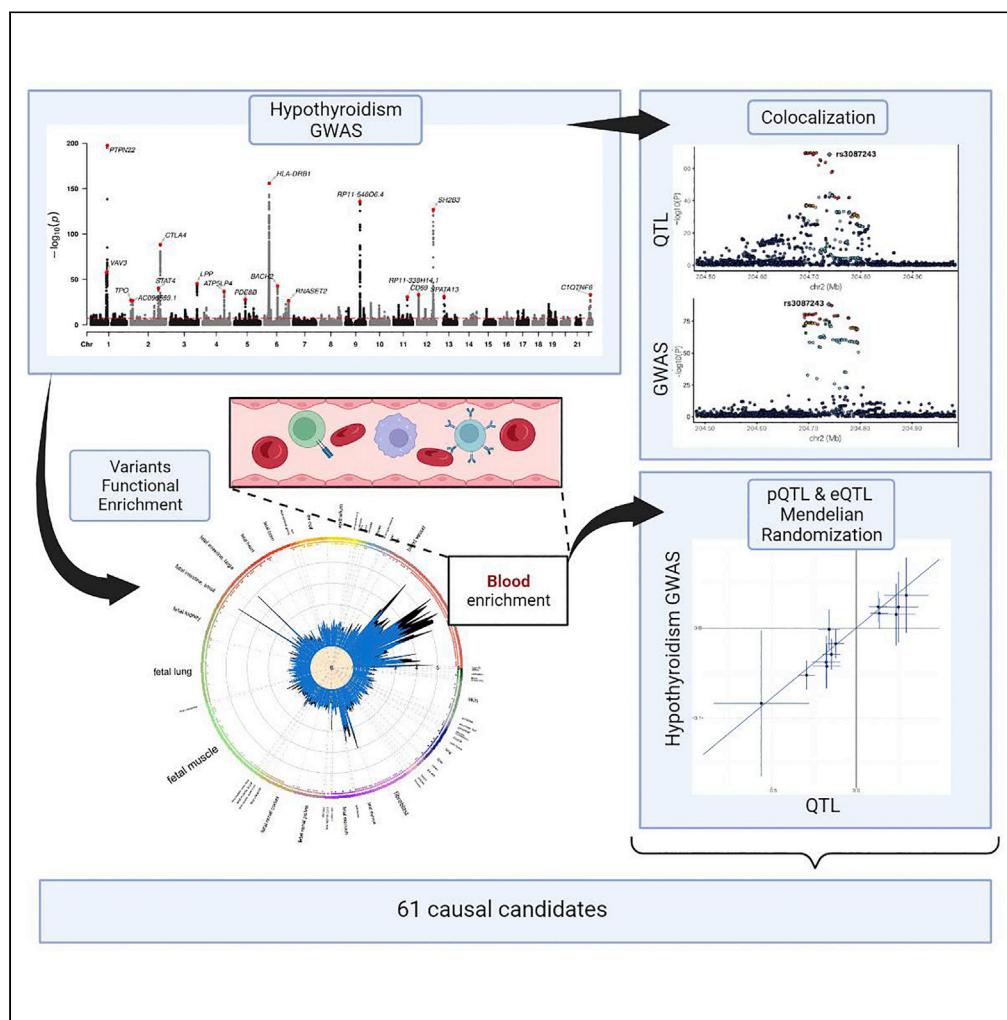


Article

# Genetic association and Mendelian randomization for hypothyroidism highlight immune molecular mechanisms



Samuel Mathieu,  
Mewen Briend,  
Erik Abner, ...,  
Tõnu Esko, Benoit  
J. Arsenault,  
Patrick Mathieu

patrick.mathieu@fmed.ulaval.ca

**Highlights**  
GWAS for hypothyroidism identified 139 risk loci including 76 novel associations

GWAS was enriched in pathways related to lymphocyte function

In total, 28 potentially deleterious missense variants were identified

Mendelian randomization and colocalization identified 61 blood causal candidate genes



## Article

## Genetic association and Mendelian randomization for hypothyroidism highlight immune molecular mechanisms

Samuel Mathieu,<sup>1,3</sup> Mewen Briend,<sup>1,3</sup> Erik Abner,<sup>2</sup> Christian Couture,<sup>1,3</sup> Zhonglin Li,<sup>3</sup> Yohan Bossé,<sup>3,4</sup> Sébastien Thériault,<sup>3,5</sup> Tõnu Esko,<sup>2</sup> Benoit J. Arsenault,<sup>1,3,6</sup> and Patrick Mathieu<sup>1,3,7,8,\*</sup>

## SUMMARY

**We carried out a genome-wide association analysis including 51,194 cases of hypothyroidism and 443,383 controls. In total, 139 risk loci were associated to hypothyroidism with genes involved in lymphocyte function. Candidate genes associated with hypothyroidism were identified by using molecular quantitative trait loci, co-localization, and enhancer-promoter chromatin looping. Mendelian randomization (MR) identified 42 blood expressed genes and circulating proteins as candidate causal molecules in hypothyroidism. Drug-gene interaction analysis provided evidence that immune checkpoint and tyrosine kinase inhibitors used in cancer therapy increase the risk of hypothyroidism. Hence, integrative mapping and MR support that expression of genes and proteins enriched in lymphocyte function are associated with the risk of hypothyroidism and provide genetic evidence for drug-induced hypothyroidism and identify actionable potential drug targets.**

## INTRODUCTION

Thyroid dysfunction including autoimmune thyroid disorder (AITD) is a frequent condition with a prevalence of 5% in the general population (Taylor et al., 2018). A majority of patients with AITD disorder present hypothyroidism, with Hashimoto disease being the most common cause (Ragusa et al., 2019). Hashimoto disease is an autoimmune condition characterized by the production of antithyroid antibodies (e.g. antithyroid peroxidase). Graves' disease is another autoimmune disorder which activates the secretion of thyroid hormone (Davies et al., 2020). However, some patients with Graves' disorder may develop hypothyroidism spontaneously or if treated with radioactive iodine or surgery (Umar et al., 2010). Other less frequent conditions leading to hypothyroidism include congenital causes, pregnancy, drugs, and iodine excess (Vanderpump, 2011). People affected by hypothyroidism need lifelong thyroid hormone replacement therapy (Biondi and Cooper, 2019).

Majority of patients are diagnosed with established hypothyroidism, whereas a significant proportion of the population is affected by subclinical hypothyroidism (Garg and Vanderpump, 2013). The pathobiology and molecular processes leading to hypothyroidism are still largely unknown. The identification of genes and molecules involved in the development of hypothyroidism could help identify at-risk individuals and develop therapy, which could be initiated before irreversible damage to the thyroid gland. Recent large-scale genome-wide association study (GWAS) for AITD including 30,234 cases of hypothyroidism and hyperthyroidism from the UK Biobank and Iceland reported 93 risk loci (Saevarsdottir et al., 2020). The latter study reported a rare mutation in *FLT3* affecting the splicing and leading to a compensatory process with increased ligand (FLT3LG) activity. Herein, we conducted a meta-analysis for hypothyroidism by using a broad definition based on the clinical and treatment status (clinical diagnosis of hypothyroidism or treatment with levothyroxine) and including 51,194 cases from UK Biobank and FinnGen. We identified 139 risk loci including 76 novel associations. Among the novel 76 risk loci, 44 were replicated in the Estonian Biobank. Mapping and Mendelian randomization (MR) identified several blood expressed genes and plasma proteins causally associated with the risk of hypothyroidism and illustrate the complex and polygenic nature of this common disorder.

## RESULTS

## Meta-analysis

In UKB, we conducted a GWAS for hypothyroidism (individuals with hypothyroidism or treatment with levothyroxine) (STAR Methods) including 25,130 cases (Table S1) and 383,471 controls (Table S2) of white

<sup>1</sup>Genomic Medicine and Molecular Epidemiology Laboratory, Quebec Heart and Lung Institute, Laval University, Québec, QC, Canada

<sup>2</sup>Institute of Genomics, University of Tartu, Tartu, Estonia

<sup>3</sup>Quebec Heart and Lung Institute, Laval University, Québec, QC, Canada

<sup>4</sup>Department of Molecular Medicine, Laval University, Québec, QC, Canada

<sup>5</sup>Department of Molecular Biology, Medical Biochemistry and Pathology, Laval University, Québec, QC, Canada

<sup>6</sup>Department of Medicine, Laval University, Québec, QC, Canada

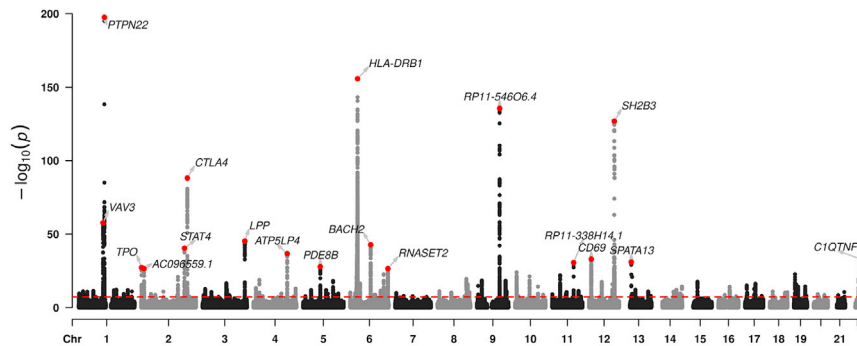
<sup>7</sup>Department of Surgery, Laval University, Québec, QC, Canada

<sup>8</sup>Lead contact

\*Correspondence: [patrick.mathieu@fmed.ulaval.ca](mailto:patrick.mathieu@fmed.ulaval.ca)

<https://doi.org/10.1016/j.isci.2022.104992>





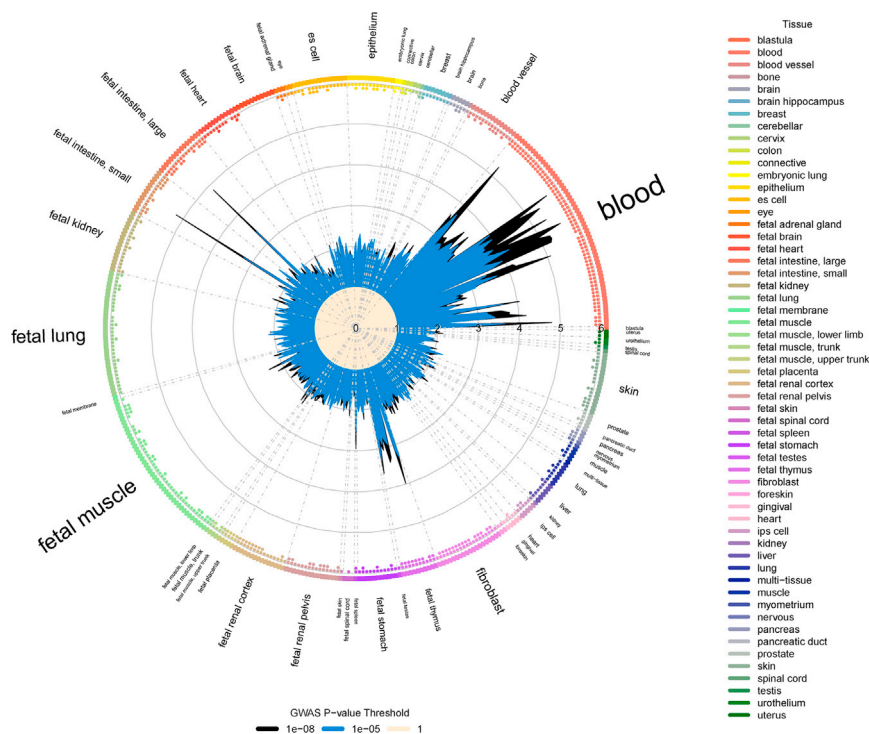
**Figure 1. Manhattan plot**

Genetic association from the meta-analysis for hypothyroidism.

British ancestry, which was meta-analyzed (fixed effect meta-analysis) with 26,064 cases and 59,912 controls from FinnGen (data freeze 5). In total, the meta-analysis included 494,577 individuals (51,194 cases and 443,383 controls) and 10,836,150 single-nucleotide polymorphisms (SNPs) were analyzed (MAF  $\geq 0.001$  and imputation info score  $>0.3$ ). [Figure S1](#) shows the QQ plot for the meta-analysis. The inflation score (lambda GC) was 1.27 and the intercept of the linkage disequilibrium (LD) score was 1.09, indicating that a majority of inflation was related to polygenicity. The heritability on the liability scale was estimated at 10.8%. In total, 22,454 variants were significantly ( $P_{\text{GWAS}} < 5.0\text{E-}08$ ) associated with hypothyroidism. [Figure 1](#) shows the Manhattan plot of the meta-analysis in which we identified 139 risk loci ( $P_{\text{GWAS}} < 5.0\text{E-}08$ ), including 76 new associations ([Data S1](#)). We implemented the Probabilistic Identification of Causal SNPs (PICS) to obtain a 95% credible set of gene variants at each risk loci ([Data S1](#)). In 45 risk loci, the index SNP was prioritized with a high confidence as the unique causal variant ([Data S1](#)). We found 63 genomic regions previously identified in AITD ([Data S1](#)). Among the previously identified loci, the strongest signal is rs2476601 ( $P_{\text{GWAS}} = 3.17\text{E-}198$ ), a missense variant in *PTPN22* and associated with several autoimmune disorders ([Stanford and Bottini, 2014](#)). We also identified the 13q12-*FLT3* (rs76428106) association with AITD recently described in 30,324 cases from UK biobank and Iceland. Among the new associations, in the MHC, rs9271365 is an intergenic variant annotated to *HLA-DRB1* and previously associated with autoimmune disorders (rheumatoid arthritis, inflammatory bowel disease) ([Liu et al., 2015](#); [Okada et al., 2014](#)) as reported in PhenoScanner. Excluding the MHC, we identified novel associations for hypothyroidism of variants previously associated with different autoimmune disorders. For instance, 10p15-*IL2RA* (rs7096384) ([Ji et al., 2017](#)), 5q31-*TCF7* (rs244686) ([Okada et al., 2014](#)), 11q13-*CCDC88B* (rs479777) ([Fischer et al., 2012](#)), 21q22-*UBASH3A* (rs12482947) ([Okada et al., 2014](#)), and 8q24-*TNFRSF11B* (rs12679857) ([Bradfield et al., 2011](#)) are known autoimmune disease risk loci, which are new associations with hypothyroidism. We also report novel associations for hypothyroidism including 1q41-*TGFB2* (rs1797071), 1q23-*NTRK1* (rs2148907), 1p13-*WDR47* (rs17565949), 2q11-*TMEM131* (rs5865), 3p13-*FOXP1* (rs11712352), 6p21-*FGD2* (rs1724087), 8q24-*AGO2* (rs11783023), 12p13-*CD69* (rs7488011), and 18q23-*NFATC1* (rs8093850), which have not been previously associated with other autoimmune disorders. In a replication phase including 17,002 cases and 178,141 controls from the Estonian Biobank, among the 76 new risk loci, we replicated 44 loci at a nominal p value ( $p < 0.05$ ), whereas 18 were replicated after Bonferroni correction ( $p < 6.58\text{E-}04$ ,  $0.05/76$ ) ([Data S2](#)).

### Annotation of coding and splice variants

We mapped variants by using Functional Mapping and Annotation of Genome-Wide Association Studies (FUMA), which identifies lead and individual significant SNPs ( $P_{\text{GWAS}} < 5\text{E-}8$  and  $r^2 < 0.6$ ) at risk loci ([STAR Methods](#)). By using FUMA, we identified 52 missense gene variants associated with hypothyroidism ([Data S3](#)). Missense variants were located in 41 different risk loci ([Data S3](#)). Among the missense variants, 27 had a Combined Annotation Dependent Depletion (CADD) score higher than 10, which represents the top 10% most deleterious variants in the human genome. The highest CADD score is rs601338 (CADD score 52), which is the lead SNP and a stop-gain mutation in *FUT2*, which encodes for a fucosyltransferase expressed in the digestive tract. Rs601338 has been previously associated with infectious and autoimmune disorders such as inflammatory bowel disease and type 1 diabetes ([Galeev et al., 2021](#)).



**Figure 2. Functional enrichment according to GARFIELD**

The radar plot shows the functional enrichment of variants associated with hypothyroidism. Radial axis shows the odds ratio for each cell type annotation. Enrichments were calculated at GWAS threshold of  $1.0E-05$  (blue) and  $1.0E-08$  (black). The dots at the edge of the plot are colored according to the tissue legend and represent significant enrichment (one dot corresponds to  $p < 1.0E-05$ ; two dots correspond to  $p < 1.0E-06$ ).

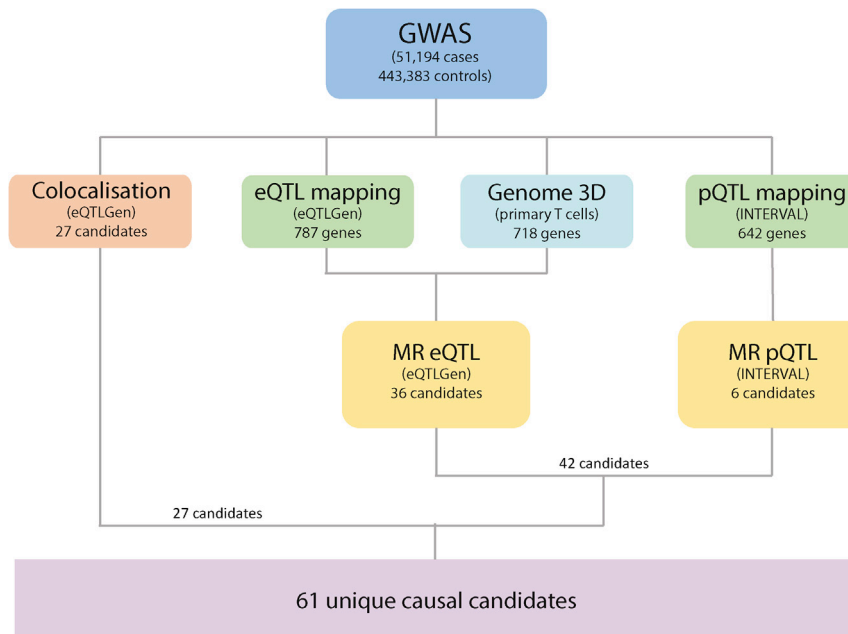
Genes previously reported and in which potentially deleterious variants were prioritized in the PICS credible set include *TYK2* (rs34536443, p.Pro1104Ala), *ADCY7* (rs78534766, p.Asp439Glu), and *PTPN22* (rs2476601, p.Trp620Arg). Among the novel associations in LD with the lead SNP ( $r^2 > 0.7$ ), *SESN1* (rs2273668, p.Leu103Ile) and *PSMB7* (rs4574, p.Val39Ala) are missense gene variants with a CADD score higher than 20. At 5p13, alternative splicing variant rs6897932 in *IL7R* ( $r^2 = 0.8$  with lead variant) has been previously associated with multiple sclerosis, primary biliary cirrhosis, and atopic disorders (Ferreira et al., 2017; International Multiple Sclerosis Genetics Consortium et al., 2011; Ji et al., 2017). The risk allele C-rs6897932 modifies the splicing process in excluding the exon 6 and results in a secreted form of IL7R (sIL7R) (Lundtoft et al., 2020).

### Cell and tissue enrichment

We have implemented GARFIELD to document the enrichment of hypothyroidism-associated gene variants in tissues and cells (lotchkova et al., 2019). GARFIELD provides a LD-corrected tissue-cell enrichment by using data from the ENCODE and Roadmap Epigenetics Project. Figure 2 shows a radar plot with a strong functional enrichment in the blood. The most significant enrichments were observed in chromatin accessibility of CD4, CD8, and CD19 primary lymphocytes. Data from GARFIELD analysis are summarized in Data S4. To buttress these findings, we implemented the stratified LD score framework including 489 tissue and cell annotations (STAR Methods) (Finucane et al., 2015). This analysis was consistent with GARFIELD and identified as most significant signal H3K27ac and H3K4me1 histone marks in primary T cells (Data S5). H3K27ac and H3K4me1 are histone marks of active chromatin and enhancer regions.

### Gene-based test and pathway enrichment

GWAS data were analyzed with MAGMA, which provides gene-based test associations (de Leeuw et al., 2015). In total, this analysis identified 430 genes that were associated with hypothyroidism ( $p < 2.60E-06$ , Bonferroni threshold for the number of genes,  $0.05/19,215$ ) (Data S6). We identified several known (*CTLA4*, *PTPN22*, *TLR3*, *CIQTNF6*, *TG*) and new associations including genes harboring deleterious



**Figure 3. Study plan**

Illustration represents the analysis pipeline for gene mapping and causal inference.

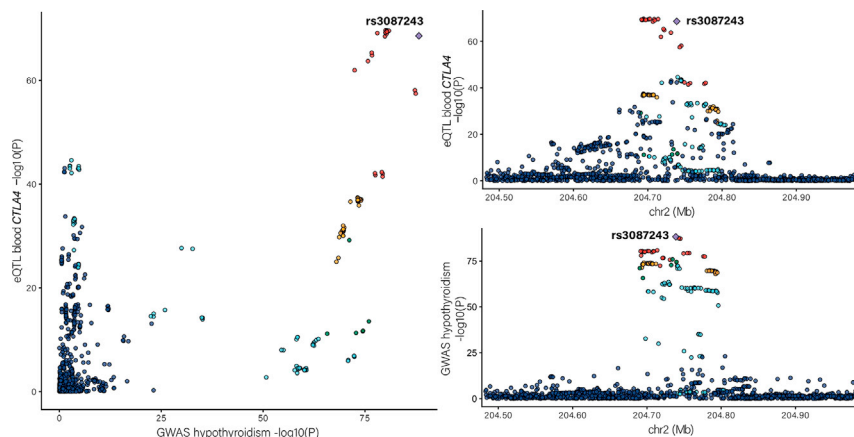
missense variants (*MYH15*, *CEP128*, *SESN1*, *AP4B1*, *PSMB7*, and *DCLRE1B*). A pathway analysis of genes mapped by MAGMA using BioCarta showed the highest enrichment in functions pertaining to T cells and antigen presentation such as costimulatory signal during T cell activation, antigen processing, interleukin 2 (IL-2) signaling, and IL-7 signal transduction (Data S7).

### Expression quantitative trait loci mapping

Figure 3 shows the post-GWAS analysis plan to assess gene mapping and to infer causality. Considering the highest enrichment of the GWAS in blood cells in both GARFIELD and the stratified LD score analyses, we mapped variants to blood expression quantitative trait loci (eQTL) in eQTLGen, a resource including 31,684 subjects (Vösa et al., 2021). Lead and individual significant SNPs ( $P_{\text{GWAS}} < 5E-08$  and  $r^2 < 0.6$ ) associated with hypothyroidism were mapped to eQTLGen. In total, 326,334 significant SNP-eQTL gene pairs (false discovery rate [FDR]  $< 0.05$ ) tagging 787 genes were identified (Data S8). Consistent with previous analyses, blood eQTL mapped genes were enriched in pathway for antigen processing and the costimulatory signal during T cell activation (Data S9). We next analyzed the shared genetic signal between the GWAS and blood eQTL signal by using Bayesian colocalization analysis. Colocalization analysis identified 27 genes with strong evidence of shared signal ( $PP > 0.8$ ) (*UBE2B*, *CAMK4*, *EDARADD*, *ACAP1*, *BEND3*, *PKN3*, *FYN*, *PSMD14*, *NTN5*, *FOXK1*, *TANK*, *AHI1*, *CD247*, *TMEM258*, *ZKSCAN8*, *ZFP36L2*, *JUND*, *DDX59*, *FADS2*, *CA8*, *INPP5B*, *ACO2*, *CD226*, *ZSCAN16*, *PHF5A*, *TOB2*, *CTLA4*) (Table S3). Figure 4 shows a LocusCompare plot at 2q33 where rs3087243 is the prioritized SNP ( $PP = 0.81$ ) between the GWAS and the blood eQTL for the expression of *CTLA4*, a gene encoding an immune checkpoint molecule (Darvin et al., 2018).

### Protein QTL mapping

We next addressed whether the GWAS for hypothyroidism was associated with blood protein QTL (pQTL) by using data from the INTERVAL study, which includes genetic association data for 2,965 different proteins (Sun et al., 2018). We identified 1,240 significant SNP-pQTL protein pairs (FDR  $< 0.05$ ), which tagged 642 blood proteins (Data S10). Among the significant blood pQTL proteins, 42 were also mapped as blood eQTL (*PCSK7*, *C4A*, *NCR3*, *TNXB*, *MICB*, *AGER*, *ICOS*, *TNFSF12*, *HAVCR2*, *STAT3*, *MAPK14*, *HSD17B14*, *COMMD1*, *CCDC134*, *PAM*, *COL11A2*, *TAPBP*, *ACACB*, *IL15RA*, *ICAM5*, *IL7R*, *B3GNT2*, *SAT2*, *STAT1*, *KREMEN1*, *RPSA*, *LY6G6C*, *DDX39B*, *GCA*, *NHP2L1*, *PTPN11*, *STAT6*, *EIF5A*, *CD226*, *PPP1R3B*, *LST1*, *IL27*, *VAV3*, *RABEPK*, *M6PR*, *FLT3*, *GDF11*).



**Figure 4. LocusCompare plot**

Left hand panel represents GWAS (x-axis) and blood eQTL for *CTLA4* (y-axis)  $-\log p$  values for the gene variants at the *CTLA4* locus. Right hand panels represent association data with the  $-\log p$  values on the y-axis for the eQTL (*CTLA4*) (upper panel) and GWAS (lower panel) according to genomic coordinates (x-axis) spanning 500 kb centered on *CTLA4*.

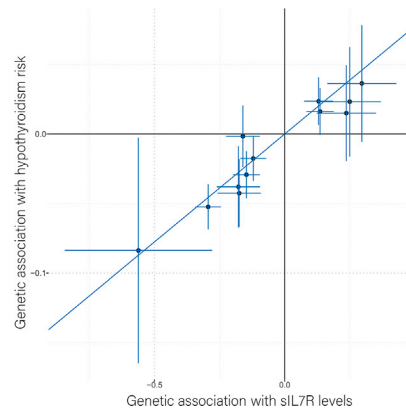
### 3D genome mapping

Our analyses showed a strong enrichment of gene variants in H3K27ac chromatin mark of primary T cells (Data S5). As gene regulation involves chromatin folding between active distant regulatory elements and gene promoters, we mapped GWAS data to H3K27ac-HiChIP carried out in primary T cells (GSE101498). H3K27ac-HiChIP, which provides a high-resolution map of enhancer promoter interactions (Mumbach et al., 2017), was analyzed by using FitHiChIP at a stringent FDR value ( $FDR < 1E-06$ ) for the identification of significant loops (STAR Methods). In total, 118,597 enhancer-promoter interactions were identified in primary T cells. This analysis showed that 718 gene promoters were connected to 174 individual significant SNPs located in distant-acting enhancers (Data S11). Among the genes mapped by enhancer-promoter interactions, 133 genes were also identified by the blood eQTL mapping (Table S4). For instance, genes such as *VAV3*, *PTPN22*, *STAT4*, *STAT3*, *CD28*, *ICOS*, *FYN*, *IL2RA*, *CD69*, *BACH2*, *CEP128*, *TYK2*, and *IRF3* were identified by blood eQTL and 3D genome mapping.

### Mendelian randomization

To identify causal blood genes and proteins associated with hypothyroidism, we performed two-sample MR. MR was performed by using at least 3 *cis*-instrumental variables ( $P < 1E-03$ ) selected within a window of  $\pm 250$  kb around the transcription start site of the candidate gene (STAR Methods). Genes identified by eQTL and 3D genome mapping were analyzed by using inverse variance weighted (IVW) MR. In eQTLGen, enough instrumental variables ( $\geq 3$ ) were available to perform 924 MR analyses. In total, 36 blood expressed genes were significant in IVW MR ( $P_{\text{Bonferroni}} < 5.41E-05$ ,  $0.05/924$ ) and without heterogeneity on the Cochran's Q test ( $P_{\text{heterogeneity}} \geq 0.05$ ) (Data S12). The most significant and with the largest effect size blood gene is *BACH2* (odds ratio [OR]: 0.57, 95% confidence interval [CI]: 0.53-0.61,  $P_{\text{IVW}} = 3.55E-55$ ), a gene also mapped by enhancer-promoter looping and encoding for a transcription factor involved in lymphocyte function. Among the causally associated genes identified in MR, 8 had shared genetic signal between the GWAS and blood eQTL in colocalization analysis (*INPP5B*, *AHI1*, *TMEM258*, *FOXK1*, *FYN*, *TANK*, *CD226*, *ZFP36L2*). As a sensitivity analysis, we performed a weighted median MR, which is less sensitive to horizontal pleiotropy (Bowden et al., 2016). This analysis showed that the causal candidate genes identified in IVW MR were also significant in weighted median MR (Data S12). Directional effects were consistent between IVW and weighted median MR analyses.

We also implemented MR analysis for the blood proteins mapped in INTERVAL (significant pQTL proteins) by using the same algorithm. In the INTERVAL cohort, there were enough instruments ( $\geq 3$ ) to perform 218 MR analyses. We identified 6 blood proteins significantly associated ( $P_{\text{Bonferroni}} < 2.29E-04$ ,  $0.05/218$ ) with hypothyroidism and without heterogeneity on the Cochran's Q test ( $P_{\text{heterogeneity}} \geq 0.05$ ) (*IL7R*, *MXRA8*, *PCSK7*, *DCBLD2*, *PAM*, *CREB3L4*) (Data S13). Sensitivity analysis with the weighted median MR was significant for 5 of the proteins (*IL7R*, *MXRA8*, *PCSK7*, *DCBLD2*, *CREB3L4*) with consistent directional effects



**Figure 5. Mendelian randomization**

Plot showing the IVW MR for blood protein sIL7R (exposure) and the GWAS for hypothyroidism (outcome).

(Data S13). The most significant blood protein is sIL7R (OR: 1.17, 95% CI: 1.13-1.21,  $P_{IVW} = 1.52E-20$ ) (Figure 5).

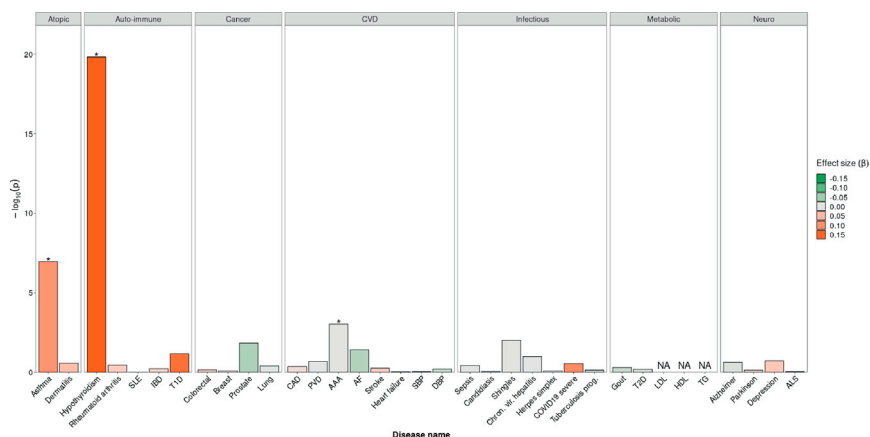
### Cross-trait analyses

To identify related traits and disorders with hypothyroidism, we performed a genetic correlation analysis by using the LD score and traits included in LD Hub (Zheng et al., 2017). After multiple test correction, 16 traits-disorders remained significantly correlated with hypothyroidism (Table S5). Hypothyroidism was positively correlated with rheumatoid arthritis ( $rg = 0.41$ ), celiac disease ( $rg = 0.37$ ), systemic lupus erythematosus ( $rg = 0.28$ ), depressive symptoms ( $rg = 0.24$ ), coronary artery disease ( $rg = 0.22$ ), primary biliary cirrhosis ( $rg = 0.22$ ), and body mass index ( $rg = 0.19$ ). It was negatively correlated with years of schooling ( $rg = -0.19$ ), high-density lipoprotein cholesterol ( $rg = -0.18$ ), age at first birth ( $rg = -0.18$ ), subjective well-being ( $rg = -0.16$ ), and age at menarche ( $rg = -0.12$ ).

As another strategy to examine traits genetically related to hypothyroidism, we analyzed GWAS data by using the interactive Cross-Phenotype Analysis of GWAS database (iCPAGdb), which provides cross-phenotype enrichment analyses from an exhaustive list of ancestry LD-specific precomputed association data from the NHGRI-EBI GWAS catalog (Wang et al., 2021). After multiple test correction (Bonferroni), this analysis identified 257 traits-disorders shared with hypothyroidism (Data S14). Traits and disorders with highest enrichments included thyroid preparation use, autoimmune disease, asthma, rheumatoid arthritis, vitiligo, type 1 diabetes, multiple sclerosis, and systemic lupus erythematosus. These data are consistent with the genetic correlations and showed shared genetic signal with several immune-related disorders.

### Drug target analysis

Genes and proteins that were associated in MR with hypothyroidism or with a shared genetic signal with the blood expression ( $PP > 0.8$ ) were considered as candidate causal genes. In total, 61 unique blood genes and proteins were deemed causally associated with hypothyroidism and were evaluated for their druggability in the Drug Gene Interaction Database (DGIdb), which provides an exhaustive list of drug-gene interactions collated from different resources (Cotto et al., 2018). In DGIdb, 82 drug-gene pairs targeting 11 genes/proteins were identified (Data S15). Results showed that several compounds were tyrosine kinase inhibitors for different indications in oncology. For instance, sunitinib inhibits PTPN12, which is negatively associated with the risk of hypothyroidism in MR (OR: 0.93, 95% CI: 0.91-0.95,  $P_{IVW} = 7.24E-10$ ). These data are consistent with report indicating an increased risk of hypothyroidism in patients treated with sunitinib (Clemons et al., 2012). CTLA4 is targeted by an FDA-approved monoclonal antibody, zalfirelimab, which is used in cancer therapy. Colocalization analysis prioritized rs3087243 at the CTLA4 locus ( $PP = 0.81$ ). In eQTLGen, A-rs3087243 is associated with an increased expression of CTLA4 ( $P_{eQTL} = 2.67E-69$ ), a gene encoding for a checkpoint molecule, and reduces the risk of hypothyroidism (OR: 0.85,  $P_{GWAS} = 5.97E-89$ ). These data are in line with the association between the treatment with checkpoint inhibitors and the development of thyroid disorder (Ferrari et al., 2018). On the other hand, novel monoclonal antibody targeting IL7R (Ellis et al., 2019), which is under development for indication in different autoimmune diseases, could reduce the risk of hypothyroidism. To this effect, in MR, we found a strong positive association



**Figure 6. Multitrait Mendelian randomization**

Bar graph showing the  $-\log_{10}(p)$  value for the MR analyses (inverse variance weighted MR) between blood protein sIL7R (exposure) and different traits and disorders (outcomes) on the x-axis. Color in bar graph represents the direction and effect size of MR as illustrated in the right-hand legend panel. \*Disease significant after Bonferroni correction. NA indicates that MR was not performed owing to the absence of at least three instruments shared between the exposure and the outcome. The upper panel represents disease categories: atopic, autoimmune, CVD (cardiovascular diseases), infectious, metabolic, and neuro (neurologic). SLE: systemic lupus erythematosus; IBD: inflammatory bowel disease; T1D: type 1 diabetes; CAD: coronary artery disease; PVD: peripheral vascular disease; AAA: abdominal aortic aneurysm, AF: atrial fibrillation; SBP: systolic blood pressure; DBP: diastolic blood pressure; T2D: type 2 diabetes; LDL: low-density lipoprotein; HDL: high-density lipoprotein; TG: triglyceride; ALS: amyotrophic lateral sclerosis.

between blood sIL7R and the risk of hypothyroidism (OR: 1.17, 95% CI: 1.13-1.21,  $P_{IVW} = 1.52E-20$ ). In order to further evaluate sIL7R as a potential therapeutic target, we leveraged 35 different diseases-traits GWAS summary statistics data sets encompassing 7 categories of disorders (atopic, autoimmune, cancer, cardiovascular, infectious, metabolic, and neurologic) and we performed a multitrait MR analysis using *cis*-instruments for blood sIL7R. After multiple test correction (Bonferroni,  $p < 1.43E-3$ ,  $0.05/35$ ), we found that in addition to hypothyroidism, sIL7R was positively associated with the risk of asthma and abdominal aortic aneurysm (Figure 6). These data suggest that inhibition of sIL7R would consistently lead to a reduction of the risk for these disorders.

### Tissue specificity score

We evaluated causally associated genes and proteins identified in MR or colocalization analyses for their tissue specificity score. Drugs that target tissue-specific genes are less likely associated with adverse side effects and are more likely to be licensed (Duffy et al., 2020). Tissue specificity score was evaluated by using the Jensen-Shannon specificity metric (Cabili et al., 2011). Tissue specificity score for each gene has been assessed for 61 tissues based on the expression from GTEx, The Human Protein Atlas and FANTOM5 (Uhlén et al., 2015). Hierarchical clustering of the tissue specificity score showed that several causally associated genes to hypothyroidism were specific to lymphoid organs and T cells (Figure S2). Pair-wise correlation for the specificity score in 61 tissues based on the expression of causally associated genes showed clustering of immune cells and lymphoid organs (Figure S3). The expression of *CAMK4* and *CD247* was highly specific to the thymus. Others such as *TMEM156*, *SKAP1*, *ACAP1*, *CTLA4*, and *IL7R* were specific to B cell, T cell, and different lymphoid organs (lymph node, tonsil, appendix, spleen) (Figure S2).

### DISCUSSION

In this work, a GWAS performed on 51,194 cases with hypothyroidism identified 139 risk loci including 28 potentially deleterious missense variants. The GWAS was enriched in functional annotation in lymphocytes. Genes mapped by molecular QTL and promoter-enhancer interactions were enriched in causal genes in MR. A comprehensive analysis for the druggable genes provided genetic evidence for drug-induced hypothyroidism and identified potentially druggable therapeutic targets.

This large-scale analysis identified 76 novel associations with hypothyroidism. Some of these novel associations are predicted deleterious missense variants of protein coding genes including for instance *SESN1*



(rs2273668, p.Leu103Ile) ( $r^2$  with lead variant = 0.95) and *PSMB7* (rs4574, p.Val39Ala) ( $r^2$  with lead variant = 0.74). *SESN1* encodes for a protein of the sestrins family, which are involved in the protection against reactive oxygen species and genotoxic stress (Wang et al., 2019). *PSMB7* encodes for a component of the proteasome, which is involved in the processing of the MHC class I (Sijts and Kloetzel, 2011).

By using two different approaches, GARFIELD and the stratified LD score, this study highlighted a strong enrichment of the GWAS for hypothyroidism in functional annotations of lymphocytes. These data are supported by the identification of missense variants located in genes encoding critical regulators of lymphocyte function (e.g. rs2476601-*PTPN22*, rs34536443-*TYK2*) (Oyamada et al., 2009). Also, MR identified several blood expressed genes involved in lymphocyte function including *BACH2* (OR: 0.57, 95% CI: 0.53-0.61,  $P_{IVW} = 3.55E-55$ ), a transcription factor with a role in B cell survival and proliferation (Miura et al., 2018). Also, blood expressed genes such as *TANK* (OR: 0.89, 95% CI: 0.86-0.92,  $P_{IVW} = 1.17E-11$ ) and *CD226* (OR: 0.91, 95% CI: 0.88-0.94,  $P_{IVW} = 1.98E-10$ ) were identified by both MR and colocalization analyses. *TANK* encodes for a negative regulator of the nuclear factor of kappa B pathway, whereas *CD226* is involved in the adhesion of cytotoxic T cell (Tahara-Hanaoka et al., 2004; Wang et al., 2015). In the blood plasma, several pQTLs were associated with hypothyroidism. MR identified 6 blood proteins (*IL7R*, *MXRA8*, *PCSK7*, *DCBLD2*, *PAM*, *CREB3L4*) significantly associated with the risk of hypothyroidism. In the blood, MR analysis showed that the level of sIL7R was positively and causally associated with the risk of hypothyroidism (OR: 1.17, 95% CI: 1.13-1.21,  $P_{IVW} = 1.52E-20$ ). In line with the latter finding, C-rs6897932 ( $r^2$  with lead variant = 0.8), an alternative splicing variant increasing sIL7R (Lundtoft et al., 2020), was associated with the risk hypothyroidism (OR: 1.05, 95% CI: 1.03-1.07,  $P_{GWAS} = 1.62E-09$ ). Taken together, these findings militate for a role of sIL7R in hypothyroidism. A functional study has previously underlined that sIL7R potentiates the bioactivity of IL7, a cytokine known to alter self-tolerance mediated by T cells (Lundström et al., 2013).

In the present work, genetic correlation using the LD score and cross-phenotype enrichment analyses consistently revealed shared genetic architecture between hypothyroidism and several immune-related disorders such as rheumatoid disorder, celiac disease, systemic lupus erythematosus, and asthma. Also, hypothyroidism was genetically correlated with other conditions such as coronary artery disease, depressive symptoms, and the body mass index. These data highlight the pleiotropy of genes involved in the development of hypothyroidism and are in line with some clinical observations. For instance, several reports underlined an association between subclinical hypothyroidism, which may affect up to 10% of the population, and coronary artery disease (Razvi et al., 2010; Rodondi et al., 2010).

Our assessment of genes causally associated with hypothyroidism and their druggability by using data from DGIdb identified 82 drug-gene pairs targeting 11 genes/proteins. We identified that checkpoint and protein kinase inhibitors targeting *CTLA4* and *PTPN12*, respectively, increased the risk of hypothyroidism. These data provide genetic evidence and support clinical observations of drug-induced hypothyroidism for checkpoint and protein kinase inhibitors. Among the potential therapeutics, anti-IL7R antibodies, which are under development (Ellis et al., 2019), could be evaluated as a therapy. A MR scan for 35 different disorders (atopic, autoimmune, cancer, cardiovascular, infectious, metabolic, and neurologic) showed positive associations for blood plasma sIL7R with the risk of hypothyroidism, asthma, and abdominal aortic aneurysm. The directional effects were the same for the three disorders (hypothyroidism, asthma, and abdominal aortic aneurysm). These data suggest that anti-IL7R-based therapy could prevent hypothyroidism, asthma, and abdominal aortic aneurysm. However, follow-up studies are needed to explore the role of sIL7R in different disorders.

In conclusion, GWAS and comprehensive mapping provided evidence that hypothyroidism is highly polygenic and illustrate the complex and pleiotropic effect of genes involved in immune regulation. Mapping and MR identified genes and proteins in the blood affecting the risk of hypothyroidism. By identifying novel mutations, expressed genes, and blood proteins, this work provides a framework for experimental follow-up studies and drug target validation.

### Limitations of study

This work has some limitations. We identified several novel risk loci including a replication stage. However, these data were obtained from a large series of individuals from European ancestry. Future studies should aim to include different populations in a transancestry association analysis. Also, though several new risk loci and causal candidate genes were identified, follow-up studies are needed to underline molecular processes involved in hypothyroidism.

**STAR★METHODS**

Detailed methods are provided in the online version of this paper and include the following:

- **KEY RESOURCES TABLE**
- **RESOURCE AVAILABILITY**
  - Lead contact
  - Materials availability
  - Data and code availability
- **EXPERIMENTAL MODEL AND SUBJECT DETAILS**
  - Study participants
- **QUANTIFICATION AND STATISTICAL ANALYSIS**
  - GWAS for hypothyroidism
  - LD score and heritability
  - Annotation of the GWAS
  - Probabilistic Identification of Causal SNPs
  - Cell and tissue enrichment
  - Gene-based and pathway analyses
  - Molecular QTL mapping
  - Colocalization
  - 3D genome mapping
  - Mendelian randomization
  - Multi-trait mendelian randomization analysis
  - Cross-trait analyses
  - Drug target analyses
  - Tissue specificity score

**SUPPLEMENTAL INFORMATION**

Supplemental information can be found online at <https://doi.org/10.1016/j.isci.2022.104992>.

**ACKNOWLEDGMENTS**

Work of the authors is supported by the Quebec Heart and Lung Institute Fund. The work of Estonian Genome Center, University of Tartu was funded by the European Union through Horizon 2020 research and innovation program under grants no. 810645 and 894987, through the European Regional Development Fund projects GENTRANSMED (2014-2020.4.01.15-0012), MOBERA5 (Norface Network project no 462.16.107), MOBEC008, MOBERA21, and Estonian Research Council Grants PUT1660 and PRG1291. Data analyses with Estonian data sets were carried out in part in the High Performance Computing Center of University of Tartu. Y.B. holds a Canada Research Chair in Genomics of Heart and Lung Diseases. S.T. and B.J.A. hold a scholarship from Fonds de Recherche du Québec-Santé (FRQS). P.M. is the recipient of the Joseph C. Edwards Foundation granted to Université Laval. The graphical abstract was created with a licence from [BioRender.com](https://BioRender.com).

**AUTHOR CONTRIBUTIONS**

S.M., M.B., Y.B., S.T., B.J.A., and P.M. designed the study plan. C.C. and Z.L. performed GWAS and meta-analysis. S.M., M.B., and P.M. conducted mapping, colocalization, MR, enrichment, and cross-trait analyses. P.M. performed 3D genome mapping. S.M. and M.B. prepared the figures and tables. S.M., M.B., and P.M. drafted the manuscript. E.A. and T.E. provided the replication stage in the Estonian Biobank. All the authors critically reviewed the manuscript and provided intellectual inputs.

**DECLARATION OF INTERESTS**

The authors declare no competing interests.

Received: October 7, 2021

Revised: May 30, 2022

Accepted: August 17, 2022

Published: September 16, 2022

**REFERENCES**

- Bhattacharyya, S., Chandra, V., Vijayanand, P., and Ay, F. (2019). Identification of significant chromatin contacts from HiChIP data by FitHiChIP. *Nat. Commun.* 10, 4221. <https://doi.org/10.1038/s41467-019-11950-y>.
- Biondi, B., and Cooper, D.S. (2019). Thyroid hormone therapy for hypothyroidism. *Endocrine* 66, 18–26. <https://doi.org/10.1007/s12020-019-02023-7>.
- Bowden, J., Davey Smith, G., Haycock, P.C., and Burgess, S. (2016). Consistent estimation in mendelian randomization with some invalid instruments using a weighted median estimator. *Genet. Epidemiol.* 40, 304–314. <https://doi.org/10.1002/gepi.21965>.
- Bradfield, J.P., Qu, H.-Q., Wang, K., Zhang, H., Sleiman, P.M., Kim, C.E., Mentch, F.D., Qiu, H., Glessner, J.T., Thomas, K.A., et al. (2011). A genome-wide meta-analysis of six type 1 diabetes cohorts identifies multiple associated loci. *PLoS Genet.* 7, e1002293. <https://doi.org/10.1371/journal.pgen.1002293>.
- Browning, S.R., and Browning, B.L. (2007). Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am. J. Hum. Genet.* 81, 1084–1097. <https://doi.org/10.1086/521987>.
- Cabili, M.N., Trapnell, C., Goff, L., Koziol, M., Tazon-Vega, B., Regev, A., and Rinn, J.L. (2011). Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev.* 25, 1915–1927. <https://doi.org/10.1101/gad.17446611>.
- Clemons, J., Gao, D., Naam, M., Breaker, K., Garfield, D., and Flaig, T.W. (2012). Thyroid dysfunction in patients treated with sunitinib or sorafenib. *Clin. Genitourin. Cancer* 10, 225–231. <https://doi.org/10.1016/j.clgc.2012.08.002>.
- Cotto, K.C., Wagner, A.H., Feng, Y.-Y., Kiwala, S., Coffman, A.C., Spies, G., Wollam, A., Spies, N.C., Griffith, O.L., and Griffith, M. (2018). DGIdb 3.0: a redesign and expansion of the drug-gene interaction database. *Nucleic Acids Res.* 46, D1068–D1073. <https://doi.org/10.1093/nar/gkx1143>.
- Darvin, P., Toor, S.M., Sasidharan Nair, V., and Elkord, E. (2018). Immune checkpoint inhibitors: recent progress and potential biomarkers. *Exp. Mol. Med.* 50, 1–11. <https://doi.org/10.1038/s12276-018-0191-1>.
- Davies, T.F., Andersen, S., Latif, R., Nagayama, Y., Barbesino, G., Brito, M., Eckstein, A.K., Stagnaro-Green, A., and Kahaly, G.J. (2020). Graves' disease. *Nat. Rev. Dis. Prim.* 6, 52. <https://doi.org/10.1038/s41572-020-0184-y>.
- de Leeuw, C.A., Mooij, J.M., Heskes, T., and Posthuma, D. (2015). MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput. Biol.* 11, e1004219. <https://doi.org/10.1371/journal.pcbi.1004219>.
- Duffy, Á., Verbanck, M., Dobbyn, A., Won, H.-H., Rein, J.L., Forrest, I.S., Nadkarni, G., Rocheleau, G., and Do, R. (2020). Tissue-specific genetic features inform prediction of drug side effects in clinical trials. *Sci. Adv.* 6, eabb6242. <https://doi.org/10.1126/sciadv.abb6242>.
- Ellis, J., van Maurik, A., Fortunato, L., Gisbert, S., Chen, K., Schwartz, A., McHugh, S., Want, A., Santos Franco, S., Oliveira, J.-J., et al. (2019). Anti-IL-7 receptor  $\alpha$  monoclonal antibody (GSK2618960) in healthy subjects - a randomized, double-blind, placebo-controlled study. *Br. J. Clin. Pharmacol.* 85, 304–315. <https://doi.org/10.1111/bcp.13748>.
- Farh, K.K.-H., Marson, A., Zhu, J., Kleinewietfeld, M., Housley, W.J., Beik, S., Shores, N., Whitton, H., Ryan, R.J.H., Shishkin, A.A., et al. (2015). Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature* 518, 337–343. <https://doi.org/10.1038/nature13835>.
- Ferrari, S.M., Fallahi, P., Galetta, F., Citi, E., Benavenga, S., and Antonelli, A. (2018). Thyroid disorders induced by checkpoint inhibitors. *Rev. Endocr. Metab. Disord.* 19, 325–333. <https://doi.org/10.1007/s11154-018-9463-2>.
- Ferreira, M.A., Vonk, J.M., Baurecht, H., Marenholz, I., Tian, C., Hoffman, J.D., Helmer, Q., Tillander, A., Ullemar, V., van Dongen, J., Lu, Y., et al. (2017). Shared genetic origin of asthma, hay fever and eczema elucidates allergic disease biology. *Nat. Genet.* 49, 1752–1757. <https://doi.org/10.1038/ng.3985>.
- Finucane, H.K., Bulik-Sullivan, B., Gusev, A., Trynka, G., Reshef, Y., Loh, P.-R., Anttila, V., Xu, H., Zang, C., Farh, K., et al. (2015). Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* 47, 1228–1235. <https://doi.org/10.1038/ng.3404>.
- Fischer, A., Schmid, B., Ellinghaus, D., Nothnagel, M., Gaede, K.I., Schürmann, M., Lipinski, S., Rosentiel, P., Zissel, G., Höhne, K., et al. (2012). A novel sarcoidosis risk locus for Europeans on chromosome 11q13.1. *Am. J. Respir. Crit. Care Med.* 186, 877–885. <https://doi.org/10.1164/rccm.201204-0708OC>.
- Foley, C.N., Staley, J.R., Breen, P.G., Sun, B.B., Kirk, P.D.W., Burgess, S., and Howson, J.M.M. (2021). A fast and efficient colocalization algorithm for identifying shared genetic risk factors across multiple traits. *Nat. Commun.* 12, 764. <https://doi.org/10.1038/s41467-020-20885-8>.
- Galeev, A., Suwandi, A., Cepic, A., Basu, M., Baines, J.F., and Grassl, G.A. (2021). The role of the blood group-related glycosyltransferases FUT2 and B4GALNT2 in susceptibility to infectious disease. *Int. J. Med. Microbiol.* 311, 151487. <https://doi.org/10.1016/j.ijmm.2021.151487>.
- Garg, A., and Vanderpump, M.P.J. (2013). Subclinical thyroid disease. *Br. Med. Bull.* 107, 101–116. <https://doi.org/10.1093/bmb/ldt024>.
- International Multiple Sclerosis Genetics Consortium; Wellcome Trust Case Control Consortium 2; Sawcer, S., Hellenthal, G., Pirinen, M., Spencer, C.C.A., Patsopoulos, N.A., Moutsianas, L., Dilthey, A., Su, Z., Freeman, C., et al. (2011). Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. *Nature* 476, 214–219. <https://doi.org/10.1038/nature10251>.
- Iotchkova, V., Ritchie, G.R.S., Geijs, M., Morganello, S., Min, J.L., Walter, K., Timpson, N.J., UK10K Consortium, Dunham, I., Birney, E., and Soranzo, N. (2019). GARFIELD classifies disease-relevant genomic features through integration of functional annotations with association signals. *Nat. Genet.* 51, 343–353. <https://doi.org/10.1038/s41588-018-0322-6>.
- Ji, S.-G., Juran, B.D., Mucha, S., Folseraas, T., Jostins, L., Melum, E., Kumasaka, N., Atkinson, E.J., Schlicht, E.M., Liu, J.Z., et al. (2017). Genome-wide association study of primary sclerosing cholangitis identifies new risk loci and quantifies the genetic relationship with inflammatory bowel disease. *Nat. Genet.* 49, 269–273. <https://doi.org/10.1038/ng.3745>.
- Kuleshov, M.V., Jones, M.R., Rouillard, A.D., Fernandez, N.F., Duan, Q., Wang, Z., Koplev, S., Jenkins, S.L., Jagodnik, K.M., Lachmann, A., et al. (2016). Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* 44, W90–W97. <https://doi.org/10.1093/nar/gkw377>.
- Leitsalu, L., Haller, T., Esko, T., Tammesoo, M.-L., Alavere, H., Snieder, H., Perola, M., Ng, P.C., Mägi, R., Milani, L., et al. (2015). Cohort profile: Estonian biobank of the Estonian genome center, university of Tartu. *Int. J. Epidemiol.* 44, 1137–1147. <https://doi.org/10.1093/ije/dyt268>.
- Liu, B., Gloude-mans, M.J., Rao, A.S., Ingelsson, E., and Montgomery, S.B. (2019). Abundant associations with gene expression complicate GWAS follow-up. *Nat. Genet.* 51, 768–769. <https://doi.org/10.1038/s41588-019-0404-0>.
- Liu, J.Z., van Sommeren, S., Huang, H., Ng, S.C., Alberts, R., Takahashi, A., Ripke, S., Lee, J.C., Jostins, L., Shah, T., et al. (2015). Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nat. Genet.* 47, 979–986. <https://doi.org/10.1038/ng.3359>.
- Loh, P.-R., Danecek, P., Palamara, P.F., Fuchsberger, C., A Reshef, Y., K Finucane, H., Schoenherr, S., Forer, L., McCarthy, S., Abecasis, G.R., et al. (2016). Reference-based phasing using the haplotype reference Consortium panel. *Nat. Genet.* 48, 1443–1448. <https://doi.org/10.1038/ng.3679>.
- Lundström, W., Highfill, S., Walsh, S.T.R., Beq, S., Morse, E., Kockum, I., Alfredsson, L., Olsson, T., Hillert, J., and Mackall, C.L. (2013). Soluble IL7R $\alpha$  potentiates IL-7 bioactivity and promotes autoimmunity. *Proc. Natl. Acad. Sci. USA* 110, E1761–E1770. <https://doi.org/10.1073/pnas.1222303110>.
- Lundtoft, C., Seyfarth, J., and Jacobsen, M. (2020). IL7R $\alpha$  genetic variants differentially affect IL-7R $\alpha$  expression and alternative splicing: a role in autoimmune and infectious diseases? *Gene Immun.* 21, 83–90. <https://doi.org/10.1038/s41435-019-0091-y>.
- Mitt, M., Kals, M., Pärn, K., Gabriel, S.B., Lander, E.S., Palotie, A., Ripatti, S., Morris, A.P., Metspalu, A., Esko, T., et al. (2017). Improved imputation accuracy of rare and low-frequency

- variants using population-specific high-coverage WGS-based imputation reference panel. *Eur. J. Hum. Genet.* 25, 869–876. <https://doi.org/10.1038/ejhg.2017.51>.
- Miura, Y., Morooka, M., Sax, N., Roychoudhuri, R., Itoh-Nakadai, A., Brydun, A., Funayama, R., Nakayama, K., Satomi, S., Matsumoto, M., et al. (2018). Bach2 promotes B cell receptor-induced proliferation of B lymphocytes and represses cyclin-dependent kinase inhibitors. *J. Immunol.* 200, 2882–2893. <https://doi.org/10.4049/jimmunol.1601863>.
- Mumbach, M.R., Satpathy, A.T., Boyle, E.A., Dai, C., Gowen, B.G., Cho, S.W., Nguyen, M.L., Rubin, A.J., Granja, J.M., Kazane, K.R., et al. (2017). Enhancer connectome in primary human cells identifies target genes of disease-associated DNA elements. *Nat. Genet.* 49, 1602–1612. <https://doi.org/10.1038/ng.3963>.
- Okada, Y., Wu, D., Trynka, G., Raj, T., Terao, C., Ikari, K., Kochi, Y., Ohmura, K., Suzuki, A., Yoshida, S., et al. (2014). Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature* 506, 376–381. <https://doi.org/10.1038/nature12873>.
- Oyamada, A., Ikebe, H., Itsumi, M., Saiwai, H., Okada, S., Shimoda, K., Iwakura, Y., Nakayama, K.I., Iwamoto, Y., Yoshikai, Y., and Yamada, H. (2009). Tyrosine kinase 2 plays critical roles in the pathogenic CD4 T cell responses for the development of experimental autoimmune encephalomyelitis. *J. Immunol.* 183, 7539–7546. <https://doi.org/10.4049/jimmunol.0902740>.
- Ragusa, F., Fallahi, P., Elia, G., Gonnella, D., Paparo, S.R., Giusti, C., Churilov, L.P., Ferrari, S.M., and Antonelli, A. (2019). Hashimoto's thyroiditis: epidemiology, pathogenesis, clinic and therapy. *Best Pract. Res. Clin. Endocrinol. Metabol.* 33, 101367. <https://doi.org/10.1016/j.beem.2019.101367>.
- Razvi, S., Weaver, J.U., Vanderpump, M.P., and Pearce, S.H.S. (2010). The incidence of ischemic heart disease and mortality in people with subclinical hypothyroidism: reanalysis of the Whickham Survey cohort. *J. Clin. Endocrinol. Metab.* 95, 1734–1740. <https://doi.org/10.1210/jc.2009-1749>.
- Rodondi, N., den Elzen, W.P.J., Bauer, D.C., Cappola, A.R., Razvi, S., Walsh, J.P., Asvold, B.O., Iervasi, G., Imaizumi, M., Collet, T.-H., et al. (2010). Subclinical hypothyroidism and the risk of coronary heart disease and mortality. *JAMA* 304, 1365–1374. <https://doi.org/10.1001/jama.2010.1361>.
- Saevarsdottir, S., Olafsdottir, T.A., Ivarsdottir, E.V., Halldorsson, G.H., Gunnarsdottir, K., Sigurdsson, A., Johannesson, A., Sigurdsson, J.K., Juliusdottir, T., Lund, S.H., et al. (2020). FLT3 stop mutation increases FLT3 ligand level and risk of autoimmune thyroid disease. *Nature* 584, 619–623. <https://doi.org/10.1038/s41586-020-2436-0>.
- Servant, N., Varoquaux, N., Lajoie, B.R., Viara, E., Chen, C.-J., Vert, J.-P., Heard, E., Dekker, J., and Barillot, E. (2015). HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol.* 16, 259. <https://doi.org/10.1186/s13059-015-0831-x>.
- Sijts, E.J.a.M., and Kloetzel, P.M. (2011). The role of the proteasome in the generation of MHC class I ligands and immune responses. *Cell. Mol. Life Sci.* 68, 1491–1502. <https://doi.org/10.1007/s00018-011-0657-y>.
- Stanford, S.M., and Bottini, N. (2014). PTPN22: the archetypal non-HLA autoimmunity gene. *Nat. Rev. Rheumatol.* 10, 602–611. <https://doi.org/10.1038/nrrheum.2014.109>.
- Stephenson, J.D., Laskowski, R.A., Nightingale, A., Hurler, M.E., and Thornton, J.M. (2019). VarMap: a web tool for mapping genomic coordinates to protein sequence and structure and retrieving protein structural annotations. *Bioinformatics* 35, 4854–4856. <https://doi.org/10.1093/bioinformatics/btz482>.
- Sun, B.B., Maranville, J.C., Peters, J.E., Stacey, D., Staley, J.R., Blackshaw, J., Burgess, S., Jiang, T., Paige, E., Surendran, P., et al. (2018). Genomic atlas of the human plasma proteome. *Nature* 558, 73–79. <https://doi.org/10.1038/s41586-018-0175-2>.
- Tahara-Hanaoka, S., Shibuya, K., Onoda, Y., Zhang, H., Yamazaki, S., Miyamoto, A., Honda, S.-I., Lanier, L.L., and Shibuya, A. (2004). Functional characterization of DNAM-1 (CD226) interaction with its ligands PVR (CD155) and nectin-2 (PRR-2/CD112). *Int. Immunol.* 16, 533–538. <https://doi.org/10.1093/intimm/dxh059>.
- Taylor, P.N., Albrecht, D., Scholz, A., Gutierrez-Buey, G., Lazarus, J.H., Dayan, C.M., and Okosieme, O.E. (2018). Global epidemiology of hyperthyroidism and hypothyroidism. *Nat. Rev. Endocrinol.* 14, 301–316. <https://doi.org/10.1038/nrendo.2018.18>.
- Uhlén, M., Fagerberg, L., Hallström, B.M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, Å., Kampf, C., Sjöstedt, E., Asplund, A., et al. (2015). Proteomics. Tissue-based map of the human proteome. *Science* 347, 1260419. <https://doi.org/10.1126/science.1260419>.
- Umar, H., Muallima, N., Adam, J.M.F., and Sanusi, H. (2010). Hashimoto's thyroiditis following Graves' disease. *Acta Med. Indones.* 42, 31–35.
- Vanderpump, M.P.J. (2011). The epidemiology of thyroid disease. *Br. Med. Bull.* 99, 39–51. <https://doi.org/10.1093/bmb/ldr030>.
- Vösa, U., Claringbould, A., Westra, H.-J., Bonder, M.J., Deelen, P., Zeng, B., Kirsten, H., Saha, A., Kreuzhuber, R., Yazar, S., et al. (2021). Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression. *Nat. Genet.* 53, 1300–1310. <https://doi.org/10.1038/s41588-021-00913-z>.
- Wang, L., Balmat, T.J., Antonia, A.L., Constantine, F.J., Henao, R., Burke, T.W., Ingham, A., McClain, M.T., Tsalik, E.L., Ko, E.R., et al. (2021). An atlas connecting shared genetic architecture of human diseases and molecular phenotypes provides insight into COVID-19 susceptibility. *Genome Med.* 13, 83. <https://doi.org/10.1186/s13073-021-00904-z>.
- Wang, L.-X., Zhu, X.-M., and Yao, Y.-M. (2019). Sestrin2: its potential role and regulatory mechanism in host immune response in diseases. *Front. Immunol.* 10, 2797. <https://doi.org/10.3389/fimmu.2019.02797>.
- Wang, W., Huang, X., Xin, H.-B., Fu, M., Xue, A., and Wu, Z.-H. (2015). TRAF family member-associated NF- $\kappa$ B activator (TANK) inhibits genotoxic nuclear factor  $\kappa$ B activation by facilitating deubiquitinase USP10-dependent deubiquitination of TRAF6 ligase. *J. Biol. Chem.* 290, 13372–13385. <https://doi.org/10.1074/jbc.M115.643767>.
- Watanabe, K., Taskesen, E., van Bochoven, A., and Posthuma, D. (2017). Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun.* 8, 1826. <https://doi.org/10.1038/s41467-017-01261-5>.
- Willer, C.J., Li, Y., and Abecasis, G.R. (2010). METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* 26, 2190–2191. <https://doi.org/10.1093/bioinformatics/btq340>.
- Yavorska, O.O., and Burgess, S. (2017). MendelianRandomization: an R package for performing Mendelian randomization analyses using summarized data. *Int. J. Epidemiol.* 46, 1734–1739. <https://doi.org/10.1093/ije/dyx034>.
- Zheng, J., Erzurumluoglu, A.M., Elsworth, B.L., Kemp, J.P., Howe, L., Haycock, P.C., Hemani, G., Tansey, K., Laurin, C., Early Genetics and Lifecourse Epidemiology EAGLE Eczema Consortium, et al. (2017). LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. *Bioinformatics* 33, 272–279. <https://doi.org/10.1093/bioinformatics/btw613>.
- Zhou, W., Nielsen, J.B., Fritsche, L.G., Dey, R., Gabrielsen, M.E., Wolford, B.N., LeFaive, J., VandeHaar, P., Gagliano, S.A., Gifford, A., et al. (2018). Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. *Nat. Genet.* 50, 1335–1341. <https://doi.org/10.1038/s41588-018-0184-y>.
- Zhu, Z., Zhang, F., Hu, H., Bakshi, A., Robinson, M.R., Powell, J.E., Montgomery, G.W., Goddard, M.E., Wray, N.R., Visscher, P.M., and Yang, J. (2016). Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat. Genet.* 48, 481–487. <https://doi.org/10.1038/ng.3538>.

## STAR★METHODS

### KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
<b>Deposited Data</b>		
GWAS summary statistics hypothyroidism	This paper	<a href="https://www.ebi.ac.uk/gwas/">https://www.ebi.ac.uk/gwas/</a>
<b>Software and Algorithms</b>		
SAIGE	Zhou et al. (2018)	<a href="https://github.com/weizhouUMICH/SAIGE">https://github.com/weizhouUMICH/SAIGE</a>
METAL	Willer et al. (2010)	<a href="http://csg.sph.umich.edu/abecasis/metal/download/">http://csg.sph.umich.edu/abecasis/metal/download/</a>
PhenoScanner	NA	<a href="http://www.phenoscanter.medschl.cam.ac.uk/">http://www.phenoscanter.medschl.cam.ac.uk/</a>
Bedtools	NA	<a href="https://bedtools.readthedocs.io/en/latest/">https://bedtools.readthedocs.io/en/latest/</a>
PLINK	NA	<a href="http://zzz.bwh.harvard.edu/plink/">http://zzz.bwh.harvard.edu/plink/</a>
GEO DataSets	GSE101498	<a href="https://www.ncbi.nlm.nih.gov/gds">https://www.ncbi.nlm.nih.gov/gds</a>
HiC-Pro	Servant et al. (2015)	<a href="https://github.com/nservant/HiC-Pro">https://github.com/nservant/HiC-Pro</a>
FitHiChIP	Bhattacharyya et al. (2019)	<a href="https://ay-lab.github.io/FitHiChIP/">https://ay-lab.github.io/FitHiChIP/</a>
LD score	Finucane et al. (2015)	<a href="https://github.com/bulik/ldsc">https://github.com/bulik/ldsc</a>
LD Hub	NA	LD Hub (broadinstitute.org)
HyPrColoc	Foley et al., (2021)	<a href="https://github.com/jrs95/hyprcoloc">https://github.com/jrs95/hyprcoloc</a>
LocusCompare	Liu et al. (2019)	<a href="http://locuscompare.com/">http://locuscompare.com/</a>
Mendelian Randomization R package	Yavorska and Burgess, 2017	<a href="https://cran.r-project.org/web/packages/MendelianRandomization/index.html">https://cran.r-project.org/web/packages/MendelianRandomization/index.html</a>
DGIdb	Cotto et al. (2018)	<a href="https://www.dgidb.org/">https://www.dgidb.org/</a>
GARFIELD	lotchkova et al. (2019)	<a href="https://www.ebi.ac.uk/birney-srv/GARFIELD/">https://www.ebi.ac.uk/birney-srv/GARFIELD/</a>
eQTLGen	Võsa et al. (2021)	<a href="https://www.eqtlgen.org/index.html">https://www.eqtlgen.org/index.html</a>
INTERVAL	Sun et al. (2018)	<a href="https://app.box.com/s/u3flbp13zjydegrxjb2uepagp1vb6bj2">https://app.box.com/s/u3flbp13zjydegrxjb2uepagp1vb6bj2</a>
Enrichr	Kuleshov et al. (2016)	<a href="https://amp.pharm.mssm.edu/Enrichr/">https://amp.pharm.mssm.edu/Enrichr/</a>
Bioplex 3.0	NA	<a href="https://bioplex.hms.harvard.edu/explorer/">https://bioplex.hms.harvard.edu/explorer/</a>
LiftOver	NA	<a href="https://genome-store.ucsc.edu/">https://genome-store.ucsc.edu/</a>
ggplot2	NA	<a href="https://ggplot2.tidyverse.org/">https://ggplot2.tidyverse.org/</a>
Morpheus	NA	<a href="https://software.broadinstitute.org/morpheus/">https://software.broadinstitute.org/morpheus/</a>
The Protein Atlas	Uhlén et al., 2015	<a href="https://www.proteinatlas.org/about/download">https://www.proteinatlas.org/about/download</a>
Tspex	NA	<a href="https://apcamargo.github.io/tspex/">https://apcamargo.github.io/tspex/</a>
PICS	Farh et al., 2015	<a href="https://pubs.broadinstitute.org/pubs/finemapping/">https://pubs.broadinstitute.org/pubs/finemapping/</a>

### RESOURCE AVAILABILITY

#### Lead contact

Further information and request for resources should be directed to and will be fulfilled by the lead contact, Patrick Mathieu ([patrick.mathieu@fmed.ulaval.ca](mailto:patrick.mathieu@fmed.ulaval.ca)).

#### Materials availability

This study did not generate new material or reagent.

#### Data and code availability

- Summary statistics for the GWAS meta-analysis have been deposited at the NHGRI-EBI GWAS catalog. Accession number is listed in the [Key resources table](#).
- This paper does not report original code.

- Any additional information required to reanalyze the data reported in this paper are available from the [lead contact](#) upon request.

## EXPERIMENTAL MODEL AND SUBJECT DETAILS

### Study participants

UK Biobank, a large open access resource including subjects aged 40–69 and prospectively enrolled, was leveraged to perform a GWAS on hypothyroidism. Cases were identified from hospital diagnosis code (ICD-9 or ICD-10) E03 for hypothyroidism or by using treatment/medication code to identify individuals taking levothyroxine. The analysis was executed under the UK Biobank application number 25205 and approval from the ethics committee of the Centre de Recherche de l'Institut Universitaire de Cardiologie et de Pneumologie de Québec. This dataset was used to conduct a meta-analysis with data from FinnGen, a large open access study prospectively enrolling subjects to study the genetics of more than 2,800 disease end-points. Replication stage was performed in EstBB, a population-based biobank with over 200,000 participants. Replication stage was approved by the Estonian Committee on Bioethics and Human Research (approval number 1.1-12/624). Individuals with hypothyroidism were identified using the ICD-10 codes from E03 category and ATC prescription H03AA01 codes ( $n = 17,002$ ). Biobank participants who did not have these diagnoses or prescriptions were considered as controls ( $n = 178,141$ ). Information on ICD codes is obtained via regular linking with the national Health Insurance Fund and other relevant databases ([Leit-salu et al., 2015](#)).

## QUANTIFICATION AND STATISTICAL ANALYSIS

### GWAS for hypothyroidism

In UK Biobank, genotyping was performed by using 25,130 cases and 383,471 controls of white British ancestry and phasing-imputation were executed by using the Haplotype Reference Consortium and merged UK10K and 1000 Genomes phase 3 reference panels.

Gender of the participants included 47.7% and 18.1% of male individuals in controls and cases respectively. Excluded samples were those with call rate <95%, outlier heterozygosity rate, sex mismatch, non-white British ancestry and excess third-degree relatives (>10). Variants with imputation score (INFO)  $\leq 0.3$  or with minor allele frequency <0.001 were excluded leaving 16,445,106 variants for the analysis (assembly GRCh37/hg37). Analysis including 25,130 cases and 383,471 controls of white British ancestry was performed by using SAIGE (Scalable and Accurate Implementation of GEneralized mixed model, version 0.36.3.1), a two-stage method implementing generalized mixed model, which is robust to unbalanced case-control ratio ([Zhou et al., 2018](#)). Analysis was performed by using age, sex, and the first 20 ancestry-based principal components and without the leave-one-chromosome-out (LOCO) scheme (LOCO = FALSE). In FinnGen, individuals have been genotyped by using Affymetrix chip arrays (Illumina Inc., San Diego, and Thermo Fisher Scientific, Santa Clara, CA, USA). Data were imputed with a population specific panel (SISu reference panel) generating datasets of 16,962,023 gene variants using the reference assembly GRCh38/hg38. Analyses were performed by using SAIGE version 0.36.3.2. Age, sex, 10 principal components and genotyping batch were added as covariates. SAIGE was run without the leave-one-chromosome-out (LOCO) scheme (LOCO = FALSE) and results were filtered to include variants with an imputation score (INFO) > 0.6. FinnGen dataset freeze 5 (as of May 2021) under the heading "Hypothyroidism, levothyroxin purchases", which included 26,064 cases and 59,912 controls, was downloaded. Data were converted to GRCh37/hg19 by using the LiftOver executable tool from UCSC. We conducted a fixed-effect meta-analysis by using data from UK Biobank and FinnGen including 51,194 cases and 443,383 controls. The analysis was performed with METAL ([Willer et al., 2010](#)) and included 10,836,151 SNPs with INFO score  $\geq 0.3$ .

All EstBB participants have been genotyped at the Core Genotyping Lab of the Institute of Genomics, University of Tartu, using Illumina Global Screening Array v3.0. Samples were genotyped and PLINK format files were created using Illumina GenomeStudio v2.0.4. Individuals were excluded from the analysis if their call-rate was <95% or if sex defined based on heterozygosity of X chromosome did not match sex in phenotype data. Before imputation, variants were filtered by call-rate <95%, HWE p value < 1e-4 (autosomal variants only), and minor allele frequency <1%. Prephasing was done using Eagle v2.3 software ([Loh et al., 2016](#)) (number of conditioning haplotypes Eagle2 uses when phasing each sample was set to: -Kpbwt = 20000) and imputation was done using Beagle v.28Sep18.793 with effective population size  $n_e = 20,000$

(Browning and Browning, 2007). Population specific imputation reference of 2297 WGS samples was used (Mitt et al., 2017). Association analysis was carried out using SAIGE (v0.43.1) software implementing mixed logistic regression model with LOCO = TRUE, using sex, age, age\_sq and ten PCs as covariates in step 1.

### LD score and heritability

The LD score was used to assess the intercept and the heritability (Finucane et al., 2015). The meta-analysis summary statistics were munged (`-munge_sumstats.py`) for processing and calculate the intercept of LD score. The heritability on the liability scale was calculated by using a population prevalence of 0.05 (`-pop-prev`) and sample prevalence (`-samp-prev`) of 0.103.

### Annotation of the GWAS

Association data were processed with Functional Mapping and Annotation of GWAS (FUMA) (Watanabe et al., 2017). Genomic risk loci were determined by using a pre-calculated LD structure of the 1000G EUR reference population. Variants in genomic loci with LD  $r^2 < 0.6$ ,  $p\text{-value} < 5E-08$  were identified as independent significant SNPs (IndSigSNPs). The IndSigSNPs independent from each other (LD  $r^2 < 0.1$ ) were identified as lead SNPs at risk loci. Genomic loci closely located ( $< 250$  kb based on the most right and left SNPs of each locus) were merged into one genomic risk locus. Variants were annotated with ANNOVAR as intergenic, intronic or exonic. The annotation of genes was based Ensembl (build 85) and entrez ID yielding identification of 19,436 protein coding genes. Exonic variants were annotated by using exonic lead and IndSigSNPs identified from FUMA and were processed with VarMap (Stephenson et al., 2019). SNPs annotated as missense or splicing SNPs by VarMap were analyzed by computing the CADD score derived from the output of FUMA.

### Probabilistic Identification of Causal SNPs

PICS (Probabilistic Identification of Causal SNPs) algorithm was implemented to fine-map risk loci. (Farh et al., 2015) PICS estimates the Bayesian probability that a variant is causal after considering the haplotype structure and the level of associations at risk loci. PICS data were used to create a 95% credible set at risk loci.

### Cell and tissue enrichment

GARFIELD uses tissue/cell-specific functional annotations (1005 features including genomic annotations, chromatin states, histone modifications, DNaseI hypersensitive sites and transcription factor binding sites), derived from ENCODE and Roadmap epigenomics data to calculate a LD corrected enrichments from GWAS data (lotchkova et al., 2019). Default settings from GARFIELD, which implements a generalized linear model, were used to calculate the functional enrichment of the GWAS. The output peaks file was used for downstream processing. As another strategy to assess the enrichment of the GWAS in functional annotations, we implemented the stratified LD score framework including 489 cell-tissues (Finucane et al., 2015). The summary statistic data of the meta-analysis was transformed by using `-munge_sumstats.py` and coefficients were obtained from multi-tissue chromatin (`cts_name = Multi_tissue_chromatin`).

### Gene-based and pathway analyses

Multi-marker Analysis of GenoMic Annotation (MAGMA) implements a multiple regression model and incorporate LD between markers to perform gene-based test (de Leeuw et al., 2015). MAGMA was run from the FUMA platform using the default settings including a window of 0kb (i.e. SNPs only assigned to the gene). In MAGMA, 19,215 genes were assessed and data were deemed significant at a Bonferroni threshold ( $p < 2.60E-06$ ,  $0.05/19,215$ ). Genes significant in MAGMA were processed to document pathway enrichment by using the BioCarta dataset and the analysis was performed through Enrichr (Kuleshov et al., 2016).

### Molecular QTL mapping

Genes mapped to eQTLGen (Vösa et al., 2021) were identified by using lead and IndSigSNPs obtained from FUMA. SNPs were assigned to *cis*-eQTLs by using a window of  $\pm 500$  kb around the transcription start site. SNP-gene pairs were filtered at false discovery rate of 5% ( $FDR < 0.05$ ). For the INTERVAL cohort (Sun et al., 2018), a large scale dataset of pQTL including data for 2,965 different blood proteins assessed with the aptamer-based multiplex protein assay from SOMAscan, lead and IndSigSNPs output from FUMA output were used to map pQTL using the same window ( $\pm 500$  kb). SNP-protein pairs with  $FDR < 0.05$  were kept for downstream analyses.

### Colocalization

Bayesian colocalization was performed to assess shared genetic signal between the GWAS and molecular QTLs by implementing HyprColoc (Foley et al., 2021). Genomic regions were defined as  $\pm 250$  kb from the transcription start site of molecular trait. Colocalization of the genetic signal was considered significant if the posterior probability was superior to 0.8 (PP > 0.8). LocusCompare has been used to illustrate the shared signal (Liu et al., 2019).

### 3D genome mapping

Public data of enhancer-promoter interactions using H3K27ac-HiChIP and carried out in primary T cells (GSE101498) were downloaded for analyses. FASTQ files were processed with HiC-Pro using the default settings (Servant et al., 2015). Loop call was performed by using FitHiChIP using a FDR < 1E-06 and a resolution of 5 kb (Bhattacharyya et al., 2019). We identified the promoters of protein coding genes as a region of  $\pm 2$  kb from the transcription start site by using data from GENCODE version 35 in build 37. The assignment of lead and IndSigSNPs to 3D mapped genes was performed by using bedtools with the intersect function.

### Mendelian randomization

We performed two-sample MR on genes mapped by eQTL, pQTL and enhancer-promoter conformation. Independent ( $r^2 < 0.1$ ) instrumental variables (SNPs) identified with PLINK1.9 based on genotypes from European populations from the 1000 Genome project, located in *cis* ( $\pm 250$  kb from the transcription start site) and with a p value < 0.001 (corresponds to  $\sim F$  statistics > 10) were selected. For eQTLGen, which reports the Z score, data were transformed to effect size (beta) and SE by using the following equation as previously described (Zhu et al., 2016):

$$\hat{\beta} = Z * \widehat{SE}, \text{ where } \widehat{SE} = 1 / \sqrt{2p(1-p)(n+Z^2)}$$

where Z-score (Z), allele frequency (p) and sample size (n).

Horizontal pleiotropy was assessed by using the Cochran's Q test and was deemed significant if  $P_{\text{heterogeneity}} < 0.05$ . Inverse variance weighted MR was performed. Sensitivity analyses were executed by using the weighted median MR, which allows the use of up to 50% of invalid instruments. Analyses were performed by using the Mendelian Randomization package (Yavorska and Burgess, 2017).

### Multi-trait mendelian randomization analysis

Multi-trait Inverse variance weighted MR were performed by using data of pQTL from INTERVAL (exposition) and 35 traits-disorders (outcome) pertaining to 7 categories (Data S16). Association from the MR analyses were deemed significant after applying the Bonferroni correction ( $p < 1.43E-03$ , 0.05/35). MR analyses were performed as described in the methods section: Mendelian randomization. Results were illustrated as bar graph generated with ggplot2 in R.

### Cross-trait analyses

The LD score was leveraged to assess genetic correlations as implemented in LD Hub (Zheng et al., 2017). Data were generated using the traits-diseases included in LD Hub and filtered at a Bonferroni threshold. GWAS data were also evaluated with the Cross-Phenotype Analysis of GWAS database (iCPAGdb) (Wang et al., 2021). iCPAGdb uses ancestry LD-specific association data across 3,793 traits-disorders from the NHGRI-EBI GWAS catalog to compute cross-phenotype enrichment analyses. iCPAGdb reports pairwise trait combination along with shared signal reported as Fisher exact test with adjustment (Benjamini-Hochsberg and Bonferroni) and the Chao-Sorenson similarity index.

### Drug target analyses

Causal genes identified from colocalization or MR were evaluated in The Drug Gene Interaction Database (DGIdb) (Cotto et al., 2018). DGIdb is a large repository of drug-gene pairs retrieved from an exhaustive list of resources. We reported drug-gene pairs by using approved and non-approved drugs collated in DGIdb.



**Tissue specificity score**

Normalized tissue gene expression for 61 tissues collated from GTEx, The Human Protein Atlas and FANTOM5 was downloaded from The Human Protein Atlas ([Uhlén et al., 2015](#)). Tissue specificity score has been calculated by using the Jensen-Shannon specificity metric. Jensen-Shannon specificity score evaluate distance between the distribution of data ([Cabili et al., 2011](#)). Jensen-Shannon specificity score was calculated by using the tspex package. Hierarchical clustering based on Euclidian distance and pairwise Pearson correlation were performed by using Morpheus from the Broad Institute.