97%), *FCLR1* (PP4 = 99%), and *TCL1A* (PP4 = 99%).

### PheWAS

We performed a meta-PheWAS for the five novel loci that colocalized between SSc and PBC. Three novel loci had significant associations with at least one phecode (**Supplemental Figure 3**). Rs10083496-G (*ANNAK2* locus) was associated with “Systemic lupus erythematosus” (OR = 1.12,  $p = 2.05 \times 10^{-7}$ ). Rs4810485-T (*CD40* locus) was associated with “Non-Hodgkins lymphoma” (OR = 1.11,  $p = 1.38 \times 10^{-6}$ ) and “Anxiety disorders” (OR = 1.04,  $p = 2.86 \times 10^{-6}$ ).

We subsequently performed an additional meta-PheWAS analysis to evaluate phenotypic associations of a genome-wide PRS based on the SSc-PBC meta-analysis statistics, excluding the HLA region (“SSc-PBC PRS”). We identified a total of 134 significant associations, the majority of which were related to immune dysregulation. As expected, SSc (OR = 1.78,  $p = 9.19 \times 10^{-30}$ ) and PBC (OR = 3.01,  $p = 5.55 \times 10^{-86}$ ) were top-ranked in their effect estimates associated with the SSc-PBC PRS, confirming that the PRS captures the risk for both diseases in the external datasets. Moreover, the PRS was associated with systemic autoimmune rheumatic diseases including “Rheumatoid arthritis” (OR = 1.24,  $p = 2.16 \times 10^{-55}$ ), “Systemic lupus erythematosus” (OR = 1.75,  $p = 3.34 \times 10^{-64}$ ), “Sicca syndrome” (OR = 1.58,  $p = 9.46 \times 10^{-43}$ ), “Polyarteritis nodosa and allied conditions” (OR = 1.18,  $p = 6.72 \times 10^{-10}$ ), as well as with non-rheumatic

autoimmune diseases including “Hypothyroidism” (OR = 1.12,  $p = 4.42 \times 10^{-55}$ ), “Multiple sclerosis” (OR = 1.30,  $p = 1.15 \times 10^{-23}$ ), “Inflammatory bowel disease and other gastroenteritis and colitis” (OR = 1.14,  $p = 3.23 \times 10^{-21}$ ) and “Idiopathic fibrosing alveolitis” (OR = 1.19,  $p = 2.00 \times 10^{-8}$ ) (**Supplemental Figure 5 and Supplemental Table 8**).

Additionally, we manually queried the top SNP from the novel loci using PheWAS results from the Open Targets Genetics webpage. Rs4810485-T (*CD40* locus) was associated with multiple autoimmune disorders, including “rheumatoid arthritis” (OR = 0.85,  $p = 5.7 \times 10^{-9}$ ), “inflammatory bowel disease” (OR = 1.08,  $p = 4.6 \times 10^{-10}$ ) and “multiple sclerosis” (OR = 1.08,  $p = 1.8 \times 10^{-5}$ ). Rs27524-G (*CAST/ERAP1* locus) was associated with “ankylosing spondylitis” (OR = 0.84,  $p = 5.4 \times 10^{-7}$ ), “iridocyclitis” (OR = 0.88,  $p = 6.9 \times 10^{-8}$ ) and “psoriasis” (OR = 0.85,  $p = 1.6 \times 10^{-6}$ ).

### **Integrative prioritization of novel candidate causal genes**

We prioritized five candidate causal genes from the novel loci using a scoring approach that integrates nine *in silico* annotation methods (**Figure 6**. MAGMA, DEPICT and ENCODE-rE2G results in the **Supplemental Table 9-11**. GPT-4 output described in the **Supplemental Note**). *CD40* received the highest priority score of 8, meeting all criteria except for credible sets containing a non-synonymous coding variant. *ERAP1* received a priority score of 6, followed by *PLD4* with a score of 5. *SPPL3* and *CCDC113* received priority scores of 4 and 3, respectively.

### **Discussion**

Our study demonstrates a strong genetic correlation between SSc and PBC, with the correlation effect estimate comparable to that between SSc and SLE. The prevalence of PBC in SSc was 2-2.5%, lower than the 8.4-14.7% prevalence of SLE in SSc.<sup>3,4,39-41</sup> Nevertheless, the prevalence of PBC in the general population, ranging from 19-402 per million persons, was also much lower than that of SLE, which ranges from 200-1500 per million persons.<sup>42-46</sup> Consequently, compared to the general population, the

relative risk (RR) of PBC and SLE in SSc likely mirrors their genetic correlation. Hence, our genetic correlation results, corroborated by the extent of phenotypic overlaps, support the existence of shared genetic susceptibilities and biological mechanisms between SSc and PBC.

We identified 44 significant non-HLA genomic loci in the fixed-model cross-phenotype GWAS meta-analysis. The robustness of our meta-analysis was supported by the PRS-meta-PheWAS analysis. In this analysis, using independent external datasets, SSc and PBC ranked among the top hits associated with the PRS, derived from our SSc-PBC meta-analysis statistics. Moreover, the SSc-PBC PRS demonstrated associations with a broad spectrum of autoimmune disorders, highlighting that the shared genetic susceptibilities between SSc and PBC captured by our cross-phenotype meta-analysis represent pleiotropic genomic regions.

We identified five novel loci that were significant in the cross-phenotype GWAS meta-analysis and colocalized between SSc and PBC: *CSNK2A2/CCDC113*, *SPPL3*, *CAST/ERAP1*, *AHNAK2*, and *CD40*. Three of the five loci were independently confirmed in a GWAS in another population: *CD40* and *AHNAK2* loci were significant in a recent SSc GWAS meta-analysis that included the Japanese population,<sup>47</sup> and the *CSNK2A2/CCDC113* locus was significant in a PBC GWAS from the Chinese population.<sup>48</sup> Moreover, we found that the lead SNP in the *CD40*, *AHNAK2*, and *CAST/ERAP1* loci was associated with other autoimmune disorders in PheWAS. The pleiotropic effects observed at these loci underscore their potential role in promoting autoimmunity. We subsequently prioritized five novel candidate causal genes for SSc and PBC based on integrating nine analytic approaches: *CD40*, *ERAP1*, *PLD4*, *SPPL3*, and *CCDC113*.

At the *CD40* locus, the SSc-PBC meta-analysis statistics colocalized with not only the reduced transcript but also reduced plasma protein levels of CD40. Such associations have also been observed in other autoimmune diseases, including inflammatory bowel disease and multiple sclerosis, as well as malignancy in our PheWAS analysis.<sup>49,50</sup> This seems paradoxical given CD40's established role in



promoting autoimmunity.<sup>51</sup> However, CD40 deficiency, a rare monogenic disorder caused by bi-allelic loss-of-function variants in *CD40*, is characterized not only by humoral immunodeficiency but also by malignancy and autoimmunity, including sclerosing cholangitis and colitis.<sup>52</sup> Thus, a causal relationship between reduced *CD40* expression due to polymorphisms and an increased risk of autoimmunity and malignancy is biologically plausible, mimicking the phenotypic manifestations of its monogenic disease counterpart. Interestingly, at the *CD40* locus, the SSc-PBC meta-analysis statistics also colocalized with increased levels of multiple plasma proteins involved in B cell functions, including *CD40L*, *FCER2*, *CD22*, *TRAF2*, and *TCL1A*. This suggests a potential compensatory response. Soluble CD40L (sCD40L), the circulating form measured in the proteomics assay, is the ligand of *CD40*, which also binds to other receptors on endothelial cells and promotes vascular pathology.<sup>53</sup> Elevated soluble CD40L levels have been found in patients with SSc and are associated with its vascular manifestations.<sup>54</sup> Overall, the complex B cell dysregulation mediating genetically determined reduced *CD40* expression in SSc warrants further investigation.

Our study also suggests that SSc and PBC may be associated with reduced major histocompatibility complex (MHC)-I-mediated immune response, potentially affecting cancer immunosurveillance. The SSc-PBC meta-analysis statistics in *CAST/ERAP1* and *SPPL3* loci colocalized with reduced transcripts of endoplasmic reticulum aminopeptidase 1 (ERAP1), increased transcripts of endoplasmic reticulum aminopeptidase 2 (ERAP2) and reduced transcripts of signal peptide peptidase-like 3 (SPPL3). These genes are implicated in the MHC-I-mediated immune response. ERAP1 and ERAP2 process and trim antigen peptides prior to their binding to MHC-I. Conditions referred to as “MHC-I-opathy”, including psoriasis and ankylosing spondylitis, were associated with *MHC-I* and *ERAP1* alleles.<sup>10,36,55-57</sup> In contrast, both SSc and PBC are associated with *MHC-II* and the risk alleles in *CAST/ERAP1* identified in our study are in the opposite direction of ERAP1 and ERAP2 expression compared to those found in the MHC-I-opathy entities.<sup>58,59</sup> Recent research has shown that reduced SPPL3 activity can dampen the MHC-I-

mediated response and promote tumor immune evasion.<sup>60</sup> Both SSc and PBC have been associated with an elevated risk of cancer, and recent findings indicate that SSc carries a higher burden of somatic mutations and genomic instability.<sup>61-64</sup> It is possible that these changes could be related to diminished immunosurveillance from a dampened MHC-I-mediated response.

Our study has several limitations. First, we used summary statistics from published GWAS meta-analyses and were unable to perform standardized quality control with individual-level genotype data.

Nevertheless, the GWAS studies included in our study were recent, comprised of large sample sizes from multiple cohorts, and led by international experts. Second, our study was performed in European ancestries; thus, our results may not be generalizable to other populations. Third, the sample size of SSc and PBC in our external datasets was limited, which precluded us from performing a replication GWAS to validate the newly discovered loci. Instead, we conducted a PRS-meta-PheWAS analysis to confirm that the PRS derived from our SSc-PBC meta-analysis captures the overall genetic risk of SSc and PBC in independent EHR-based datasets. Fourth, the sample size of GTEx and CEDAR was smaller than that of eQTLGen, which could limit statistical power in our eQTL colocalization analyses for the relevant tissues and cells. Finally, for our meta-PheWAS analyses, the diagnoses in the EHR-based datasets rely on administrative codes, which may have non-random missingness and low sensitivity for phenotype detection.<sup>65</sup>

In conclusion, our study revealed a strong genetic correlation between SSc and PBC and provided insights into their shared genetic susceptibility. We prioritized seven novel genes that were potentially involved in the common causal mechanisms between SSc and PBC. These discoveries prioritize therapeutic targets for both SSc and PBC. Moreover, our study advocates for heightened awareness among rheumatologists about the possibility of concurrent PBC in patients with SSc.

## Acknowledgements and affiliations:

Dr. Luo's work was supported by the Rheumatology Research Foundation Scientist Development Award.

Dr. Bernstein's work was supported by the National Institute of Arthritis and Musculoskeletal and Skin Diseases (grant K23-AR-075112), the National Heart, Lung, and Blood Institute (grant R01-HL-164758), and the Department of Defense (grant W81XWH2210163).

Dr. Liu was supported by the National Institute of Diabetes and Digestive and Kidney Diseases (grant 1K01DK137031).

The work was previously presented at the 2023 ACR Convergence: Luo Y, Khan A, Perreault G, Liu L, Lee C, Gourh P, Pomenti S, Kiryluk K, Bernstein E. Shared Genetic Susceptibility Between Systemic Sclerosis and Primary Biliary Cholangitis: Analyses from Genome-Wide Association Studies [abstract]. *Arthritis Rheumatol.* 2023; 75 (suppl 9).

## References

1. McMahan, Z.H., Kulkarni, S., Chen, J., Chen, J.Z., Xavier, R.J., Pasricha, P.J., and Khanna, D. (2023). Author Correction: Systemic sclerosis gastrointestinal dysmotility: risk factors, pathophysiology, diagnosis and management. *Nat Rev Rheumatol* 19, 191. 10.1038/s41584-023-00929-1.
2. Lindor, K.D., Bowlus, C.L., Boyer, J., Levy, C., and Mayo, M. (2019). Primary Biliary Cholangitis: 2018 Practice Guidance from the American Association for the Study of Liver Diseases. *Hepatology* 69, 394-419. 10.1002/hep.30145.
3. Rigamonti, C., Shand, L.M., Feudjo, M., Bunn, C.C., Black, C.M., Denton, C.P., and Burroughs, A.K. (2006). Clinical features and prognosis of primary biliary cirrhosis associated with systemic sclerosis. *Gut* 55, 388-394. 10.1136/gut.2005.075002.
4. Assassi, S., Fritzler, M.J., Arnett, F.C., Norman, G.L., Shah, K.R., Gourh, P., Manek, N., Perry, M., Ganesh, D., Rahbar, M.H., and Mayes, M.D. (2009). Primary biliary cirrhosis (PBC), PBC autoantibodies, and hepatic parameter abnormalities in a large population of systemic sclerosis patients. *J Rheumatol* 36, 2250-2256. 10.3899/jrheum.090340.
5. Lopez-Isac, E., Acosta-Herrera, M., Kerick, M., Assassi, S., Satpathy, A.T., Granja, J., Mumbach, M.R., Beretta, L., Simeon, C.P., Carreira, P., et al. (2019). GWAS for systemic sclerosis identifies multiple risk loci and highlights fibrotic and vasculopathy pathways. *Nat Commun* 10, 4955. 10.1038/s41467-019-12760-y.
6. Cordell, H.J., Fryett, J.J., Ueno, K., Darlay, R., Aiba, Y., Hitomi, Y., Kawashima, M., Nishida, N., Khor, S.S., Gervais, O., et al. (2021). An international genome-wide meta-analysis of primary biliary cholangitis: Novel risk loci and candidate drugs. *J Hepatol* 75, 572-581. 10.1016/j.jhep.2021.04.055.

7. Ellinghaus, D., Jostins, L., Spain, S.L., Cortes, A., Bethune, J., Han, B., Park, Y.R., Raychaudhuri, S., Pouget, J.G., Hubenthal, M., et al. (2016). Analysis of five chronic inflammatory diseases identifies 27 new associations and highlights disease-specific patterns at shared loci. *Nat Genet* 48, 510-518. 10.1038/ng.3528.
8. Ishigaki, K., Sakaue, S., Terao, C., Luo, Y., Sonehara, K., Yamaguchi, K., Amariuta, T., Too, C.L., Laufer, V.A., Scott, I.C., et al. (2022). Multi-ancestry genome-wide association analyses identify novel genetic mechanisms in rheumatoid arthritis. *Nat Genet* 54, 1640-1651. 10.1038/s41588-022-01213-w.
9. Bentham, J., Morris, D.L., Graham, D.S.C., Pinder, C.L., Tomblason, P., Behrens, T.W., Martin, J., Fairfax, B.P., Knight, J.C., Chen, L., et al. (2015). Genetic association analyses implicate aberrant regulation of innate and adaptive immunity genes in the pathogenesis of systemic lupus erythematosus. *Nat Genet* 47, 1457-1464. 10.1038/ng.3434.
10. Vosa, U., Claringbould, A., Westra, H.J., Bonder, M.J., Deelen, P., Zeng, B., Kirsten, H., Saha, A., Kreuzhuber, R., Yazar, S., et al. (2021). Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression. *Nat Genet* 53, 1300-1310. 10.1038/s41588-021-00913-z.
11. Momozawa, Y., Dmitrieva, J., Theatre, E., Deffontaine, V., Rahmouni, S., Charlotteaux, 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. 10.1038/s41467-018-04365-8.
12. Consortium, G.T. (2020). The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* 369, 1318-1330. 10.1126/science.aaz1776.
13. Sun, B.B., Chiou, J., Traylor, M., Benner, C., Hsu, Y.H., Richardson, T.G., Surendran, P., Mahajan, A., Robins, C., Vasquez-Grinnell, S.G., et al. (2023). Plasma proteomic associations with genetics and health in the UK Biobank. *Nature* 622, 329-338. 10.1038/s41586-023-06592-6.
14. 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., and Abecasis, G.R. (2015). A global reference for human genetic variation. *Nature* 526, 68-74. 10.1038/nature15393.
15. Murphy, A.E., Schilder, B.M., and Skene, N.G. (2021). MungeSumstats: a Bioconductor package for the standardization and quality control of many GWAS summary statistics. *Bioinformatics* 37, 4593-4596. 10.1093/bioinformatics/btab665.
16. Bulik-Sullivan, B., Finucane, H.K., Anttila, V., Gusev, A., Day, F.R., Loh, P.R., ReproGen, C., Psychiatric Genomics, C., Genetic Consortium for Anorexia Nervosa of the Wellcome Trust Case Control, C., Duncan, L., et al. (2015). An atlas of genetic correlations across human diseases and traits. *Nat Genet* 47, 1236-1241. 10.1038/ng.3406.
17. Bulik-Sullivan, B.K., Loh, P.R., Finucane, H.K., Ripke, S., Yang, J., Schizophrenia Working Group of the Psychiatric Genomics, C., Patterson, N., Daly, M.J., Price, A.L., and Neale, B.M. (2015). LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet* 47, 291-295. 10.1038/ng.3211.
18. Willer, C.J., Li, Y., and Abecasis, G.R. (2010). METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* 26, 2190-2191. 10.1093/bioinformatics/btq340.
19. Devlin, B., and Roeder, K. (1999). Genomic control for association studies. *Biometrics* 55, 997-1004. 10.1111/j.0006-341x.1999.00997.x.
20. Chang, X., and Wang, K. (2012). wANNOVAR: annotating genetic variants for personal genomes via the web. *J Med Genet* 49, 433-436. 10.1136/jmedgenet-2012-100918.
21. Lee, C.H., Shi, H., Pasaniuc, B., Eskin, E., and Han, B. (2021). PLEIO: a method to map and interpret pleiotropic loci with GWAS summary statistics. *Am J Hum Genet* 108, 36-48. 10.1016/j.ajhg.2020.11.017.

22. Yang, Z., Wang, C., Liu, L., Khan, A., Lee, A., Vardarajan, B., Mayeux, R., Kiryluk, K., and Ionita-Laza, I. (2023). CARMA is a new Bayesian model for fine-mapping in genome-wide association meta-analyses. *Nat Genet* 55, 1057-1065. 10.1038/s41588-023-01392-0.
23. Weissbrod, O., Hormozdiari, F., Benner, C., Cui, R., Ulirsch, J., Gazal, S., Schoech, A.P., van de Geijn, B., Reshef, Y., Marquez-Luna, C., et al. (2020). Functionally informed fine-mapping and polygenic localization of complex trait heritability. *Nat Genet* 52, 1355-1363. 10.1038/s41588-020-00735-5.
24. de Leeuw, C.A., Mooij, J.M., Heskes, T., and Posthuma, D. (2015). MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput Biol* 11, e1004219. 10.1371/journal.pcbi.1004219.
25. Watanabe, K., Taskesen, E., van Bochoven, A., and Posthuma, D. (2017). Functional mapping and annotation of genetic associations with FUMA. *Nat Commun* 8, 1826. 10.1038/s41467-017-01261-5.
26. Pers, T.H., Karjalainen, J.M., Chan, Y., Westra, H.J., Wood, A.R., Yang, J., Lui, J.C., Vedantam, S., Gustafsson, S., Esko, T., et al. (2015). Biological interpretation of genome-wide association studies using predicted gene functions. *Nat Commun* 6, 5890. 10.1038/ncomms6890.
27. Wallace, C. (2021). A more accurate method for colocalisation analysis allowing for multiple causal variants. *PLoS Genet* 17, e1009440. 10.1371/journal.pgen.1009440.
28. Gottesman, O., Kuivaniemi, H., Tromp, G., Faucett, W.A., Li, R., Manolio, T.A., Sanderson, S.C., Kannry, J., Zinberg, R., Basford, M.A., et al. (2013). The Electronic Medical Records and Genomics (eMERGE) Network: past, present, and future. *Genet Med* 15, 761-771. 10.1038/gim.2013.72.
29. All of Us Research Program, I., Denny, J.C., Rutter, J.L., Goldstein, D.B., Philippakis, A., Smoller, J.W., Jenkins, G., and Dishman, E. (2019). The "All of Us" Research Program. *N Engl J Med* 381, 668-676. 10.1056/NEJMs1809937.
30. Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L.T., Sharp, K., Motyer, A., Vukcevic, D., Delaneau, O., O'Connell, J., et al. (2018). The UK Biobank resource with deep phenotyping and genomic data. *Nature* 562, 203-209. 10.1038/s41586-018-0579-z.
31. Liu, L., Khan, A., Sanchez-Rodriguez, E., Zandoni, F., Li, Y., Steers, N., Balderes, O., Zhang, J., Krithivasan, P., LeDesma, R.A., et al. (2022). Genetic regulation of serum IgA levels and susceptibility to common immune, infectious, kidney, and cardio-metabolic traits. *Nat Commun* 13, 6859. 10.1038/s41467-022-34456-6.
32. Kiryluk, K., Sanchez-Rodriguez, E., Zhou, X.J., Zandoni, F., Liu, L., Mladkova, N., Khan, A., Marasa, M., Zhang, J.Y., Balderes, O., et al. (2023). Genome-wide association analyses define pathogenic signaling pathways and prioritize drug targets for IgA nephropathy. *Nat Genet*. 10.1038/s41588-023-01422-x.
33. Khan, A., Shang, N., Nestor, J.G., Weng, C., Hripcsak, G., Harris, P.C., Gharavi, A.G., and Kiryluk, K. (2023). Polygenic risk affects the penetrance of monogenic kidney disease. medRxiv. 10.1101/2023.05.07.23289614.
34. Ge, T., Chen, C.Y., Ni, Y., Feng, Y.A., and Smoller, J.W. (2019). Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nat Commun* 10, 1776. 10.1038/s41467-019-09718-5.
35. Open Targets Genetics. <https://genetics.opentargets.org/>.
36. Ghousaini, M., Mountjoy, E., Carmona, M., Peat, G., Schmidt, E.M., Hercules, A., Fumis, L., Miranda, A., Carvalho-Silva, D., Buniello, A., et al. (2021). Open Targets Genetics: systematic identification of trait-associated genes using large-scale genetics and functional genomics. *Nucleic Acids Res* 49, D1311-D1320. 10.1093/nar/gkaa840.
37. Gschwind, A.R., Mualim, K.S., Karbalayghareh, A., Sheth, M.U., Dey, K.K., Jagoda, E., Nurtdinov, R.N., Xi, W., Tan, A.S., Jones, H., et al. (2023). An encyclopedia of enhancer-gene regulatory interactions in the human genome. bioRxiv. 10.1101/2023.11.09.563812.

38. Shringarpure, S.S., Wang, W., Karagounis, S., Wang, X., Reisetter, A.C., Auton, A., and Khan, A.A. (2024). Large language models identify causal genes in complex trait GWAS. *medRxiv*, 2024.2005.2030.24308179. [10.1101/2024.05.30.24308179](https://doi.org/10.1101/2024.05.30.24308179).
39. Pakozdi, A., Nihtyanova, S., Moynzadeh, P., Ong, V.H., Black, C.M., and Denton, C.P. (2011). Clinical and serological hallmarks of systemic sclerosis overlap syndromes. *J Rheumatol* **38**, 2406-2409. [10.3899/jrheum.101248](https://doi.org/10.3899/jrheum.101248).
40. Moynzadeh, P., Aberer, E., Ahmadi-Simab, K., Blank, N., Distler, J.H., Fierlbeck, G., Genth, E., Guenther, C., Hein, R., Henes, J., et al. (2015). Disease progression in systemic sclerosis-overlap syndrome is significantly different from limited and diffuse cutaneous systemic sclerosis. *Ann Rheum Dis* **74**, 730-737. [10.1136/annrheumdis-2013-204487](https://doi.org/10.1136/annrheumdis-2013-204487).
41. Foocharoen, C., Netwijitpan, S., Mahakkanukrauh, A., Suwannaroj, S., and Nanagara, R. (2016). Clinical characteristics of scleroderma overlap syndromes: comparisons with pure scleroderma. *Int J Rheum Dis* **19**, 913-923. [10.1111/1756-185X.12884](https://doi.org/10.1111/1756-185X.12884).
42. Kim, W.R., Lindor, K.D., Locke, G.R., 3rd, Therneau, T.M., Homburger, H.A., Batts, K.P., Yawn, B.P., Petz, J.L., Melton, L.J., 3rd, and Dickson, E.R. (2000). Epidemiology and natural history of primary biliary cirrhosis in a US community. *Gastroenterology* **119**, 1631-1636. [10.1053/gast.2000.20197](https://doi.org/10.1053/gast.2000.20197).
43. Sood, S., Gow, P.J., Christie, J.M., and Angus, P.W. (2004). Epidemiology of primary biliary cirrhosis in Victoria, Australia: high prevalence in migrant populations. *Gastroenterology* **127**, 470-475. [10.1053/j.gastro.2004.04.064](https://doi.org/10.1053/j.gastro.2004.04.064).
44. Tian, J., Zhang, D., Yao, X., Huang, Y., and Lu, Q. (2023). Global epidemiology of systemic lupus erythematosus: a comprehensive systematic analysis and modelling study. *Ann Rheum Dis* **82**, 351-356. [10.1136/ard-2022-223035](https://doi.org/10.1136/ard-2022-223035).
45. Izmirly, P.M., Parton, H., Wang, L., McCune, W.J., Lim, S.S., Drenkard, C., Ferucci, E.D., Dall'Era, M., Gordon, C., Helmick, C.G., and Somers, E.C. (2021). Prevalence of Systemic Lupus Erythematosus in the United States: Estimates From a Meta-Analysis of the Centers for Disease Control and Prevention National Lupus Registries. *Arthritis Rheumatol* **73**, 991-996. [10.1002/art.41632](https://doi.org/10.1002/art.41632).
46. Stojan, G., and Petri, M. (2018). Epidemiology of systemic lupus erythematosus: an update. *Curr Opin Rheumatol* **30**, 144-150. [10.1097/BOR.0000000000000480](https://doi.org/10.1097/BOR.0000000000000480).
47. Terao, C., Ohmura, K., Kawaguchi, Y., Nishimoto, T., Kawasaki, A., Takehara, K., Furukawa, H., Kochi, Y., Ota, Y., Ikari, K., et al. (2013). PLD4 as a novel susceptibility gene for systemic sclerosis in a Japanese population. *Arthritis Rheum* **65**, 472-480. [10.1002/art.37777](https://doi.org/10.1002/art.37777).
48. Qiu, F., Tang, R., Zuo, X., Shi, X., Wei, Y., Zheng, X., Dai, Y., Gong, Y., Wang, L., Xu, P., et al. (2017). A genome-wide association study identifies six novel risk loci for primary biliary cholangitis. *Nat Commun* **8**, 14828. [10.1038/ncomms14828](https://doi.org/10.1038/ncomms14828).
49. 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. [10.1038/ng.3359](https://doi.org/10.1038/ng.3359).
50. International Multiple Sclerosis Genetics, C. (2019). Multiple sclerosis genomic map implicates peripheral immune cells and microglia in susceptibility. *Science* **365**. [10.1126/science.aav7188](https://doi.org/10.1126/science.aav7188).
51. Karnell, J.L., Rieder, S.A., Ettinger, R., and Kolbeck, R. (2019). Targeting the CD40-CD40L pathway in autoimmune diseases: Humoral immunity and beyond. *Adv Drug Deliv Rev* **141**, 92-103. [10.1016/j.addr.2018.12.005](https://doi.org/10.1016/j.addr.2018.12.005).
52. Banday, A.Z., Nisar, R., Patra, P.K., Kaur, A., Sadanand, R., Chaudhry, C., Bukhari, S.T.A., Banday, S.Z., Bhattarai, D., and Notarangelo, L.D. (2023). Clinical and Immunological Features, Genetic Variants, and Outcomes of Patients with CD40 Deficiency. *J Clin Immunol* **44**, 17. [10.1007/s10875-023-01633-1](https://doi.org/10.1007/s10875-023-01633-1).

53. Michel, N.A., Zirlik, A., and Wolf, D. (2017). CD40L and Its Receptors in Atherothrombosis-An Update. *Front Cardiovasc Med* 4, 40. 10.3389/fcvm.2017.00040.
54. Allanore, Y., Borderie, D., Meune, C., Lemarechal, H., Weber, S., Ekindjian, O.G., and Kahan, A. (2005). Increased plasma soluble CD40 ligand concentrations in systemic sclerosis and association with pulmonary arterial hypertension and digital ulcers. *Ann Rheum Dis* 64, 481-483. 10.1136/ard.2003.020040.
55. Kuiper, J.J., Prinz, J.C., Stratikos, E., Kusnierczyk, P., Arakawa, A., Springer, S., Mintoff, D., Padjen, I., Shumnalieva, R., Vural, S., et al. (2023). EULAR study group on 'MHC-I-opathy': identifying disease-overarching mechanisms across disciplines and borders. *Ann Rheum Dis* 82, 887-896. 10.1136/ard-2022-222852.
56. International Genetics of Ankylosing Spondylitis, C., Cortes, A., Hadler, J., Pointon, J.P., Robinson, P.C., Karaderi, T., Leo, P., Cremin, K., Pryce, K., Harris, J., et al. (2013). Identification of multiple risk variants for ankylosing spondylitis through high-density genotyping of immune-related loci. *Nat Genet* 45, 730-738. 10.1038/ng.2667.
57. 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. 10.1038/ncomms15382.
58. Gourh, P., Safran, S.A., Alexander, T., Boyden, S.E., Morgan, N.D., Shah, A.A., Mayes, M.D., Doumatey, A., Bentley, A.R., Shriner, D., et al. (2020). HLA and autoantibodies define scleroderma subtypes and risk in African and European Americans and suggest a role for molecular mimicry. *Proc Natl Acad Sci U S A* 117, 552-562. 10.1073/pnas.1906593116.
59. Mulinacci, G., Palermo, A., Gerussi, A., Asselta, R., Gershwin, M.E., and Invernizzi, P. (2022). New insights on the role of human leukocyte antigen complex in primary biliary cholangitis. *Front Immunol* 13, 975115. 10.3389/fimmu.2022.975115.
60. Jongsma, M.L.M., de Waard, A.A., Raaben, M., Zhang, T., Cabukusta, B., Platzer, R., Blomen, V.A., Xagara, A., Verkerk, T., Bliss, S., et al. (2021). The SPPL3-defined glycosphingolipid repertoire orchestrates HLA class I-mediated immune responses. *Immunity* 54, 387. 10.1016/j.immuni.2021.01.016.
61. Onishi, A., Sugiyama, D., Kumagai, S., and Morinobu, A. (2013). Cancer incidence in systemic sclerosis: meta-analysis of population-based cohort studies. *Arthritis Rheum* 65, 1913-1921. 10.1002/art.37969.
62. Liang, Y., Yang, Z., and Zhong, R. (2012). Primary biliary cirrhosis and cancer risk: a systematic review and meta-analysis. *Hepatology* 56, 1409-1417. 10.1002/hep.25788.
63. Ricard, L., Hirsch, P., Largeaud, L., Deswarte, C., Jachiet, V., Mohty, M., Riviere, S., Malard, F., Tenon, M., de Vassogne, F., et al. (2020). Clonal haematopoiesis is increased in early onset in systemic sclerosis. *Rheumatology (Oxford)* 59, 3499-3504. 10.1093/rheumatology/keaa282.
64. Gniadecki, R., Iyer, A., Hennessey, D., Khan, L., O'Keefe, S., Redmond, D., Storek, J., Durand, C., Cohen-Tervaert, J.W., and Osman, M. (2022). Genomic instability in early systemic sclerosis. *J Autoimmun* 131, 102847. 10.1016/j.jaut.2022.102847.
65. Gianfrancesco, M.A., and Goldstein, N.D. (2021). A narrative review on the validity of electronic health record-based research in epidemiology. *BMC Med Res Methodol* 21, 234. 10.1186/s12874-021-01416-5.

Autoimmune Disease GWAS Datasets					
Phenotype	N <sub>cases</sub>	N <sub>controls</sub>	Ancestry	Publication year	Reference
SSc	9,095	17,584	European	2019	<sup>5</sup>
PBC	8,021	16,489	European	2021	<sup>6</sup>
RA	22,350	78,823	European	2022	<sup>8</sup>
SLE	7,219	15,991	European	2015	<sup>9</sup>

eQTL datasets					
Dataset	Tissue	N	Ancestry	Publication year	Reference
eQTLGen	blood	31,864	Predominantly European	2020	<sup>10</sup>
GTE <sub>x</sub> V8	skin, lung, liver B cells, CD4+ T cells,	508 (skin), 436 (lung), 178 (liver)	European	2021	<sup>12</sup>
CEDAR	CD8+ T cells, monocytes, neutrophils	323	European	2018	<sup>11</sup>

pQTL datasets					
Dataset	Tissue	N	Ancestry	Publication year	Reference
UK Biobank	plasma	54,219	Predominantly European	2023	<sup>13</sup>

**Table 1** GWAS summary statistics included in this study.

SSc: systemic sclerosis. PBC: primary biliary cholangitis. RA: Rheumatoid arthritis. SLE: systemic lupus erythematosus. eQTL: expression quantitative trait loci. pQTL: protein quantitative trait loci. GTE<sub>x</sub>: Genotype-Tissue Expression. CEDAR: Correlated Expression and Disease Association Research



**Table 2** Genomic loci significant in the cross-phenotype GWAS, colocalized between SSc and PBC and do not have evidence of heterogeneity

Locus	Lead SNP	CHR	BP	A1	A2	FE OR	FE p-values	PLEIO p-values	Colocalization PP4	Novel loci
<i>TNPO3</i>	rs17338998	7	128618559	T	C	1.52	2.37E-40	4.16E-73	98%	No
<i>CSNK2A2/CCDC113</i>	rs2731783	16	58253460	A	G	1.15	7.79E-11	1.50E-09	96%	Yes
<i>IL12RB1</i>	rs2305743	19	18193191	A	G	0.86	1.95E-14	7.42E-14	96%	No
<i>LOC100506023</i>	rs2022449	1	173238736	T	G	1.12	1.69E-10	2.82E-09	92%	No
<i>SPPL3</i>	rs551125	12	121203427	C	T	0.91	1.65E-09	2.59E-08	91%	Yes
<i>CAST/ERAP1</i>	rs27524	5	96101944	A	G	0.9	4.89E-10	6.04E-09	85%	Yes
<i>HEMGN/ANP32B</i>	rs4743150	9	100740124	T	C	0.88	8.29E-11	2.80E-10	85%	No
<i>AHNAK2</i>	rs10083496	14	105402786	G	A	1.09	2.56E-08	7.37E-08	80%	Yes
<i>CD40</i>	rs4810485	20	44747947	T	G	1.13	1.44E-08	7.37E-08	73%	Yes

SSc: systemic sclerosis. PBC: primary biliary cholangitis. PLEIO: Pleiotropic Locus Exploration and Interpretation using Optimal test. FE: fixed-effect model. PP4: posterior probability of hypothesis 4 that there is one variant causal to both traits. A1: effect allele. A2: non-effect allele.

## Figure Legends

**Figure 1** Flowchart for the overview of the study.

We performed multiple cross-phenotype GWAS analyses to identify the shared genetic susceptibility between SSc and PBC. Additionally, we also performed a single-center retrospective chart review to evaluate the prevalence of PBC, including potentially undiagnosed cases, in patients with SSc.

SSc: systemic sclerosis. PBC: primary biliary cholangitis. GWAS: genome-wide association studies. PLEIO: Pleiotropic Locus Exploration and Interpretation using Optimal test. eQTL: expression quantitative trait loci. PheWAS: phenome-wide association studies.

**Figure 2** Pairwise genetic correlation in SSc, PBC, RA and SLE.

We performed genetic correlation analyses using linkage disequilibrium score regression pairwise among SSc, PBC, RA and SLE. The effect estimates of genetic correlation between SSc and PBC, and between SSc and SLE, were both 0.84, higher than other pairwise estimates.

**Figure 3** Manhattan plot of the cross-phenotype GWAS meta-analysis in SSc and PBC using the fixed-effect model.

We performed a cross-phenotype GWAS meta-analysis in SSc and PBC using the fixed-effect model. We identified 44 significant genomic loci ( $p < 5 \times 10^{-8}$ ). We found 9 loci that colocalized between SSc and PBC and did not show evidence of heterogeneity (indicated in red), among which five were novel (indicated with an asterisk). SNPs with evidence of heterogeneity ( $P_{\text{het}} < 0.05$ ) were excluded.

**Figure 4** Tissue and pathway enrichment analyses.

**a** Tissue enrichment analysis using MAGMA prioritized tissues related to the immune system (spleen, whole blood and EBV-transformed lymphocytes), respiratory system (lung) and digestive system

(terminal ileum). **b** Tissue enrichment analysis using DEPICT prioritized multiple tissues and cells related to the immune system, respiratory system and musculoskeletal system. **c** Significant enrichment of multiple immune-related pathways associated with SSc and PBC using MAGMA.

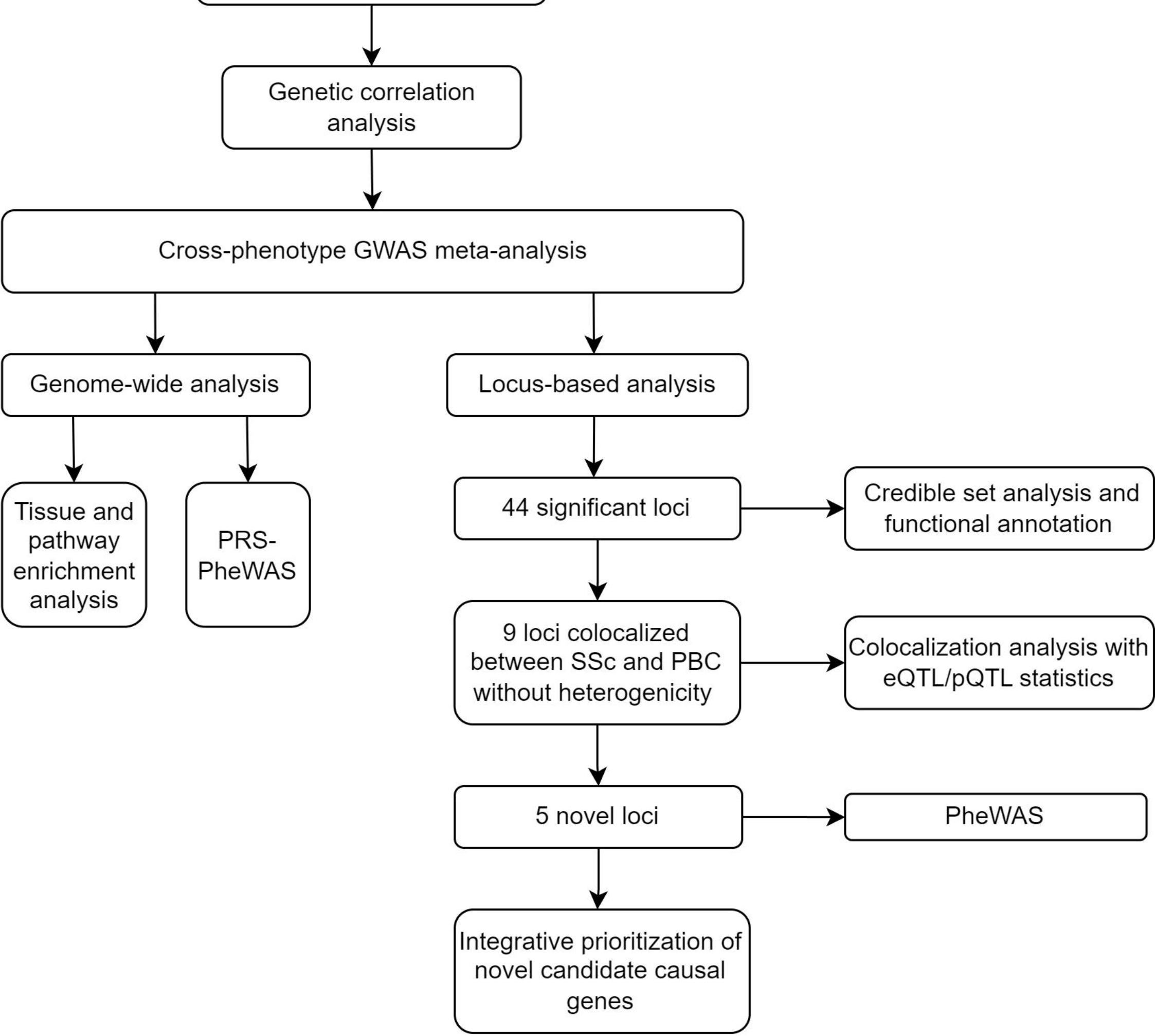
**Figure 5** Colocalization analysis with expression quantitative trait loci (eQTL) in blood, skin, lung, liver and immune cells.

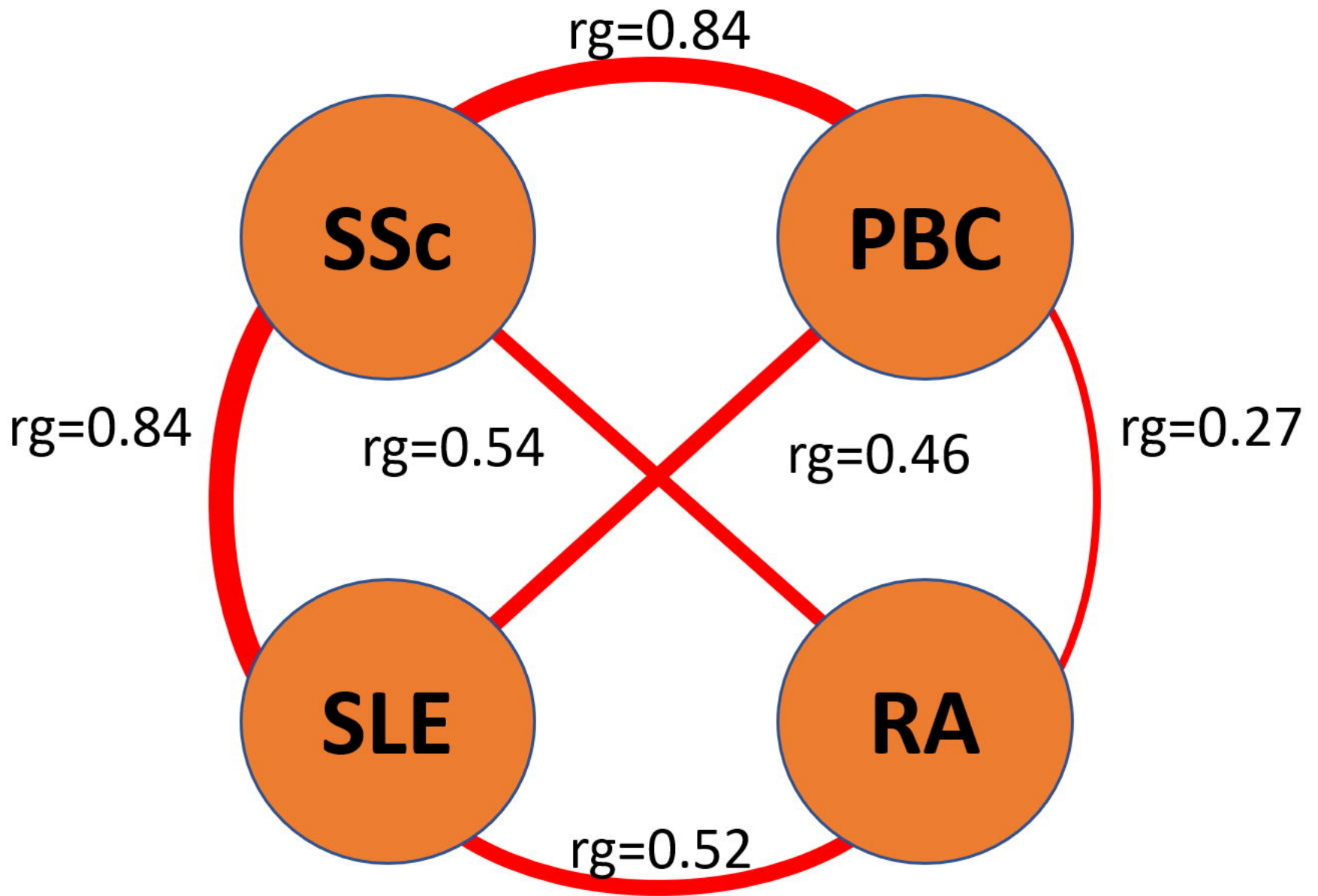
Out of the nine loci that were significant in the fixed-model cross-phenotype meta-analysis, showed no evidence of heterogeneity, and colocalized between SSc and PBC, seven colocalized with expressed genes in at least one of the examined tissues.

**Figure 6** Integrative prioritization of novel candidate causal genes

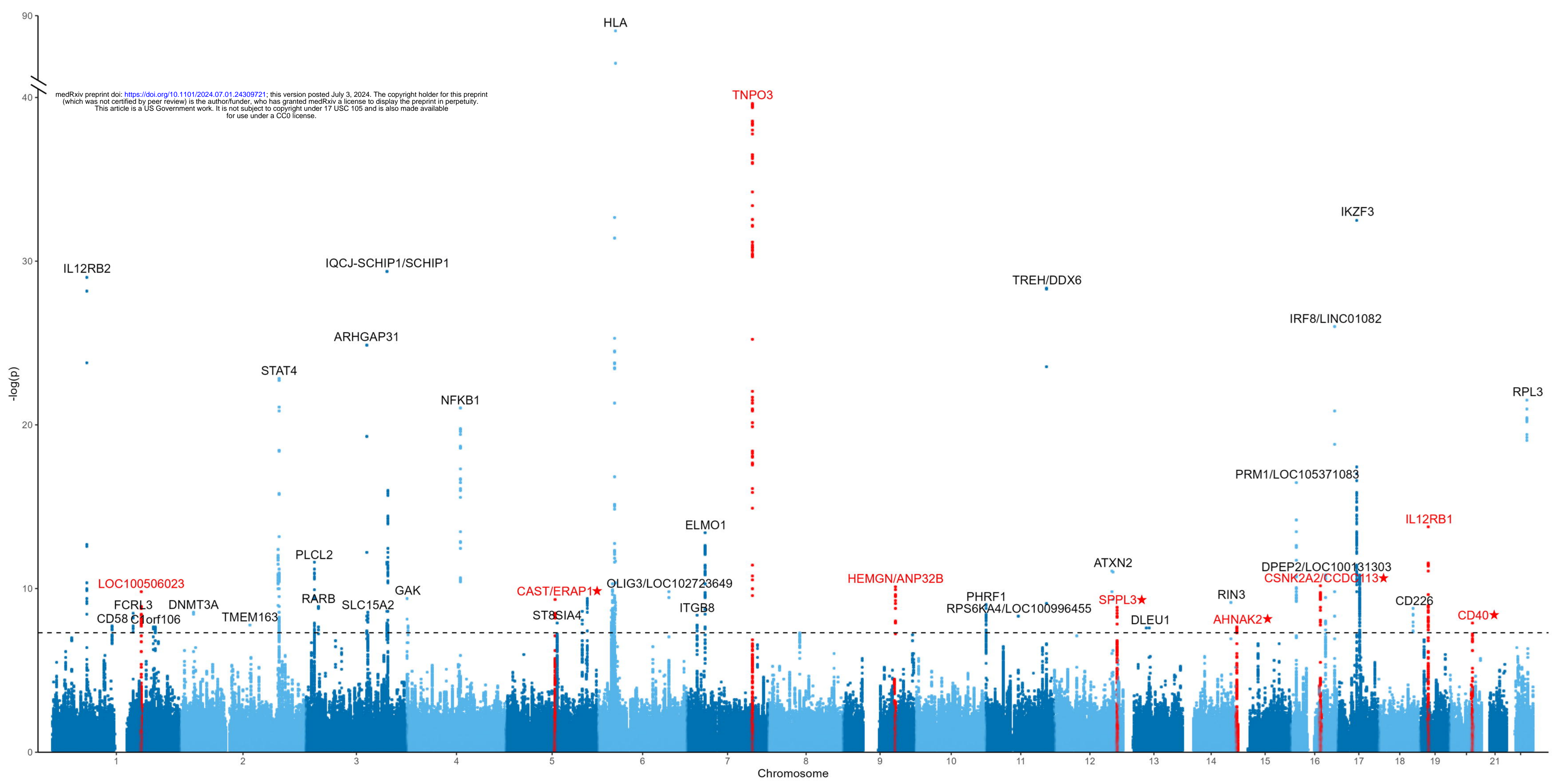
We prioritized five novel candidate causal genes based on a priority score integrating nine criteria: *CD40*, *ERAP1*, *PLD4*, *SPPL3* and *CCDC113*

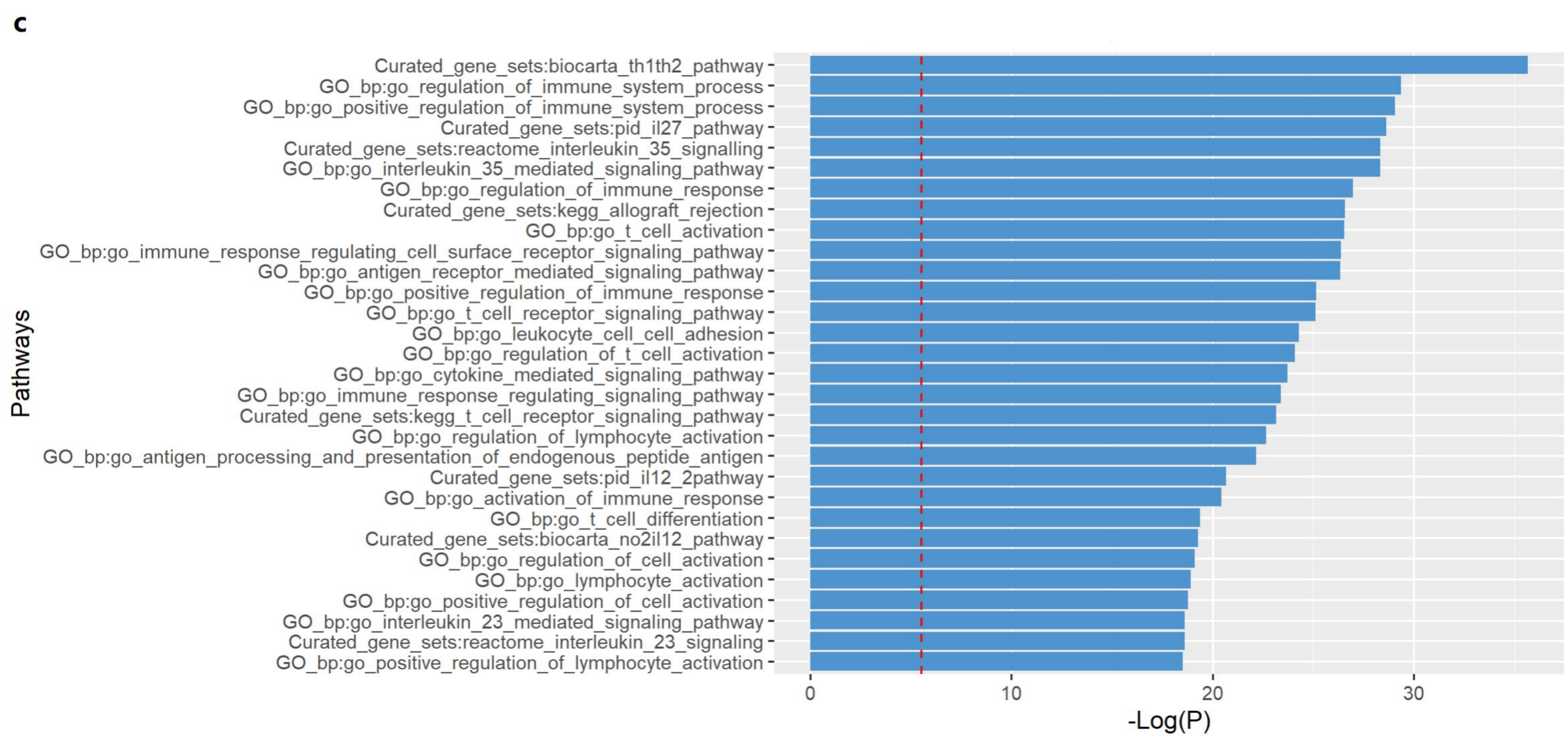
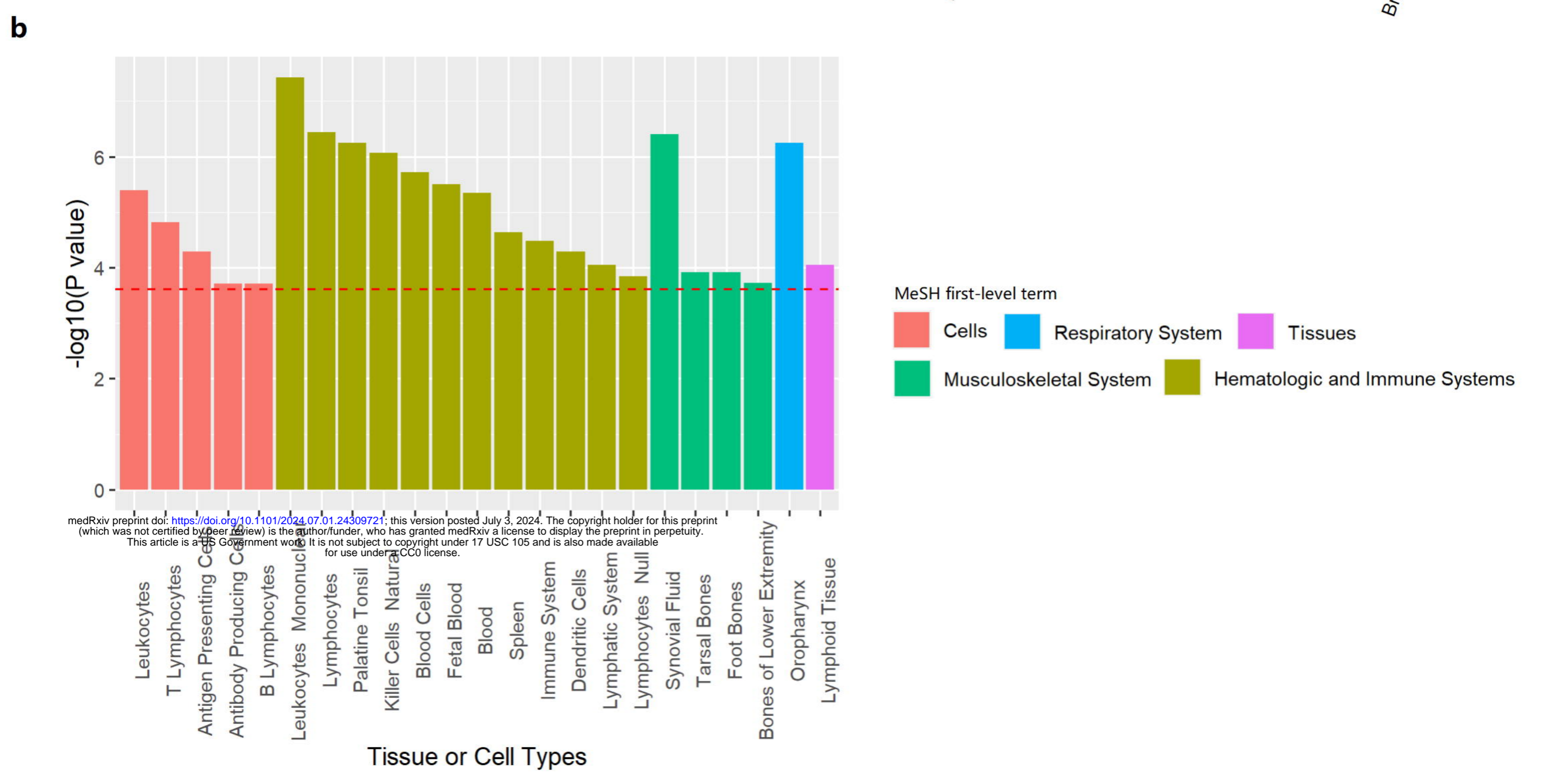
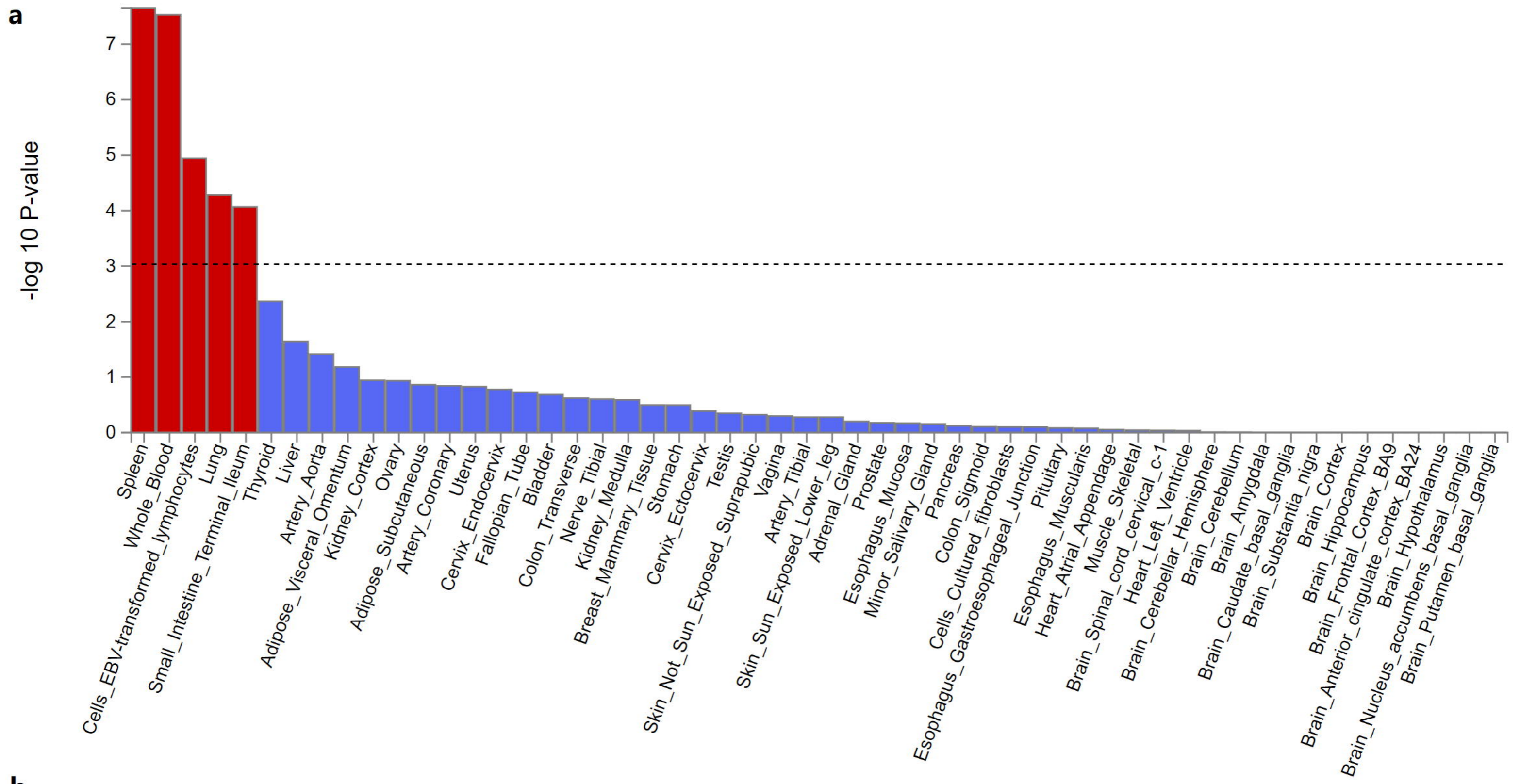
# GWAS summary statistics of SSc and PBC

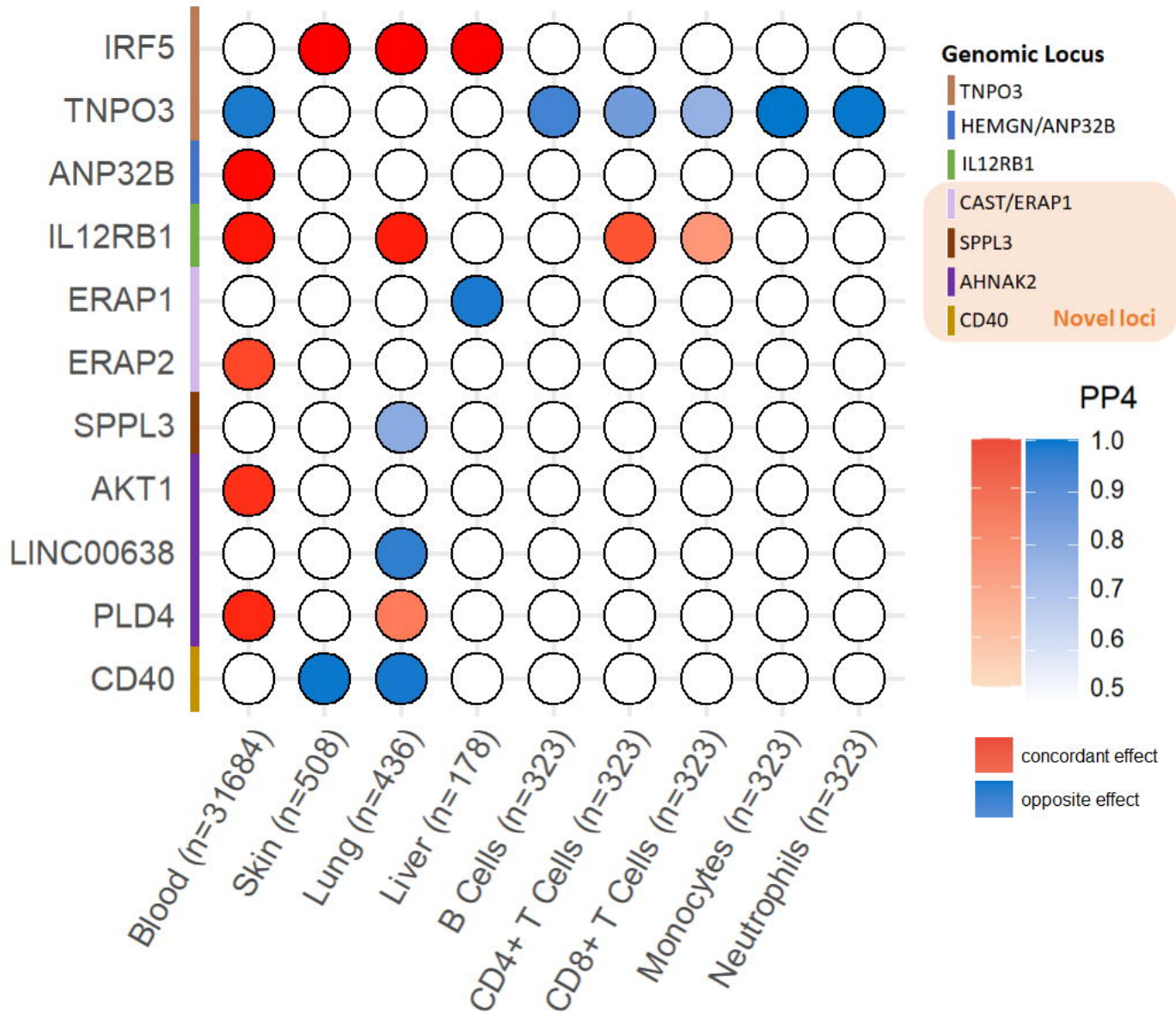




SSc: systemic sclerosis; PBC: primary biliary cholangitis; SLE: systemic lupus erythematosus; RA: rheumatoid arthritis; rg: genetic correlation









Locus	Lead SNP	Candidate gene	Nesrest gene	Cis-eQTL colocalization	Cis-pQTL colocalization	Nonsynonymous coding variant	MAGMA	DEPICT	V2G	ENCODE-rE2G	LLM	Priority score
CD40	rs4810485	CD40										8
		MMP9										2
		CDH22										1
		NCOA5										1
		SLC12A5										1
		WFDC10A										1
		CTSA										1
ERAP1 /CAST	rs27524	ERAP1										6
		ERAP2										3
		LNPEP										1
		CAST										1
AHNAK2	rs10083496	PLD4										5
		AHNAK2										3
		AKT1										2
		CLBA1										2
		CDCA4										2
		CEP170B										1
		GPR132										1
SPPL3	rs551125	SPPL3										4
		OASL										2
		ACADS										1
		HNF1A										1
		P2RX7										1
CCDC113 /CSNK2A2	rs2731783	CCDC113										3
		CSNK2A2										2
		ZBTB42										1
		MMP15										1
		NDRG4										1

medRxiv preprint doi: <https://doi.org/10.1101/2024.07.01.24309721>; this version posted July 3, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.