PheWAS and cross-disorder analysis reveal genetic architecture, pleiotropic loci and phenotypic correlations across 11 autoimmune disorders

Apostolia Topaloudi[1], Pritesh Jain[1], Melanie B. Martinez[1], Josephine K. Bryant[1], Grace Reynolds[1,2], Zoi Zagoriti[3], George Lagoumintzis[3], Eleni Zamba-Papanicolaou[4], John Tzartos[5], Konstantinos Poulas[3], Kleopas A. Kleopa[6], Socrates Tzartos[3,7], Marianthi Georgitsi[8], Petros Drineas[9] and Peristera Paschou[1]*

[1]Department of Biological Sciences, Purdue University, West Lafayette, IN, United States, [2]West Lafayette High School, West Lafayette, IN, United States, [3]Department of Pharmacy, University of Patras, Rio, Greece, [4]Department of Neuroepidemiology and Centre for Neuromuscular Disorders, The Cyprus Institute of Neurology and Genetics, Nicosia, Cyprus, [5]B' Neurology Department, School of Medicine, National & Kapodistrian University of Athens, "Attikon" University Hospital., Athens, Greece, [6]Department of Neuroscience and Centre for Neuromuscular Disorders, The Cyprus Institute of Neurology and Genetics, Nicosia, Cyprus, [7]Tzartos NeuroDiagnostics, Athens, Greece, [8]Department of Molecular Biology and Genetics, School of Health Sciences, Democritus University of Thrace, Alexandroupoli, Greece, [9]Department of Computer Science, Purdue University, West Lafayette, IN, United States

Introduction: Autoimmune disorders (ADs) are a group of about 80 disorders that occur when self-attacking autoantibodies are produced due to failure in the self-tolerance mechanisms. ADs are polygenic disorders and associations with genes both in the human leukocyte antigen (HLA) region and outside of it have been described. Previous studies have shown that they are highly comorbid with shared genetic risk factors, while epidemiological studies revealed associations between various lifestyle and health-related phenotypes and ADs.

Methods: Here, for the first time, we performed a comparative polygenic risk score (PRS) - Phenome Wide Association Study (PheWAS) for 11 different ADs (Juvenile Idiopathic Arthritis, Primary Sclerosing Cholangitis, Celiac Disease, Multiple Sclerosis, Rheumatoid Arthritis, Psoriasis, Myasthenia Gravis, Type 1 Diabetes, Systemic Lupus Erythematosus, Vitiligo Late Onset, Vitiligo Early Onset) and 3,254 phenotypes available in the UK Biobank that include a wide range of socio-demographic, lifestyle and health-related outcomes. Additionally, we investigated the genetic relationships of the studied ADs, calculating their genetic correlation and conducting cross-disorder GWAS meta-analyses for the observed AD clusters.

Results: In total, we identified 508 phenotypes significantly associated with at least one AD PRS. 272 phenotypes were significantly associated after excluding

variants in the HLA region from the PRS estimation. Through genetic correlation and genetic factor analyses, we identified four genetic factors that run across studied ADs. Cross-trait meta-analyses within each factor revealed pleiotropic genome-wide significant loci.

Discussion: Overall, our study confirms the association of different factors with genetic susceptibility for ADs and reveals novel observations that need to be further explored.

1 Introduction

Autoimmune disorders (ADs) are a group of about 80 (1) disorders that occur when self-attacking autoantibodies are produced due to failure in the self-tolerance mechanisms (2). The estimated overall prevalence is 3% in the United States (3), and recent studies report an increasing trend (4–6). Additionally, ADs are often comorbid and cluster within families (7, 8). The majority of ADs are polygenic and previous studies revealed associations with genes in the human leukocyte antigen (HLA) region (9, 10). However, multiple additional associations with genes outside of the HLA are found with various ADs, and many times they are implicated in more than one disorder (10). The genetic correlation across multiple ADs has not been fully explored (11, 12). So far, cross-disorder GWAS meta-analyses have only focused on a few ADs, usually three to seven at a time (12–15), while others have only focused on pairwise meta-analyses (16, 17). Given the wide comorbidities observed in epidemiological studies and the evidence for sharing common genetic background across multiple ADs, a systematic large-scale analysis is warranted.

In ADs, like other complex disorders, environmental factors are also involved in disease development along with genetic predisposition. Multiple studies have reported associations between viral infections and specific autoimmune diseases (18). For instance, a recent study (19) is suggesting that infection of Epstein-Barr virus could be the leading cause of Multiple Sclerosis. Additional associations between ADs and environmental factors such as smoking, and UV exposure have also been reported (20, 21). Epidemiological studies have reported a high comorbidity across different ADs (8) as well as links to other traits, including psychotic disorders (22), allergies (23), and obesity (24).

Given the complex genetic background of ADs, Polygenic Risk Scores (PRS) which are an estimate of an individual's genetic predisposition for a trait, are an important tool to help understand disease correlations. They are usually calculated as the total of the risk alleles an individual carries, weighted by their effect sizes as measured in a previous genome-wide association study (GWAS) (25). This genetic risk can then become the basis of a Phenome-wide association study (PheWAS), with a goal to explore whether risk variants identified by a GWAS or disease PRS are associated with a wide variety of phenotypes (26). Biobanks that combine genetic data with Electronic Health Records (EHR) are essential for the PheWAS approach, as they are the source of the phenotypes used in the analysis (27). Since the PheWAS is a hypothesis-free analysis, it can be used to generate new hypotheses about novel associations that might have not been uncovered through hypothesis-driven approaches.

Here, for the first time, using summary statistics data of 11 different ADs and genetic and phenotypic data from the UK Biobank, we performed a PRS-PheWAS, interrogating associations of AD PRS with a wide range of socio-demographic, lifestyle, and health related phenotypes. Additionally, we investigated the genetic relationships across the studied ADs and conducted cross-trait GWAS meta-analyses for the identified AD sub-groups. Our findings present an overview of the phenotypic and genetic architecture and relationships of ADs.

2 Methods

2.1 Study population

The UK Biobank is a large-scale, population-based, prospective cohort that recruited between 2006 and 2010 over 500,000 participants from the UK aged 40–69 years old. The participants provided blood, urine, and saliva samples for biochemical tests and genotyping, as well as self-reported information which was then linked to their health-related records. The phenotypic and genetic data we used in this study were obtained from UK Biobank under application number #61553.

The initial UK Biobank dataset included 488,377 individuals genotyped on the Affymetrix UK BiLEVE Axiom array or the Affymetrix UK Biobank Axiom array. We performed standard quality control on individuals and genetic markers (info>=0.9, maf>=0.01, geno<=0.02, hwe >= 10^{-6}) with PLINK 1.9 (28). Initially, participants with withdrawn consent, sex mismatch, sex aneuploidy, self-reported non-white British ancestry, and with kinship coefficient <0.0625 (third-degree relatedness (29)) were excluded. Additional Principal Component Analysis (PCA) with 1000 Genomes data as

reference was performed using TeraPCA (30) to exclude individuals with non-European ancestry. The final dataset included 330,841 individuals and 7,634,371 SNPs. 53.98% of the selected participants are females, the average age is 56.8 (sd=8) years. Table S1 provides a breakdown of the participants' age and the percentage of the selected autoimmune diagnoses present in the UK Biobank.

2.2 PRS-pheWAS

2.2.1 Polygenic Risk Scores

Publicly available and in-house GWAS summary statistics for 11 ADs performed on datasets of European ancestry and no UK Biobank participants were collected. For the PRS calculations we used PRSice2 (31). The independent SNPs with p-values<10^{-5}, after clumping using a window of 500kb and an r^2 threshold of 0.1, were included in the PRS calculations and the score was yielded as the weighted, standardized sum of the effect (score-std option). Additionally, we normalized the PRS for age, sex, genotyping batch, and the first ten PCs. We repeated the PRS calculations excluding the extended HLA region (hg19, chr6 25-33 Mb). Table 1 shows the studied autoimmune datasets and the number of SNPs included in the PRS calculations. PRS performance was evaluated using Nagelkerke's pseudo-R2 metric for each AD. We used the AD summary statistics as the base and UK Biobank participants as the target data. We defined the individuals with the ICD10 code diagnosis for the studied AD as cases, and the UK Biobank individuals with no reported ICD10 diagnoses were defined as controls. Age, sex, genotyping batch and first ten PCs were included as covariates.

2.2.2 Phenotypes

We included 3,254 phenotypes from UK Biobank that were assigned to seven broad categories: Biomarkers, Cognition and Mental Health, Disease Diagnoses, Health and Medical History, Physical Measures, Lifestyle, and Sociodemographics. Specifically for the Disease Diagnoses category, we included only the ICD10 codes and used the R PheWAS tool (32) to map similar diagnoses into one phecode. The breakdown of data fields in each category is shown in the Supplementary Materials (Figures S1, S2).

2.2.3 PheWAS

For the PheWAS analyses, we used the tool PHESANT (33) to test the association of each disease PRS with each UK Biobank outcome. PHESANT, which is described in detail in (33), is commonly used in PheWAS analyses and automatically removes the instances with missing values from the UK Biobank Data-Codings. Age, sex, the first 10 principal components to correct for population stratification, and the genotyping batch were included as covariates in all regression models. To account for multiple testing, we used the R function p.adjust to calculate the FDR adjusted p-values and set the significance threshold at pFDR<0.05.

2.3 Cross-Disorder GWAS Meta-analysis

Pairwise genetic correlation analyses were performed for all 11 ADs after removing the extended HLA region (hg19, chr6 25-33 Mb) using LDSC (34). Only SNPs present in the HapMap3 reference panel were included in analyses and we used precalculated LD scores from 1000 Genomes European data. Datasets with less than 200,000 SNPs overlap with the LDSC reference data or heritability z-score <1.5 [as defined in (35)], were excluded from downstream analyses, namely CEL, PSO, and JIA datasets were removed.

To further explore the architecture and correlations of the studied disorders, we performed exploratory factor analysis (EFA) on the genetic correlation matrix using the R tool GenomicSEM

TABLE 1 Autoimmune Disease datasets used in this study.

Autoimmune Disorder	Abbreviation	Cases	Controls	SNPs in sumstats	SNPs in PRS	SNPs in PRS (no HLA)	PMID
Rheumatoid Arthritis	RA	14,361	43,923	8,747,962	309	132	24390342
Systemic Lupus Erythematosus	SLE	4,036	6,959	7,915,251	200	144	26502338
Vitiligo Late Onset	VITL	1,467	19,156	7,552,975	77	49	30674883
Vitiligo Early Onset	VITE	704	9,031	8,020,475	84	60	30674883
Type 1 Diabetes	T1D	9,358	15,705	6,621,966	236	198	32005708
Primary Sclerosing Cholangitis	PSC	2,871	12,019	7,891,602	157	50	27992413
Psoriasis	PSO	2,997	9,183	161,173	191	121	23143594
Multiple Sclerosis	MS	9,772	17,376	472,086	147	86	21833088
Celiac Disease*	CEL	12,041	12,228	139,553	122	100	22057235
Juvenile Idiopathic Arthritis	JIA	2,816	13,056	122,330	45	5	23603761
Myasthenia Gravis**	MG	1,401	3,508	5,755,778	21	14	34400559

*We included the summary statistics only from the European ancestry individuals in this study.
**For the PRS calculations we used the summary statistics after excluding the UK Biobank samples, while for the rest of the analyses we included the full dataset described in the study.
The number of SNPs in the PRS calculations corresponds to the independent SNPs with p<10^{-5}.

(36). We further used a confirmatory factor analysis (CFA) to validate our model. For groups of disorders within each of the factors, we performed a cross-disorder GWAS meta-analysis with ReACt (37) and corrected for sample overlap between the datasets. In order to identify potentially pleiotropic SNPs, in each meta-analysis we estimated the posterior probability (m-value) using METASOFT (38) to identify SNPs with high m-values (m-value>0.9) for all studies in the meta-analysis. Then, using the pleiotropic SNPs, we identified the LD independent regions (r^2<0.1) from the index SNPs with p<5×10^{-8}. We used the default LD clumping window (250kb) and mapped into the regions the genes located no more than 20kb away. As reference for the LD estimation, we used the European samples from 1000 Genomes. Additionally, we merged into one overlapping genomic regions using bedtools (39). All genes that mapped to the identified LD independent regions for each meta-analysis after clumping, were submitted to g:Profiler (40) to perform functional enrichment analysis for Gene Ontology terms (GO : BP, GO : CC, GO : MF, released 2021-12-15), Reactome (REAC, released 2022-1-3) and Kyoto Encyclopedia of Genes and Genomes (KEGG FTP, released 2021-12-27). For all experiments we performed the recommended multiple hypothesis correction (g:SCS) method with the significance threshold of p = 0.05. We repeated the analysis after excluding the electronic GO annotations (Inferred from Electronic Annotation [IEA]) to have higher confidence in the enrichment analysis.

3 Results

3.1 Individual disorder PheWAS

First, we investigated the potential association of AD genetic risk to other phenotypes, including socioeconomic factors, lifestyle, biomarkers, disease diagnoses, health history and mental health, performing PRS-PheWAS. We used the LD-independent SNPs with p<10^{-5} to calculate PRS for each of the studied ADs in each individual, and tested the association of the normalized -for age, sex, the first 10 principal components, and genotyping batch-autoimmune PRS, with 3,254 phenotypes in 330,841 UK Biobank samples (Table S1). The PheWAS analysis was adjusted for age, sex, the first 10 principal components, and genotyping batch. We found a large number of associations with each disorder which differed depending on whether the HLA region was included in the analysis (Figure 1, Table 2, Figures S3, S4 and Tables S2, S3). For SLE PRS with HLA region included in the analysis, we found the highest number of associations to different phenotypes (n=263). On the other hand, analysis for SLE PRS without the HLA region included, was associated with only 38 phenotypes. Interestingly, for CEL, T1D and RA, more PRS associations to phenotypes were actually found when the HLA region was excluded from the calculations. For Psoriasis, genetic risk was found associated with other phenotypes only when HLA was included in the genetic risk calculations (significant association with 79 phenotypes).

In the following, we describe in detail patterns that emerge across all studied disorders and highlight significant results for phenotype associations to genetic risk with at least three ADs.

3.1.1 Disease diagnoses

For six of the studied ADs (CEL, RA, MS, SLE, T1D, VITE), we observed a significant positive association of PRS to the same disease diagnosis (Table S3). These results indicate a good predictive power of the respective PRS. We should note that for PSC, the disease diagnosis phenotype was not available in the dataset. Additionally, we estimated the Nagelkerke's pseudo-R2 for all ADs (Table S4).

Celiac disease was found significantly associated with genetic risk for all 11 ADs that we studied. We observed that higher PRS for RA, VITE, VITL, JIA and PSO is associated with lower risk for the "Celiac disease" diagnosis phenotype. On the other hand, higher PRS for MS, MG, PSC, SLE, T1D and CEL was associated with
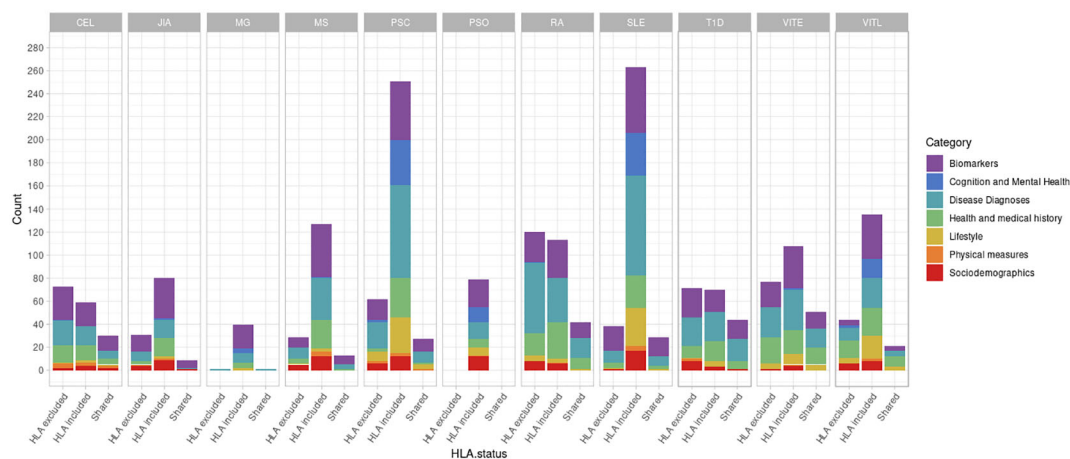
FIGURE 1

Number of significant phenotypes associated with autoimmune polygenic risk scores (p<10^{-5}). The different colors represent the general UK Biobank categories. The "HLA excluded" bar shows the number of significant associations with the phenotypes when HLA was excluded from the AD PRS calculations. The "HLA included" bar shows the number of significant associations with the phenotypes when HLA was included in the AD PRS calculations. The "Shared" bar shows the number of significant associations with the phenotypes for both HLA included or excluded AD PRSs.

TABLE 2 The most significant associations ($p_{FDR}<0.05$) of each AD PRS and the UK Biobank phenotypes.

Autoimmune Disorder	HLA included				HLA excluded			
	Top phenotype	β	95% Interval	Pfrd	Top phenotype	β	95% Interval	Pfdr
Celiac Disease	Celiac disease	0.650	[0.617 — 0.683]	<10-300	Celiac disease	0.536	[0.494 — 0.578]	7.17E-134
Juvenile Idiopathic Arthritis	Celiac disease	-0.364	[-0.407 — -0.32]	2.4E-57	Hypothyroidism NOS	0.111	[0.096 — 0.125]	1.52E-44
Multiple Sclerosis	Multiple sclerosis	0.498	[0.452 — 0.544]	1.58E-97	Multiple sclerosis	0.406	[0.352 — 0.460]	3.89E-46
Myasthenia Gravis	White blood cell (leukocyte) count	-0.029	[-0.033 — -0.026]	2.76E-59	Hypothyroidism NOS	0.036	[0.021 — 0.051]	7.85E-03
Primary Sclerosing Cholangitis	Celiac disease	0.699	[0.667 — 0.731]	<10-300	Eosinophill count	0.03	[0.027 — 0.034]	8.41E-63
Psoriasis	Psoriasis vulgaris	0.372	[0.336 — 0.407]	5.04E-90	no significant outcome			
Rheumatoid Arthritis	Rheumatoid arthritis	0.312	[0.288 — 0.337]	5.09E-131	Hypothyroidism NOS	0.203	[0.188 — 0.218]	1.99E-156
Systemic Lupus Erythematosus	Celiac disease	0.74	[0.707 — 0.772]	<10-300	Cystatin C	0.022	[0.018 — 0.025]	3.44E-37
Type 1 Diabetes	Type 1 diabetes	0.323	[0.286 — 0.360]	1.93E-62	Hypothyroidism NOS	0.151	[0.136 — 0.166]	3.63E-82
Vitiligo Early Onset	Skin colour	0.168	[0.161 — 0.176]	<10-300	Skin colour	0.256	[0.247 — 0.262]	<10-300
Vitiligo Early Onset	Ease of skin tanning	-0.136	[-0.143 — -0.13]	<10-300	Ease of skin tanning	-0.215	[-0.221 — -0.21]	<10-300
Vitiligo Late Onset	Skin colour	0.149	[0.141 — 0.156]	<10-300	Skin colour	0.264	[0.257 — 0.272]	<10-300
Vitiligo Late Onset	Ease of skin tanning	-0.123	[-0.129 — -0.117]	<10-300	Ease of skin tanning	-0.231	[-0.238 — -0.225]	<10-300

The table shows the strongest associated phenotypes with each AD PRS with and without HLA, the beta, the 95% CI and the FDR adjusted p-value.

higher risk with the "Celiac disease" diagnosis in the UK Biobank. The association with CEL, T1D, PSC, and RA remained significant even after excluding the HLA, although with an opposite effect direction for the last one (Figure 2; Table S2).

"Ulcerative colitis" diagnosis was the second digestive phenotype most commonly found associated with the autoimmune PRS, and high MS, RA(no-HLA), JIA(HLA), PSC(no-HLA), and CEL(no-HLA) PRS were associated with higher risk for the diagnosis (Figure 2, Table S2).

In the endocrine diagnoses, most autoimmune PRS were associated with "Hypothyroidism" followed by "Type 1 diabetes". RA, VITE, PSC, SLE, T1D and MG were associated with higher risk of "Hypothyroidism", even after HLA was excluded. JIA, VITL, MS and CEL association with Hypothyroidism was significant only after HLA was excluded (Figure 2; Table S2).

In dermatologic diagnoses, the autoimmune "Sicca syndrome" was the most associated phenotype with the autoimmune PRS. We observed a positive association of the "Sicca syndrome" diagnosis with SLE, CEL, RA(no-HLA), MS(HLA), PSC(HLA), and MG(HLA) PRS. In contrast, there was a negative association with PRS VITE, PSO, and VITL (Figure 2; Table S2).

In the neoplasms category, high PRS for VITL and VITE was associated with lower risk for skin cancer outcomes, including Non-Hodgkins lymphoma, other non-epithelial cancer of skin and melanomas of skin (Figure 2; Table S2).

3.1.2 Cognition and mental health

For PSC and SLE PRS, we found the largest overlap (n=20) of traits associated in the same direction. These associations included lower risk for phenotypes such as addictions, depression, and "low/worse" mental health, while they were positively associated with phenotypes describing higher cognitive function (Figure 3; Table S2). For PRS of MS, and MG, we also found an association with lower risk for phenotypes describing poor mental health (Figure 3; Table S2). On the contrary, higher PRS for VITL was associated with phenotypes describing poor mental health and depression (n=10), and had a negative association with phenotypes describing cognitive function (n=4) (Figure 3; Table S2). PSO and VITE associated with higher risk with phenotypes describing poor mental health and anxiety (Figure 3; Table S2).
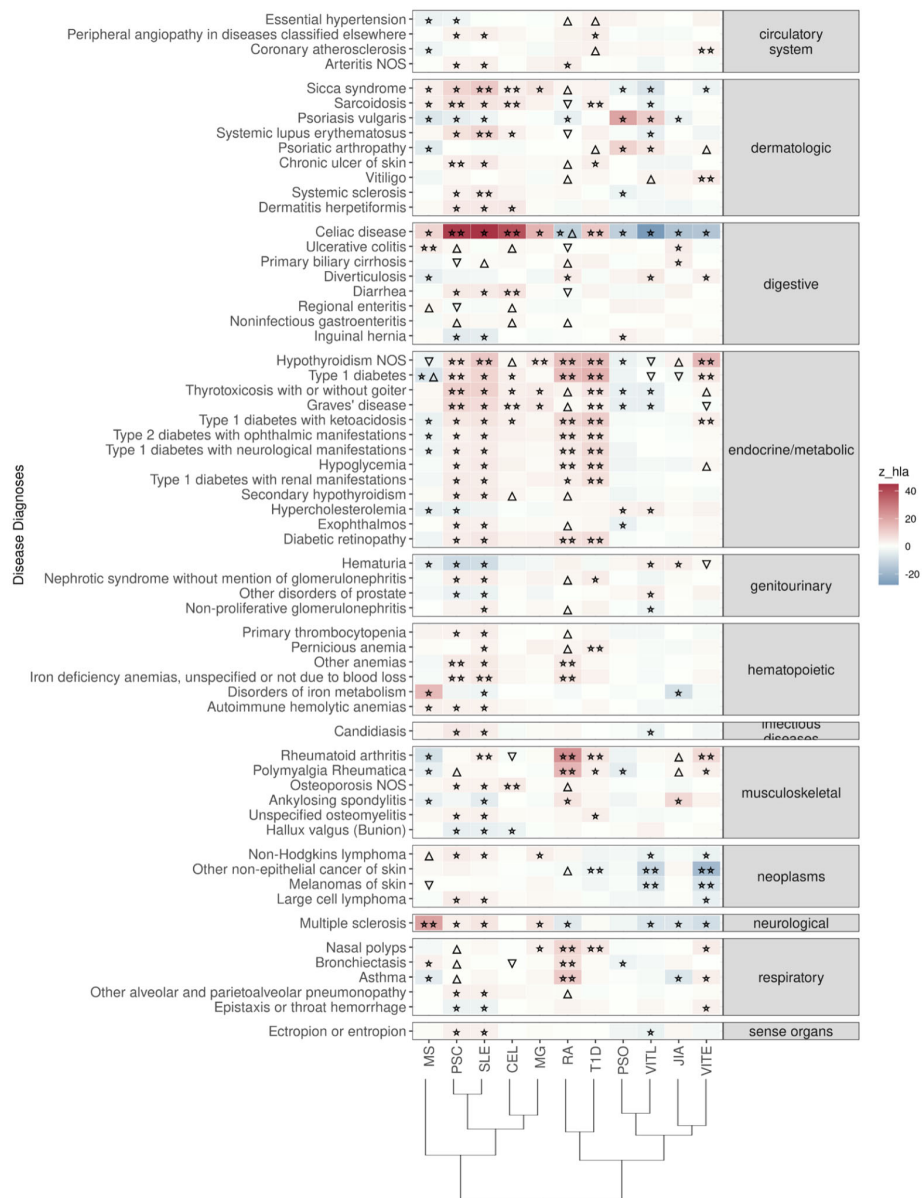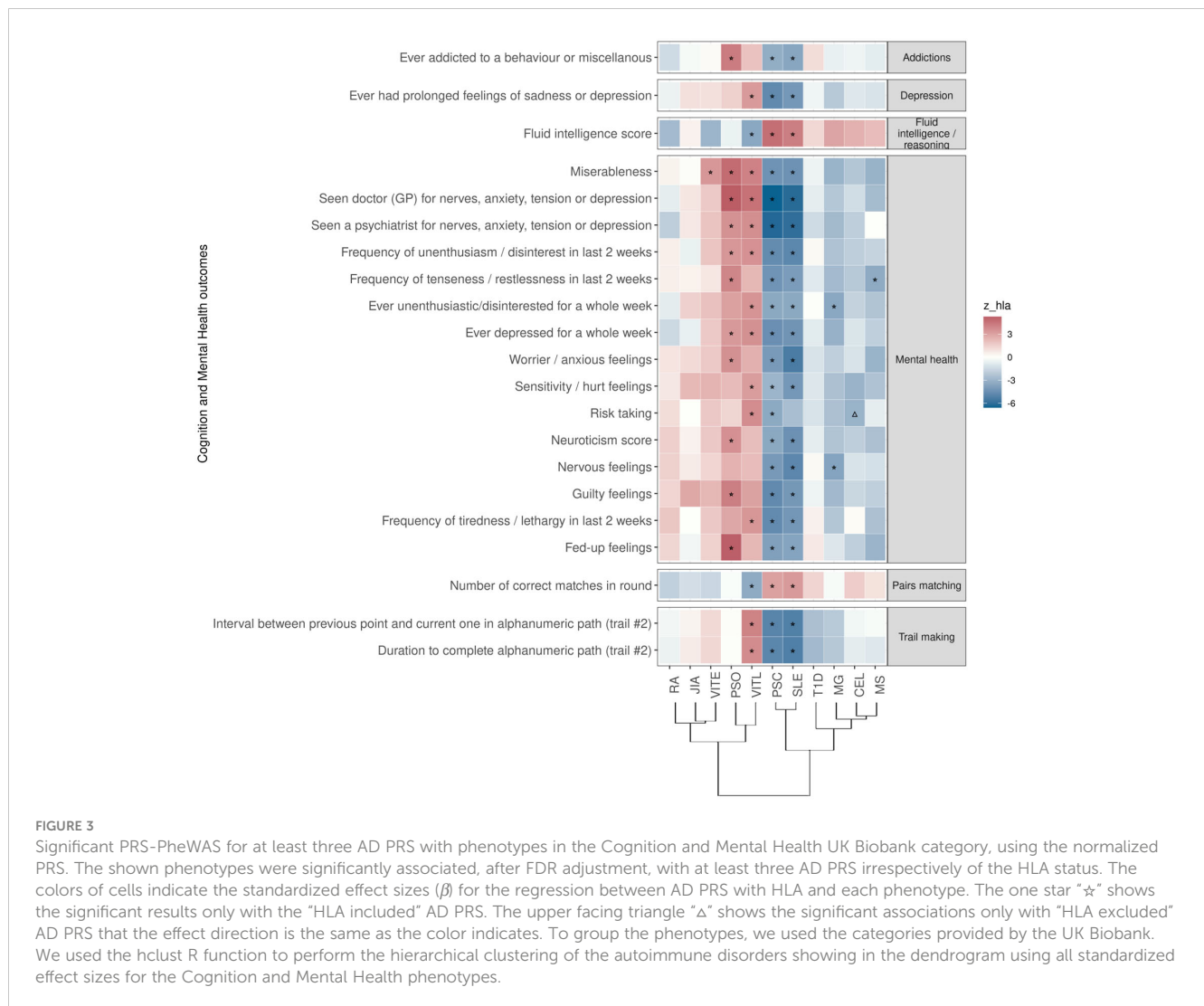
FIGURE 2
Significant PRS-PheWAS for at least three AD PRS with phenotypes in the Disease Diagnoses UK Biobank category, using the normalized PRS. The shown phenotypes were significantly associated, after FDR adjustment, with at least three AD PRS irrespectively of the HLA status. The colors of cells indicate the standardized effect sizes (β) for the regression between AD PRS with HLA and each phenotype. The one star "☆" shows the significant results only with the "HLA included" AD PRS. The two stars "☆☆" show the significant associations with both "HLA included or excluded" AD PRS with the same effect direction. The star and the upper facing triangle "☆△" show the significant associations with both "HLA included or excluded" AD PRS but with opposite effect directions. The upper facing triangle "△" shows the significant associations only with "HLA excluded" AD PRS that the effect direction is the same as the color indicates. The down-facing triangle "▽" shows the significant associations only with "HLA excluded" AD PRS that the effect direction is the opposite of what the color indicates. To group the disease diagnoses phenotypes, we used the R PheWAS tool and collapsed similar ICD-10 codes into one phecode. We used the hclust R function to perform the hierarchical clustering of the autoimmune disorders shown in the dendrogram using all standardized effect sizes for the disease diagnoses phenotypes.

3.1.3 Lifestyle

In this category the trait "Never eat eggs, dairy, wheat, sugar: Wheat products" was associated with PRS for ten ADs when HLA was included in analysis, suggesting susceptibility to food allergies; VITL, VITE, PSO, JIA, RA are negatively associated with the phenotype, while PSC, SLE, MG, CEL (irrespectively of HLA) and T1D were positively associated with the phenotype (Figure 4; Table S2).

Again, same as for the previous category of traits, PSC and SLE PRS had the largest overlap of associated phenotypes (n=20) in the same effect direction. They were negatively associated with phenotypes related to dietary habits (higher intake of dried fruit, salad/raw vegetable, non-oily fish), cannabis usage, exercise, smoking status (Figure 4; Table S2).

Additionally, VITL and VITE PRS (irrespectively of HLA) were positively associated with darker skin color, and negatively

FIGURE 3

Significant PRS-PheWAS for at least three AD PRS with phenotypes in the Cognition and Mental Health UK Biobank category, using the normalized PRS. The shown phenotypes were significantly associated, after FDR adjustment, with at least three AD PRS irrespectively of the HLA status. The colors of cells indicate the standardized effect sizes (β) for the regression between AD PRS with HLA and each phenotype. The one star "☆" shows the significant results only with the "HLA included" AD PRS. The upper facing triangle "△" shows the significant associations only with "HLA excluded" AD PRS that the effect direction is the same as the color indicates. To group the phenotypes, we used the categories provided by the UK Biobank. We used the hclust R function to perform the hierarchical clustering of the autoimmune disorders showing in the dendrogram using all standardized effect sizes for the Cognition and Mental Health phenotypes.

associated with higher risk of "ease of skin tanning", "childhood sunburn occasions", "use of sun/uv protection" and "facial aging" (Figure 4; Table S2). We observed the opposite associations between PSC and SLE PRS and these sun exposure phenotypes (Figure 4; Table S2).

3.1.4 Health and medical history

In this category the self-reported phenotype "Diagnosed with coeliac disease or gluten sensitivity" was significantly associated with 11 autoimmune PRS (Figure S5; Table S2). We observed a positive association with PSC, SLE, T1D, CEL, MG, and a negative association with RA, JIA, VITL, VITE, PSO, these results are consistent with similar phenotypes, such as the "Celiac disease" diagnosis and the "Never eat eggs, dairy, wheat, sugar: Wheat products" phenotype in the Lifestyle category.

Additionally, high PRS T1D, VITL, and VITE were associated with lower risk for "Basal cell carcinoma" phenotype under the Cancer register sub-category. Specifically, for VITE and VITL we observed a negative association with self-reported basal cell carcinoma (Figure S5; Table S2).

We also observed significant associations of autoimmune PRS with phenotypes in the Sociodemographics (Figure S6), Biomarkers (Figures S7-S9) and Physical Measures (Figure S10) categories, without any patterns emerging across disorders. Results are shown in the supplement.

3.2 AD Genetic Architecture and Cross-Disorder GWAS Meta-analysis

Driven by the known comorbidity across AD [based on epidemiological studies (8)] and the overlap in phenotypic associations with autoimmune PRS that we described above, we proceeded to perform cross-disorder genetic correlation and GWAS summary statistics meta-analyses to explore the genetic relationship and genetic architecture of ADs and identify potentially pleiotropic loci. Such pleiotropic loci would drive pathophysiology across multiple ADs.

Initially, we performed analysis for all 11 ADs (Figure S11), however, given the limited SNP overlap of our datasets for CEL,
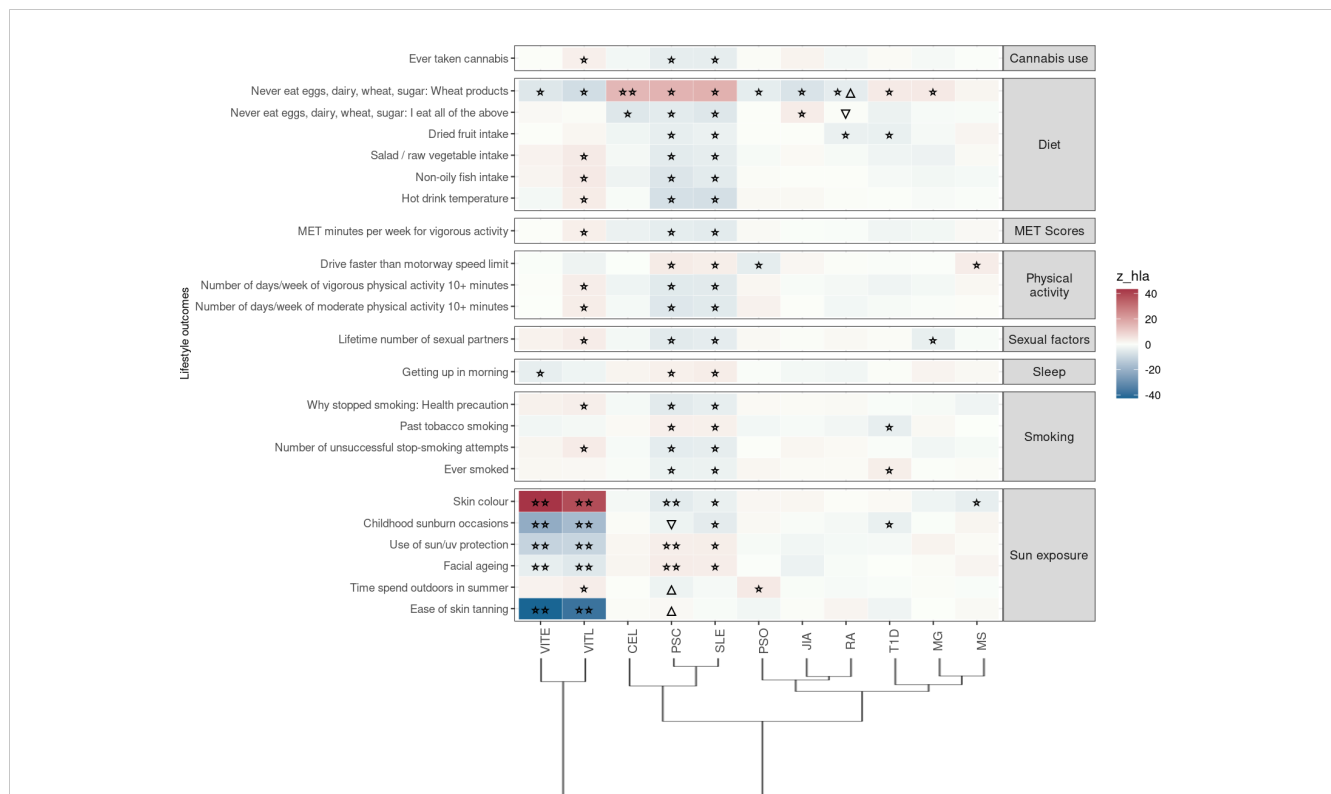
FIGURE 4
Significant PRS–PheWAS for at least three AD PRS with phenotypes in the Lifestyle UK Biobank category, using the normalized PRS. The shown phenotypes were significantly associated, after FDR adjustment, with at least three AD PRS irrespectively of the HLA status. The colors of cells indicate the standardized effect sizes (β) for the regression between AD PRS with HLA and each phenotype. The one star "☆" shows the significant results only with the "HLA included" AD PRS. The two stars "☆☆" show the significant associations with both "HLA included or excluded" AD PRS with the same effect direction. The star and the upper facing triangle "☆△" show the significant associations with both "HLA included or excluded" AD PRS but with opposite effect directions. The upper facing triangle "△" shows the significant associations only with "HLA excluded" AD PRS that the effect direction is the same as the color indicates. To group the phenotypes, we used the categories provided by the UK Biobank. We used the hclust R function to perform the hierarchical clustering of the autoimmune disorders shown in the dendrogram using all standardized effect sizes for the Lifestyle category phenotypes.

PSO and JIA we excluded them from further analyses. After correction for multiple testing, we observed significant positive correlations of RA with T1D (rg_{no-HLA}:0.52), SLE (rg_{no-HLA}:0.51), and MG (rg_{no-HLA}:0.47). VITL and VITE were also significantly correlated (rg_{no-HLA}:0.64). Additional autoimmune correlations with p<0.05 are shown in Figures 5A, B, including the pairwise correlations after excluding HLA.

Since the pairwise genetic correlation analysis showed a complicated correlation pattern among the studied disorders, we performed exploratory factor analysis (EFA) followed by a confirmatory factor analysis (CFA) to dissect the AD relationships. We used the four-factor model identified in EFA and included the disorders with loadings greater than |0.3| in each factor. The CFA analysis in GenomicSEM showed a good fit of the model to the data (χ^2 (12) =16.1; AIC =64.1; CFI = 0.98; SRMR = 0.07). The first factor included VITE, VITL and MG. MG, RA and SLE were included in the second factor, while the third factor consisted of T1D, PSC and MG (with a negative loading). Lastly, factor four consisted of PSC and MS (Figure 5C).

In the cross-disorder meta-analysis on the first factor, that includes VITL, VITE and MG, we identified nine significant pleiotropic (m-value>0.9 in all studies) LD independent regions

(Table 3); two of them mapping on the *Cytotoxic T-Lymphocyte Associated Protein 4* (CTLA4) - *Inducible T Cell Costimulator* (ICOS) and Fli-1 Proto-Oncogene, ETS Transcription Factor (FLI1) genes, were not significant in the individual GWAS studies included here. However in the GWAS catalog (41) FLI1 has a significant association with Vitiligo, when the onset age is not taken into account (42), and CTLA4 is found associated with different GWASs (not studied here) for both Myasthenia gravis (43) and Vitiligo (42). CTLA4 was also significant in the gene-based analysis (44) of the data we used in this meta-analysis. The gene set enrichment analysis including the genes in the significant and pleiotropic regions identified four significantly enriched GO : BP terms; bone cell development, immune system development, myeloid cell development, and immune system process (Figure 6A; Table S5).

When we performed the meta-analysis of MG, RA and SLE, we identified 17 genome-wide significant pleiotropic loci. Three of these loci mapping to *Protein Tyrosine Phosphatase Receptor Type C* (PTPRC), *Interleukin 12 Receptor Subunit Beta 2* (IL12RB2) and LINC00824 were not genome-wide significant in the GWAS studies we analyzed (focusing on European ancestry), however, they were reported as significant associations in GWAS of higher power that

were multi-ethnic (Table S6). The gene set enrichment analysis of genome-wide significant and pleiotropic regions identified 22 significantly enriched terms. Among them, six (DN2 thymocyte differentiation, regulation of MAP kinase activity, stress granule assembly, B cell proliferation, side of membrane, GRB7 events in ERBB2 signaling) were significant even after excluding the IEA GO terms (see Methods) (Figure 6B; Table S7).

In the meta-analysis of MG, T1D and PSC, we observed seven pleiotropic and genome-wide significant loci. One of them, found on chr4:10,709,726-10,726,520 (closest gene *Cytokine Dependent Hematopoietic Cell Linker* (CLNK), 23Kb downstream) has not been previously found to be associated with either of the three

disorders (Table S8). The gene-set enrichment analysis identified six significantly enriched terms after multiple testing correction, with "RUNX1 and FOXP3 control the development of regulatory T lymphocytes (Tregs)" and "T cell receptor signaling pathway" as the two top terms (Table S9). These two were the only significant terms when we repeated the analysis after excluding the IEA GO terms (see Methods) (Figure 6C; Table S9).

Finally, for the cross-disorder meta-analysis of PSC and MS, we identified two genome-wide significant and pleiotropic loci, mapping to the previously associated *Interleukin 2 Receptor Subunit Alpha* (IL2RA) and *BTB Domain And CNC Homolog 2* (BACH2) genes (Table S10). The gene-set enrichment analysis
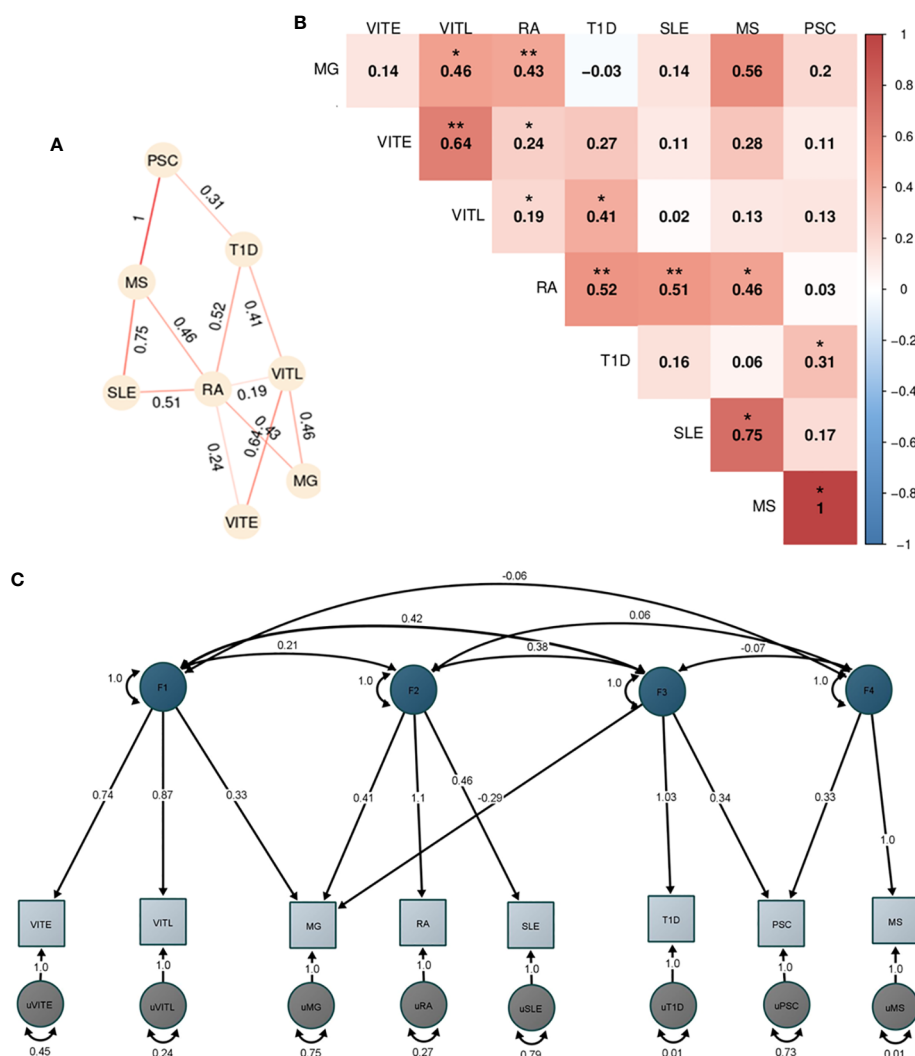

FIGURE 5
Genetic correlation and factor analysis for 8 autoimmune disorders. The figure shows the analyses of the 8 autoimmune disorders with enough overlap (>200.000 SNPs) with HapMap3 data provided by LDSC after excluding the HLA locus (hg19, chr6 25-33 Mb). **(A)** Heatmap of the pairwise LDSC genome wide genetic correlations of the 8 autoimmune disorders after excluding the SNPs in the HLA region. The red color reflects more positive correlation coefficients while blue reflects more negative coefficients, and the numbers within each cell are the correlation coefficients. The correlations with p<0.05 are denoted with one asterisk (*), while two asterisks (**) show the correlations that are significant after Bonferroni correction. **(B)** Network representation of the genetic correlation between the autoimmune disorders with p<0.05. The numbers show the correlation coefficient and the stronger the line color shows a higher coefficient. **(C)** Path graph of the confirmatory factor model estimated using the Genomic SEM. Four factors were identified. The factor loadings for each trait are depicted by arrows between the trait and the factor, with the standardized loading value and the standard error in the parentheses. Correlation between factors is indicated by arrows between them. Residual variance for each trait is indicated by the two-headed arrow connecting the variable to itself.

TABLE 3 Genome-wide significant (p<5x10^{-8}) LD independent loci from the VITL-VITE-MG meta-analysis.

SNP	Locus	Nearest genes (<20kb)	P-meta	OR	SE	Top SNP VITE	P-VITE	Top SNP VITL	P-VITL	Top SNP MG	P-MG
rs9981704	chr21:43831955-43867059	TMPRSS3 UBASH3A	6.09x10^{-19}	1.31	0.03	–	4.30x10^{-4}	rs12482396	4.06x10^{-20}	–	3.60x10^{-3}
rs60946162	chr3:188084682-188133518	LPP	1.78x10^{-15}	0.8	0.03	–	2.43x10^{-7}	rs13098877	2.62x10^{-11}	–	1.17x10^{-3}
rs7137828	chr12:111833788-112037526	ATXN2 SH2B3	1.70x10^{-13}	0.82	0.03	–	7.31x10^{-5}	rs10774624	8.28x10^{-9}	–	1.26x10^{-3}
rs8088891	chr18:60007263-60029292	TNFRSF11A	3.51x10^{-13}	1.22	0.03	–	9.89x10^{-4}	–	1.14x10^{-3}	rs4369774	1.09x10^{-13}
rs1951459	chr6:167370353-167455629	FGFR1OP MIR3939 RNASET2	3.61x10^{-13}	0.81	0.03	rs2247315	3.48x10^{-9}	rs366938	1.14x10^{-8}	–	9.53x10^{-4}
rs64547	chr22:37575469-37595156	C1QTNF6 SSTR3	8.14x10^{-13}	0.82	0.03	rs229528	2.09x10^{-8}	rs229527	1.55x10^{-12}	–	7.33x10^{-3}
rs831071	chr3:71426419-71430389	FOXP1	1.01x10^{-10}	1.2	0.03	–	2.11x10^{-3}	rs60135207	3.14x10^{-9}	–	3.40x10^{-6}
rs3116513	chr2:204694611-204792732	CTLA4 ICOS	1.42x10^{-10}	1.19	0.03	–	2.25x10^{-2}	–	5.64x10^{-5}	–	2.65x10^{-6}
rs644515	chr11:128589472-128617231	FLI1	1.44x10^{-9}	1.17	0.03	–	2.89x10^{-5}	–	2.89x10^{-5}	–	1.24x10^{-2}

The column SNP contains the top SNP in each locus. The columns P, OR and SE correspond to the top SNP in each locus. The autoimmune disorder specific Top SNP column contains the top genome-wide significant SNP in the locus that was available in the input dataset. The autoimmune disorder specific P column contains the lowest p-value in the locus that was available in the input dataset.

identified three significant terms, primary adaptive immune response, primary adaptive immune response involving T cells and B cells, and interleukin-2 receptor complex. All of them remained significant even after we excluded the IEA GO terms (Table S11).

3.3 PheWAS findings shared across ADs in the same genetic factor

Finally, we explored whether the ADs belonging to each of the four factors that were identified by EFA, also share associations with phenotypes detected in the PheWAS. For MG, VITE and VITL which make up the first identified factor, we detected 29 shared phenotypes across all three ADs (Figure S12). However, there was no phenotype with the same effect direction for all three ADs. For MG, RA and SLE, which make up the second factor, we detected 26 shared phenotypes, across five categories (Figure S13). The Hypothyroidism disease diagnosis and the health related phenotype of "other serious medical conditions/disability diagnosed by doctor", were the two phenotypes, associated with all three ADs with the same effect direction, for the same HLA status. For the ADs of the third factor (PSC, T1D and MG), we identified 16 shared phenotypes in four categories (Figure S14). Nine of them had the same effect direction for all three AD for the same HLA status. Lastly, for factor four (PSC and MS), we identified 76 shared associations, 43 of them had the same effect direction for both ADs (Figure S15).

4 Discussion

We report results on the first PRS-PheWAS analysis exploring the association of genetic risk for 11 autoimmune disorders and 3,254 phenotypes on 330,841 individuals of European ancestry from the UK Biobank. Additionally, we explored the genetic relationship between the studied ADs seeking to dissect the genetic architecture of these highly correlated and often comorbid phenotypes.

We were able to recover previously identified associations between ADs based on epidemiological or genetic studies. For instance, a study in a Taiwanese population showed higher risk of first-degree relatives with Sicca to develop other autoimmune disorders including SLE, MS, MG, and RA (45); and we also observed here a positive association between the PRS of these four ADs and Sicca syndrome outcome. Other autoimmune-related diagnosis outcomes associated with higher risk for the studied ADs, included hypothyroidism and Graves disease. A link between those disorders and RA, Vitiligo, SLE, T1D, CEL, and MG, is also supported by the literature (46–50). Additionally, as reported in previous studies (51–53), we also observed a negative association between Vitiligo risk and skin cancer.

Interestingly, we observed many associations with environmental and lifestyle factors. Diet and specifically the consumption of non-wheat products was the outcome that we found to be associated with the risk for most of the studied ADs pointing to gluten intolerance and food allergies. We observed a significant positive association between PSC, SLE, CEL, MG, T1D, and RA (when HLA was excluded from PRS calculations) and not
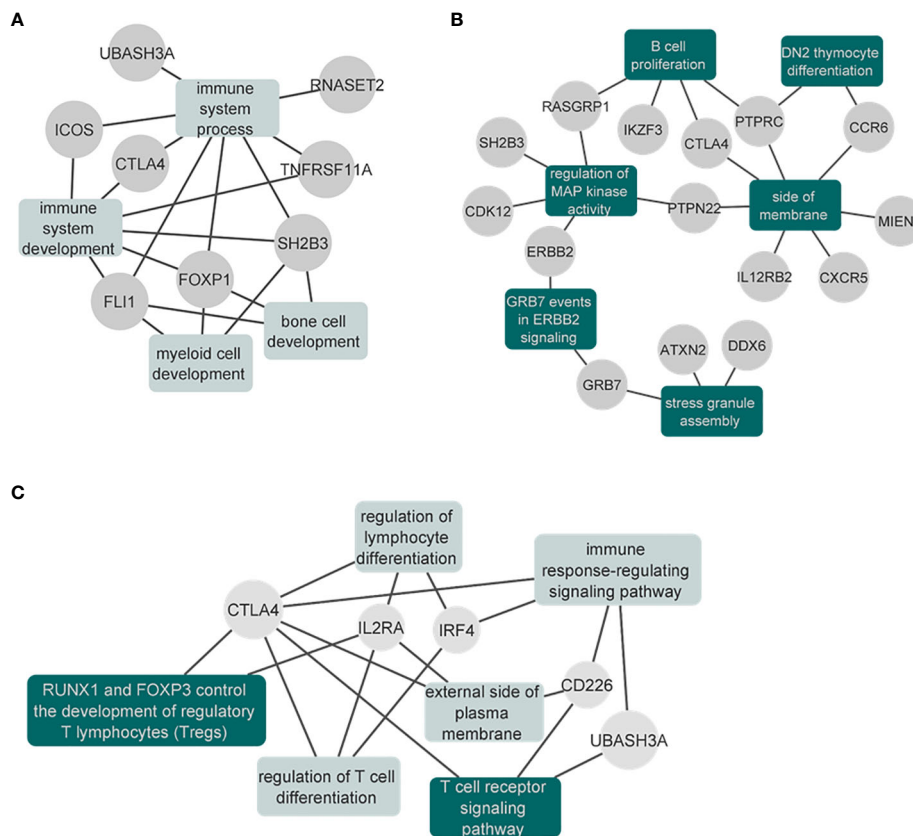
FIGURE 6
Network plots of the enrichment analysis for the cross-disorder meta-analyses. **(A)** Results of the significantly enriched terms from the genes identified in the VITE-VITL-MG meta-analysis. Results are also shown in Supplementary Table 5. **(B)** Results of the significantly enriched terms (after excluding the IEA terms) from the genes identified in the SLE-MG-RA meta-analysis. The full results are also shown in the Supplementary Table 7. **(C)** Results of the significantly enriched terms from the genes identified in the T1D-MG-PSC meta-analysis. The full results are also shown in the Supplementary Table 9. Enriched gene sets that remained significant after excluding the IEA GO terms are shown in dark green.

consuming wheat, while the association was negative for VITE, VITL, PSO, JIA and RA. We also observed the same pattern of AD associations with the Celiac disease diagnosis phenotype, which could be indicative of the connection between gluten intolerance and Celiac disease. There is prior evidence suggesting that a gluten-free diet could be beneficial not only for patients with Celiac disease but also for T1D, RA, MS, autoimmune hepatitis, and PSO (54).

Smoking was another factor that we found significantly associated with SLE and PSC genetic risk. Indeed, it has been previously suggested that smoking is associated with higher risk for double-stranded DNA seropositivity, a marker used for SLE diagnosis, in SLE patients (55), while for PSC, there is some evidence to suggest that smoking is associated with lower risk for developing the disease (56, 57), although not always consistently supported (58). Interestingly, a previous study found that severe sunburn incidents and higher tanning ability in women are associated with higher risk of developing Vitiligo (59), however, we actually observed the opposite association for Vitiligo genetic risk, perhaps indicating different behavior towards sun exposure based on genetic risk. We also observed a negative association between CEL genetic risk and the weight of the first child. Previous studies have shown that women with undiagnosed or untreated

celiac disease have higher risk to deliver a baby with reduced birthweight (60).

The link between autoimmune disorders and mental health has been previously described. For instance, exposure to stress-related disorders was found to be associated with higher risk for ADs (61), and both positive and negative associations of ADs with psychotic disorders have been summarized elsewhere (22). In our study, we observed that risk for VITL and PSO was positively associated with self-reported outcomes describing poor mental health, which is in line with previous works (22, 61–63). We also observed that risk for SLE and PSC was associated with better mental health outcomes. Epidemiological studies have reported higher psychological distress in patients with SLE and PSC (61, 64–67). However, in a study exploring the genetic correlation between immune and psychiatric related phenotypes using GWAS summary statistics, SLE was found to be significantly positively correlated only with Schizophrenia and no other psychiatric phenotypes (68). Additionally, a study using Mendelian Randomization between SLE and depression showed SLE genetic variants mildly reduce the odds of depression, suggesting that the observed association between SLE and depression might not be attributed to genetic factors (69). Thus, further analyses could be useful to explore the gap between the

associations between SLE and mental health phenotypes observed in epidemiological studies, but not when using genetic data.

Exploring closer the phenotypes that are associated with ADs belonging to the same genetic factor (as identified by EFA), we found that for factor three (PSC, T1D and MG) there is a clear pattern of associations with same direction of effect with phenotypes belonging in several disease diagnoses, health and medical history, and lifestyle categories as well as biomarkers. For two different factors, two and three, we observed the same direction of association with the Hypothyroidism diagnosis phenotype. Specifically for the ADs of factor two we detected a positive association with additional AD diagnoses, which we discussed in more details earlier here. Interestingly, for disorders in the first factor (VITE, VITL and MG), we did not observe any phenotype associated with the same effect direction with all three ADs. This suggests that although Vitiligo and MG are genetically correlated, their PRSs are associated with the opposite direction with the studied phenotypes. The only exception to this was the association that we observed with the Hypothyroidism disease diagnosis where VITE and MG PRS show association in the same direction. On the other hand, for PSC and MS (factor four), we observed positive associations with other ADs such as Sicca syndrome, Celiac disease, but not with Hypothyroidsm. Overall, the shared phenotypes in each factor reveal patterns whose link to ADs warrants further exploration.

It is well demonstrated that ADs are often comorbid and share both HLA and non-HLA genetic loci (8–10, 15). In a recent study (11), where the genetic correlation between 13 (7 of them are also studied here) autoimmune and inflammatory disorders was also explored, the authors observed correlations across ADs and similar patterns to what we also found. Furthermore, we also provide here a more detailed analysis to understand the genetic architecture of ADs including EFA to reveal subgroups of disorders and cross-disorder GWAS to reveal pleiotropic loci that could underlie multiple disorders and drive comorbidities. Indeed, in line with the existing notion of shared genetic background across ADs, we detected numerous genome-wide significant and pleiotropic loci in each meta-analysis. All except one had already been previously associated with at least one of the ADs included in the meta-analysis, or were associated with the traits in studies of different ancestries or larger sample GWAS which we could not analyze here because summary statistics data were not available. Importantly, we identify one novel genome-wide significant and pleiotropic locus in the meta-analysis of T1D-MG-PSC. This is a previously unknown locus that could play a role in the etiology of all three disorders and is found 23Kb downstream of *CLNK* gene that encodes Clnk, an adapter of the SLP76 family, is involved in the regulation of immunoreceptor signaling (70).

This study comes with both strengths and limitations. The PheWAS analysis allowed us to detect significant associations between AD risk and multiple phenotypes, even after excluding the HLA region. Additionally, we were able to detect pleiotropic loci in the autoimmune subgroups that are involved in immune-related processes as the gene-set enrichment analysis revealed. However, there are limitations in this study that should be considered when interpreting the results. For the PRS calculations, although we used

the largest AD summary statistics data available, there were differences in power regarding their sample size and number of SNPs. Also, as the number of UK Biobank participants with AD diagnoses is limited, we were not able to calculate the optimal p-value threshold for SNPs to be included in PRS calculations, but rather set as threshold the p-value 10^{-5}.

In conclusion, in this study we observed ADs PRS to be associated with multiple health-related and environmental factors, even after excluding the HLA region, and explored the genetic relationships of the selected ADs by estimating their genetic correlation and identifying pleiotropic genetic regions that underlie genetic risk across multiple ADs. Overall, our analyses indicate potential factors associated with genetic risk for ADs, some of which have been reported previously, and novel observations that need further exploration. These results suggest that the assessment of additional exposures related to lifestyle, mental and physical health risks by clinicians, could be beneficial for individuals with higher risk for autoimmune disorders.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding author.

Author contributions

PP designed the study, all authors contributed to the data analysis and drafting the manuscript.

Conflict of interest

KP is Founder and Head of the Institute for Research and Innovation. Based at Patras Science Park.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the

reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2023.1147573/full#supplementary-material

SUPPLEMENTARY TABLE 1
Demographic information of the 330,841 UK Biobank participants included in the analysis. The ICD10 Disease Diagnoses category includes the diagnoses for the autoimmune disorders that are included in this study and were present in the UK Biobank.

SUPPLEMENTARY TABLE 2
PRS-PheWAS results for association of genetic risk of 11 autoimmune disorders with 3,254 phenotypes in UK Biobank. Genetic risk scores were calculated as the weighted standardized sum of the effect of independent SNPs with p-values$<10^{-5}$ for each disorder. The estimation of the genetic risk scores was repeated after excluding the extended HLA region (hg19, chr6 25-33 Mb). Estimates were generated by PHESANT.

SUPPLEMENTARY TABLE 3
Associations of AD PRS with disease diagnosis phenotype for the same disorder. The table shows the associations of the disease diagnosis phenotypes and the same AD PRS with and without HLA, only for the ADs that the same diagnosis was available. For the Vitiligo early and late onset we used the general Vitiligo diagnosis phenotype that was available in the UK Biobank.

SUPPLEMENTARY TABLE 4
Nagelkerke's pseudo-R2 values for the studied ADs. The table shows the Nagelkerke's pseudo-R2 values and the liability scale Nagelkerke's pseudo-R2, using the population prevalence as listed in the Prevalence column. The PMID column refers to the studies reporting the prevalence values included in the analysis.

SUPPLEMENTARY TABLE 5
Significantly enriched gene sets genes identified in the VITL-VITE-MG meta-analysis. The p.Val column is the adjusted p-value using the suggested g:SCS

method by the g:Prolifer tool used for the analysis. The Genes column contains the provided genes that are in each identified gene set.

SUPPLEMENTARY TABLE 6
Genome-wide significant (p<5x10-8) LD independent loci from the SLE-RA-MG meta-analysis. The column SNP contains the top SNP in each locus. The columns P, OR and SE correspond to the top SNP in each locus. The autoimmune disorder specific Top SNP column contains the top genome-wide significant SNP in the locus that was available in the input dataset. The autoimmune disorder specific P column contains the lowest p-value in the locus that was available in the input dataset.

SUPPLEMENTARY TABLE 7
Significantly enriched gene sets genes identified in the SLE-MG-RA meta-analysis. The p.Val column is the adjusted p-value using the suggested g:SCS method by the g:Prolifer tool used for the analysis. The Genes column contains the provided genes that are in each identified gene set.

SUPPLEMENTARY TABLE 8
Genome-wide significant (p<5x10-8) LD independent loci from the T1D-MG-PSC meta-analysis. The column SNP contains the top SNP in each locus. The columns P, OR and SE correspond to the top SNP in each locus. The autoimmune disorder specific Top SNP column contains the top genome-wide significant SNP in the locus that was available in the input dataset. The autoimmune disorder specific P column contains the lowest p-value in the locus that was available in the input dataset.

SUPPLEMENTARY TABLE 9
Significantly enriched gene sets genes identified in the T1D-MG-PSC meta-analysis. The p.Val column is the adjusted p-value using the suggested g:SCS method by the g:Prolifer tool used for the analysis. The Genes column contains the provided genes that are in each identified gene set.

SUPPLEMENTARY TABLE 10
Genome-wide significant (p<5x10-8) LD independent loci from the PSC-MS meta-analysis. The column SNP contains the top SNP in each locus. The columns P, OR and SE correspond to the top SNP in each locus. The autoimmune disorder specific Top SNP column contains the top genome-wide significant SNP in the locus that was available in the input dataset. The autoimmune disorder specific P column contains the lowest p-value in the locus that was available in the input dataset.

SUPPLEMENTARY TABLE 11
Significantly enriched gene sets genes identified in the PSC-MS meta-analysis. The p.Val column is the adjusted p-value using the suggested g:SCS method by the g:Prolifer tool used for the analysis. The Genes column contains the provided genes that are in each identified gene set.

References

1. Hayter SM, Cook MC. Updated assessment of the prevalence, spectrum and case definition of autoimmune disease. *Autoimmun Rev* (2012) 11:754–65. doi: 10.1016/j.autrev.2012.02.001

2. Wang L, Wang F-S, Gershwin ME. Human autoimmune diseases: a comprehensive update. *J Intern Med* (2015) 278:369–95. doi: 10.1111/joim.12395

3. Cooper GS, Stroehla BC. The epidemiology of autoimmune diseases. *Autoimmun Rev* (2003) 2:119–25. doi: 10.1016/S1568-9972(03)00006-5

4. Mayer-Davis EJ, Lawrence JM, Dabelea D, Divers J, Isom S, Dolan L, et al. Incidence trends of type 1 and type 2 diabetes among youths, 2002-2012. *N Engl J Med* (2017) 376:1419–29. doi: 10.1056/NEJMoa1610187

5. Fatoye F, Gebrye T, Svenson LW. Real-world incidence and prevalence of systemic lupus erythematosus in Alberta, Canada. *Rheumatol Int* (2018) 38:1721–6. doi: 10.1007/s00296-018-4091-4

6. Dinse GE, Parks CG, Weinberg CR, Co CA, Wilkerson J, Zeldin DC, et al. Increasing prevalence of antinuclear antibodies in the United States. *Arthritis Rheumatol* (2020) 72:1026–35. doi: 10.1002/art.41214

7. Wu X, Chen H, Xu H. The genomic landscape of human immune-mediated diseases. *J Hum Genet* (2015) 60:675–81. doi: 10.1038/jhg.2015.99

8. Somers EC, Thomas SL, Smeeth L, Hall AJ. Are individuals with an autoimmune disease at higher risk of a second autoimmune disorder? *Am J Epidemiol* (2009) 169:749–55. doi: 10.1093/aje/kwn408

9. Klein J, Sato A. The HLA system. Second of two parts. *N Engl J Med* (2000) 343:782–6. doi: 10.1056/NEJM200009143431106

10. Cho JH, Gregersen PK. Genomics and the multifactorial nature of human autoimmune disease. *N Engl J Med* (2011) 365:1612–23. doi: 10.1056/NEJMra1100030

11. Lincoln MR, Connally N, Axisa P-P, Gasperi C, Mitrovic M, van Heel D, et al. Joint analysis reveals shared autoimmune disease associations and identifies common mechanisms. *bioRxiv* (2021). doi: 10.1101/2021.05.13.21257044

12. Shirai Y, Nakanishi Y, Suzuki A, Konaka H, Nishikawa R, Sonehara K, et al. Multi-trait and cross-population genome-wide association studies across autoimmune and allergic diseases identify shared and distinct genetic component. *Ann Rheum Dis* (2022) 81:1301–12. doi: 10.1136/annrheumdis-2022-222460

13. Márquez A, Kerick M, Zhernakova A, Gutierrez-Achury J, Chen W-M, Onengut-Gumuscu S, et al. Meta-analysis of Immunochip data of four autoimmune diseases reveals novel single-disease and cross-phenotype associations. *Genome Med* (2018) 10:97. doi: 10.1186/s13073-018-0604-8

14. Cotsapas C, Voight BF, Rossin E, Lage K, Neale BM, Wallace C, et al. Pervasive sharing of genetic effects in autoimmune disease. *PloS Genet* (2011) 7:e1002254. doi: 10.1371/journal.pgen.1002254

15. Li YR, Li J, Zhao SD, Bradfield JP, Mentch FD, Maggadottir SM, et al. Meta-analysis of shared genetic architecture across ten pediatric autoimmune diseases. *Nat Med* (2015) 21:1018–27. doi: 10.1038/nm.3933

16. González-Serna D, Ochoa E, López-Isac E, Julià A, Degenhardt F, Ortego-Centeno N, et al. A cross-disease meta-GWAS identifies four new susceptibility loci shared between systemic sclerosis and Crohn's disease. *Sci Rep* (2020) 10:1862. doi: 10.1038/s41598-020-58741-w

17. Ellinghaus D, Jostins L, Spain SL, Cortes A, Bethune J, Han B, et al. Analysis of five chronic inflammatory diseases identifies 27 new associations and highlights disease-specific patterns at shared loci. *Nat Genet* (2016) 48:510–8. doi: 10.1038/ng.3528

18. Ercolini AM, Miller SD. The role of infections in autoimmune disease. *Clin Exp Immunol* (2009) 155:1–15. doi: 10.1111/j.1365-2249.2008.03834.x

19. Bjornevik K, Cortese M, Healy BC, Kuhle J, Mina MJ, Leng Y, et al. Longitudinal analysis reveals high prevalence of Epstein-Barr virus associated with multiple sclerosis. *Science* (2022) 375:296–301. doi: 10.1126/science.abj8222

20. Rosenblum MD, Remedios KA, Abbas AK. Mechanisms of human autoimmunity. *J Clin Invest* (2015) 125:2228–33. doi: 10.1172/JCI78088

21. Harris-Tryon TA, Bel S. Gene/environment interaction and autoimmune disease. In: Teperino R, editor. *Beyond Our Genes: Pathophysiology of Gene and Environment Interaction and Epigenetic Inheritance*. Cham: Springer International Publishing (2020). p. 139–56.

22. Jeppesen R, Benros ME. Autoimmune diseases and psychotic disorders. *Front Psychiatry* (2019) 10:131. doi: 10.3389/fpsyt.2019.00131

23. Krishna MT, Subramanian A, Adderley NJ, Zemedikun DT, Gkoutos GV, Nirantharakumar K. Allergic diseases and long-term risk of autoimmune disorders: longitudinal cohort study and cluster analysis. *Eur Respir J* (2019) 54. doi: 10.1183/13993003.00476-2019

24. Versini M, Jeandel P-Y, Rosenthal E, Shoenfeld Y. Obesity in autoimmune diseases: not a passive bystander. *Autoimmun Rev* (2014) 13:981–1000. doi: 10.1016/j.autrev.2014.07.001

25. Choi SW, Mak TS-H, O'Reilly PF. Tutorial: a guide to performing polygenic risk score analyses. *Nat Protoc* (2020) 15:2759–72. doi: 10.1038/s41596-020-0353-1

26. Pendergrass SA, Brown-Gentry K, Dudek SM, Torstenson ES, Ambite JL, Avery CL, et al. The use of phenome-wide association studies (PheWAS) for exploration of novel genotype-phenotype relationships and pleiotropy discovery. *Genet Epidemiol* (2011) 35:410–22. doi: 10.1002/gepi.20589

27. Denny JC, Ritchie MD, Basford MA, Pulley JM, Bastarache L, Brown-Gentry K, et al. PheWAS: demonstrating the feasibility of a phenome-wide scan to discover gene-disease associations. *Bioinformatics* (2010) 26:1205–10. doi: 10.1093/bioinformatics/btq126

28. 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. doi: 10.1186/s13742-015-0047-8

29. Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen W-M. Robust relationship inference in genome-wide association studies. *Bioinformatics* (2010) 26:2867–73. doi: 10.1093/bioinformatics/btq559

30. Bose A, Kalantzis V, Kontopoulou E-M, Elkady M, Paschou P, Drineas P. TeraPCA: a fast and scalable software package to study genetic variation in tera-scale genotypes. *Bioinformatics* (2019) 35:3679–83. doi: 10.1093/bioinformatics/btz157

31. Choi SW, O'Reilly PF. PRSice-2: Polygenic Risk Score software for biobank-scale data. *Gigascience* (2019) 8. doi: 10.1093/gigascience/giz082

32. Carroll RJ, Bastarache L, Denny JC. R PheWAS: data analysis and plotting tools for phenome-wide association studies in the R environment. *Bioinformatics* (2014) 30:2375–6. doi: 10.1093/bioinformatics/btu197

33. Millard LAC, Davies NM, Gaunt TR, Smith GD, Tilling K. Software Application Profile: PHESANT: a tool for performing automated phenome scans in UK Biobank. *Int J Epidemiol* (2018) 47:29–35. doi: 10.1093/ije/dyx204

34. Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh P-R, et al. An atlas of genetic correlations across human diseases and traits. *Nat Genet* (2015) 47:1236–41. doi: 10.1038/ng.3406

35. Zheng J, Erzurumluoglu AM, Elsworth BL, Kemp JP, Howe L, Haycock PC, et al. LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. *Bioinformatics* (2017) 33:272–9. doi: 10.1093/bioinformatics/btw613

36. Grotzinger AD, Rhemtulla M, de Vlaming R, Ritchie SJ, Mallard TT, Hill WD, et al. Genomic structural equation modelling provides insights into the multivariate genetic architecture of complex traits. *Nat Hum Behav* (2019) 3:513–25. doi: 10.1038/s41562-019-0566-x

37. Yang Z, Paschou P, Drineas P. Reconstructing SNP allele and genotype frequencies from GWAS summary statistics. *bioRxiv* (2021) 2021:4. doi: 10.1101/2021.04.02.438281

38. Han B, Eskin E. Interpreting meta-analyses of genome-wide association studies. *PLoS Genet* (2012) 8:e1002555. doi: 10.1371/journal.pgen.1002555

39. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* (2010) 26:841–2. doi: 10.1093/bioinformatics/btq033

40. Raudvere U, Kolberg L, Kuzmin I, Arak T, Adler P, Peterson H, et al. g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res* (2019) 47:W191–8. doi: 10.1093/nar/gkz369

41. Buniello A, MacArthur JAL, Cerezo M, Harris LW, Hayhurst J, Malangone C, et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res* (2019) 47:D1005–12. doi: 10.1093/nar/gky1120

42. Jin Y, Andersen G, Yorgov D, Ferrara TM, Ben S, Brownson KM, et al. Genome-wide association studies of autoimmune vitiligo identify 23 new risk loci and highlight key pathways and regulatory variants. *Nat Genet* (2016) 48:1418–24. doi: 10.1038/ng.3680

43. Renton AE, Pliner HA, Provenzano C, Evoli A, Ricciardi R, Nalls MA, et al. A genome-wide association study of myasthenia gravis. *JAMA Neurol* (2015) 72:396–404. doi: 10.1001/jamaneurol.2014.4103

44. Topaloudi A, Zagoriti Z, Flint AC, Martinez MB, Yang Z, Tsetsos F, et al. Myasthenia gravis genome-wide association study implicates AGRN as a risk locus. *J Med Genet* (2021) 59:801–9. doi: 10.1136/jmedgenet-2021-107953

45. Kuo C-F, Grainge MJ, Valdes AM, See L-C, Luo S-F, Yu K-H, et al. Familial risk of sjögren's syndrome and co-aggregation of autoimmune diseases in affected families: A nationwide population study. *Arthritis Rheumatol* (2015) 67:1904–12. doi: 10.1002/art.39127

46. Conigliaro P, D'Antonio A, Pinto S, Chimenti MS, Triggianese P, Rotondi M, et al. Autoimmune thyroid disorders and rheumatoid arthritis: A bidirectional interplay. *Autoimmun Rev* (2020) 19:102529. doi: 10.1016/j.autrev.2020.102529

47. Ferrari SM, Elia G, Virili C, Centanni M, Antonelli A, Fallahi P. Systemic lupus erythematosus and thyroid autoimmunity. *Front Endocrinol* (2017) 8:138. doi: 10.3389/fendo.2017.00138

48. Frommer L, Kahaly GJ. Type 1 diabetes and autoimmune thyroid disease-the genetic link. *Front Endocrinol* (2021) 12:618213. doi: 10.3389/fendo.2021.618213

49. Lopomo A, Berrih-Aknin S. Autoimmune thyroiditis and myasthenia gravis. *Front Endocrinol* (2017) 8:169. doi: 10.3389/fendo.2017.00169

50. Baldini E, Odorisio T, Sorrenti S, Catania A, Tartaglia F, Carbotta G, et al. Vitiligo and autoimmune thyroid disorders. *Front Endocrinol* (2017) 8:290. doi: 10.3389/fendo.2017.00290

51. Ban L, Labbouz S, Grindlay D, Batchelor JM, Ratib S. Risk of skin cancer in people with vitiligo: a systematic review and meta-analysis. *Br J Dermatol* (2018) 179:971–2. doi: 10.1111/bjd.16703

52. Weng Y-C, Ho HJ, Chang Y-L, Chang Y-T, Wu C-Y, Chen Y-J. Reduced risk of skin cancer and internal Malignancies in vitiligo patients: a retrospective population-based cohort study in Taiwan. *Sci Rep* (2021) 11:20195. doi: 10.1038/s41598-021-99786-9

53. Wu W23 and Me Research Team, , Amos CI, Lee JE, Wei Q, Sarin KY, et al. Inverse relationship between vitiligo-related genes and skin cancer risk. *J Invest Dermatol* (2018) 138:2072–5. doi: 10.1016/j.jid.2018.03.1511

54. Lerner A, Shoenfeld Y, Matthias T. Adverse effects of gluten ingestion and advantages of gluten withdrawal in nonceliac autoimmune disease. *Nutr Rev* (2017) 75:1046–58. doi: 10.1093/nutrit/nux054

55. Freemer MM, King TE Jr, Criswell LA. Association of smoking with dsDNA autoantibody production in systemic lupus erythematosus. *Ann Rheum Dis* (2006) 65:581–4. doi: 10.1136/ard.2005.039438

56. Andersen IM, Tengesdal G, Lie BA, Boberg KM, Karlsen TH, Hov JR. Effects of coffee consumption, smoking, and hormones on risk for primary sclerosing cholangitis. *Clin Gastroenterol Hepatol* (2014) 12:1019–28. doi: 10.1016/j.cgh.2013.09.024

57. Boonstra K, de Vries EMG, van Geloven N, van Erpecum KJ, Spanier M, Poen AC, et al. Risk factors for primary sclerosing cholangitis. *Liver Int* (2016) 36:84–91. doi: 10.1111/liv.12894

58. Eaton JE, Juran BD, Atkinson EJ, Schlicht EM, Xie X, de Andrade M, et al. A comprehensive assessment of environmental exposures among 1000 North American patients with primary sclerosing cholangitis, with and without inflammatory bowel disease. *Aliment Pharmacol Ther* (2015) 41:980–90. doi: 10.1111/apt.13154

59. Dunlap R, Wu S, Wilmer E, Cho E, Li W-Q, Lajevardi N, et al. Pigmentation traits, sun exposure, and risk of incident vitiligo in women. *J Invest Dermatol* (2017) 137:1234–9. doi: 10.1016/j.jid.2017.02.004

60. Butler MM, Kenny LC, McCarthy FP. Coeliac disease and pregnancy outcomes. *Obstet Med* (2011) 4:95–8. doi: 10.1258/om.2011.110007

61. Song H, Fang F, Tomasson G, Arnberg FK, Mataix-Cols D, Fernández de la Cruz L, et al. Association of stress-related disorders with subsequent autoimmune disease. *JAMA* (2018) 319:2388–400. doi: 10.1001/jama.2018.7028

62. Hedemann TL, Liu X, Kang CN, Husain MI. Associations between psoriasis and mental illness: an update for clinicians. *Gen Hosp Psychiatry* (2022) 75:30–7. doi: 10.1016/j.genhosppsych.2022.01.006

63. Patel KR, Singam V, Rastogi S, Lee HH, Silverberg NB, Silverberg JI. Association of vitiligo with hospitalization for mental health disorders in US adults. *J Eur Acad Dermatol Venereol* (2019) 33:191–7. doi: 10.1111/jdv.15255

64. Cheung AC, Patel H, Meza-Cardona J, Cino M, Sockalingam S, Hirschfield GM. Factors that influence health-related quality of life in patients with primary sclerosing cholangitis. *Dig Dis Sci* (2016) 61:1692–9. doi: 10.1007/s10620-015-4013-1

65. Ranieri V, Kennedy E, Walmsley M, Thorburn D, McKay K. The Primary Sclerosing Cholangitis (PSC) Wellbeing Study: Understanding psychological distress in those living with PSC and those who support them. *PLoS One* (2020) 15:e0234624. doi: 10.1371/journal.pone.0234624

66. Tisseverasinghe A, Peschken C, Hitchon C. Anxiety and mood disorders in systemic lupus erythematosus: current insights and future directions. *Curr Rheumatol Rep* (2018) 20:85. doi: 10.1007/s11926-018-0797-2

67. Zhang L, Fu T, Yin R, Zhang Q, Shen B. Prevalence of depression and anxiety in systemic lupus erythematosus: a systematic review and meta-analysis. *BMC Psychiatry* (2017) 17:70. doi: 10.1186/s12888-017-1234-1

68. Tylee DS, Sun J, Hess JL, Tahir MA, Sharma E, Malik R, et al. Genetic correlations among psychiatric and immune-related phenotypes based on genome-wide association data. *Am J Med Genet B Neuropsychiatr Genet* (2018) 177:641–57. doi: 10.1002/ajmg.b.32652

69. Chen J, Xu T, Wu M. Depression in systemic lupus erythematosus: Modifiable or inheritable? a two-sample mendelian randomization study. *Front Genet* (2022) 13:988022. doi: 10.3389/fgene.2022.988022

70. Cao MY, Davidson D, Yu J, Latour S, Veillette A. Clnk, a novel SLP-76-related adaptor molecule expressed in cytokine-stimulated hemopoietic cells. *J Exp Med* (1999) 190:1527–34. doi: 10.1084/jem.190.10.1527