medRxiv preprint doi: https://doi.org/10.1101/2024.04.02.24305135; this version posted April 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 The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Shared genetic etiology between hypothyroidism and complex diseases: a large-scale genome-wide cross-trait analysis

Shifang Li, Meijiao Gong

Laboratory of Immunology and Vaccinology, FARAH, ULiège, Liège 4000, Belgium.

Correspondence:

Shifang Li, fruceslee@gmail.com

Laboratory of Immunology and Vaccinology, FARAH, ULiège, Liège 4000, Belgium

Abstract

Hypothyroidism is a common condition of thyroid hormone insufficiency, and there is growing evidence of its link with additional diseases. It remains unclear whether these associations share a common genetic architecture. To address this gap, by leveraging summary-level genetic data from the UK Biobank of hypothyroidism and the FinnGen study of three complex diseases (sarcoidosis, chronic sinusitis, and interstitial lung disease (ILD) endpoints), we evaluated their shared genetic etiology. A significant genetic correlation was found between hypothyroidism and the three diseases. Cross-trait analyses utilizing the MTAG and CPASSOC models revealed 12, 2, and 12 shared loci between hypothyroidism and chronic sinusitis, ILD endpoints, and sarcoidosis, respectively. The SNP heritability enrichment analysis across 37 tissues and 136 cell types at the single-cell level identified candidate tissues and cell types that were shared by the diseases. Interestingly, we found a positive genetic relationship between these four diseases and central memory CD4+ T cells in the blood, supported by strong colocalization evidence (posterior probability >0.9). Mendelian randomization and colocalization analysis showed a link between hypothyroidism and sarcoidosis with two genes (DOCK6 and CD226) in the blood. Furthermore, among the hypothyroidism-driven plasma proteins, RIPK2 was identified as a potentially actionable mediator of hypothyroidism's effect on ILD endpoints. Overall, our findings contribute to improving our understanding of the molecular basis of these diseases' intricate relationships, as well as providing insights toward disease prevention and comorbidity management.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2024.04.02.24305135; this version posted April 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. It is made available under a CC-BY-NC-ND 4.0 International license .

Keywords: Hypothyroidism; cross-trait analyses; central memory CD4+ T cell; genetics; comorbidity; colocalization

Introduction

Hypothyroidism is a disorder caused by inadequate synthesis, secretion, or biological effects of thyroid-stimulating hormone (TSH) [1]. TSH exhibits clear polygenicity based on all 42 significant associations collected from a recent large-scale genome-wide association study (GWAS), which accounted for 33% of the genetic diversity in TSH levels, despite the heritability of TSH levels being estimated at 65% [2]. In addition to genetic determinants, intrinsic factors (eg., age and gender) and environmental risk factors, such as smoking and BMI, have been shown to influence thyroid function. For example, women are more likely to be affected by hypothyroidism than men, and the prevalence of hypothyroidism in women significantly climbs to 7% in people aged 85-89 years [3–5]. However, age, gender, smoking, body mass index, thyroid peroxidase antibody levels, and alcohol usage only explain around 7% of TSH and 5% of free thyroxine variation [6, 7]. In addition to determining what factors are responsible for the pathophysiology of hypothyroidism, the consequences of hypothyroidism for complicated diseases are gaining attention. Among older patients (aged over 65 years), a large-scale case-control study found that a history of hypothyroidism increased the likelihood of being diagnosed with dementia [5]. Clinical observation studies showed that hypothyroidism was associated with sarcoidosis and chronic rhinitis [8, 9]. Furthermore, studies of two common interstitial lung diseases (ILD), idiopathic pulmonary fibrosis and fibrosing hypersensitivity pneumonitis, also suggested that the presence of hypothyroidism increased mortality [10, 11]. However, it remains unclear whether these relationships arise by chance or have a causal connection. As is well known, observational studies are susceptible to unmeasured confounding bias, which restricts the ability to draw causal findings. One solution is to use Mendelian randomization (MR) studies, which use genetic polymorphisms as a tool for causal inference, to yield unconfounded estimates in an observational situation even in the

medRxiv preprint doi: https://doi.org/10.1101/2024.04.02.24305135; this version posted April 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. It is made available under a CC-BY-NC-ND 4.0 International license .

face of unmeasured confounding [12]. Furthermore, GWASs have been conducted for hypothyroidism and other diseases. However, their shared genomic architecture is largely unexplored. Undoubtedly, figuring out their shared genetic architecture might help us better understand the complicated molecular pathways that underpin hypothyroidism and other complex diseases.

To address such a gap, we performed a large-scale genome-wide cross-trait analysis to assess the genetic correlations, causal links, and shared genetic components between hypothyroidism and three complex diseases, which gave insights into their comorbidity.

Material and methods

GWAS data source

Summary-level data for the associations of hypothyroidism-associated SNPs were derived from the MRC-IEU Consortium (access id: ukb-b-4226), which involved 9,674 cases and 453,336 controls [13]. Summary-level data for interstitial lung disease endpoints (4,572 cases and 407,609 controls), sarcoidosis (4,399 cases and 405,620), chronic sinusitis (17,987 cases and 308,457 controls) were obtained from FinnGen study (R10 release) [14]. Summary statistics for 4,907 plasma proteins in 35,559 Icelanders were obtained from the Collaborative Analysis of Diagnostic Criteria in Europe project (deCODE) [15]. Summary data for 14 CD4+ memory-related T cells were retrieved from a sample of 3,757 Sardinians [16]. All *cis*-eQTL used in the present study were collected from both GTEx and the eQTLGen Consortium [17, 18]. The UCSC tool liftOver was used to coordinate the genomic position of SNPs in the GWAS.

Heritability and overall genetic correlation analysis

The 1000 Genomes project calculated linkage disequilibrium (LD) scores for approximately 1.2 million common SNPs in the HapMap3 reference panel. Single-trait SNP heritability for hypothyroidism and three other complex diseases was estimated using stratified linkage disequilibrium score regression (SLDSC) with the baseline-LD model [19, 20]. SNP heritability estimates were transformed to the liability scale based on the observed sample prevalence and population prevalence,

medRxiv preprint doi: https://doi.org/10.1101/2024.04.02.24305135; this version posted April 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. It is made available under a CC-BY-NC-ND 4.0 International license .

assuming population prevalence for hypothyroidism, ILD endpoints, sarcoidosis, and chronic sinusitis were 0.046, 0.0003, 0.000145, and 0.08, respectively [21–24]. Meanwhile, a pair-wise genetic correlation analysis was conducted using LDSC without restricting the intercept [25]. Sensitivity analyses have been carried out using LDSC with the single-trait heritability intercept restricted. Because there was no sample overlap in the MRC-IEU Consortium and FinnGen studies, we set all single-trait intercepts to 1 and all cross-trait intercepts to 0. A *p*-value of 0.05 was used to indicate statistical significance.

Compared to LDSC, a genetic covariance analyzer (GNOVA) generates higher estimation accuracy for genetic correlations and a more powerful statistical inference [26]. As a result, GNOVA was used to evaluate the SNP-based heritability and genetic association between hypothyroidism and the other three complex diseases, with default settings. In short, GNOVA calculates genetic covariance using all genetic variants shared by two GWAS summary statistics. Calculations were performed using the 1000 Genomes Project's European population-derived reference data and default parameters. In addition, sample overlap correction between two independent datasets of GWAS summary data was statistically determined.

Local genetic correlation analysis

To determine whether regions of the genome contribute to diseases, ρ-HESS (heritability estimation from summary statistics) was employed with the default setting to estimate pair-wise local genetic correlation [26, 27]. In brief, this approach divides the genome into 1,703 predefined 1.5 Mb LD-independent regions and properly evaluates genetic association within each region. Statistical significance was determined at a Bonferroni-corrected *p*-value threshold of 0.05/1,703, whereas suggestive significance was defined as $p<0.05$.

Cross-trait GWAS meta-analysis

Given the probability of a significant connection between hypothyroidism and the other three complicated diseases, we utilized cross-trait GWAS meta-analysis to identify the risk SNPs involved. We used two complementary cross-trait GWAS meta-analysis methods: MTAG (multitrait analysis of GWAS) and CPASSOC (cross

medRxiv preprint doi: https://doi.org/10.1101/2024.04.02.24305135; this version posted April 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. It is made available under a CC-BY-NC-ND 4.0 International license .

phenotype association) [27, 28]. MTAG was a multi-trait genome-wide analysis strategy that increased statistical power when compared to traditional single-trait GWAS analyses. The upper bound for the false discovery rate ("maxFDR") was determined for evaluating the assumptions on the equal variance-covariance of shared SNP impact sizes that underpin the diseases [29]. CPASSOC combines GWAS summary statistics from several correlated characteristics to discover variants associated with at least one trait while correcting for population structure or cryptic relatedness, and so served as a sensitivity study to determine the divergence from MTAG's assumption. The analysis used Shet, a test statistic given by CPASSOC that allows for heterogeneous effects of a characteristic across multiple study designs. Independent SNPs that were genome-wide significant (p<5e-08) in cross-trait meta-analyses (e.g. hypothyroidism-sarcoidosis) using both MTAG and CPASSOC, but not identified in the original single-trait GWAS (e.g. hypothyroidism or sarcoidosis) were prioritized, as identified by LD clumping (parameters: --clump-p1 5e-8 --clump-p2 1e-5 --clump-r2 0.2 --clump-kb 1000) in PLINK [30]. Novel SNPs were classified as shared SNPs not driven by a single trait or in LD with index SNPs discovered in single-trait GWASs (LD $r^2 < 0.2$). The Ensembl Variant Effect Predictor (VEP) was utilized to provide detailed functional annotation of the identified variants [31].

Summary data-based Mendelian randomization analysis

Using version 1.03 of the SMR software tool [31], we conducted an analysis using summary data from GWAS of four diseases and eQTL studies from GTEx (blood, lung, spleen, and small intestine) or eQTLgene (blood) to investigate the relationship between gene expression and four diseases. A trait-wise Bonferroni-corrected SMR p-value<0.05/number of examined genes indicated significant gene expression due to causation. Heterogeneity in dependent instruments (HEIDI) tests were used to determine whether the observed connections were influenced by linkage effects. A P_{HEIDI} score of less than 0.05 suggests that the observed relationships are the result of independent genetic variants in LD.

Colocalization analysis and Bayesian fine-mapping analysis

medRxiv preprint doi: https://doi.org/10.1101/2024.04.02.24305135; this version posted April 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. It is made available under a CC-BY-NC-ND 4.0 International license .

Colocalization analysis was performed to verify the shared genetic variants between traits using the "colco.abf" function from the coloc R package, as previously described [32, 33]. This method uses a Bayesian algorithm to compute posterior probabilities for five mutually incompatible hypotheses on the sharing of causal variants in a genomic region. Posterior probabilities (PH4) greater than 0.8 were considered to be co-localized [32]. Meanwhile, we used SuSiE (v.0.11.42) ("susie_rss" function of the susieR R package) to identify a 95%-credible list of SNPs for each independent shared genetic variant within 500-kb [34].

Tissue enrichment analysis

To identify the tissues most related to clinical diseases, we scanned the GTEx tissue (v8) enrichment analysis utilizing LDSC and multimarker analysis of genomic annotation (MAGMA), as described by Bryois and others [35]. In brief, using the GTEx dataset, we employed the pre-computed median expression across subjects and excluded tissues that were not sampled in at least 100 individuals, non-natural tissues, and testis tissues. Following that, the expression of tissues by organ (with the exception of brain tissues) was averaged, yielding gene expression profiles for 37 tissues, with the 10% most specific genes in each tissue being used for the subsequent tissue enrichment studies. For LDSC, SNPs from 100-kb regions surrounding the 10% most specific genes in each tissue were included in the baseline model independently for each tissue, and the coefficient z-score *p*-value was chosen as a measure of the cell type's associations with the traits. For MAGMA, gene-level association statistics were computed using a window of 35-kb upstream to 10-kb downstream of gene coordinates. The European reference panel from phase 3 of the 1000 Genomes Project was utilized as the reference population. For all traits, we utilized MAGMA to determine whether the 10% most specific gene in each tissue was linked to gene-level genetic associations with the trait. Across all tissues per trait, we set the significance threshold for both LDSC and MAGMA at a 5% false discovery rate (FDR).

Cell-type enrichment analyses using scRNA-seq datasets

To identify cell types underlying complex traits, the scRNA-seq data from four tissues (whole blood [36], spleen [37], small intestinal [38], and lung [39]) and GWAS

medRxiv preprint doi: https://doi.org/10.1101/2024.04.02.24305135; this version posted April 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. It is made available under a CC-BY-NC-ND 4.0 International license .

summary statistics of four diseases we studied were integrated using three different genetic prioritization models: LDSC applied to specifically expressed genes (LDSC-SEG), MAGMA, and single-cell disease relevance score (scDRS) [40–42]. The first two methods were implemented using the CELLECT snakemake workflow [43]. In brief, the CELLEX method was used for calculating a single ES estimate (ESm) score for each subpopulation in each cell type. Annotations were constructed using 1000 Genomes Project SNPs, as in the default S-LDSC baseline model [44]. The Hapmap3 SNPs and the excluded major histocompatibility complex (MHC) region were utilized to calculate LD scores. Following that, LDSC-SEG and MAGMA regression analyses were conducted. Meanwhile, scDRS, which integrates gene expression patterns from scRNA-seq with polygenic disease information from GWASs, was used to discover the cell subpopulations driving GWAS enrichments. Putative disease gene sets were identified from the top 1,000 MAGMA genes weighted by their Z-scores. The normalized disease scores were generated in scDRS by the CLI with "scdrs compute-score" using a covariate matrix that included assay, gender, age, ethnicity, and the time of cold storage (if available). For multiple hypothesis testing, *p*-values were FDR corrected using the Benjamini-Hochberg methods across all tissues and diseases.

Mendelian randomization and mediation analysis

Two-sample MR analysis was used to infer the probable causality effect, using the R package "TwoSampleMR" V.0.5.6 and CAUSE, as previously reported [33]. In brief, the genome-wide significant (p<5e-08) SNPs located outside the MHC region (chromosome 6: 28,477,797-33,448,354 (GRCh37)) were first extracted. Independent SNPs ($r^2 < 0.001$ and clump window>10,000 kb) were used as instrumental variables (IVs) in MR analysis through LD based on the 1000 Genomes European reference panel. Our analysis focused on *cis*-pQTLs, located within a 500-kb area upstream and downstream of the gene body, to find plasma proteins related to diseases. If two or more SNPs were available, the inverse variance weighted (IVW) approach was employed as the primary analysis, and if only one IV was available, the Wald ratio was used. Additionally, pleiotropy was assessed using MR-Egger's intercept, and

medRxiv preprint doi: https://doi.org/10.1101/2024.04.02.24305135; this version posted April 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. It is made available under a CC-BY-NC-ND 4.0 International license .

instrument heterogeneity was estimated using the Cochran Q test and I^2 statistics with the "Isq()" method. Leave-one-out analysis was conducted to determine whether the observed correlation was caused by any single IV. To rule out any pleiotropic effects, we evaluated each selected instrument SNP in Open Targets Genetics (https://genetics.opentargets.org/) databases for previously reported associations. Associations with $p < 5e\text{-}08$ were deemed statistically significant. To account for multiple testings, we utilized Benjamini-Hochberg correction and a significance criterion of FDR < 0.05 to evaluate statistical significance. Furthermore, the CAUSE model, which accounts for correlated and uncorrelated horizontal pleiotropic effects through a multivariate linear model adjusted by a joint distribution of instrumental SNPs, was used to avoid more false positives caused by correlated horizontal pleiotropy than previous methods [45]. To estimate the causal mediation effects ($\beta_{mediated}$), network MR with a product of coefficients method was employed, as described by Yoshiji and others [46]. In brief, we first estimated the effect of protein levels on hypothyroidism ($\beta_{protein\text{-}to\text{-}hypothyroidism}$) and the effect of hypothyroidism on chronic sinusitis ($\beta_{hypothyroidism\text{-}to\text{-}chronic\ sinusitis}$). After that, we multiplied these values ($\beta_{mediated} = \beta_{protein\text{-}to\text{-}hypothyroidism} \times \beta_{hypothyroidism\text{-}to\text{-}chronic\ sinusitis}$) and divided $\beta_{mediated}$ by $\beta_{protein\text{-}to\text{-}chronic\ sinusitis}$ to estimate the proportion mediated.

Results

Hypothyroidism shows a significant genetic correlation with complex diseases

Bivariate LDSC was used to evaluate the genetic association (without a constrained intercept) between hypothyroidism and the three complex diseases (**Figure 1A**). Using unconstrained LDSC, three complex diseases were identified as having a substantial genetic correlation with hypothyroidism (ILD endpoints, Rg=0.2425, p=0.0063; sarcoidosis, Rg=0.2937, p=2.05e-05; chronic sinusitis, Rg=0.1776, p=0.0004). The liability-scale SNP heritability estimates for hypothyroidism, ILD endpoints, sarcoidosis, and chronic sinusitis were 0.0469, 0.0042, 0.0126, and 0.0241, respectively. The computed intercept of genetic covariance ranged from 0.018 to 0.040, indicating a slight sample overlap between hypothyroidism and the selected complex diseases. Given the limited sample overlap

medRxiv preprint doi: https://doi.org/10.1101/2024.04.02.24305135; this version posted April 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. It is made available under a CC-BY-NC-ND 4.0 International license .

between hypothyroidism and the three complex diseases, we further confined the intercepts of genetic covariance estimates to 0, allowing LDSC to provide better power with slightly lower standard errors (SE) (**Supplementary Table S1**). As a result, the estimated genetic association was marginally reduced while remaining significant. GNOVA and ρ-HESS analyses revealed a strong genetic link between hypothyroidism and three complex diseases (**Figure 1A**).

Given the significant global genetic association, we investigated whether distinct genomic regions conferred local genetic correlation at genomic regions with GWAS sites relevant to each trait. A total of 13 suggestively significant region pairs were discovered (uncorrected $p<0.05$, ρ-HESS, **Supplementary Table S2**) . There are 4 regions for hypothyroidism and chronic sinusitis, 6 regions for hypothyroidism and sarcoidosis, and 3 regions for hypothyroidism and ILD endpoints. The average local genetic association was nearly the same in regions with hypothyroidism-specific loci or the three complex disease-specific loci (**Supplementary Figure S1**). No other common region with a substantial local genetic association was discovered (**Supplementary Figure S1**). Interestingly, it was found that the correlation at hypothyroidism-specific regions (local Rg=0.2, SE=0.058) is significantly greater than sarcoidosis-specific (local $Rg_{hypothyroidism}$=0.2, $SE_{hypothyroidism}$=0.058 versus local $Rg_{sarcoidosis}$=0.19, $SE_{sarcoidosis}$=0.61) as well as chronic sinusitis-specific loci (local $Rg_{hypothyroidism}$=0.12, $SE_{hypothyroidism}$=0.041 versus local $Rg_{chronic\ sinusitis}$=-0.16, $SE_{chronic\ sinusitis}$=0.15), indicating that loci that increase hypothyroidism tend to consistently increase the risk of sarcoidosis and chronic sinusitis (**Figure 1B**). Overall, these results suggest hypothyroidism and complex diseases are likely linked due to the sharing of genetic variants across the entire genome rather than in specific genomic regions.

Hypothyroidism has a causal effect on sarcoidosis and chronic sinusitis

Bidirectional MR was used to investigate the probable causative effect and whether the shared genetic basis of hypothyroidism and complex diseases supported pleiotropy. Genetically predicted hypothyroidism was linked to increased risk of three complex diseases (IVW beta=3.49 for ILD endpoints, SE=0.68, *p*-value=2.63e-07;

medRxiv preprint doi: https://doi.org/10.1101/2024.04.02.24305135; this version posted April 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. It is made available under a CC-BY-NC-ND 4.0 International license .

IVW beta=4.26 for sarcoidosis, SE=0.87, *p*-value=8.83e-07; IVW beta=1.78 for chronic sinusitis, SE=0.47, *p*-value=1.87e-04) (**Supplementary Table S3**). The MR-Egger intercept test revealed no directional pleiotropy, supporting the validity of the findings. The positive relationships between hypothyroidism and these three diseases were not driven by outliers, as confirmed by leave-one analysis. In the reverse direction, no significant association was found between these three complex diseases and the risk of hypothyroidism ((odds ratio) OR=1.0026 for ILD endpoints, 95% confidence interval (CI):0.998-1.0064, *p*-value=0.19; OR=1.0062 for sarcoidosis, 95%CI: 0.99-1.012, *p*-value=0.061; OR=1.012 for chronic sinusitis, 95%CI: 0.99-1.03, *p*-value=0.209). These findings are consistent across several MR approaches (**Supplementary Table S3**). To validate these results, the largest-to-date GWASs of hypothyroidism, involving 51,194 cases of hypothyroidism and 443,383 controls from Finland and the UK Biobank study, were utilized to evaluate the causation between hypothyroidism and the other three diseases (**Figure 1C** and **Supplementary Table S4**) [46]. All of these causal associations between hypothyroidism and complex disease could be repeated (**Figure 1C** and **Supplementary Table S4**). However, the CAUSE model shows that, aside from ILD endpoints, only chronic sinusitis and sarcoidosis are genetically affected by hypothyroidism.

Cross-trait meta-analysis and pleiotropic loci

Given the considerable genetic link between hypothyroidism and three complex diseases, we performed a cross-trait meta-analysis to increase our capacity to uncover genetic SNPs shared across the traits. A total of 26 genome-wide significant SNPs (p<5e-08) were found in both MTAG and CPASSOC (**Figure 2** and **Supplementary Table S5**). There were 12 loci related to hypothyroidism and chronic sinusitis; 2 loci associated with hypothyroidism and ILD endpoints; and 12 loci associated with hypothyroidism and sarcoidosis. Importantly, we identified two novel SNPs: one related to hypothyroidism and chronic sinusitis (rs174573, mapped gene: FADS2, $P_{\text{hypothyroidism}}$=1.30e-07, $P_{\text{chronic sinusitis}}$=1.04e-07, $P_{\text{CPASSOC\&MTAG}}$<5e-08), and another with hypothyroidism and sarcoidosis (rs12806363, mapped gene: RPS6KA4,

medRxiv preprint doi: https://doi.org/10.1101/2024.04.02.24305135; this version posted April 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. It is made available under a CC-BY-NC-ND 4.0 International license .

$P_{hypothyroidism}$=6.4e-08, $P_{sarcoidosis}$=1.52e-06, $P_{CPASSOC\&MTAG}$<5e-08). The MTAG studies on hypothyroidism, sarcoidosis, chronic sinusitis, and interstitial lung disease yielded maxFDR values of 0.00044, 0.00825, 0.026, and 0.27, respectively. Notably, the MTAG results were highly comparable with those generated by CPASSOC, implying that the MTAG results are reliable and that bias in MTAG assumptions is unlikely to be significant. Among these 26 loci, 11 risk SNPs were consistently significant (p<0.05) when examined by ρ-HESS methods (**Supplementary Table S2**), and 16 loci had colocalization probabilities (PH4) over 0.8, confirming that hypothyroidism and the selected three diseases share the same genetic variants (**Supplementary Table S5**). Finally, two (rs11066320 and rs3184504), one (rs653178), and four (rs12349571, rs11065784, rs11066320, and rs3184504) common loci between hypothyroidism and sarcoidosis, ILD endpoints, and chronic sinusitis, respectively, were verified by different approaches (PPA4>0.8, **Supplementary Table S6**). Interestingly, the two independent loci shared by hypothyroidism and sarcoidosis (rs11066320 and rs3184504) were also found in the cross-traits study of hypothyroidism and chronic sinusitis (**Supplementary Table S5**). Notably, rs653178, the common risk locus for hypothyroidism and ILD endpoints, is located on an intron of ATXN2, which also serves as a TF-binding site for XBP1, CREB3, BATF3, JDP2, FOS, ATF7, and JUN.

In order to infer causative variants at each of the pleiotropic loci, a credible set of variants that were 95% probable were identified, based on posterior probability (**Supplementary Table S7**). In total, we discovered 147 candidate causative SNPs across all shared loci. Five pleiotropic SNPs (rs11066320, rs2847259, rs3184504, rs6679677, and rs7705526) also had posterior probabilities greater than 0.99. Overall, this credible set of variants offers targets for further experimental studies.

SNP heritability enrichment at the tissue and cell type level

To determine which tissues affect by genetics factors, we used the LDSC and MAGMA methods to identify human tissues with enrichment for genetic associations using human GTEx (v8) resources. For these two approaches (LDSC and MAGMA), we evaluated whether the 10% most specific genes in each tissue had more genetic associations with each of the traits. After adjusting for the baseline model and

medRxiv preprint doi: https://doi.org/10.1101/2024.04.02.24305135; this version posted April 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. It is made available under a CC-BY-NC-ND 4.0 International license .

performing multiple corrections, we observed significant SNP heritability enrichment for hypothyroidism across four tissues (**Figure 3A** and **Supplementary Table S8**), with hypothyroidism having a greater enrichment in the spleen, small intestine, blood, and lung. Following multiple corrections, no significant enrichment was detected for sarcoidosis, chronic sinusitis, or ILD endpoints (**Figure 3A** and **Supplementary Table S8**).

Given tissue heterogeneity, we used publicly available scRNA-seq datasets from four tissues: whole blood (50,115 cells), spleen (94,256 cells), small intestinal (36,359 cells), and lung (71,752 cells), to assess the genetic association with cell type expression specificity for hypothyroidism and the other three diseases (**Figure 3B** and **Supplementary Table S9**). To robustly identify the tissues implied by these traits, three approaches (LDSC, MAGMA, and scDRS) were implemented, each with a different assumption and procedure. Specifically, scDRSs, an orthogonal approach to other methodologies, were utilized to detect GWAS trait enrichment at the single-cell level. This approach assesses not only the associations between cell types and GWAS traits but also the heterogeneity of the associations within cell types. Following the heritability enrichment analysis, we uncovered significant enrichment for hypothyroidism with sarcoidosis and chronic sinusitis in CD4 positive alpha-beta memory T cells, CD4 positive alpha-beta T cells, and CD8 positive alpha-beta T cells in blood, indicating that the polygenic risks associated with these three diseases were enriched in the cells responsible for adaptive immunity (**Figure 3B** and **Supplementary Table S9**). Hypothyroidism was also found to share group 2 innate lymphoid cells, group 3 innate lymphoid cells, CD4 positive alpha beta T cells, CD8 positive alpha beta T cell regulatory T cells, and T helper 17 cells in the lung with sarcoidosis and chronic sinusitis. Furthermore, the activated CD8+ alpha beta T cells in the spleen showed a greater heritability enrichment for hypothyroidism with sarcoidosis and chronic sinusitis, with significant heterogeneity. Interestingly, both hypothyroidism and sarcoidosis showed considerable heritability enrichment in T cells and dendritic cells in the small intestine (**Figure 3C** and **Supplementary Table S9**), shedding light on the biology of comorbidity via the same cell type.

medRxiv preprint doi: https://doi.org/10.1101/2024.04.02.24305135; this version posted April 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. It is made available under a CC-BY-NC-ND 4.0 International license .

Of note, the significant heritability does not necessarily indicate a causal relationship between cell types and diseases. To validate the causative influence of CD4 positive alpha-beta memory T cells in the blood associated with diseases (**Figure 3B**), 14 CD4+ memory-related T cells from a cohort of 3,757 Sardinians were used in MR analysis (**Supplementary Table S10-S11**). After multiple corrections, 9 pairs of cell-disease associations were identified. Interestingly, it was found that central memory CD4+ T cell absolute count was significantly genetically associated with hypothyroidism (OR=1.008, 95%CI=1.0029-1.013; FDR=1.16e-02), ILD endpoints (OR=1.94, 95%CI=1.49-2.51; FDR=9.44e-06), sarcoidosis (OR=2.49, 95%CI=1.92-3.25; FDR=1.41e-10), and chronic sinusitis (OR=1.51, 95%CI=1.32-1.73; FDR=2.86e-08), respectively (**Figure 3C** and **Supplementary Table S10-S11**). Furthermore, colocalization analysis showed that central memory CD4+ T cell absolute count may share the same genetic variants with hypothyroidism (PH4=0.994), ILD endpoints (PH4=0.992), sarcoidosis (PH4=0.992), and chronic sinusitis (PH4=0.991) (**Figure 3D**). Altogether, these findings indicate that hypothyroidism appears to influence the link between central memory CD4+ T cell absolute count and sarcoidosis or chronic sinusitis.

Identification of shared functional genes for hypothyroidism and complex diseases

To infer causality and identify putative functional genes for hypothyroidism and complex diseases, we utilized GWAS summary data with small intestine, spleen, and lung eQTL summary data from GTEx, as well as whole blood eQTL summary data from eQTLGen and GTEx resources (v8) (**Figure 4A** and **Supplementary Table S12**). After multiple corrections (Bonferroni-corrected $p<0.05$), two substantial risk gene pairs (HLA-DPB2 and HLA-DQB1-AS1) in the small intestine were identified for hypothyroidism and related diseases. Hypothyroidism and sarcoidosis share two genes including AP003774.4 and HLA-DPB2 in the spleen. In addition, PPP1R18 in the lung is genetically shared by hypothyroidism and sarcoidosis (SMR adjusted $p<0.05$ and $P_{HEIDI}>0.05$). Interestingly, some other common risk genes were discovered, although they were not statistically significant in both hypothyroidism

medRxiv preprint doi: https://doi.org/10.1101/2024.04.02.24305135; this version posted April 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. It is made available under a CC-BY-NC-ND 4.0 International license .

and other diseases, they were significant in at least one of the diseases. For example, CUTALP, which is strongly associated with hypothyroidism, has also been linked to ILD endpoints and sarcoidosis in the small intestine, blood, and lungs. Similarly, DOCK6 was shared by hypothyroidism and three additional diseases of the small intestine and blood. To further validate these associations, colocalization analyses for the 48 common gene-tissues pairs were conducted using human GTEx resources (v7) (**Figure 4B** and **Supplementary Table S13**). DOCK6 and CD226 reveal substantial colocalization evidence with hypothyroidism and sarcoidosis in the blood. Interestingly, CCDC88B has significant colocalization evidence with hypothyroidism and sarcoidosis in blood as both share the same genetic variant, rs479777, one of the pleiotropic loci identified (**Figure 3, Figure 4C**, and **Supplementary Table S13**). However, no clear causation evidence has been established for the connection of CCDC88B with sarcoidosis (SMR p=6.84e-09, P_{HEIDI}=6.85e-03). It is worth noting that CUTALP, which is generally related to the ILD endpoint and shows strong colocalization evidence with hypothyroidism, also showed medium colocalization evidence with the ILD endpoint in the small intestine, blood, and lung.

Identification of shared functional plasma proteins for hypothyroidism and diseases

Because plasma proteome is an abundant resource of potential drug targets for diseases, putative shared functional circulation proteins for hypothyroidism and complex diseases were identified using large-scale GWASs of 4,907 circulating proteins in 35,559 Icelanders (**Figure 5A** and **Supplementary Table S14**). Given that *cis*-pQTLs were anticipated to have a more direct and specific biological influence on the protein (relative to *trans*-pQTLs), MR studies with only *cis*-pQTLs as IVs and diseases as outcomes were carried out [47]. After filtering for heterogeneity ($I^2 < 50\%$ for all), directional pleiotropy ($P_{\text{Egger intercept}} > 0.05$), reverse causation ($P_{\text{Steiger test}} < 0.05$), and leave-one-out analysis, 34 causality proteins were observed (FDR<0.05) (**Figure 5A** and **Supplementary Table S15**). Interestingly, AIF1 was found to be negatively associated with a decreased risk of hypothyroidism (OR[95%CI]=0.981[0.974, 0.988]; p=1.65e-07) and chronic sinusitis (OR[95%CI]=0.655[0.550, 0.780]; p=2.00e-06),

medRxiv preprint doi: https://doi.org/10.1101/2024.04.02.24305135; this version posted April 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. It is made available under a CC-BY-NC-ND 4.0 International license .

despite a positive correlation with sarcoidosis (OR[95%CI]=10.33[7.40, 14.43]; *p*=6.32e-43). To evaluate the indirect effect of proteins on chronic sinusitis via hypothyroidism, we performed a mediation analysis with effect estimates from two-step MR and the overall effect from primary MR. The mediation analysis revealed that AIF1 had a modest mediation effect on chronic sinusitis via hypothyroidism (7.83%). In addition to AIF1, no other proteins were identified to be statistically consistently associated with both hypothyroidism and complex diseases. However, at the nominal level, 19 protein-disease associations are shown (**Figure 5B** and **Supplementary Table S15**), and in particular, 12 pairs of associations ($\beta_{\text{protein to diseases}} \times \beta_{\text{protein to hypothyroidism}} > 0$) have shown a consistent direction of effect between hypothyroidism and three selected complex diseases ($\beta_{\text{hypothyroidism to diseases}} > 0$). For example, a genetic tendency to elevated ALDH2 was found to be substantially linked with an increased risk of hypothyroidism (OR[95%CI]=1.053[1.040, 1.066]; *p*=5.26e-17), as well as an increased risk of ILD endpoints, sarcoidosis, and chronic sinusitis. Furthermore, higher genetically predicted levels of both ANTXR1 (OR[95%CI]=0.316[0.128, 0.782]; *p*=0.0127) and NBL (OR[95%CI]=2.704[1.149, 6.36]; *p*=0.022) were linked with decreased risk of ILD endpoints. Notably, only B3GALT6, which had >80% of the posterior probability of colocalization of the genetic association with hypothyroidism, was explained by the same genetic variant (rs67492154), whereas no strong colocalization evidence was found for other proteins with the corresponding diseases (**Figure 5B** and **Supplementary Table S16**).

Identification of the causal effect of hypothyroidism-driven proteins on diseases

Alternatively, plasma proteins may act as mediators for the effect of hypothyroidism on complex diseases, providing additional insight into underlying comorbidities. To accomplish this, we first estimate the causal effect of hypothyroidism on plasma protein levels in a two-sample MR analysis, with hypothyroidism as the exposure and 4,907 plasma protein levels as the outcomes (**Figure 5C** and **Supplementary Table S17**). The F-statistic, which measures the strength of the association between genetic variants and hypothyroidism, was 73.37, indicating no weak instrument bias. After filtering by heterogeneity test ($I^2<0.5$) and

medRxiv preprint doi: https://doi.org/10.1101/2024.04.02.24305135; this version posted April 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. It is made available under a CC-BY-NC-ND 4.0 International license .

directional pleiotropy test ($P_{\text{Egger intercept}}$>0.05), 1,315 proteins were estimated to be impacted by hypothyroidism (FDR<0.05). 207 proteins that failed leave-one-out analysis and 6 proteins with bidirectional effects were also removed from further investigation. As a result, a total of 1,102 protein levels, including RIPK2, were identified as hypothyroidism-driven proteins, with no apparent heterogeneity ($I^2 < 50\%$), directional pleiotropy ($P_{\text{Egger intercept}} > 0.05$), or reverse causation (**Figure 5D** and **Supplementary Table S18**). The weighted median, weighted mode, and MR-Egger slope methods all produced directionally consistent results with the IVW method for RIPK2 (**Supplementary Table S18**). Among these hypothyroidism-driven proteins, proteins that have a positive effect from hypothyroidism were significantly enriched in the toll-like receptor signaling pathway, chemokine signaling pathway, NOD-like receptor signaling pathway, and MAPK signaling pathway, while negatively affected proteins were negatively associated with histidine metabolism, tryptophan metabolism, and neurotrophin signaling pathway, indicating the imprint impact of hypothyroidism on the immune response (**Figure 5E**).

We next used two-sample MR to assess the causative effects of the identified hypothyroidism-driven proteins on three complex diseases. We again employed *cis*-pQTLs for these proteins as IVs, and GWASs from three diseases as outcome variables. Following the *cis*-pQTL search and data harmonization, 130 proteins were analyzed in MR to assess their estimated causal effect on three diseases. After multiple testing corrections (FDR<0.05), no significant association was identified between hypothyroidism-driven proteins and the three diseases. However, at the nominal significance level, 24 protein-disease association pairs were identified (**Figure 5F**). Particularly, 13 pairs of associations ($\beta_{\text{protein to diseases}} \times \beta_{\text{hypothyroidism to protein}} > 0$) reveal a consistent direction of hypothyroidism's effect on diseases ($\beta_{\text{hypothyroidism to diseases}} > 0$). For example, higher genetically predicted RIPK2 levels (OR[95%CI]=2.88[1.55, 5.38]; p=8.5e-04) were related to an increased risk of ILD endpoints (**Figure 5F**). To check out the premise of a lack of directional pleiotropy, which could reintroduce confounding, we used the Open Target Genetics (https://genetics.opentargets.org/) databases to determine whether the RIPK2

medRxiv preprint doi: https://doi.org/10.1101/2024.04.02.24305135; this version posted April 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. It is made available under a CC-BY-NC-ND 4.0 International license .

cis-pQTLs (rs160438) were related to any traits. The deCODE study's lead *cis*-pQTL for RIPK2 (rs160438) did not disclose any additional relationship at the genome-wide significance threshold of p<5e-08. Colocalization analysis revealed that only RIPK2 had medium colocalization with the diseases (**Figure 5F** and **Supplementary Table S16**). To better understand the origin of RIPK2 in the lung at the scRNA-seq level, we evaluated single-cell RIPK2 expression in human lung tissues from six healthy people and five systemic sclerosis-associated interstitial lung disease (SSC-ILD) patients, as described previously [48, 49]. RIPK2 was mainly expressed in monocytes, macrophages, and monocyte-derived macrophages (monocyte-derived Mph) (**Figure 5G**). Overall, these findings suggested that RIPK2 may be a potential mediator of the effect of hypothyroidism on ILD endpoints.

Discussion

In this study, we provide evidence of causality and shared genetic etiology between hypothyroidism and three complex diseases. Our findings provide additional insight into their comorbidity and may lead to a better understanding of their pathophysiology and the development of treatments.

Both GONVA and LDSC analyses in our study revealed a genetic correlation between hypothyroidism and three complex diseases, providing genetic evidence of a significant positive genetic correlation and supporting the hypothesis that genetic factors play an important role in their comorbidity. Although there was a significant overall genetic correlation between hypothyroidism and three complex diseases, no shared region was discovered for their local genetic correlation, implying that these associations are most likely correlated due to genetic variants shared across the entire genome rather than in specific genomic regions. Importantly, to our knowledge, our study showed for the first time that hypothyroidism increases the risk of sarcoidosis and chronic sinusitis. Despite the fact that hypothyroidism was found to have a statistical significance on ILD endpoints using IVW methods, no significant evidence was revealed when using the CAUSE model, and their causality association should be interpreted with caution because we cannot rule out the possibility that this positive association was caused by correlated pleiotropy.

medRxiv preprint doi: https://doi.org/10.1101/2024.04.02.24305135; this version posted April 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. It is made available under a CC-BY-NC-ND 4.0 International license .

Notably, we discovered several shared genetic loci between hypothyroidism and three complex diseases using cross-trait meta-analysis and colocalization analysis. Specifically, by using distinct methods (MTAG and CPASSOC), we may minimize potential bias due to sample overlap. Furthermore, consistently significant loci were identified in both analyses, increasing the reliability of our findings. Notably, cross-trait GWAS meta-analyses discovered two novel SNPs shared between hypothyroidism and sarcoidosis (rs12806363) and chronic sinusitis (rs174573), implying that these SNPs are likely engaged in regulating similar pathways shared by their comorbidities. rs12806363 is an intron variant in RPS6KA4, which encodes a member of the RSK (ribosomal S6 kinase) family of serine/threonine kinases, whereas rs174573 is a missense variant in FADS2, a member of the fatty acid desaturase gene family that is involved in alpha-linolenic acid metabolism and arachidonate biosynthesis III. A study found higher expression of FADS2 in CD4+ cells in asthmatic patients [50]. Additionally, long-chain polyunsaturated fatty acids produced by FADS2 have anti-inflammatory effects and may regulate immunological function [51, 52]. CRISPR-based technology is required to functionally confirm and define the regulatory effects of genetic variants underlying diseases.

Furthermore, results from heritability enrichment analysis indicate that the putative tissues and cell types may be involved in the shared etiology. It was discovered that GWAS findings for hypothyroidism were more prevalent in the spleen, blood, small intestine, and lungs. The detailed cell types across these four tissues were subsequently utilized in the enrichment analysis to identify shared cell types. Central memory CD4+ T cells in the blood were found to be genetically related to these four diseases, with substantial colocalization evidence (posterior probability>0.9). Consistent with our findings, a recent study also showed that the proportions of Th1 cell-like effector memory T cells and CD4+ tissue-resident memory T cells were significantly increased in nasal polyps from patients with uncontrolled severe chronic rhinosinusitis with nasal polyps (CRSwNP) compared to inferior turbinates from 4 healthy subjects [53], further supporting a role that cannot be ignored in central memory CD4+ T cells on the diseases. Overall, enriched tissues supply additional

medRxiv preprint doi: https://doi.org/10.1101/2024.04.02.24305135; this version posted April 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. It is made available under a CC-BY-NC-ND 4.0 International license .

evidence for disease comorbidity. It should be noted that enrichment analysis results do not represent causality; therefore, in addition to blood, further analysis should be used to evaluate the causative effect of cell types in tissues on diseases when GWASs of cell types in these tissues are available.

In addition to cross-trait meta-analysis, we evaluated eQTL and pQTL data to investigate whether their association might be explained by shared risk genes or proteins. Using SMR, HEIDI, and colocalization analysis, we discovered that CD226 and DOCK6 may act as an association between hypothyroidism and sarcoidosis. CD226, a member of the immunoglobulin superfamily, is a functional protein originally produced on natural killer and T cells [54]. A recent study has reported that CD226 expression in T cells from sarcoidosis [55]. Furthermore, it was proposed that CD4 T cells moving to the lungs via CXCR3 could be activated by CD226, potentially contributing to the pathogenesis of sarcoidosis. DOCK6, a member of the DOCK-C subfamily that displays GEF activity for both Rac1 and CDC42 through its Dock Homology Region-2 (DHR-2) domain, was also shared by hypothyroidism and sarcoidosis, as validated by two distinct datasets (GTEx and eQTLgene) [56]. Recent research indicates that DOCK6 is linked to various diseases, including cancers like gastric cancer and acute myeloid leukemia, and functions as an oncogene [57–60]. However, as far as we know, the biological association between DOCK6 and sarcoidosis and hypothyroidism is largely unknown; more research is needed to understand the mechanisms of DOCK6 in both of these diseases. Meanwhile, we also conducted a systematic investigation into the causes and consequences of plasma proteins associated with hypothyroidism, as well as their relationship to all three diseases we have studied (**Figure 5H**). Interestingly, we showed that AIF1 may act as a mediator in the association between hypothyroidism and chronic sinusitis, despite the fact that no strong colocalization analysis was detected in either trait. We explored the causal influence of hypothyroidism-driven proteins on the three analyzed diseases while maintaining the cause direction of hypothyroidism in account. Interestingly, hypothyroidism was found to affect about 1,102 protein levels, indicating that hypothyroidism has a substantial effect on plasma protein levels. Although no

medRxiv preprint doi: https://doi.org/10.1101/2024.04.02.24305135; this version posted April 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. It is made available under a CC-BY-NC-ND 4.0 International license .

statistically significant association was observed between hypothyroidism-driven proteins and the three studied diseases, RIPK2 was identified as a potential mediator of hypothyroidism's effect on ILD endpoints, as indicated by colocalization analysis with medium evidence. Additional research employing *cis*-pQTL of RIPK2 from diverse cohorts is required to corroborate these findings. Furthermore, despite the fact that no significant relationship was found in our study, it provides insight for further studying the mechanisms underlying the association between hypothyroidism and other diseases employing hypothyroidism-driven proteins in the future.

We also note that our study has severe limitations. First, the GWAS data for hypothyroidism and the three complex diseases utilized in this study were sourced from the European population; therefore, the results may not be applicable to other ancestries. Second, the availability of summary-level GWAS rather than individual data for four diseases limits our ability to analyze sex and age-specific genetic effects. Third, while our findings show a genetic link and overlap between hypothyroidism, sarcoidosis, and chronic sinusitis, the underlying biological mechanisms remain unexplained. For example, the mechanism of central memory CD4+ T cell absolute count and causal genes (DOCK6 and CD226) involved in disease crosstalk remain largely unknown; therefore, functional experimental validation of the enriched tissues and cells is required. Finally, the study's GWAS sample size for hypothyroidism was not the largest. Mathieu and colleagues recently conducted a GWAS meta-analysis of hypothyroidism that included 51,194 cases and 443,383 controls, and 139 risk loci from the UK Biobank and FinnGen studies were discovered [61]. However, due to differences in research design, we employed hypothyroidism GWASs from the UK Biobank in our study, which minimizes the possibility of our discovery being false positive.

In conclusion, by leveraging GWAS summary data, we unearthed significant genetic correlations and uncovered the shared loci between hypothyroidism and three complex diseases. The shared genes, proteins, and cell types that are associated with their comorbidity were also identified. Furthermore, MR analysis revealed a causal effect of hypothyroidism on chronic sinusitis and sarcoidosis. Overall, these findings

medRxiv preprint doi: https://doi.org/10.1101/2024.04.02.24305135; this version posted April 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. It is made available under a CC-BY-NC-ND 4.0 International license .

shed light on an understanding of disease pathogenesis and point to possible therapy targets.

Acknowledgments

We thank Consortium des Equipements de Calcul Intensif en Fédération Wallonie Bruxelles (CECI) funded by the Fonds de la Recherche Scientifique de Belgique (FRS-FNRS) for providing the supercomputing facilities utilized for the analysis. The authors also thank all of the researchers involved in the associated genetic consortia and GWAS for making the data available to the public.

Data availability and code availability

All the datasets analyzed during the current study are publicly available. Custom codes will be made available on GitHub (https://github.com/frucelee/Shared-genetic-etiology-between-hypothyroidism-and-complex-diseases) upon publication of the manuscript.

Contributions

SF conceived and designed the study. SF and MJ performed the data analysis. SF and MJ conducted the interpretation of analysis results. SF wrote the manuscript. All authors read and approved the manuscript.

Funding

Not applicable.

Declarations

Ethics approval and consent to participate

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References

1. Bianco AC. Emerging Therapies in Hypothyroidism. Annu Rev Med. 2024;75:307–19.

2. Teumer A, Chaker L, Groeneweg S, Li Y, Di Munno C, Barbieri C, et al. Genome-wide analyses identify a role for SLC17A4 and AADAT in thyroid hormone regulation. Nat Commun. 2018;9:1–14.

medRxiv preprint doi: https://doi.org/10.1101/2024.04.02.24305135; this version posted April 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. It is made available under a CC-BY-NC-ND 4.0 International license .

3. Vanderpump MPJ. The epidemiology of thyroid disease. Br Med Bull. 2011;99:39–51.

4. Gussekloo J, Van Exel E, De Craen AJM, Meinders AE, Frölich M, Westendorp RGJ. Thyroid status, disability and cognitive function, and survival in old age. Jama. 2004;292:2591–9.

5. Wieland DR, Wieland JR, Wang H, Chen YH, Lin CH, Wang JJ, et al. Thyroid Disorders and Dementia Risk: A Nationwide Population-Based Case-Control Study. Neurology. 2022;99:E679–87.

6. Chaker L, Razvi S, Bensenor IM, Azizi F, Pearce EN, Peeters RP. Hypothyroidism. Nat Rev Dis Prim. 2022;8.

7. Chaker L, Korevaar TIM, Medici M, Uitterlinden AG, Hofman A, Dehghan A, et al. Thyroid Function Characteristics and Determinants: The Rotterdam Study. Thyroid. 2016;26:1195–204.

8. Kašiković Lečić S, Javorac J, Lovrenski A, Đokić J, Sazdanić Velikić D, Živanović D. Case report: Thyroid sarcoidosis as a rare localization of the disease: Report of two cases and review of the literature. Front Med. 2023;10 March:1–8.

9. Choi HG, Kim TJ, Hong SK, Min C, Yoo DM, Kim H, et al. Thyroid Diseases and Chronic Rhinosinusitis: A Nested Case–Control Study Using a National Health Screening Cohort. Int J Environ Res Public Health. 2022;19.

10. O'Dwyer DN, Wang BR, Nagaraja V, Flaherty KR, Khanna D, Murray S, et al. Hypothyroidism Is Associated with Increased Mortality in Interstitial Pneumonia with Autoimmune Features. Ann Am Thorac Soc. 2022;19:1772–6.

11. Oldham JM, Kumar D, Lee C, Patel SB, Takahashi-Manns S, Demchuk C, et al. Thyroid disease is prevalent and predicts survival in patients with idiopathic pulmonary fibrosis. Chest. 2015;148:692–700.

12. Skrivankova VW, Richmond RC, Woolf BAR, Davies NM, Swanson SA, Vanderweele TJ, et al. Strengthening the reporting of observational studies in epidemiology using mendelian randomisation (STROBE-MR): Explanation and elaboration. BMJ. 2021;375.

13. Elsworth B, Lyon M, Alexander T, Liu Y, Matthews P, Hallett J, et

medRxiv preprint doi: https://doi.org/10.1101/2024.04.02.24305135; this version posted April 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. It is made available under a CC-BY-NC-ND 4.0 International license .

al. The MRC IEU OpenGWAS data infrastructure. bioRxiv. 2020;:2020.08.10.244293.

14. Kurki MI, Karjalainen J, Palta P, Sipilä TP, Kristiansson K, Donner KM, et al. FinnGen provides genetic insights from a well-phenotyped isolated population. Nature. 2023;613:508–18.

15. Ferkingstad E, Sulem P, Atlason BA, Sveinbjornsson G, Magnusson MI, Styrmisdottir EL, et al. Large-scale integration of the plasma proteome with genetics and disease. Nat Genet. 2021;53:1712–21.

16. Orrù V, Steri M, Sidore C, Marongiu M, Serra V, Olla S, et al. Complex genetic signatures in immune cells underlie autoimmunity and inform therapy. Nat Genet. 2020;52:1036–45.

17. Võsa U, Claringbould A, Westra HJ, Bonder MJ, Deelen P, Zeng B, et al. Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression. Nat Genet. 2021;53:1300–10.

18. Aguet F, Brown AA, Castel SE, Davis JR, He Y, Jo B, et al. Genetic effects on gene expression across human tissues. Nature. 2017;550:204–13.

19. Finucane HK, Bulik-Sullivan B, Gusev A, Trynka G, Reshef Y, Loh PR, et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. Nat Genet. 2015;47:1228–35.

20. Gazal S, Finucane HK, Furlotte NA, Loh PR, Palamara PF, Liu X, et al. Linkage disequilibrium-dependent architecture of human complex traits shows action of negative selection. Nat Genet. 2017;49:1421–7.

21. Wyne KL, Nair L, Schneiderman CP, Pinsky B, Antunez Flores O, Guo D, et al. Hypothyroidism Prevalence in the United States: A Retrospective Study Combining National Health and Nutrition Examination Survey and Claims Data, 2009-2019. J Endocr Soc. 2023;7:1–11.

22. Ye Y, Sing CW, Hubbard R, Lam DCL, Li HL, Li GHY, et al. Prevalence, incidence, and survival analysis of interstitial lung diseases in Hong Kong: a 16-year population-based cohort study. Lancet Reg Heal - West Pacific. 2024;42:100871.

23. Sikjær MG, Hilberg O, Ibsen R, Løkke A. Sarcoidosis: A nationwide

medRxiv preprint doi: https://doi.org/10.1101/2024.04.02.24305135; this version posted April 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. It is made available under a CC-BY-NC-ND 4.0 International license .

registry-based study of incidence, prevalence and diagnostic work-up. Respir Med. 2021;187 March.

24. Dietz de Loos D, Lourijsen ES, Wildeman MAM, Freling NJM, Wolvers MDJ, Reitsma S, et al. Prevalence of chronic rhinosinusitis in the general population based on sinus radiology and symptomatology. J Allergy Clin Immunol. 2019;143:1207–14.

25. Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh PR, et al. An atlas of genetic correlations across human diseases and traits. Nat Genet. 2015;47:1236–41.

26. Lu Q, Li B, Ou D, Erlendsdottir M, Powles RL, Jiang T, et al. A Powerful Approach to Estimating Annotation-Stratified Genetic Covariance via GWAS Summary Statistics. Am J Hum Genet. 2017;101:939–64.

27. Liu Q, Tang B, Zhu Z, Kraft P, Deng Q, Stener-Victorin E, et al. A genome-wide cross-trait analysis identifies shared loci and causal relationships of type 2 diabetes and glycaemic traits with polycystic ovary syndrome Abbreviations 2hGlu adj BMI 2h glucose after an oral glucose challenge adjusted for BMI CPASSOC Cross. 2022;:1483–94.

28. Zhu X, Feng T, Tayo BO, Liang J, Young JH, Franceschini N, et al. Meta-analysis of correlated traits via summary statistics from GWASs with an application in hypertension. Am J Hum Genet. 2015;96:21–36.

29. Yang Y, Musco H, Simpson-Yap S, Zhu Z, Wang Y, Lin X, et al. Investigating the shared genetic architecture between multiple sclerosis and inflammatory bowel diseases. Nat Commun. 2021;12:1–12.

30. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: A tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007;81:559–75.

31. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, et al. The Ensembl Variant Effect Predictor. Genome Biol. 2016;17:1–14.

32. Giambartolomei C, Vukcevic D, Schadt EE, Franke L, Hingorani AD, Wallace C, et al. Bayesian Test for Colocalisation between Pairs of Genetic Association Studies Using Summary Statistics. PLoS Genet. 2014;10.

medRxiv preprint doi: https://doi.org/10.1101/2024.04.02.24305135; this version posted April 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. It is made available under a CC-BY-NC-ND 4.0 International license .

33. Li SF, Gong MJ. Drug targets for haemorrhoidal disease: Proteome-wide Mendelian randomisation and colocalisation analyses. Gut. 2024;0:1–2.

34. Zou Y, Carbonetto P, Wang G, Stephens M. Fine-mapping from summary data with the "Sum of Single Effects" model. PLoS Genet. 2022;18:1–24.

35. Bryois J, Skene NG, Hansen TF, Kogelman LJA, Watson HJ, Liu Z, et al. Genetic identification of cell types underlying brain complex traits yields insights into the etiology of Parkinson's disease. Nat Genet. 2020;52:482–93.

36. Jones RC, Karkanias J, Krasnow MA, Pisco AO, Quake SR, Salzman J, et al. The Tabula Sapiens: A multiple-organ, single-cell transcriptomic atlas of humans. Science (80-). 2022;376.

37. Madissoon E, Wilbrey-Clark A, Miragaia RJ, Saeb-Parsy K, Mahbubani KT, Georgakopoulos N, et al. ScRNA-seq assessment of the human lung, spleen, and esophagus tissue stability after cold preservation. Genome Biol. 2019;21:1–16.

38. Egozi A, Llivichuzhca-Loja D, McCourt BT, Bahar Halpern K, Farack L, An X, et al. Insulin is expressed by enteroendocrine cells during human fetal development. Nat Med. 2021;27:2104–7.

39. He P, Lim K, Sun D, Pett JP, Jeng Q, Polanski K, et al. A human fetal lung cell atlas uncovers proximal-distal gradients of differentiation and key regulators of epithelial fates. Cell. 2022;185:4841-4860.e25.

40. Finucane HK, Reshef YA, Anttila V, Slowikowski K, Gusev A, Byrnes A, et al. Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. Nat Genet. 2018;50:621–9.

41. de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: Generalized Gene-Set Analysis of GWAS Data. PLoS Comput Biol. 2015;11:1–19.

42. Zhang MJ, Hou K, Dey KK, Sakaue S, Jagadeesh KA, Weinand K, et al. Polygenic enrichment distinguishes disease associations of individual cells in single-cell RNA-seq data. Nat Genet. 2022;54:1572–80.

43. Timshel PN, Thompson JJ, Pers TH. Genetic mapping of etiologic brain cell types for obesity. Elife. 2020;9:1–45.

44. Altshuler DM, Durbin RM, Abecasis GR, Bentley DR, Chakravarti A, Clark AG,

medRxiv preprint doi: https://doi.org/10.1101/2024.04.02.24305135; this version posted April 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. It is made available under a CC-BY-NC-ND 4.0 International license .

et al. An integrated map of genetic variation from 1,092 human genomes. Nature. 2012;491:56–65.

45. Morrison J, Knoblauch N, Marcus JH, Stephens M, He X. Mendelian randomization accounting for correlated and uncorrelated pleiotropic effects using genome-wide summary statistics. Nat Genet. 2020;52:740–7.

46. Yoshiji S, Butler-Laporte G, Lu T, Willett JDS, Su CY, Nakanishi T, et al. Proteome-wide Mendelian randomization implicates nephronectin as an actionable mediator of the effect of obesity on COVID-19 severity. Nat Metab. 2023;5:248–64.

47. Zheng J, Haberland V, Baird D, Walker V, Haycock PC, Hurle MR, et al. Phenome-wide Mendelian randomization mapping the influence of the plasma proteome on complex diseases. Nat Genet. 2020;52:1122–31.

48. Papazoglou A, Huang M, Bulik M, Lafyatis A, Tabib T, Morse C, et al. Epigenetic Regulation of Profibrotic Macrophages in Systemic Sclerosis–Associated Interstitial Lung Disease. Arthritis Rheumatol. 2022;74:2003–14.

49. Li SF, Gong MJ. Integrative single-cell analysis of epigenomic and transcriptomic states in patients with systemic sclerosis-associated interstitial lung disease. medRxiv. 2023.

50. Sharma S, Zhou X, Thibault DM, Himes BE, Liu A, Szefler SJ, et al. A genome-wide survey of CD4+ lymphocyte regulatory genetic variants identifies novel asthma genes. J Allergy Clin Immunol. 2014;134:1153–62.

51. Zhao Y, Yang G, Wu N, Cao X, Gao J. Integrated transcriptome and phosphoproteome analyses reveal that fads2 is critical for maintaining body LC-PUFA homeostasis. J Proteomics. 2020;229 April:103967.

52. Philippou E, Nikiphorou E. Are we really what we eat? Nutrition and its role in the onset of rheumatoid arthritis. Autoimmun Rev. 2018;17:1074–7.

53. Xu Z, Huang Y, Meese T, Van Nevel S, Holtappels G, Vanhee S, et al. The multi-omics single-cell landscape of sinus mucosa in uncontrolled severe chronic rhinosinusitis with nasal polyps. Clin Immunol. 2023;256 July.

54. Huang Z, Qi G, Miller JS, Zheng SG. CD226: An Emerging Role in Immunologic Diseases. Front Cell Dev Biol. 2020;8 July:1–9.

medRxiv preprint doi: https://doi.org/10.1101/2024.04.02.24305135; this version posted April 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. It is made available under a CC-BY-NC-ND 4.0 International license .

55. Hata K, Yanagihara T, Matsubara K, Kunimura K, Suzuki K, Tsubouchi K, et al. Mass cytometry identifies characteristic immune cell subsets in bronchoalveolar lavage fluid from interstitial lung diseases. Front Immunol. 2023;14 March:1–13.

56. Miyamoto Y, Yamauchi J, Sanbe A, Tanoue A. Dock6, a Dock-C subfamily guanine nucleotide exchanger, has the dual specificity for Rac1 and Cdc42 and regulates neurite outgrowth. Exp Cell Res. 2007;313:791–804.

57. Hassed S, Li S, Mulvihill J, Aston C, Palmer S. Adams–Oliver syndrome review of the literature: Refining the diagnostic phenotype. Am J Med Genet Part A. 2017;173:790–800.

58. Jones KM, Silfvast-Kaiser A, Leake DR, Diaz LZ, Levy ML. Adams–Oliver Syndrome Type 2 in Association with Compound Heterozygous DOCK6 Mutations. Pediatr Dermatol. 2017;34:e249–53.

59. Li X, Jiang M, Chen D, Xu B, Wang R, Chu Y, et al. MiR-148b-3p inhibits gastric cancer metastasis by inhibiting the Dock6/Rac1/Cdc42 axis. J Exp Clin Cancer Res. 2018;37:1–15.

60. Niemöller C, Renz N, Bleul S, Blagitko-Dorfs N, Greil C, Yoshida K, et al. Single cell genotyping of exome sequencing-identified mutations to characterize the clonal composition and evolution of inv(16) AML in a CBL mutated clonal hematopoiesis. Leuk Res. 2016;47:41–6.

61. Mathieu S, Briend M, Abner E, Couture C, Li Z, Bossé Y, et al. Genetic association and Mendelian randomization for hypothyroidism highlight immune molecular mechanisms. iScience. 2022;25.

Figure Legend

Figure 1. Genetic correlation and causality between hypothyroidism and three complex diseases. (**A**) The heritability, genetic covariance, and global genetic association of hypothyroidism with three complex diseases. (**B**) The average local genetic association between hypothyroidism and three complex diseases. The error bars show the 95% confidence intervals of the estimates. (**C**) Summary of bidirectional MR analyses of hypothyroidism and three complex diseases. The

medRxiv preprint doi: https://doi.org/10.1101/2024.04.02.24305135; this version posted April 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. It is made available under a CC-BY-NC-ND 4.0 International license .

abbreviation IVW refers to inverse variance weighting. The dot represents -log10 *p*-values in MR analysis. The red dotted line represents the -log10(0.05). LDSC, linkage disequilibrium score regression; GNOVA, genetic covariance analyzer.

Figure 2. Manhattan plot for genome-wide association study using cross-trait GWAS meta-analysis. The x-axis displays the number of autosomes, while the y-axis displays the -log10 *p*-values for statistical significance derived from the CPASSOC model. The dots represent SNPs. The red dotted line represents the genome-wide significance threshold (p<5e-08). Pleiotropic loci that are significant in single-trait GWASs, including MTAG and CPASSOC, were indicated by text.

Figure 3. Tissue/cell type-specific enrichment of SNP heritability for four diseases. (**A**) The heritability enrichment of tissues for diseases using LDSC and MAGMA. The x-axis displays -log10 *p*-values for each individual test. (**B**) The heritability enrichment of cell types for four diseases using LDSC, MAGMA, and scDRS. An expression specificity matrix created with CELLEX was used to perform LDSC-SEG and MAGMA regression analyses. For scDRS analysis, a co-variable matrix containing assay, gender, age, ethnicity, and time of cold storage (only for the spleen scRNA dataset) was used. Dot colors and point size indicate considerable cell type-disease associations. scDRS: single-cell disease relevance score. FDR indicates that the *p*-value was adjusted using the Benjamini-Hochberg method for all cell types across diseases. (**C**) UMAP embeddings of cell types in the small intestinal with normalized scDRS scores for the four diseases. Pink represents cells that are enriched for the aforementioned disease, while gray implies non-relevant cells. (**D**) The causal association between the listed cell types and four diseases. The dot size shows -log10 *p*-values from MR analysis using the IVW approach. (**E**) Colocalization analyses of cell type QTL for central memory CD4+ T cell absolute count in four diseases. We utilized colocalization to investigate whether the QTL for central memory CD4+ T cell absolute count shared the same causative variant as four other diseases. Each dot indicates a single nucleotide polymorphism, while colors indicate linkage

medRxiv preprint doi: https://doi.org/10.1101/2024.04.02.24305135; this version posted April 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. It is made available under a CC-BY-NC-ND 4.0 International license .

disequilibrium (LD; r^2) with the most likely causative variant, rs3184504. The posterior probability (PH4) was calculated using the "colco.abf" function from the coloc R package. FDR, false discovery rate.

Figure 4. Prioritized genes related to four diseases. (**A**) SMR prioritized genes linked to four diseases. To study the association between gene expression and four diseases, *cis*-eQTL from GTEx v8 (blood, lung, spleen, and small intestine) or eQTLgene (blood) were used. Prioritized genes were statistically related to at least one disease (Bonferroni-corrected $p<0.05$), and P_{HEIDI} scores < 0.05 were used for plotting. The point size and color show the size of associations between genes and diseases. (**B**) Colocalization analysis of potential genes associated with four diseases. We used colocalization to determine whether the putatively shared genes identified in (A) could share the same causative variants with diseases. Colocalization analyses were performed using *cis*-eQTL from GTEx v7 (blood, lung, spleen, and small intestine) or eQTLgene (blood). (**C**) Colocalization analysis of the CCDC88B *cis*-eQTL with hypothyroidism and sarcoidosis. Each dot indicates a single nucleotide polymorphism, while colors indicate linkage disequilibrium (LD; r^2) with the most likely causative variant, rs479777. The posterior probability (PH4) was calculated using the "colco.abf" function in coloc R package.

Figure 5. Prioritized plasma proteins associated with four diseases. (**A**) Volcano graphic depicting the effect of plasma protein on four diseases based on MR analyses. Color dots indicate significant relationships (uncorrected $p<0.05$). The labeled dot represents a significant association that passes multiple test corrections (FDR< 0.05). Only proteins with no significant heterogeneity ($I^2 < 50\%$), directional pleiotropy ($P_{\mathrm{Egger\ intercept}} > 0.05$), reverse causation, and passing leave-one-out analysis were presented in the plotting. (**B**) Shared the associations between 29 causal plasma proteins (statistically significant for at least one disease) and 4 diseases based on MR analysis. (Left) MR analysis reveals the causative effect of plasma proteins on four diseases; (right) colocalization analyses of *cis*-pQTL with four diseases. (**C**) Volcano

medRxiv preprint doi: https://doi.org/10.1101/2024.04.02.24305135; this version posted April 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. It is made available under a CC-BY-NC-ND 4.0 International license .

plot depicting the effect of hypothyroidism on plasma protein based on MR analyses. Color dots indicate significant relationships after multiple testing corrections (FDR<0.05). Proteins were plotted with no significant heterogeneity ($I^2 < 50\%$), directional pleiotropy ($P_{\text{Egger intercept}} > 0.05$), reverse causation, leave-one-out analysis, and multiple testing adjustments (FDR<0.05). (**D**) An MR scatter-plot depicting the effect of hypothyroidism on plasma RIPK2 levels. Using the inverse variance weighted method, an increase in hypothyroidism predicted by one standard deviation was linked to higher levels of RIPK2 (β=0.62, 95%CI=1.20-2.89, *p*=0.0052). (**E**) Enrichment analysis of biological functions for hypothyroidism-influenced proteins, both positively and negatively. (**F**) Shared the associations between 23 hypothyroidism-driven plasma proteins and diseases identified by MR analysis. (Left) Use of MR analysis to determine the causative influence of hypothyroidism-driven plasma proteins on diseases; (right) Colocalization analyses of *cis*-pQTL and four diseases. (**G**) Single-cell resolution analysis of RIPK2 expression in lung cell types from six healthy individuals and five systemic sclerosis-associated interstitial lung disease (SSC-ILD) patients. (Up) Sixteen annotated lung cell types. (Down) RIPK2 expression patterns in six healthy individuals and five SSC-ILD patients. UMAP, Uniform Manifold Approximation and Projection. EC, endothelial cell; AT, Alveolar Type. (**H**) A schematic illustration of the causes and consequences of proteins associated with hypothyroidism, as well as their relationship to the three diseases investigated. Figures generated with BioRender (https://biorender.com/).

Supplementary Figure

Figure S1. Local genetic correlations between hypothyroidism and complex diseases revealed by ρ-HESS (Heritability Estimation from Summary Statistics). Local genetic correlations between hypothyroidism and chronic sinusitis (A), ILD endpoints (B), and sarcoidosis (C), respectively. Significant local genetic correlation estimates are highlighted in red and blue for even and odd chromosomes, respectively.

Supplementary Table

medRxiv preprint doi: https://doi.org/10.1101/2024.04.02.24305135; this version posted April 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. It is made available under a CC-BY-NC-ND 4.0 International license .

Supplementary Table S1. Genetic correlations between hypothyroidism and complex diseases.

Supplementary Table S2. Summary of localized genetic correlations between hypothyroidism and complex diseases.

Supplementary Table S3. Summary of Mendelian randomization results between hypothyroidism and complex diseases using GWASs of hypothyroidism from MRC-IEU Consortium.

Supplementary Table S4. Summary of Mendelian randomization replication results between hypothyroidism and complex diseases using GWASs meta-analysis of hypothyroidism.

Supplementary Table S5. Genome-wide significant loci in a genome-wide cross-trait analysis approach.

Supplementary Table S6. Shared risk regions in different analyses.

Supplementary Table S7. List of 95% credible set SNPs for each independent locus using fine mapping analysis.

Supplementary Table S8. Genetic enrichment for hypothyroidism and complex diseases in specific tissues using LDSC and MAGMA.

Supplementary Table S9. Genetic enrichment for hypothyroidism and complex diseases in specific tissues using LDSC, MAGMA, and scDRS.

Supplementary Table S10. Summary of Mendelian randomization results for causal effects of CD4+ memory-related cells on diseases.

Supplementary Table S11. Summary of Mendelian randomization results for the causal effect of diseases on CD4+ memory-related cells.

Supplementary Table S12. SMR-prioritized genes associated with hypothyroidism and three complex diseases.

Supplementary Table S13. Co-localization analysis of shared SMR priority genes in Table S12.

Supplementary Table S14. Analyzing the causal effect of plasma proteins on disease using MR (unfiltered).

Supplementary Table S15. Significant causal effects of plasma proteins on disease.

medRxiv preprint doi: https://doi.org/10.1101/2024.04.02.24305135; this version posted April 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. It is made available under a CC-BY-NC-ND 4.0 International license .

Supplementary Table S16. Co-localization analysis of etiological proteins that have a causal impact on disease.

Supplementary Table S17. Causal effect of hypothyroidism on plasma proteins using MR (unfiltered).

Supplementary Table S18. Significant causal effect of hypothyroidism on plasma proteins (hypothyroidism-driven proteins).

Figure 1

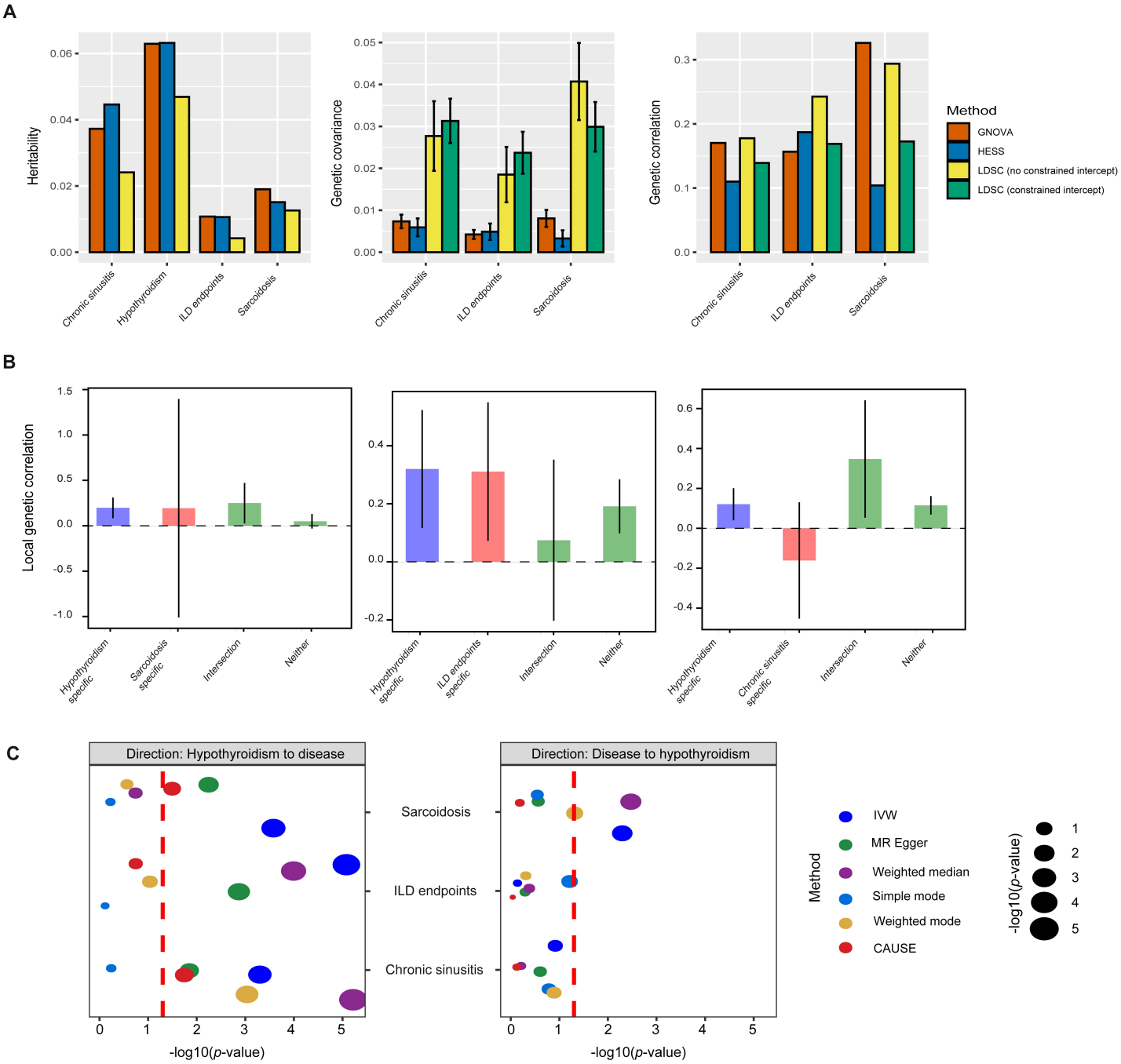

A
Heritability
Genetic covariance
Genetic correlation
Chronic sinusitis
Hypothyroidism
ILD endpoints
Sarcoidosis
Method
GNOVA
HESS
LDSC (no constrained intercept)
LDSC (constrained intercept)
B
Local genetic correlation
Hypothyroidism specific
Sarcoidosis specific
Intersection
Neither
ILD endpoints specific
Chronic sinusitis specific
C
Direction: Hypothyroidism to disease
Direction: Disease to hypothyroidism
Sarcoidosis
ILD endpoints
Chronic sinusitis
-log10(p-value)
Method
IVW
MR Egger
Weighted median
Simple mode
Weighted mode
CAUSE
-log10(p-value)

Figure 2

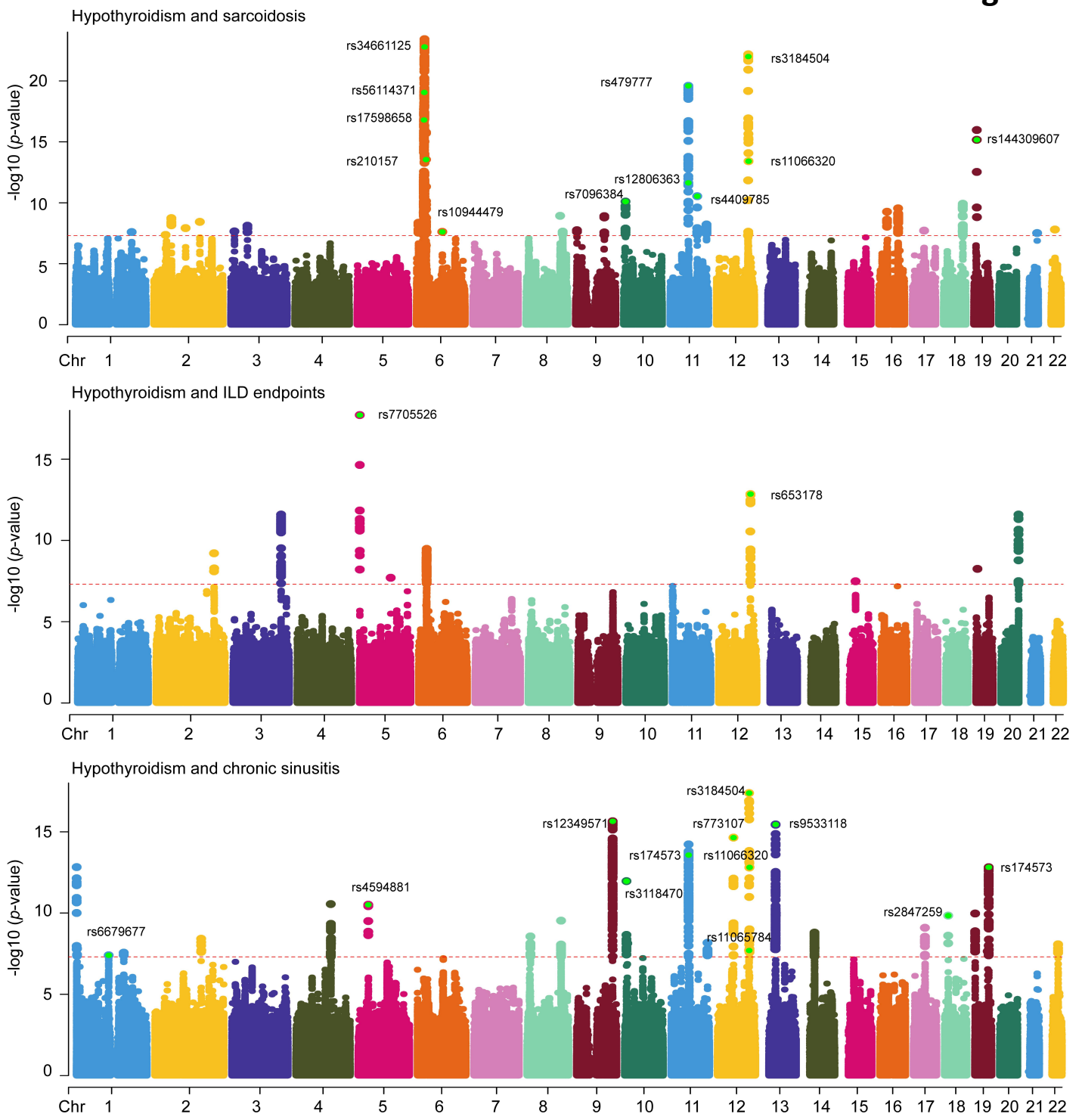

Hypothyroidism and sarcoidosis
-log10 (p-value)
rs34661125
rs56114371
rs17598658
rs210157
rs10944479
rs7096384
rs12806363
rs479777
rs4409785
rs3184504
rs11066320
rs144309607
Chr 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22
Hypothyroidism and ILD endpoints
-log10 (p-value)
rs7705526
rs653178
Chr 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22
Hypothyroidism and chronic sinusitis
-log10 (p-value)
rs6679677
rs4594881
rs12349571
rs174573
rs3118470
rs3184504
rs773107
rs11066320
rs11065784
rs9533118
rs2847259
rs174573
Chr 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22

Figure 3

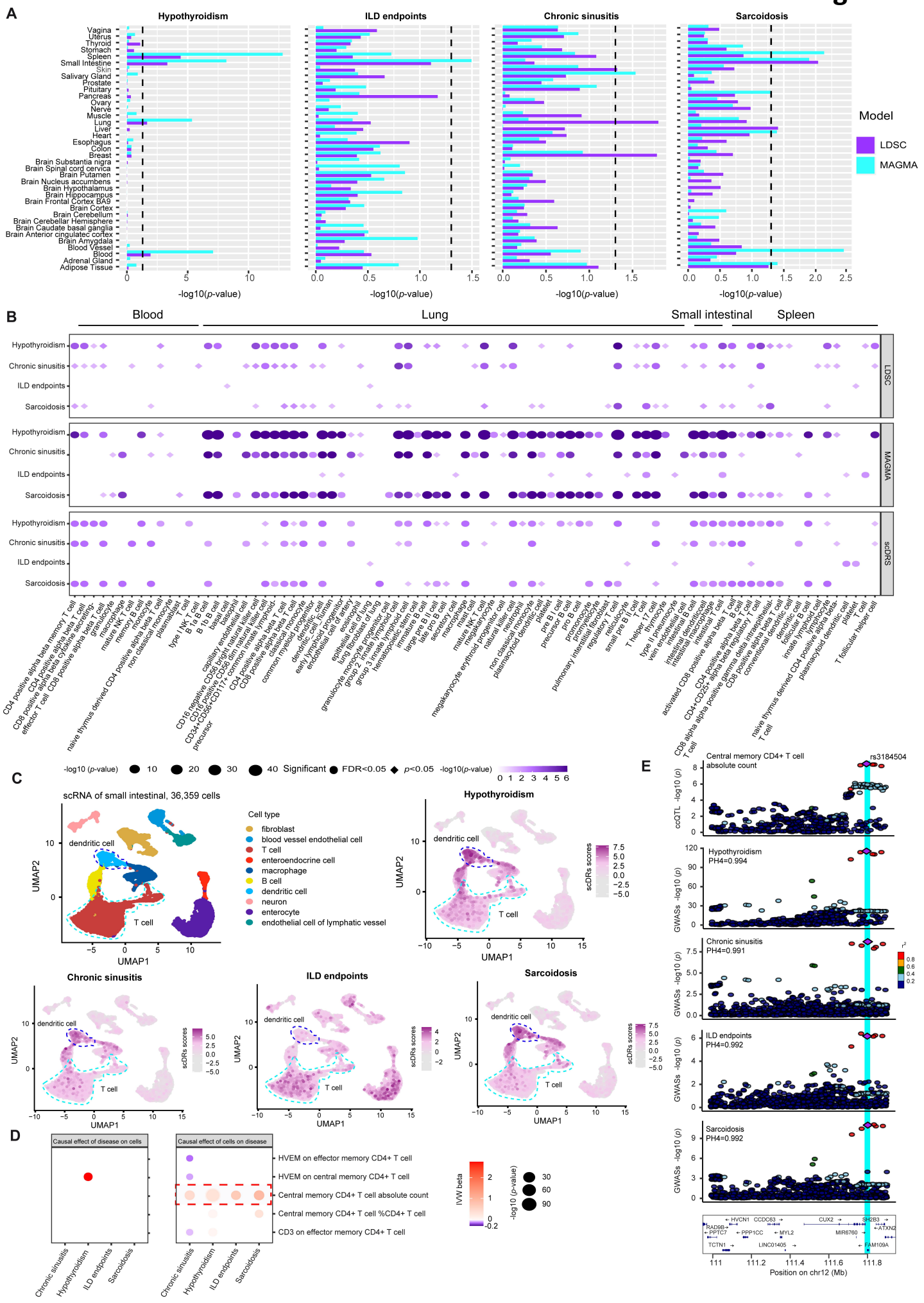

A
Hypothyroidism
ILD endpoints
Chronic sinusitis
Sarcoidosis
Vagina
Uterus
Thyroid
Stomach
Spleen
Small Intestine
Skin
Salivary Gland
Prostate
Pituitary
Pancreas
Ovary
Nerve
Muscle
Lung
Liver
Heart
Esophagus
Colon
Breast
Brain Substantia nigra
Brain Spinal cord cervica
Brain Putamen
Brain Nucleus accumbens
Brain Hypothalamus
Brain Hippocampus
Brain Frontal Cortex BA9
Brain Cortex
Brain Cerebellum
Brain Cerebellar Hemisphere
Brain Caudate basal ganglia
Brain Anterior cingulatec cortex
Brain Amygdala
Blood Vessel
Blood
Adrenal Gland
Adipose Tissue
-log10(p-value)
Model
LDSC
MAGMA
B
Blood
Lung
Small intestinal
Spleen
Hypothyroidism
Chronic sinusitis
ILD endpoints
Sarcoidosis
LDSC
MAGMA
scDRS
-log10 (p-value)
10
20
30
40
Significant
FDR<0.05
p<0.05
-log10(p-value)
C
scRNA of small intestinal, 36,359 cells
Cell type
fibroblast
blood vessel endothelial cell
T cell
enteroendocrine cell
macrophage
B cell
dendritic cell
neuron
enterocyte
endothelial cell of lymphatic vessel
dendritic cell
T cell
UMAP1
UMAP2
Hypothyroidism
Chronic sinusitis
ILD endpoints
Sarcoidosis
scDRs scores
D
Causal effect of disease on cells
Causal effect of cells on disease
HVEM on effector memory CD4+ T cell
HVEM on central memory CD4+ T cell
Central memory CD4+ T cell absolute count
Central memory CD4+ T cell %CD4+ T cell
CD3 on effector memory CD4+ T cell
Chronic sinusitis
Hypothyroidism
ILD endpoints
Sarcoidosis
IVW beta
-log10 (p-value)
E
Central memory CD4+ T cell
absolute count
rs3184504
ccQTL -log10 (p)
Hypothyroidism
PH4=0.994
Chronic sinusitis
PH4=0.991
ILD endpoints
PH4=0.992
Sarcoidosis
PH4=0.992
GWASs -log10 (p)
r²
HVCN1
CCDC63
CUX2
SH2B3
RAD9B
ATXN2
PPTC7
PPP1CC
MYL2
MIR6780
TCTN1
LINC01405
FAM109A
Position on chr12 (Mb)

Figure 4

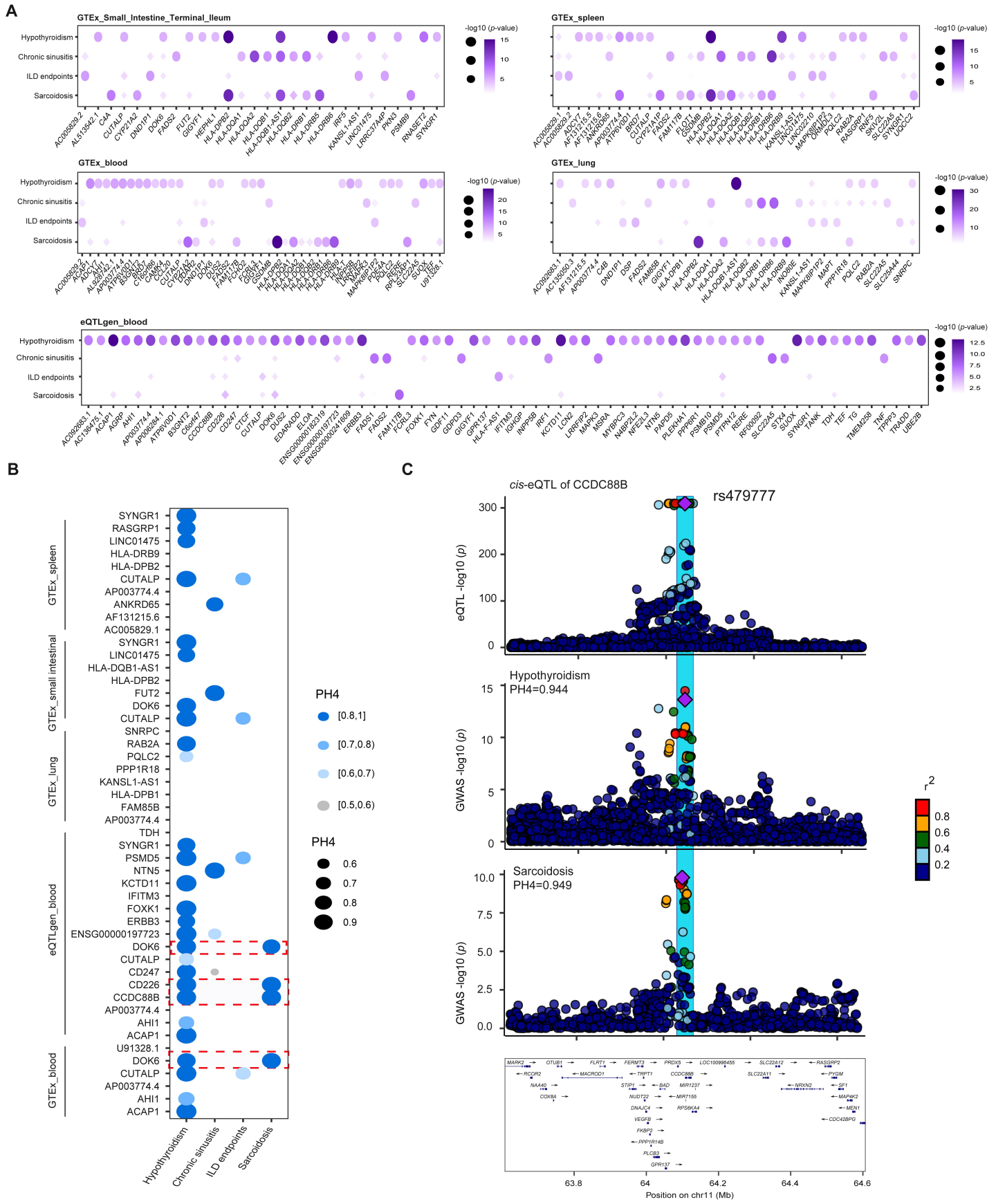
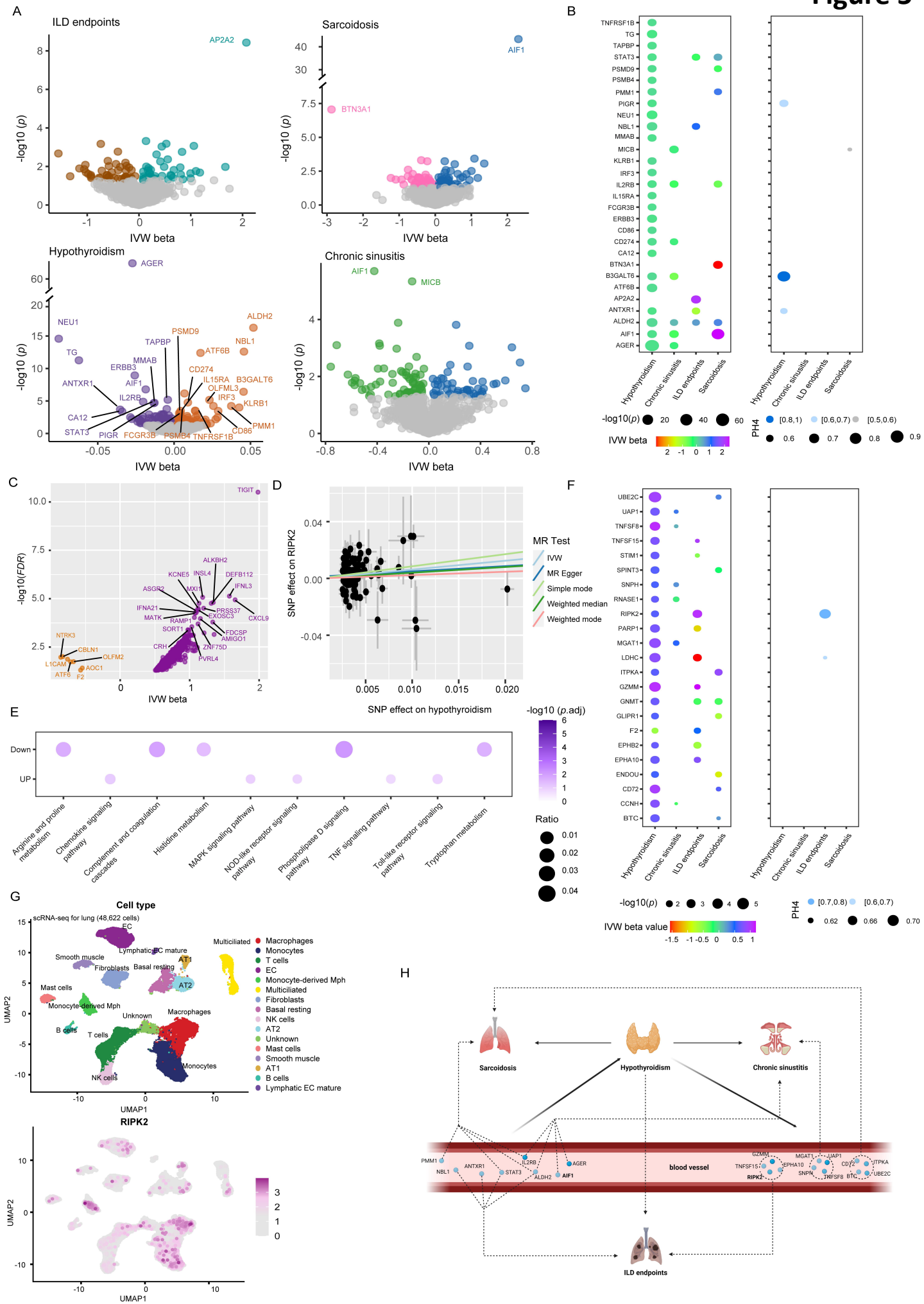

A
GTEx_Small_Intestine_Terminal_Ileum
GTEx_spleen
GTEx_blood
GTEx_lung
eQTLgen_blood
Hypothyroidism
Chronic sinusitis
ILD endpoints
Sarcoidosis
-log10 (p-value)
B
GTEx_spleen
GTEx_small intestinal
GTEx_lung
eQTLgen_blood
GTEx_blood
PH4
[0.8,1]
[0.7,0.8)
[0.6,0.7)
[0.5,0.6)
PH4
0.6
0.7
0.8
0.9
C
cis-eQTL of CCDC88B
rs479777
eQTL -log10 (p)
Hypothyroidism
PH4=0.944
GWAS -log10 (p)
Sarcoidosis
PH4=0.949
r2
0.8
0.6
0.4
0.2
Position on chr11 (Mb)

Figure 5


A
ILD endpoints
AP2A2
Sarcoidosis
AIF1
BTN3A1
Hypothyroidism
AGER
Chronic sinusitis
AIF1
MICB
-log10 (p)
IVW beta
B
Hypothyroidism
Chronic sinusitis
ILD endpoints
Sarcoidosis
-log10(p)
IVW beta
PH4
C
-log10(FDR)
IVW beta
D
SNP effect on RIPK2
SNP effect on hypothyroidism
MR Test
IVW
MR Egger
Simple mode
Weighted median
Weighted mode
E
Down
UP
Arginine and proline metabolism
Chemokine signaling pathway
Complement and coagulation cascades
Histidine metabolism
MAPK signaling pathway
NOD-like receptor signaling pathway
Phospholipase D signaling pathway
TNF signaling pathway
Toll-like receptor signaling pathway
Tryptophan metabolism
-log10 (p.adj)
Ratio
F
Hypothyroidism
Chronic sinusitis
ILD endpoints
Sarcoidosis
-log10(p)
IVW beta value
PH4
G
Cell type
scRNA-seq for lung (48,622 cells)
Macrophages
Monocytes
T cells
EC
Monocyte-derived Mph
Multiciliated
Fibroblasts
Basal resting
NK cells
AT2
Unknown
Mast cells
Smooth muscle
AT1
B cells
Lymphatic EC mature
UMAP1
UMAP2
RIPK2
H
Sarcoidosis
Hypothyroidism
Chronic sinusitis
blood vessel
ILD endpoints