

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

3 Shifang Li, Meijiao Gong

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

5 Correspondence:

6 Shifang Li, fruceslee@gmail.com

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

8
9 **Abstract**

10 Hypothyroidism is a common condition of thyroid hormone insufficiency, and
11 there is growing evidence of its link with additional diseases. It remains unclear
12 whether these associations share a common genetic architecture. To address this gap,
13 by leveraging summary-level genetic data from the UK Biobank of hypothyroidism
14 and the FinnGen study of three complex diseases (sarcoidosis, chronic sinusitis, and
15 interstitial lung disease (ILD) endpoints), we evaluated their shared genetic etiology.
16 A significant genetic correlation was found between hypothyroidism and the three
17 diseases. Cross-trait analyses utilizing the MTAG and CPASSOC models revealed 12,
18 2, and 12 shared loci between hypothyroidism and chronic sinusitis, ILD endpoints,
19 and sarcoidosis, respectively. The SNP heritability enrichment analysis across 37
20 tissues and 136 cell types at the single-cell level identified candidate tissues and cell
21 types that were shared by the diseases. Interestingly, we found a positive genetic
22 relationship between these four diseases and central memory CD4⁺ T cells in the
23 blood, supported by strong colocalization evidence (posterior probability >0.9).
24 Mendelian randomization and colocalization analysis showed a link between
25 hypothyroidism and sarcoidosis with two genes (DOCK6 and CD226) in the blood.
26 Furthermore, among the hypothyroidism-driven plasma proteins, RIPK2 was
27 identified as a potentially actionable mediator of hypothyroidism's effect on ILD
28 endpoints. Overall, our findings contribute to improving our understanding of the
29 molecular basis of these diseases' intricate relationships, as well as providing insights
30 toward disease prevention and comorbidity management.

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

33

34 **Introduction**

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

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

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

70 **Material and methods**

71 **GWAS data source**

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

84 **Heritability and overall genetic correlation analysis**

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

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

99 Compared to LDSC, a genetic covariance analyzer (GNOVA) generates higher
100 estimation accuracy for genetic correlations and a more powerful statistical inference
101 [26]. As a result, GNOVA was used to evaluate the SNP-based heritability and genetic
102 association between hypothyroidism and the other three complex diseases, with
103 default settings. In short, GNOVA calculates genetic covariance using all genetic
104 variants shared by two GWAS summary statistics. Calculations were performed using
105 the 1000 Genomes Project's European population-derived reference data and default
106 parameters. In addition, sample overlap correction between two independent datasets
107 of GWAS summary data was statistically determined.

108 **Local genetic correlation analysis**

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

116 **Cross-trait GWAS meta-analysis**

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

121 phenotype association) [27, 28]. MTAG was a multi-trait genome-wide analysis
122 strategy that increased statistical power when compared to traditional single-trait
123 GWAS analyses. The upper bound for the false discovery rate ("maxFDR") was
124 determined for evaluating the assumptions on the equal variance-covariance of shared
125 SNP impact sizes that underpin the diseases [29]. CPASSOC combines GWAS
126 summary statistics from several correlated characteristics to discover variants
127 associated with at least one trait while correcting for population structure or cryptic
128 relatedness, and so served as a sensitivity study to determine the divergence from
129 MTAG's assumption. The analysis used Shet, a test statistic given by CPASSOC that
130 allows for heterogeneous effects of a characteristic across multiple study designs.
131 Independent SNPs that were genome-wide significant ($p < 5e-08$) in cross-trait
132 meta-analyses (e.g. hypothyroidism-sarcoidosis) using both MTAG and CPASSOC,
133 but not identified in the original single-trait GWAS (e.g. hypothyroidism or
134 sarcoidosis) were prioritized, as identified by LD clumping (parameters: --clump-p1
135 5e-8 --clump-p2 1e-5 --clump-r2 0.2 --clump-kb 1000) in PLINK [30]. Novel SNPs
136 were classified as shared SNPs not driven by a single trait or in LD with index SNPs
137 discovered in single-trait GWASs ($LD\ r^2 < 0.2$). The Ensembl Variant Effect Predictor
138 (VEP) was utilized to provide detailed functional annotation of the identified variants
139 [31].

140 **Summary data-based Mendelian randomization analysis**

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

150 **Colocalization analysis and Bayesian fine-mapping analysis**

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

159 **Tissue enrichment analysis**

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

178 **Cell-type enrichment analyses using scRNA-seq datasets**

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

181 summary statistics of four diseases we studied were integrated using three different
182 genetic prioritization models: LDSC applied to specifically expressed genes
183 (LDSC-SEG), MAGMA, and single-cell disease relevance score (scDRS) [40–42].
184 The first two methods were implemented using the CELLECT snakemake workflow
185 [43]. In brief, the CELLEX method was used for calculating a single ES estimate
186 (ESm) score for each subpopulation in each cell type. Annotations were constructed
187 using 1000 Genomes Project SNPs, as in the default S-LDSC baseline model [44].
188 The Hapmap3 SNPs and the excluded major histocompatibility complex (MHC)
189 region were utilized to calculate LD scores. Following that, LDSC-SEG and
190 MAGMA regression analyses were conducted. Meanwhile, scDRS, which integrates
191 gene expression patterns from scRNA-seq with polygenic disease information from
192 GWASs, was used to discover the cell subpopulations driving GWAS enrichments.
193 Putative disease gene sets were identified from the top 1,000 MAGMA genes
194 weighted by their Z-scores. The normalized disease scores were generated in scDRS
195 by the CLI with "scdrs compute-score" using a covariate matrix that included assay,
196 gender, age, ethnicity, and the time of cold storage (if available). For multiple
197 hypothesis testing, *p*-values were FDR corrected using the Benjamini-Hochberg
198 methods across all tissues and diseases.

199 **Mendelian randomization and mediation analysis**

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

211 instrument heterogeneity was estimated using the Cochran Q test and I^2 statistics with
212 the "Isq()" method. Leave-one-out analysis was conducted to determine whether the
213 observed correlation was caused by any single IV. To rule out any pleiotropic effects,
214 we evaluated each selected instrument SNP in Open Targets Genetics
215 (<https://genetics.opentargets.org/>) databases for previously reported associations.
216 Associations with $p < 5e-08$ were deemed statistically significant. To account for
217 multiple testings, we utilized Benjamini-Hochberg correction and a significance
218 criterion of $FDR < 0.05$ to evaluate statistical significance. Furthermore, the CAUSE
219 model, which accounts for correlated and uncorrelated horizontal pleiotropic effects
220 through a multivariate linear model adjusted by a joint distribution of instrumental
221 SNPs, was used to avoid more false positives caused by correlated horizontal
222 pleiotropy than previous methods [45]. To estimate the causal mediation effects
223 (β_{mediated}), network MR with a product of coefficients method was employed, as
224 described by Yoshiji and others [46]. In brief, we first estimated the effect of protein
225 levels on hypothyroidism ($\beta_{\text{protein-to-hypothyroidism}}$) and the effect of hypothyroidism on
226 chronic sinusitis ($\beta_{\text{hypothyroidism-to-chronic sinusitis}}$). After that, we multiplied these values
227 ($\beta_{\text{mediated}} = \beta_{\text{protein-to-hypothyroidism}} \times \beta_{\text{hypothyroidism-to-chronic sinusitis}}$) and divided β_{mediated} by
228 $\beta_{\text{protein-to-chronic sinusitis}}$ to estimate the proportion mediated.

229 **Results**

230 **Hypothyroidism shows a significant genetic correlation with complex diseases**

231 Bivariate LDSC was used to evaluate the genetic association (without a
232 constrained intercept) between hypothyroidism and the three complex diseases
233 (**Figure 1A**). Using unconstrained LDSC, three complex diseases were identified as
234 having a substantial genetic correlation with hypothyroidism (ILD endpoints,
235 $R_g=0.2425$, $p=0.0063$; sarcoidosis, $R_g=0.2937$, $p=2.05e-05$; chronic sinusitis,
236 $R_g=0.1776$, $p=0.0004$). The liability-scale SNP heritability estimates for
237 hypothyroidism, ILD endpoints, sarcoidosis, and chronic sinusitis were 0.0469,
238 0.0042, 0.0126, and 0.0241, respectively. The computed intercept of genetic
239 covariance ranged from 0.018 to 0.040, indicating a slight sample overlap between
240 hypothyroidism and the selected complex diseases. Given the limited sample overlap

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

247 Given the significant global genetic association, we investigated whether distinct
248 genomic regions conferred local genetic correlation at genomic regions with GWAS
249 sites relevant to each trait. A total of 13 suggestively significant region pairs were
250 discovered (uncorrected $p < 0.05$, ρ -HESS, **Supplementary Table S2**). There are 4
251 regions for hypothyroidism and chronic sinusitis, 6 regions for hypothyroidism and
252 sarcoidosis, and 3 regions for hypothyroidism and ILD endpoints. The average local
253 genetic association was nearly the same in regions with hypothyroidism-specific loci
254 or the three complex disease-specific loci (**Supplementary Figure S1**). No other
255 common region with a substantial local genetic association was discovered
256 (**Supplementary Figure S1**). Interestingly, it was found that the correlation at
257 hypothyroidism-specific regions (local $R_g = 0.2$, $SE = 0.058$) is significantly greater
258 than sarcoidosis-specific (local $R_{g_{\text{hypothyroidism}}} = 0.2$, $SE_{\text{hypothyroidism}} = 0.058$ versus local
259 $R_{g_{\text{sarcoidosis}}} = 0.19$, $SE_{\text{sarcoidosis}} = 0.61$) as well as chronic sinusitis-specific loci (local
260 $R_{g_{\text{hypothyroidism}}} = 0.12$, $SE_{\text{hypothyroidism}} = 0.041$ versus local $R_{g_{\text{chronic sinusitis}}} = -0.16$, $SE_{\text{chronic sinusitis}} = 0.15$), indicating that loci that increase hypothyroidism tend to consistently
261 increase the risk of sarcoidosis and chronic sinusitis (**Figure 1B**). Overall, these
262 results suggest hypothyroidism and complex diseases are likely linked due to the
263 sharing of genetic variants across the entire genome rather than in specific genomic
264 regions.
265

266 **Hypothyroidism has a causal effect on sarcoidosis and chronic sinusitis**

267 Bidirectional MR was used to investigate the probable causative effect and
268 whether the shared genetic basis of hypothyroidism and complex diseases supported
269 pleiotropy. Genetically predicted hypothyroidism was linked to increased risk of three
270 complex diseases (IVW $\beta = 3.49$ for ILD endpoints, $SE = 0.68$, $p\text{-value} = 2.63 \times 10^{-7}$;

271 IVW beta=4.26 for sarcoidosis, SE=0.87, p -value=8.83e-07; IVW beta=1.78 for
272 chronic sinusitis, SE=0.47, p -value=1.87e-04) (**Supplementary Table S3**). The
273 MR-Egger intercept test revealed no directional pleiotropy, supporting the validity of
274 the findings. The positive relationships between hypothyroidism and these three
275 diseases were not driven by outliers, as confirmed by leave-one analysis. In the
276 reverse direction, no significant association was found between these three complex
277 diseases and the risk of hypothyroidism ((odds ratio) OR=1.0026 for ILD endpoints,
278 95% confidence interval (CI):0.998-1.0064, p -value=0.19; OR=1.0062 for sarcoidosis,
279 95%CI: 0.99-1.012, p -value=0.061; OR=1.012 for chronic sinusitis, 95%CI:
280 0.99-1.03, p -value=0.209). These findings are consistent across several MR
281 approaches (**Supplementary Table S3**). To validate these results, the largest-to-date
282 GWASs of hypothyroidism, involving 51,194 cases of hypothyroidism and 443,383
283 controls from Finland and the UK Biobank study, were utilized to evaluate the
284 causation between hypothyroidism and the other three diseases (**Figure 1C** and
285 **Supplementary Table S4**) [46]. All of these causal associations between
286 hypothyroidism and complex disease could be repeated (**Figure 1C** and
287 **Supplementary Table S4**). However, the CAUSE model shows that, aside from ILD
288 endpoints, only chronic sinusitis and sarcoidosis are genetically affected by
289 hypothyroidism.

290 **Cross-trait meta-analysis and pleiotropic loci**

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

301 $P_{\text{hypothyroidism}}=6.4\text{e-}08$, $P_{\text{sarcoidosis}}=1.52\text{e-}06$, $P_{\text{CPASSOC\&MTAG}}<5\text{e-}08$). The MTAG studies
302 on hypothyroidism, sarcoidosis, chronic sinusitis, and interstitial lung disease yielded
303 maxFDR values of 0.00044, 0.00825, 0.026, and 0.27, respectively. Notably, the
304 MTAG results were highly comparable with those generated by CPASSOC, implying
305 that the MTAG results are reliable and that bias in MTAG assumptions is unlikely to
306 be significant. Among these 26 loci, 11 risk SNPs were consistently significant
307 ($p<0.05$) when examined by ρ -HESS methods (**Supplementary Table S2**), and 16
308 loci had colocalization probabilities (PH4) over 0.8, confirming that hypothyroidism
309 and the selected three diseases share the same genetic variants (**Supplementary Table**
310 **S5**). Finally, two (rs11066320 and rs3184504), one (rs653178), and four (rs12349571,
311 rs11065784, rs11066320, and rs3184504) common loci between hypothyroidism and
312 sarcoidosis, ILD endpoints, and chronic sinusitis, respectively, were verified by
313 different approaches ($\text{PPA4}>0.8$, **Supplementary Table S6**). Interestingly, the two
314 independent loci shared by hypothyroidism and sarcoidosis (rs11066320 and
315 rs3184504) were also found in the cross-traits study of hypothyroidism and chronic
316 sinusitis (**Supplementary Table S5**). Notably, rs653178, the common risk locus for
317 hypothyroidism and ILD endpoints, is located on an intron of ATXN2, which also
318 serves as a TF-binding site for XBP1, CREB3, BATF3, JDP2, FOS, ATF7, and JUN.

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

325 **SNP heritability enrichment at the tissue and cell type level**

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

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

337 Given tissue heterogeneity, we used publicly available scRNA-seq datasets from
338 four tissues: whole blood (50,115 cells), spleen (94,256 cells), small intestinal (36,359
339 cells), and lung (71,752 cells), to assess the genetic association with cell type
340 expression specificity for hypothyroidism and the other three diseases (**Figure 3B** and
341 **Supplementary Table S9**). To robustly identify the tissues implied by these traits,
342 three approaches (LDSC, MAGMA, and scDRS) were implemented, each with a
343 different assumption and procedure. Specifically, scDRSs, an orthogonal approach to
344 other methodologies, were utilized to detect GWAS trait enrichment at the single-cell
345 level. This approach assesses not only the associations between cell types and GWAS
346 traits but also the heterogeneity of the associations within cell types. Following the
347 heritability enrichment analysis, we uncovered significant enrichment for
348 hypothyroidism with sarcoidosis and chronic sinusitis in CD4 positive
349 alpha-beta memory T cells, CD4 positive alpha-beta T cells, and CD8 positive
350 alpha-beta T cells in blood, indicating that the polygenic risks associated with these
351 three diseases were enriched in the cells responsible for adaptive immunity (**Figure**
352 **3B** and **Supplementary Table S9**). Hypothyroidism was also found to share group 2
353 innate lymphoid cells, group 3 innate lymphoid cells, CD4 positive alpha beta T cells,
354 CD8 positive alpha beta T cell regulatory T cells, and T helper 17 cells in the lung
355 with sarcoidosis and chronic sinusitis. Furthermore, the activated CD8+ alpha beta T
356 cells in the spleen showed a greater heritability enrichment for hypothyroidism with
357 sarcoidosis and chronic sinusitis, with significant heterogeneity. Interestingly, both
358 hypothyroidism and sarcoidosis showed considerable heritability enrichment in T
359 cells and dendritic cells in the small intestine (**Figure 3C** and **Supplementary Table**
360 **S9**), shedding light on the biology of comorbidity via the same cell type.

361 Of note, the significant heritability does not necessarily indicate a causal
362 relationship between cell types and diseases. To validate the causative influence of
363 CD4 positive alpha-beta memory T cells in the blood associated with diseases (**Figure**
364 **3B**), 14 CD4+ memory-related T cells from a cohort of 3,757 Sardinians were used in
365 MR analysis (**Supplementary Table S10-S11**). After multiple corrections, 9 pairs of
366 cell-disease associations were identified. Interestingly, it was found that central
367 memory CD4+ T cell absolute count was significantly genetically associated with
368 hypothyroidism (OR=1.008, 95%CI=1.0029-1.013; FDR=1.16e-02), ILD endpoints
369 (OR=1.94, 95%CI=1.49-2.51; FDR=9.44e-06), sarcoidosis (OR=2.49,
370 95%CI=1.92-3.25; FDR=1.41e-10), and chronic sinusitis (OR=1.51,
371 95%CI=1.32-1.73; FDR=2.86e-08), respectively (**Figure 3C** and **Supplementary**
372 **Table S10-S11**). Furthermore, colocalization analysis showed that central memory
373 CD4+ T cell absolute count may share the same genetic variants with hypothyroidism
374 (PH4=0.994), ILD endpoints (PH4=0.992), sarcoidosis (PH4=0.992), and chronic
375 sinusitis (PH4=0.991) (**Figure 3D**). Altogether, these findings indicate that
376 hypothyroidism appears to influence the link between central memory CD4+ T cell
377 absolute count and sarcoidosis or chronic sinusitis.

378 **Identification of shared functional genes for hypothyroidism and complex** 379 **diseases**

380 To infer causality and identify putative functional genes for hypothyroidism and
381 complex diseases, we utilized GWAS summary data with small intestine, spleen, and
382 lung eQTL summary data from GTEx, as well as whole blood eQTL summary data
383 from eQTLGen and GTEx resources (v8) (**Figure 4A** and **Supplementary Table**
384 **S12**). After multiple corrections (Bonferroni-corrected $p < 0.05$), two substantial risk
385 gene pairs (HLA-DPB2 and HLA-DQB1-AS1) in the small intestine were identified
386 for hypothyroidism and related diseases. Hypothyroidism and sarcoidosis share two
387 genes including AP003774.4 and HLA-DPB2 in the spleen. In addition, PPP1R18 in
388 the lung is genetically shared by hypothyroidism and sarcoidosis (SMR adjusted
389 $p < 0.05$ and $P_{\text{HEIDI}} > 0.05$). Interestingly, some other common risk genes were
390 discovered, although they were not statistically significant in both hypothyroidism

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

407 **Identification of shared functional plasma proteins for hypothyroidism and** 408 **diseases**

409 Because plasma proteome is an abundant resource of potential drug targets for
410 diseases, putative shared functional circulation proteins for hypothyroidism and
411 complex diseases were identified using large-scale GWASs of 4,907 circulating
412 proteins in 35,559 Icelanders (**Figure 5A** and **Supplementary Table S14**). Given that
413 *cis*-pQTLs were anticipated to have a more direct and specific biological influence on
414 the protein (relative to *trans*-pQTLs), MR studies with only *cis*-pQTLs as IVs and
415 diseases as outcomes were carried out [47]. After filtering for heterogeneity ($I^2 < 50\%$
416 for all), directional pleiotropy ($P_{\text{Egger intercept}} > 0.05$), reverse causation ($P_{\text{Steiger test}} < 0.05$),
417 and leave-one-out analysis, 34 causality proteins were observed (FDR < 0.05) (**Figure**
418 **5A** and **Supplementary Table S15**). Interestingly, AIF1 was found to be negatively
419 associated with a decreased risk of hypothyroidism (OR[95%CI]=0.981[0.974, 0.988];
420 $p=1.65e-07$) and chronic sinusitis (OR[95%CI]=0.655[0.550, 0.780]; $p=2.00e-06$),

421 despite a positive correlation with sarcoidosis (OR[95%CI]=10.33[7.40, 14.43];
422 $p=6.32e-43$). To evaluate the indirect effect of proteins on chronic sinusitis via
423 hypothyroidism, we performed a mediation analysis with effect estimates from
424 two-step MR and the overall effect from primary MR. The mediation analysis
425 revealed that AIF1 had a modest mediation effect on chronic sinusitis via
426 hypothyroidism (7.83%). In addition to AIF1, no other proteins were identified to be
427 statistically consistently associated with both hypothyroidism and complex diseases.
428 However, at the nominal level, 19 protein-disease associations are shown (**Figure 5B**
429 **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
430 hypothyroidism and three selected complex diseases ($\beta_{\text{hypothyroidism to diseases}} > 0$). For
431 example, a genetic tendency to elevated ALDH2 was found to be substantially linked
432 with an increased risk of hypothyroidism (OR[95%CI]=1.053[1.040, 1.066];
433 $p=5.26e-17$), as well as an increased risk of ILD endpoints, sarcoidosis, and chronic
434 sinusitis. Furthermore, higher genetically predicted levels of both ANTXR1
435 (OR[95%CI]=0.316[0.128, 0.782]; $p=0.0127$) and NBL (OR[95%CI]=2.704[1.149,
436 6.36]; $p=0.022$) were linked with decreased risk of ILD endpoints. Notably, only
437 B3GALT6, which had >80% of the posterior probability of colocalization of the
438 genetic association with hypothyroidism, was explained by the same genetic variant
439 (rs67492154), whereas no strong colocalization evidence was found for other proteins
440 with the corresponding diseases (**Figure 5B and Supplementary Table S16**).

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

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

451 directional pleiotropy test ($P_{\text{Egger intercept}} > 0.05$), 1,315 proteins were estimated to be
452 impacted by hypothyroidism ($\text{FDR} < 0.05$). 207 proteins that failed leave-one-out
453 analysis and 6 proteins with bidirectional effects were also removed from further
454 investigation. As a result, a total of 1,102 protein levels, including RIPK2, were
455 identified as hypothyroidism-driven proteins, with no apparent heterogeneity
456 ($I^2 < 50\%$), directional pleiotropy ($P_{\text{Egger intercept}} > 0.05$), or reverse causation (**Figure**
457 **5D** and **Supplementary Table S18**). The weighted median, weighted mode, and
458 MR-Egger slope methods all produced directionally consistent results with the IVW
459 method for RIPK2 (**Supplementary Table S18**). Among these hypothyroidism-driven
460 proteins, proteins that have a positive effect from hypothyroidism were significantly
461 enriched in the toll-like receptor signaling pathway, chemokine signaling pathway,
462 NOD-like receptor signaling pathway, and MAPK signaling pathway, while
463 negatively affected proteins were negatively associated with histidine metabolism,
464 tryptophan metabolism, and neurotrophin signaling pathway, indicating the imprint
465 impact of hypothyroidism on the immune response (**Figure 5E**).

466 We next used two-sample MR to assess the causative effects of the identified
467 hypothyroidism-driven proteins on three complex diseases. We again employed
468 *cis*-pQTLs for these proteins as IVs, and GWASs from three diseases as outcome
469 variables. Following the *cis*-pQTL search and data harmonization, 130 proteins were
470 analyzed in MR to assess their estimated causal effect on three diseases. After
471 multiple testing corrections ($\text{FDR} < 0.05$), no significant association was identified
472 between hypothyroidism-driven proteins and the three diseases. However, at the
473 nominal significance level, 24 protein-disease association pairs were identified
474 (**Figure 5F**). Particularly, 13 pairs of associations ($\beta_{\text{protein to diseases}} \times \beta_{\text{hypothyroidism to protein}} > 0$)
475 reveal a consistent direction of hypothyroidism's effect on diseases ($\beta_{\text{hypothyroidism to$
476 $\text{diseases}} > 0$). For example, higher genetically predicted RIPK2 levels
477 ($\text{OR} [95\% \text{CI}] = 2.88 [1.55, 5.38]; p = 8.5e-04$) were related to an increased risk of ILD
478 endpoints (**Figure 5F**). To check out the premise of a lack of directional pleiotropy,
479 which could reintroduce confounding, we used the Open Target Genetics
480 (<https://genetics.opentargets.org/>) databases to determine whether the RIPK2

481 *cis*-pQTLs (rs160438) were related to any traits. The deCODE study's lead *cis*-pQTL
482 for RIPK2 (rs160438) did not disclose any additional relationship at the genome-wide
483 significance threshold of $p < 5e-08$. Colocalization analysis revealed that only RIPK2
484 had medium colocalization with the diseases (**Figure 5F** and **Supplementary Table**
485 **S16**). To better understand the origin of RIPK2 in the lung at the scRNA-seq level, we
486 evaluated single-cell RIPK2 expression in human lung tissues from six healthy people
487 and five systemic sclerosis-associated interstitial lung disease (SSC-ILD) patients, as
488 described previously [48, 49]. RIPK2 was mainly expressed in monocytes,
489 macrophages, and monocyte-derived macrophages (monocyte-derived Mph) (**Figure**
490 **5G**). Overall, these findings suggested that RIPK2 may be a potential mediator of the
491 effect of hypothyroidism on ILD endpoints.

492 **Discussion**

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

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

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

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

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

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

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

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

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

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

603 **Acknowledgments**

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

609 **Data availability and code availability**

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

614 **Contributions**

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

618 **Funding**

619 Not applicable.

620 **Declarations**

621 *Ethics approval and consent to participate*

622 Not applicable.

623 *Competing interests*

624 The authors declare that they have no competing interests.

625 **References**

- 626 1. Bianco AC. Emerging Therapies in Hypothyroidism. *Annu Rev Med.*
627 2024;75:307–19.
- 628 2. Teumer A, Chaker L, Groeneweg S, Li Y, Di Munno C, Barbieri C, et al.
629 Genome-wide analyses identify a role for SLC17A4 and AADAT in thyroid hormone
630 regulation. *Nat Commun.* 2018;9:1–14.

- 631 3. Vanderpump MPJ. The epidemiology of thyroid disease. *Br Med Bull.*
632 2011;99:39–51.
- 633 4. Gussekloo J, Van Exel E, De Craen AJM, Meinders AE, Frölich M, Westendorp
634 RGJ. Thyroid status, disability and cognitive function, and survival in old age. *Jama.*
635 2004;292:2591–9.
- 636 5. Wieland DR, Wieland JR, Wang H, Chen YH, Lin CH, Wang JJ, et al. Thyroid
637 Disorders and Dementia Risk: A Nationwide Population-Based Case-Control Study.
638 *Neurology.* 2022;99:E679–87.
- 639 6. Chaker L, Razvi S, Bensenor IM, Azizi F, Pearce EN, Peeters RP. Hypothyroidism.
640 *Nat Rev Dis Prim.* 2022;8.
- 641 7. Chaker L, Korevaar TIM, Medici M, Uitterlinden AG, Hofman A, Dehghan A, et al.
642 Thyroid Function Characteristics and Determinants: The Rotterdam Study. *Thyroid.*
643 2016;26:1195–204.
- 644 8. Kašiković Lečić S, Javorac J, Lovrenski A, Đokić J, Sazdanić Velikić D, Živanović
645 D. Case report: Thyroid sarcoidosis as a rare localization of the disease: Report of two
646 cases and review of the literature. *Front Med.* 2023;10 March:1–8.
- 647 9. Choi HG, Kim TJ, Hong SK, Min C, Yoo DM, Kim H, et al. Thyroid Diseases and
648 Chronic Rhinosinusitis: A Nested Case–Control Study Using a National Health
649 Screening Cohort. *Int J Environ Res Public Health.* 2022;19.
- 650 10. O’Dwyer DN, Wang BR, Nagaraja V, Flaherty KR, Khanna D, Murray S, et al.
651 Hypothyroidism Is Associated with Increased Mortality in Interstitial Pneumonia with
652 Autoimmune Features. *Ann Am Thorac Soc.* 2022;19:1772–6.
- 653 11. Oldham JM, Kumar D, Lee C, Patel SB, Takahashi-Manns S, Demchuk C, et al.
654 Thyroid disease is prevalent and predicts survival in patients with idiopathic
655 pulmonary fibrosis. *Chest.* 2015;148:692–700.
- 656 12. Skrivankova VW, Richmond RC, Woolf BAR, Davies NM, Swanson SA,
657 Vanderweele TJ, et al. Strengthening the reporting of observational studies in
658 epidemiology using mendelian randomisation (STROBE-MR): Explanation and
659 elaboration. *BMJ.* 2021;375.
- 660 13. Elsworth B, Lyon M, Alexander T, Liu Y, Matthews P, Hallett J, et

661 al. The MRC IEU OpenGWAS data infrastructure. *bioRxiv*.
662 2020;:2020.08.10.244293.

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

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

669 16. Orrù V, Steri M, Sidore C, Marongiu M, Serra V, Olla S, et al. Complex genetic
670 signatures in immune cells underlie autoimmunity and inform therapy. *Nat Genet*.
671 2020;52:1036–45.

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

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

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

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

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

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

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

691 registry-based study of incidence, prevalence and diagnostic work-up. *Respir Med.*
692 2021;187 March.

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

696 25. Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh PR, et al. An
697 atlas of genetic correlations across human diseases and traits. *Nat Genet.*
698 2015;47:1236–41.

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

702 27. Liu Q, Tang B, Zhu Z, Kraft P, Deng Q, Stener-Victorin E, et al. A genome-wide
703 cross-trait analysis identifies shared loci and causal relationships of type 2 diabetes
704 and glycaemic traits with polycystic ovary syndrome Abbreviations 2hGlu adj BMI
705 2h glucose after an oral glucose challenge adjusted for BMI CPASSOC Cross.
706 2022;:1483–94.

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

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

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

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

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

- 721 33. Li SF, Gong MJ. Drug targets for haemorrhoidal disease: Proteome-wide
722 Mendelian randomisation and colocalisation analyses. *Gut*. 2024;0:1–2.
- 723 34. Zou Y, Carbonetto P, Wang G, Stephens M. Fine-mapping from summary data
724 with the “Sum of Single Effects” model. *PLoS Genet*. 2022;18:1–24.
- 725 35. Bryois J, Skene NG, Hansen TF, Kogelman LJA, Watson HJ, Liu Z, et al. Genetic
726 identification of cell types underlying brain complex traits yields insights into the
727 etiology of Parkinson’s disease. *Nat Genet*. 2020;52:482–93.
- 728 36. Jones RC, Karkanias J, Krasnow MA, Pisco AO, Quake SR, Salzman J, et al. The
729 *Tabula Sapiens*: A multiple-organ, single-cell transcriptomic atlas of humans. *Science*
730 (80-). 2022;376.
- 731 37. Madissoon E, Wilbrey-Clark A, Miragaia RJ, Saeb-Parsy K, Mahbubani KT,
732 Georgakopoulos N, et al. ScRNA-seq assessment of the human lung, spleen, and
733 esophagus tissue stability after cold preservation. *Genome Biol*. 2019;21:1–16.
- 734 38. Egozi A, Llivichuzhca-Loja D, McCourt BT, Bahar Halpern K, Farack L, An X, et
735 al. Insulin is expressed by enteroendocrine cells during human fetal development. *Nat*
736 *Med*. 2021;27:2104–7.
- 737 39. He P, Lim K, Sun D, Pett JP, Jeng Q, Polanski K, et al. A human fetal lung cell
738 atlas uncovers proximal-distal gradients of differentiation and key regulators of
739 epithelial fates. *Cell*. 2022;185:4841-4860.e25.
- 740 40. Finucane HK, Reshef YA, Anttila V, Slowikowski K, Gusev A, Byrnes A, et al.
741 Heritability enrichment of specifically expressed genes identifies disease-relevant
742 tissues and cell types. *Nat Genet*. 2018;50:621–9.
- 743 41. de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: Generalized
744 Gene-Set Analysis of GWAS Data. *PLoS Comput Biol*. 2015;11:1–19.
- 745 42. Zhang MJ, Hou K, Dey KK, Sakaue S, Jagadeesh KA, Weinand K, et al.
746 Polygenic enrichment distinguishes disease associations of individual cells in
747 single-cell RNA-seq data. *Nat Genet*. 2022;54:1572–80.
- 748 43. Timshel PN, Thompson JJ, Pers TH. Genetic mapping of etiologic brain cell types
749 for obesity. *Elife*. 2020;9:1–45.
- 750 44. Altshuler DM, Durbin RM, Abecasis GR, Bentley DR, Chakravarti A, Clark AG,

751 et al. An integrated map of genetic variation from 1,092 human genomes. *Nature*.
752 2012;491:56–65.

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

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

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

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

765 49. Li SF, Gong MJ. Integrative single-cell analysis of epigenomic and transcriptomic
766 states in patients with systemic sclerosis-associated interstitial lung disease. *medRxiv*.
767 2023.

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

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

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

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

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

- 781 55. Hata K, Yanagihara T, Matsubara K, Kunimura K, Suzuki K, Tsubouchi K, et al.
782 Mass cytometry identifies characteristic immune cell subsets in bronchoalveolar
783 lavage fluid from interstitial lung diseases. *Front Immunol.* 2023;14 March:1–13.
- 784 56. Miyamoto Y, Yamauchi J, Sanbe A, Tanoue A. Dock6, a Dock-C subfamily
785 guanine nucleotide exchanger, has the dual specificity for Rac1 and Cdc42 and
786 regulates neurite outgrowth. *Exp Cell Res.* 2007;313:791–804.
- 787 57. Hased S, Li S, Mulvihill J, Aston C, Palmer S. Adams–Oliver syndrome review
788 of the literature: Refining the diagnostic phenotype. *Am J Med Genet Part A.*
789 2017;173:790–800.
- 790 58. Jones KM, Silfvast-Kaiser A, Leake DR, Diaz LZ, Levy ML. Adams–Oliver
791 Syndrome Type 2 in Association with Compound Heterozygous DOCK6 Mutations.
792 *Pediatr Dermatol.* 2017;34:e249–53.
- 793 59. Li X, Jiang M, Chen D, Xu B, Wang R, Chu Y, et al. MiR-148b-3p inhibits gastric
794 cancer metastasis by inhibiting the Dock6/Rac1/Cdc42 axis. *J Exp Clin Cancer Res.*
795 2018;37:1–15.
- 796 60. Niemöller C, Renz N, Bleul S, Blagitko-Dorfs N, Greil C, Yoshida K, et al. Single
797 cell genotyping of exome sequencing-identified mutations to characterize the clonal
798 composition and evolution of inv(16) AML in a CBL mutated clonal hematopoiesis.
799 *Leuk Res.* 2016;47:41–6.
- 800 61. Mathieu S, Briend M, Abner E, Couture C, Li Z, Bossé Y, et al. Genetic
801 association and Mendelian randomization for hypothyroidism highlight immune
802 molecular mechanisms. *iScience.* 2022;25.

803

804 **Figure Legend**

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

811 abbreviation IVW refers to inverse variance weighting. The dot represents $-\log_{10}$
812 p -values in MR analysis. The red dotted line represents the $-\log_{10}(0.05)$. LDSC,
813 linkage disequilibrium score regression; GNOVA, genetic covariance analyzer.

814

815 **Figure 2. Manhattan plot for genome-wide association study using cross-trait**
816 **GWAS meta-analysis.** The x-axis displays the number of autosomes, while the y-axis
817 displays the $-\log_{10}$ p -values for statistical significance derived from the CPASSOC
818 model. The dots represent SNPs. The red dotted line represents the genome-wide
819 significance threshold ($p < 5e-08$). Pleiotropic loci that are significant in single-trait
820 GWASs, including MTAG and CPASSOC, were indicated by text.

821

822 **Figure 3. Tissue/cell type-specific enrichment of SNP heritability for four**
823 **diseases.** (A) The heritability enrichment of tissues for diseases using LDSC and
824 MAGMA. The x-axis displays $-\log_{10}$ p -values for each individual test. (B) The
825 heritability enrichment of cell types for four diseases using LDSC, MAGMA, and
826 scDRS. An expression specificity matrix created with CELLEX was used to perform
827 LDSC-SEG and MAGMA regression analyses. For scDRS analysis, a co-variable
828 matrix containing assay, gender, age, ethnicity, and time of cold storage (only for the
829 spleen scRNA dataset) was used. Dot colors and point size indicate considerable cell
830 type-disease associations. scDRS: single-cell disease relevance score. FDR indicates
831 that the p -value was adjusted using the Benjamini-Hochberg method for all cell
832 types across diseases. (C) UMAP embeddings of cell types in the small intestinal with
833 normalized scDRS scores for the four diseases. Pink represents cells that are enriched
834 for the aforementioned disease, while gray implies non-relevant cells. (D) The causal
835 association between the listed cell types and four diseases. The dot size shows $-\log_{10}$
836 p -values from MR analysis using the IVW approach. (E) Colocalization analyses of
837 cell type QTL for central memory CD4⁺ T cell absolute count in four diseases. We
838 utilized colocalization to investigate whether the QTL for central memory CD4⁺ T
839 cell absolute count shared the same causative variant as four other diseases. Each dot
840 indicates a single nucleotide polymorphism, while colors indicate linkage

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

844

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

860

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

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

892

893 **Supplementary Figure**

894 **Figure S1. Local genetic correlations between hypothyroidism and complex** 895 **diseases revealed by ρ -HESS (Heritability Estimation from Summary Statistics).**

896 Local genetic correlations between hypothyroidism and chronic sinusitis (A), ILD
897 endpoints (B), and sarcoidosis (C), respectively. Significant local genetic correlation
898 estimates are highlighted in red and blue for even and odd chromosomes, respectively.

899

900 **Supplementary Table**

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

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

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

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

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

913 Supplementary Table S6. Shared risk regions in different analyses.

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

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

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

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

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

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

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

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

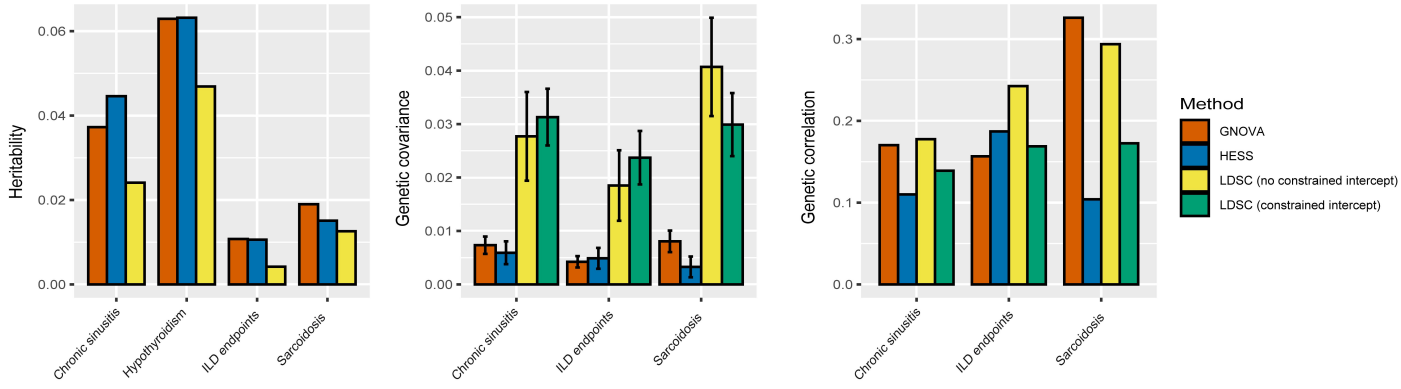
930 Supplementary Table S15. Significant causal effects of plasma proteins on disease.

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

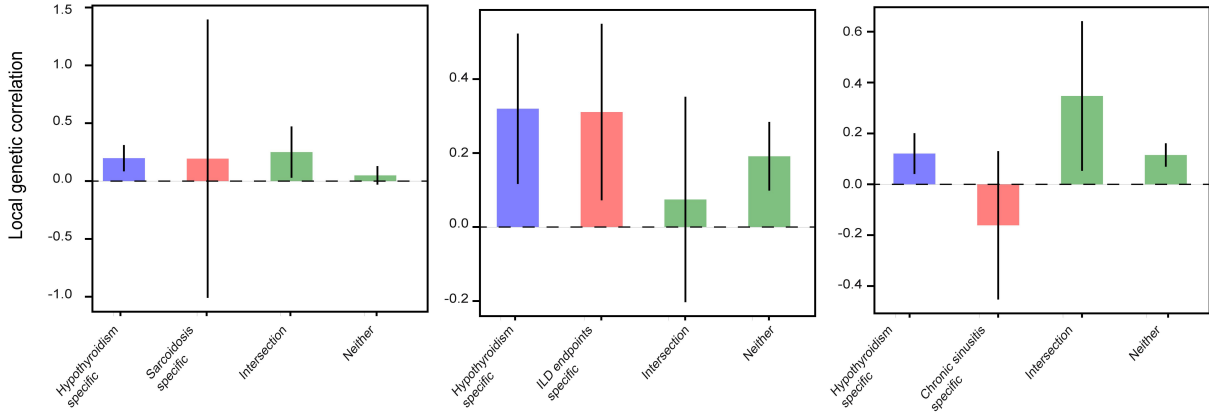
933 Supplementary Table S17. Causal effect of hypothyroidism on plasma proteins using
934 MR (unfiltered).

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

A



B



C

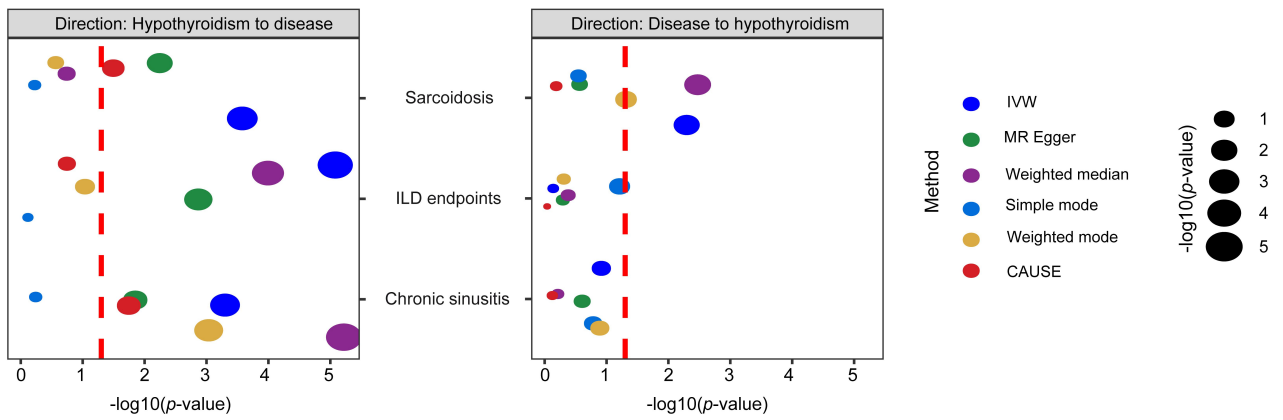


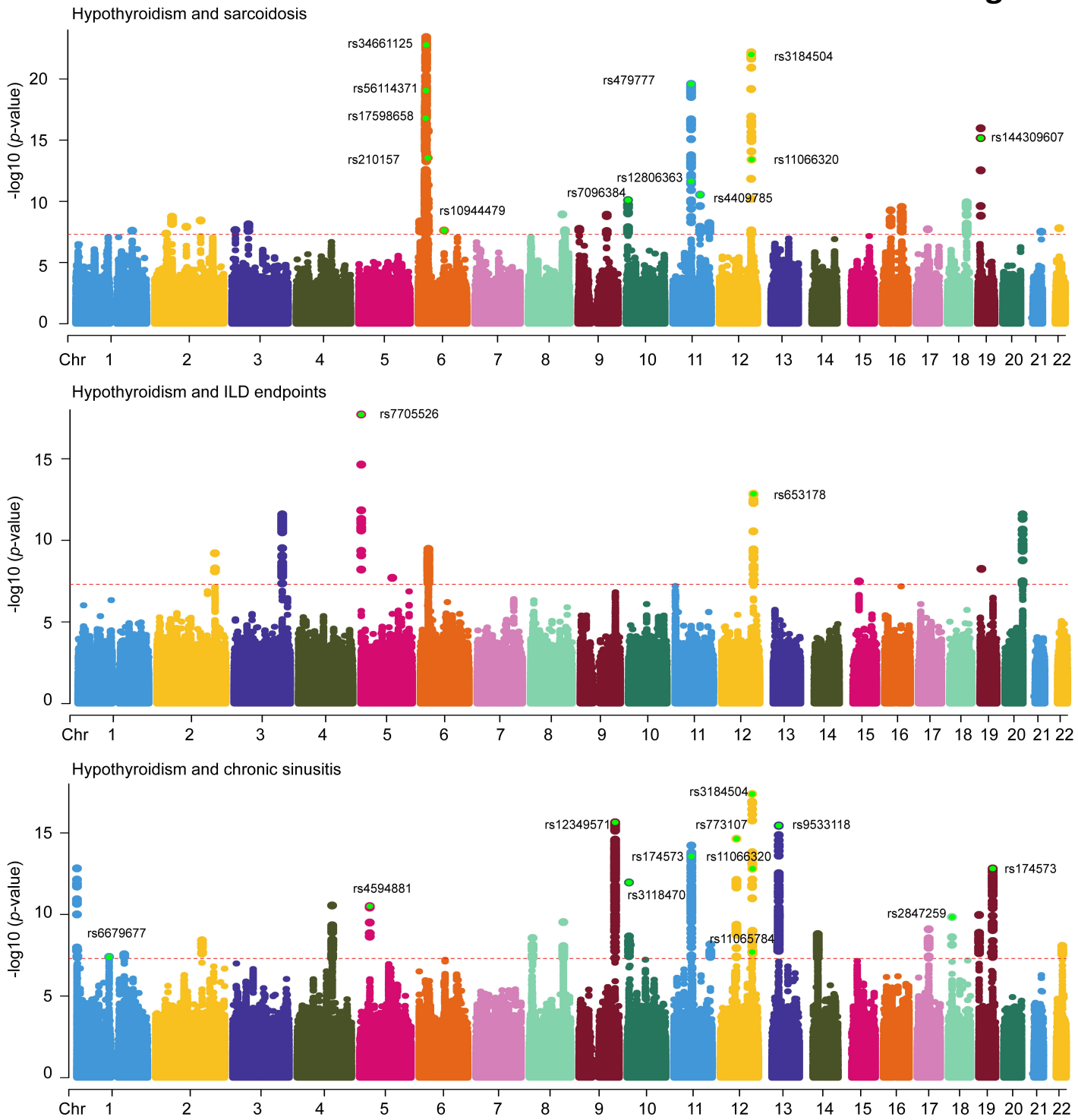
Figure 2

Figure 5

