scientific reports

Check for updates

OPEN

Proteome-wide Mendelian randomization reveals causal associations between plasma proteins and autoimmune thyroid disease

Yang Li[1,2], Weixi Zhu[1,2], Yijing Chen[1,2], Qingqing Kang[1], Ying Zhang[1], Pan Yang[1], Shumin Wang[1], Chao Liu[1]✉, Yi Zhang[1]✉ & Qiu Zhang[1]✉

Autoimmune thyroid diseases (AITD) are the most common autoimmune disorders. Identifying new biomarkers and therapeutic targets in plasma proteins is crucial. We conducted a proteome-wide Mendelian randomization (MR) and colocalization analysis to determine plasma proteins causally associated with AITD. Proteome-wide summary-level genome-wide association studies (GWAS) were collected from the UK Biobank Pharma Proteomics Project (UKB-PPP) and deCODE genetics, encompassing 2922 and 4719 plasma proteins, respectively. Genetic associations with AITD were derived from an AITD GWAS meta-analysis study (30,234 cases and 725,172 controls) and the FinnGen database (40,926 cases and 274,069 controls). MR analysis, including summary-data-based Mendelian randomization (SMR), Wald Ratio, and IVW methods, was employed to estimate the causal effects between plasma proteins and AITD. Colocalization analysis was used to assess whether identified proteins and AITD shared the common causal variants. Genetically predicted levels of 11 plasma proteins were found to have a causal association with AITD. Colocalization analysis revealed that five of these proteins had evidence of colocalization, including leukemia inhibitory factor (LIF), interleukin-7 receptor subunit alpha (IL7RA), CD226, tumor necrosis factor ligand superfamily member 11 (TNF11), and transcription factor junD (JUND). Genetically predicted levels of LIF and IL7RA were associated with an increased risk of AITD, whereas CD226, TNF11, and JUND were inversely related to AITD risk. This study has identified multiple candidate plasma proteins causally associated with AITD. Among these proteins, LIF, IL7RA, CD226, TNF11, and JUND are considered to have potential as disease biomarkers and therapeutic targets, but further clinical and experimental validation is still necessary in the future.

Keywords Autoimmune thyroid diseases, Plasma protein, Mendelian randomization, Colocalization analysis, Causal association

Autoimmune thyroid diseases (AITD), encompassing primarily Grave's disease and Hashimoto's thyroiditis, represent a group of conditions characterized by immune-mediated damage to the thyroid gland[1], with a combined prevalence of approximately 5% in the general population, making them the most common among autoimmune disorders[2]. The concordance of AITD in monozygotic twins exceeds 50%, highlighting its genetic susceptibility[3]. The diagnosis of AITD mainly relies on serological markers, such as thyroid peroxidase antibodies (TPOAb), anti-thyroglobulin antibodies (TgAb) and thyrotropin receptor antibodies (TRAb). However, the sensitivity and specificity of these diagnostic methods still have limitations, leading to challenges in early detection and differentiation from other thyroid disorders[4]. Furthermore, the presence of antibodies does not always correlate with disease activity, complicating the assessment of disease progression and response to treatment[2,5]. Additionally, the common treatment of AITD focuses on managing symptoms and normalizing thyroid hormone levels, such as antithyroid drugs, radioactive iodine therapy, and thyroidectomy for Graves'

[1]Department of Endocrinology, First Affiliated Hospital of Anhui Medical University, Hefei, China. [2]Yang Li, Weixi Zhu and Yijing Chen contributed equally to this work and shared first authorship. ✉email: lc_ahmu@163.com; zy18356056506@163.com; zhangqiu@ahmu.edu.cn

disease and levothyroxine for Hashimoto's thyroiditis. However, the current therapeutic strategies for AITD are limited by their lack of specificity, potential side effects, and inability to prevent disease recurrence or progression[1,2,6]. Therefore, it is essential to identify refined biomarkers to enhance early diagnosis and specific therapeutic targets to modulate the immune response.

Plasma proteins play a crucial role in the onset, progression, diagnosis, and treatment of diseases, particularly in autoimmune disorders[7]. The exploration of plasma proteins in the context of AITD represents a significant advancement in understanding the disease markers, potential therapeutic targets and molecular mechanisms. For instance, proteins involved in the immune response, inflammation, and thyroid hormone transport have been found to be altered in AITD patients[8,9]. These findings suggest that plasma proteins may play crucial roles in the initiation and perpetuation of autoimmune responses against thyroid tissues. Many targeted therapies have been utilized in the treatment of Graves' disease, such as rituximab targeting CD20 and imatinib targeting Tyrosine kinase[10,11]. However, targeted therapies for autoimmune hypothyroidism are still under investigation.

Mendelian randomization (MR) employs genetic variations as instrumental variables (IVs) for exposures to infer causal relationships between exposure and outcome. It effectively overcomes issues of reverse causation and confounders, which are common in observational studies. This is because genetic variations are randomly assorted at conception, thereby balancing confounding factors and reverse causation[12]. Recently, large-scale studies linking plasma proteins with genetics have provided opportunities for comprehensive investigation into the causal effects of plasma proteins on AITD[13,14]. These proteomic studies not only deepen our understanding of disease mechanisms but also help reveal potential biomarkers and therapeutic targets, as many plasma proteins are key regulators in critical pathways[15]. The use of MR analysis to explore causal relationships between plasma proteins and disease phenotypes has been widely adopted[16,17]. It has been found that using cis-variants as IVs for the protein-encoding gene can possibly balance horizontal pleiotropy[18].

Based on summary-level genome-wide association studies (GWAS) of plasma proteins from two large-scale studies, we conducted a proteome-wide MR study complemented with colocalization analysis, exploring the causal associations between plasma proteins and AITD. The study design overview can be found in Fig. 1.

Materials and methods

Ethical statement

In conducting this study, we utilized publicly accessible, de-identified data that had already received approval from an ethical standards committee in the original articles. Consequently, no further ethical approval was required for this research.

Data source

The summary-level GWAS statistics on plasma proteins were derived from the UK Biobank Pharma Proteomics Project (UKB-PPP) and the deCODE genetics[13,14]. We collected the European population discovery cohort from the UKB-PPP (https://www.synapse.org/#!Synapse:syn51364943/files/), which included GWAS data of 2,922 plasma proteins, measured in 34,557 participants using the Olink platform. The obtained data were matched with the corresponding SNPs according to the official annotation document. Meanwhile, we retrieved data from the deCODE genetics (https://www.decode.com/summarydata/), which included summary-level GWAS data of 4,719 plasma proteins in 35,559 Icelanders, measured using the SomaScan platform. This data was processed to exclude low-quality SNPs and map the allele information based on the official annotation files. Detailed information can be found in Supplementary Table S1.

The summary-level GWAS datasets for AITD primarily gathered from a GWAS meta-analysis of AITD and the FinnGen database[19,20]. The GWAS meta-analysis dataset (https://www.decode.com/summarydata/) is a comprehensive analysis combining data from the UK Biobank and Iceland, encompassing 30,234 cases and 725,172 controls. This dataset focuses on three AITD-related diseases: Grave's disease (n = 2,400; ICD-10: E05.9), Hashimoto's thyroiditis (n = 397; ICD-10: E06.3), and hypothyroidism (n = 27,437; ICD-10: E03.9). The GWAS statistics about AITD from the FinnGen database (R9 version; https://r9.finngen.fi/) include 40,926 cases and 274,069 controls, mainly concerning hypothyroidism (ICD-10: E03.9). There is no sample overlap between the two outcome datasets. Further information is available in Supplementary Table S1.

The Saedis Saevarsdottir et al. AITD meta-GWAS cohort had minor sample overlap with the exposure datasets, approximately 4.6% with the UKB-PPP cohort and 4.7% with the deCODE cohort, while no overlap existed with the FinnGen cohort. Given the low degree of overlap, the impact on causal inference in two-sample MR analyses is expected to be minimal.[21].

Genetic instruments

To identify IVs related to plasma proteins for inverse variance weighted (IVW) and Wald ratio methods, we selected SNPs within a 1 MB cis-regulatory region around the transcription start site (TSS)[18]. These candidate IVs had to meet the following criteria for MR analysis: (1) SNPs must be associated with the exposure at the genome-wide significance ($p < 5 \times 10^{-8}$). (2) There should be no significant linkage disequilibrium between SNPs based on the European 1000 Genomes Project Phase 3 ($r^2 < 0.001$, https://www.internationalgenome.org/data/). (3) SNPs must have an *F*-statistic greater than 10 to minimize weak instrument bias. (4) SNPs should not be associated with the outcome ($p > 5 \times 10^{-8}$). (5) SNPs that are palindromic or incompatible between the exposure and outcome were excluded from further analysis. (6) To minimize bias due to rare variants, only SNPs with a minor allele frequency (MAF) greater than 1% were retained. For the selection of IVs in the summary-data-based Mendelian randomization (SMR) method, we chose the SNP with the lowest p-value in the same genetic location. Detailed information about the selected IVs is displayed in Supplementary Tables S2-5.

In the UKB-PPP and deCODE cohorts, 4907 probes were used to measure 4719 plasma proteins, and 2940 probes were utilized to assess 2922 plasma proteins, respectively. Plasma proteins were excluded from MR

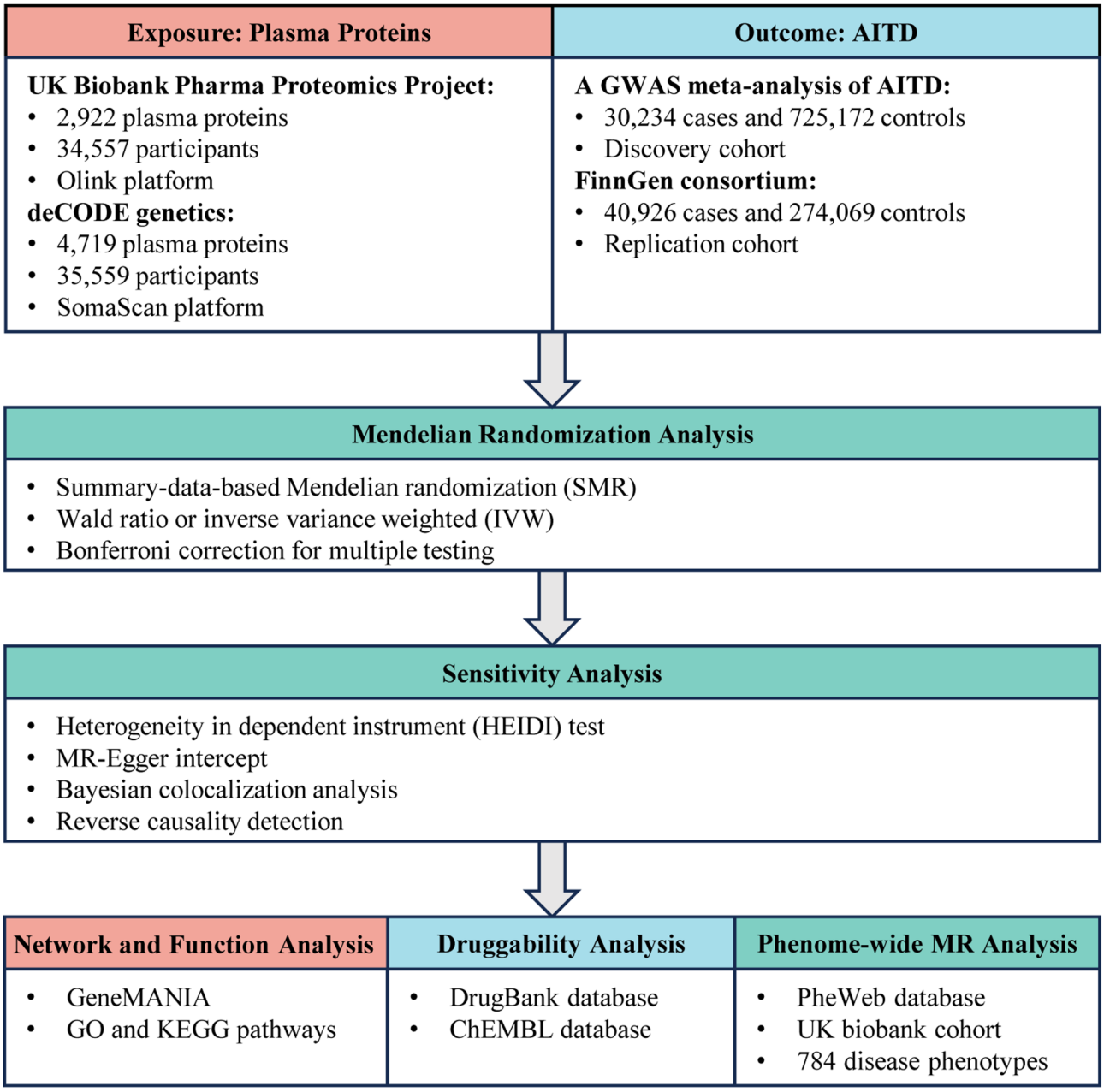

Fig. 1. Study design overview for identification of plasma proteins causally associated with AITD.

analysis under the following conditions: (1) the gene encoding the plasma protein is located on the X or Y chromosome; (2) there are no available IVs, indicating no SNP significantly associated with the trait at the genome-wide level ($p > 5 \times 10^{-8}$); (3) the available IVs do not exist in the outcome data.

Mendelian randomization analysis

In this study, we utilized summary-level GWAS statistics of plasma proteins from UKB-PPP and deCODE genetics as the exposure, with two independent datasets on AITD serving as the outcomes for MR analysis. A GWAS meta-analysis of AITD was designated as the discovery cohort, while a GWAS dataset from the FinnGen database was set as the replication cohort. Given the technical and methodological differences between the Olink and SomaScan platforms, such as epitope effects, our MR analysis for shared proteins required significance on at least one platform with consistent effect direction across both platforms. Three statistical methods, namely Wald ratio, IVW[22], and SMR[23], were employed to estimate causal effects between the exposures and outcomes. Sensitivity analyses were also conducted to assess the pleiotropy of the genetic instruments.

The SMR approach was utilized to estimate causal associations by using the top cis-SNP within the 1 MB region around the TSS, with the heterogeneity in dependent instrument (HEIDI) test assessing the presence of pleiotropy (HEIDI p-value < 0.05 indicating pleiotropy)[23]. The Wald ratio method was used when only one

cis-SNP was present. When the number of cis-SNPs exceeded two, the IVW method was performed to estimate causal effects, accompanied by Cochran's Q test for heterogeneity[24] and MR-Egger intercept for pleiotropy[25]. Heterogeneity among IVs was considered if the Cochran's Q test *p*-value was less than 0.05. Pleiotropy was detected if the MR-Egger intercept *p*-value was less than 0.05. The IVW random-effects model was applied when significant heterogeneity was identified. Causal effects between plasma proteins and AITD were scaled to one standard deviation (SD) increase in genetically predicted levels of plasma protein. The Bonferroni method was applied to correct for multiple testing.

We established that a causal association between plasma proteins and AITD must meet the following criteria: (1) The causal effect must demonstrate significance after Bonferroni correction in both the discovery and replication cohort. (2) Horizontal pleiotropy should not be observed in any cohort, including MR-Egger intercept and HEIDI test. (3) The direction of the causal associations for the same protein must be consistent, even when employing different statistical methods and different data sources.

Colocalization analysis

In our research, we utilized colocalization analysis to ascertain if the observed correlations between certain protein and AITD were attributable to linkage disequilibrium. This process was underpinned by a Bayesian framework, which evaluated five distinct scenarios: (1) neither trait is associated; (2) only the first trait is associated; (3) only the second trait is associated; (4) both traits are associated but through separate causal variants; and (5) both traits are associated via a common causal variant[26]. Each scenario was assigned a posterior probability (PP), denoted as H0 through H4. The analysis parameters included prior probabilities of the SNP's association with the first trait at 1×10^{-4}, with the second trait at 1×10^{-4}, and with both traits at 1×10^{-5}. A strong evidence of colocalization was defined as a posterior probability for H4 (PP.H4) greater than 0.8, while a PP.H4 between 0.5 and 0.8 was considered to indicate moderate evidence of colocalization. Recognizing the limitations of traditional colocalization methods in identifying situations with multiple causal variants affecting both traits, we adopted the Sum of Single Effects (SuSiE) colocalization approach[27]. This method integrates GWAS summary data and genetic correlation matrix to detect multiple causal variants. We defined that within any of the colocalization algorithm, if a plasma protein and AITD demonstrated a colocalization in both the discovery and the replication cohort, then a colocalization effect was confirmed.

Reverse causality detection

Following the same IVs selection criteria mentioned earlier, we identified IVs in both the discovery and the replication cohort to investigate the reverse causal association between plasma proteins and AITD. The IVW, weighted median[28], and MR-Egger methods were utilized to assess the causal effect, with the IVW method being the primary approach. Additionally, we conducted Cochran's Q test and MR-Egger intercept to evaluate the heterogeneity and pleiotropy, respectively. A reverse causal relationship between plasma proteins and AITD was considered when the causal effect estimated by the IVW method had a Bonferroni-corrected *p*-value, and the IVs showed no pleiotropy (*p*-value of MR-Egger intercept greater than 0.05), and the direction of the causal association was consistent across all three MR analyses.

Functional enrichment and drug targets analysis

The GeneMANIA website (https://genemania.org/) was employed to construct protein–protein interaction (PPI) networks of AITD-associated plasma proteins[29]. Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways was applied to perform functional enrichment analysis based on the latest online database. Enriched function was considered significant when the false discovery rate (FDR) was less than 0.05. Additionally, to assess the druggability of causal proteins, we searched the DrugBank (https://go.drugbank.com/)[30] and ChEMBL (https://www.ebi.ac.uk/chembl/)[31] databases for drugs that target or may target these causal proteins.

Phenome-wide MR (Phe-MR) analysis of the causal plasma proteins

To investigate the causal associations of the identified causal plasma proteins on other diseases, we conducted Phe-MR analysis[32]. The acquisition of IVs for causal proteins was as described previously. SNP-outcome effects were obtained from the PheWeb database (https://www.leelabsg.org/resources)[33]. This database employs the Scalable and Accurate Implementation of Generalized mixed model (SAIGE) approach to analyze GWAS data from the UK Biobank cohort and identifies 1403 disease phenotypes based on ICD-9/10 codes[34]. In this study, we excluded disease phenotypes with fewer than 500 cases due to statistical power. A total of 784 diseases were included in the Phe-MR analysis. The MR analysis was similar as described above.

Software and statistical analysis

"R" (version 4.1.2), "TwoSampleMR" (version 0.5.6)[35], "coloc" (version 5.2.3)[26,27], "SMR" (version 1.3.1)[23], "clusterProfiler" (version 4.8.3)[36], and "PLINK" (version 1.9.0)[37] were used for statistical analysis. All tests were two-tailed. The Bonferroni method was applied for multiple hypothesis testing. When the Bonferroni-corrected *p*-value was satisfied, the association was significant. When the *p*-value is less than 0.05 but not significant after Bonferroni correction, a suggestive association was considered.

Results

Selection of cis-genetic instruments

When estimating causal effects using the IVW or Wald ratio methods, a total of 2878 plasma proteins were included in the MR analysis. In the UKB-PPP cohort, cis-SNPs were used as IVs for plasma proteins in both the discovery (6407 cis-SNPs for 2001 probes) and replication cohorts (5665 cis-SNPs for 1975 probes) (Supplementary Table

S2). In the deCODE genetics cohort, 6629 cis-SNPs were used as IVs for 1780 plasma proteins (Supplementary Table S3). There were 912 plasma proteins shared between the UKB-PPP and deCODE genetics cohorts. All IVs had suitable *F*-statistics, ranging from 13 to 28,634, effectively balancing weak instrument bias. When estimating causal associations using the SMR method, we analyzed the causal relationship between 2880 plasma proteins and AITD, with 915 proteins being found in both cohorts. In the UKB-PPP cohort, there were 1974 and 2022 probes for discovery and replication respectively, which were enrolled for further MR analysis. Detailed information on cis-SNPs used for the SMR analysis can be found in Supplementary Tables S4-5.

Causal associations between plasma proteins and AITD identified by SMR

Utilizing SMR analysis to explore the causal association between plasma proteins and AITD in the UKB-PPP database, we identified 37 and 24 proteins that remained significant after Bonferroni correction for discovery cohort ($p < 2.53 \times 10^{-5}$; 0.05/1974) and replication cohort ($p < 2.47 \times 10^{-5}$; 0.05/2022), respectively (Fig. 2A,B; Supplementary Table S4). Through pleiotropy testing using the HEIDI test ($p > 0.05$), a total of 13 plasma proteins were ultimately determined as causal proteins. Among these, 4 plasma proteins showed significance after Bonferroni correction in both AITD outcomes, including CD226, leukemia inhibitory factor (LIF), complement C1q tumor necrosis factor-related protein 6 (C1QT6), and tumor necrosis factor ligand superfamily member 11 (TNF11). Similarly, for plasma proteins in deCODE genetics, 15 causal proteins were screened out after Bonferroni correction and HEIDI test (Fig. 2C,D; Supplementary Table S5). Among these, transcription factor junD (JUND), valyl-tRNA synthetase 1 (SYVC), and proteasome subunit beta type-9 (PSB9) were validated in the replication cohort. In the SMR-identified causal proteins, for each one SD increase in protein levels predicted by the gene, the odds ratios (ORs) of the causal effects ranged from 0.090 (95% CI 0.056–0.144) for PSB9 to 4.291 (95% CI 3.135–5.873) for LIF (Fig. 2E,F). For proteins from different data sources, the direction of the causal effect was consistent (Supplementary Table S6).

Causal associations between plasma proteins and AITD identified by IVW or Wald ratio

In addition to using the SMR to assess causal effects, we also applied the IVW and Wald ratio methods. For the UKB-PPP cohort, 13 plasma proteins met the significance criteria after Bonferroni correction in the discovery cohort ($p < 2.50 \times 10^{-5}$; 0.05/2001), and 10 in the replication cohort ($p < 2.53 \times 10^{-5}$; 0.05/1975) (Fig. 3A,B; Supplementary Table S7). Following the assessment of pleiotropy by utilizing the MR-Egger intercept ($p > 0.05$), no proteins were excluded. However, we observed that most identified plasma proteins had only one cis-SNP, making it impossible to evaluate their pleiotropy using the MR-Egger intercept. Therefore, the HEIDI test from the SMR analysis was employed to examine the pleiotropy of these proteins when the cis-SNP was shared in both SMR and Wald ratio methods (Supplementary Tables S4-5). For other proteins whose pleiotropy could not be measured, they were included in subsequent analysis. Finally, 9 proteins in the discovery cohort and 7 in the replication cohort were determined as causal proteins. Among these, only one protein, ribonuclease T2 (RNT2), was present in both outcomes. Similarly, for the deCODE genetics cohort, a total of 5 causal proteins were ultimately screened out (Fig. 3C,D; Supplementary Table S8). However, no common causal proteins were identified in the two outcomes. In the causal proteins identified by IVW and Wald ratio, for each one SD increase in the predicted protein levels by the gene, the ORs of the causal effects ranged from 0.580 (95% CI 0.471–0.715) for Elongin-A (ELOA1) to 2.507 (95% CI 1.715–3.665) for Phosphomannomutase 1 (PMM1) (Fig. 3E,F).

In summary, when using the same MR method, a total of 8 proteins (CD226, LIF, C1QT6, TNF11, JUND, SYVC, PSB9 and RNT2) satisfied the criteria of significance after Bonferroni correction, absence of significant pleiotropy, and consistency in the direction of causal effect in both AITD datasets. Furthermore, when considering the SMR, IVW, and Wald ratio approaches collectively, we discovered that thyrotroph embryonic factor (TEF), interleukin-7 receptor subunit alpha (IL7RA), and 3-hydroxyisobutyryl-CoA hydrolase, mitochondrial (HIBCH) also met the aforementioned criteria, despite being identified by different MR methods in the two outcomes (Supplementary Tables S4-8). Therefore, 11 plasma proteins were considered to have a causal association with AITD (Table 1).

Identification of proteins with colocalization effect

To investigate whether the identified plasma proteins and AITD might be influenced by the same genetic variations, we conducted a colocalization analysis. Among the 11 causal proteins, only 5 were proven to have a colocalization relationship with AITD (Supplementary Table S9). During the colocalization analysis between UKB-PPP and AITD outcomes, CD226, LIF, and TNF11 demonstrated strong colocalization effects in both the discovery and replication cohorts, sharing the same causal variant (PP.H4 > 0.8), with the SuSiE method suggesting that TNF11 shares multiple causal variants with AITD (Fig. 4A). Meanwhile, in the colocalization analysis using deCODE genetics and AITD data, only 2 plasma proteins (IL7RA, JUND) were identified to have colocalization effects with the outcomes, with JUND considered to share multiple causal variants (Fig. 4B).

Reverse MR analysis of AITD liability and identified causal plasma proteins

To explore whether a reverse causal relationship between the identified plasma proteins and AITD was existed, we performed the IVW, weighted median, and MR-Egger methods to estimate the causal effects, and the IVW served as the primary analysis. The selection of IVs was as described previously. Ultimately, no proteins were found to have a reverse causal association after Bonferroni correction ($p = 0.0045$; Supplementary Table S10). However, a potential reverse causal association in RNT2 was found not only in the discovery cohort (OR = 1.023, 95% CI 1.001–1.045, $p = 0.036$) but also in the replication cohort (OR = 1.032, 95% CI 1.009–1.057, $p = 0.006$). Moreover, the direction of the causal effects was consistent across all three MR methods.

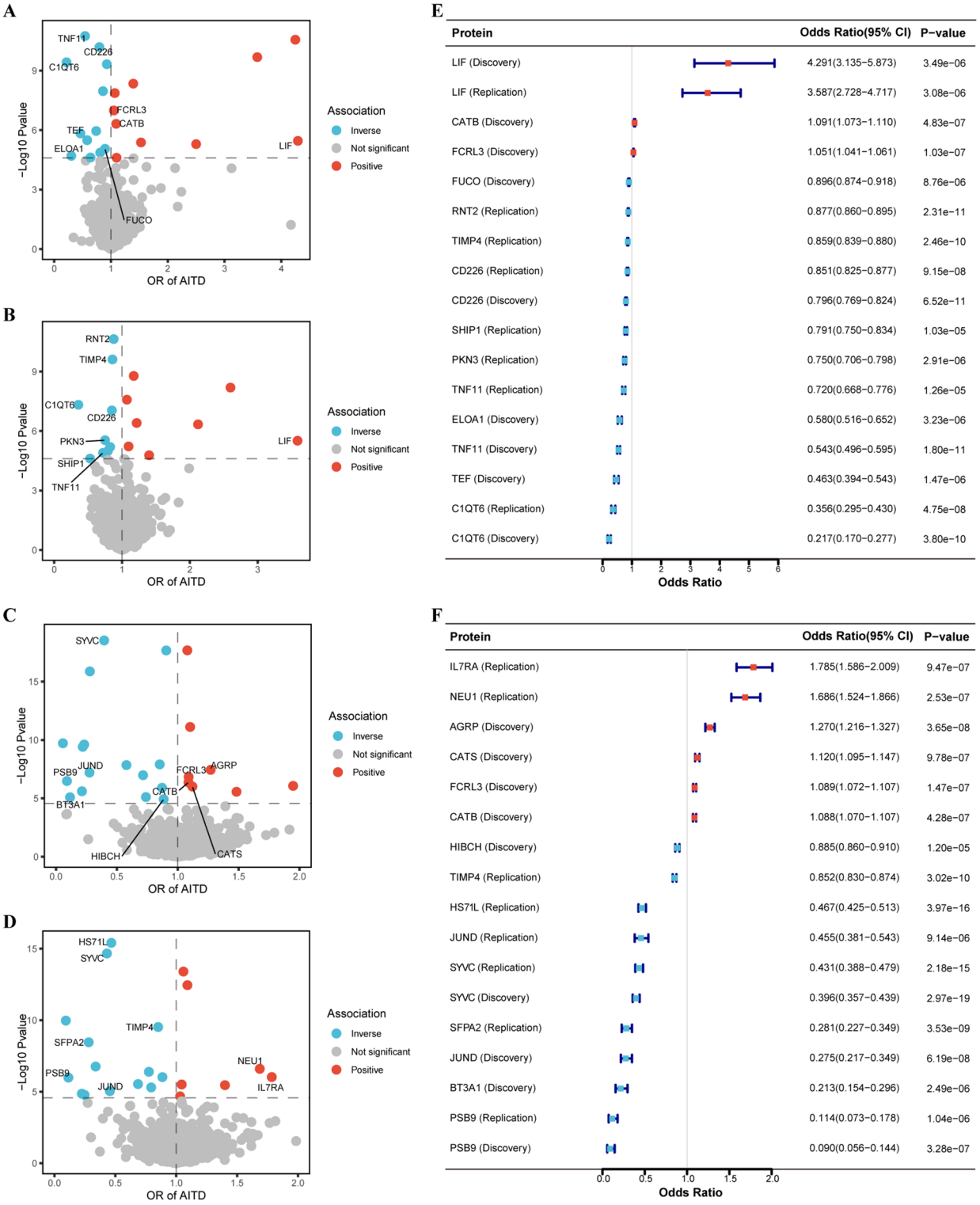

Fig. 2. Causal associations between plasma proteins and AITD identified by SMR. (**A**,**B**) Volcano plots displayed the causal plasma proteins identified in the discovery and replication cohort of AITD, using data from UKB-PPP. (**C**,**D**) Volcano plots displayed the causal plasma proteins identified in the discovery and replication cohort of AITD, using data from deCODE genetics. (**E**,**F**) Forest plots showed the odds ratio, 95% confidence interval and *p*-value of the causal plasma proteins for UKB-PPP and deCODE genetics, respectively.

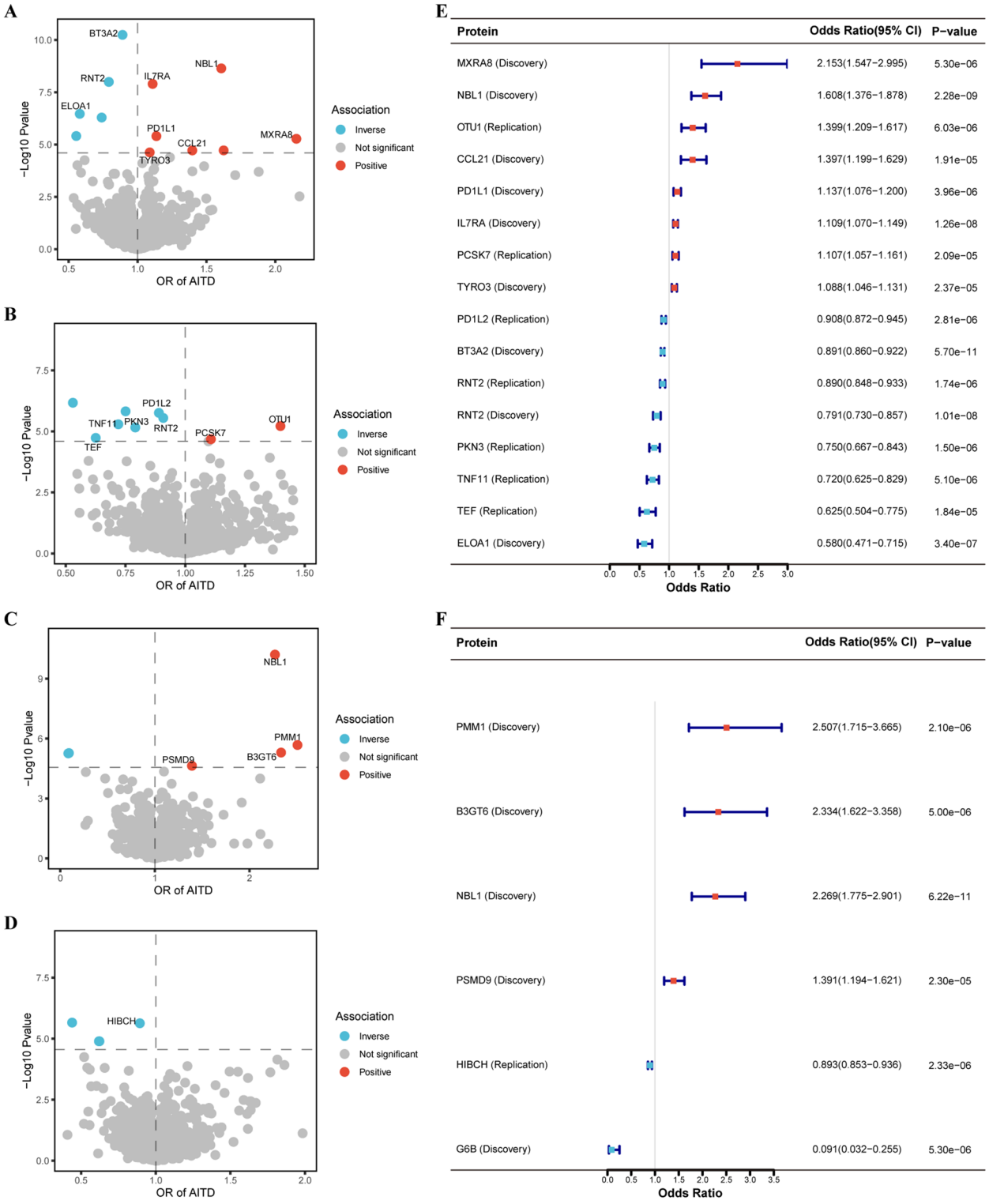

Fig. 3. Causal associations between plasma proteins and AITD identified by IVW or Wald ratio. (**A**,**B**) Volcano plots displayed the causal plasma proteins identified in the discovery and replication cohort of AITD, using data from UKB-PPP. (**C**,**D**) Volcano plots displayed the causal plasma proteins identified in the discovery and replication cohort of AITD, using data from deCODE genetics. (**E**,**F**) Forest plots showed the odds ratio, 95% confidence interval and *p*-value of the causal plasma proteins for UKB-PPP and deCODE genetics, respectively.

Gene	Protein	UniProt ID	Mendelian randomization in discovery cohort				Mendelian randomization in replication cohort			
			OR (95% CI)	P value	*P* value for HEIDI test	*P* value for MR-Egger intercept	OR (95% CI)	*P* value	*P* value for HEIDI	*P* value for MR-Egger
CD226	CD226	Q15762	0.796 (0.769, 0.824)	6.52×10^{-11}	0.591	NA	0.851 (0.825, 0.877)	9.15×10^{-8}	0.514	NA
LIF	LIF	P15018	4.291 (3.135, 5.873)	3.49×10^{-6}	0.412	NA	3.587 (2.728, 4.717)	3.08×10^{-6}	0.232	NA
C1QTNF6	C1QT6	Q9BXI9	0.217 (0.170, 0.277)	3.80×10^{-10}	0.327	NA	0.356 (0.295, 0.430)	4.75×10^{-8}	0.101	NA
TNFSF11	TNF11	O14788	0.543 (0.496, 0.595)	1.80×10^{-11}	0.175	NA	0.720 (0.668, 0.776)	1.26×10^{-5}	0.736	NA
JUND	JUND	P17535	0.275 (0.217, 0.349)	6.19×10^{-8}	0.537	NA	0.455 (0.381, 0.543)	9.14×10^{-6}	0.386	NA
VARS1	SYVC	P26640	0.396 (0.357, 0.439)	2.97×10^{-19}	0.073	NA	0.431 (0.388, 0.479)	2.18×10^{-15}	0.114	NA
PSMB9	PSB9	P28065	0.090 (0.056, 0.144)	3.28×10^{-7}	0.086	NA	0.114 (0.073, 0.178)	1.04×10^{-6}	0.189	NA
RNASET2	RNT2	O00584	0.791 (0.730, 0.857)	1.01×10^{-8}	NA	0.914	0.890 (0.848, 0.933)	1.74×10^{-6}	NA	0.476
TEF	TEF	Q10587	0.463 (0.394, 0.543)	1.47×10^{-6}	0.851	NA	0.625 (0.504, 0.775)	1.84×10^{-5}	0.206	NA
IL7R	IL7RA	P16871	1.109 (1.070, 1.149)	1.26×10^{-8}	NA	0.620	1.785 (1.586, 2.009)	9.47×10^{-7}	0.437	NA
HIBCH	HIBCH	Q6NVY1	0.885 (0.860, 0.910)	1.20×10^{-5}	NA	0.063	0.893 (0.853, 0.936)	2.33×10^{-6}	NA	0.263

Table 1. The summary results of MR analysis for 11 causal plasma proteins.

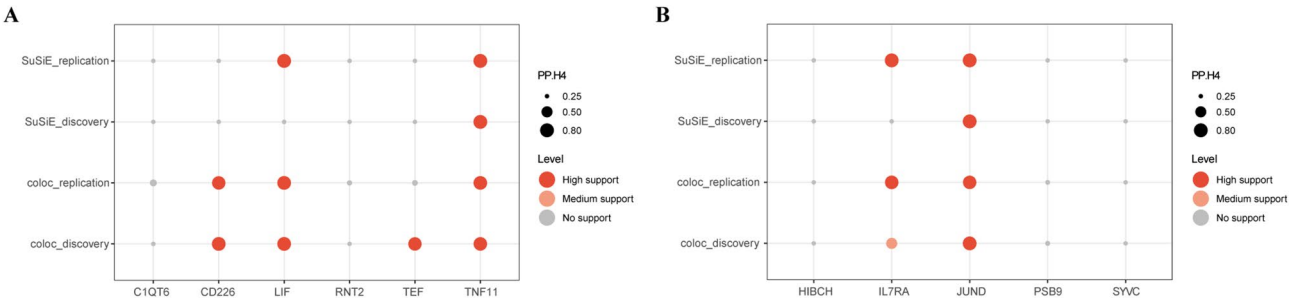

Fig. 4. Colocalization analysis between the causal plasma proteins and AITD in the discovery and replication cohorts. (**A**) Utilizing data from UKB-PPP, high support evidence for colocalization were determined in both AITD cohorts. (**B**) Utilizing data from deCODE genetics, high support evidence for colocalization were determined in both AITD cohorts.

Network and function analysis of the AITD-associated plasma proteins

Using genes from all plasma proteins that satisfied Bonferroni correction, a PPI network was constructed employing the GeneMANIA website tools (Supplementary Fig. S1A). Next, three significant PPI subunits, which had physical interactions, were identified. The PPI subunit centered by tumor necrosis factor (TNF) was associated with antigen processing and presentation (Supplementary Fig. S1B). The PPI subunit formed by butyrophilin subfamily 3 member A1 (BTN3A1), BTN3A2, and BTN3A3 was mainly related to the T cell receptor signaling pathway (Supplementary Fig. S1C). The PPI subunit comprising STAT3, IL7R, and LIF was primarily associated with the JAK-STAT signaling pathway and leukocyte differentiation (Supplementary Fig. S1D). GO enrichment analysis suggested that these AITD-associated plasma proteins were comprehensively involved in immunity, including activation of immune response, leukocyte mediated immunity, adaptive immune response, and cytokine receptor binding (Fig. 5A). The top KEGG enriched pathways were antigen processing and presentation, cytokine-cytokine receptor interaction, and natural killer cell mediated cytotoxicity (Fig. 5B), further demonstrating the relevance of AITD-associated plasma proteins to the functions of human immune system.

Druggability of the causal plasma proteins

By searching the Drugbank and ChEMBL databases, we investigated whether there are targeted drugs for the identified causal plasma proteins (Supplementary Table S11). Denosumab and Lenalidomide, targeting TNF11, have been approved for the treatment of osteoporosis and multiple myeloma, respectively. However, these drugs act by inhibiting TNF11 expression, which contradicts TNF11 being a protective factor for AITD. This suggested that the use of these drugs might increase the risk of AITD. Valine, a substrate of SYVC, is an approved health supplement. Genetically predicted levels of SYVC in plasma is a protective factor for AITD, hence the intake of Valine may be beneficial for AITD by accelerating the catalytic action of SYVC. Carfilzomib, targeting the PSB9 protein, is an approved drug for treating multiple myeloma, but its inhibition of PSB9 could increase AITD risk. Unfortunately, other causal plasma proteins do not yet have related targeted drugs.

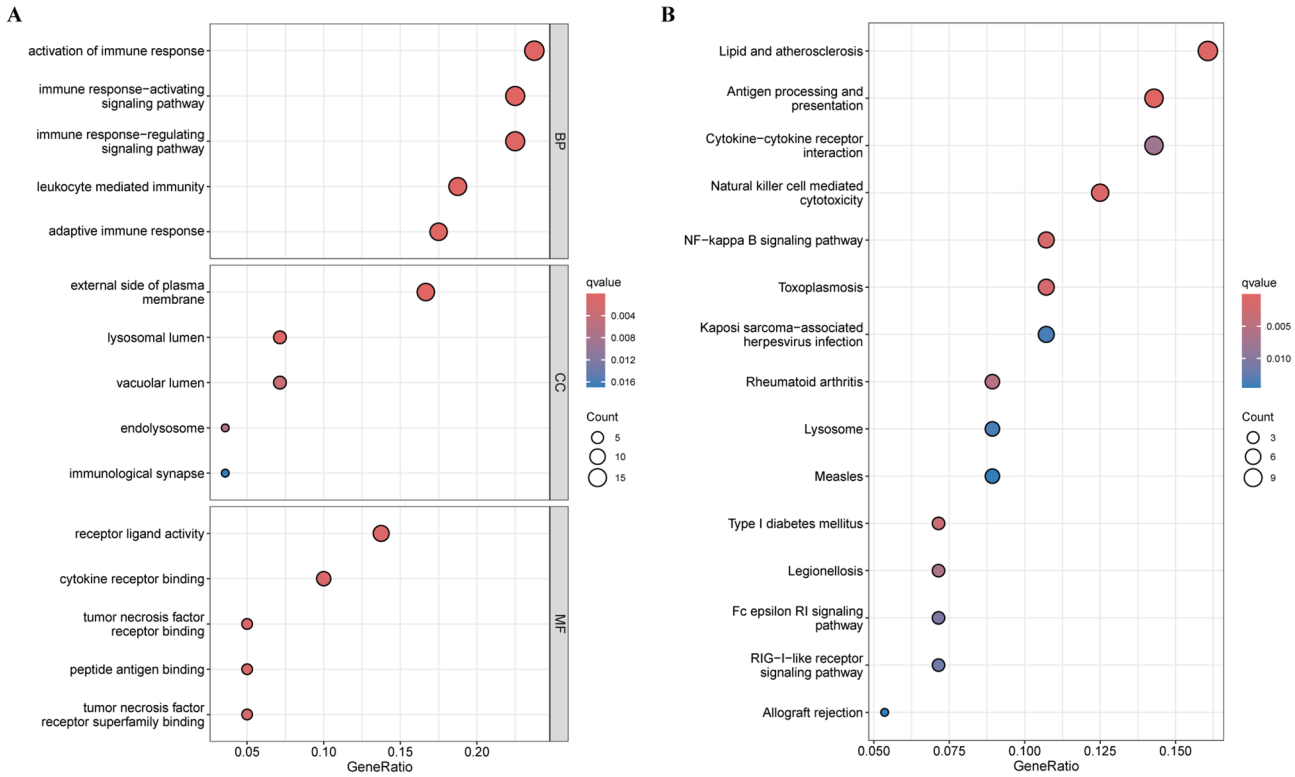

Fig. 5. Functional enrichment analysis of the AITD-associated plasma proteins. (**A**) The GO analysis showed the top enriched functions of biological process, cellular components and molecular functions. (**B**) The top 15 KEGG pathways related to the AITD-associated plasma proteins.

Phe-MR analysis of the causal plasma proteins

To evaluate whether the causal plasma proteins have beneficial or detrimental effects on other indications, we conducted a Phe-MR analysis on 784 disease phenotypes from the UK Biobank (Supplementary Table S12). Genetically predicted high levels of CD226 in plasma was associated with reduced risks of hypothyroidism, inflammatory bowel diseases, asthma, and osteoarthrosis. However, these also contributed to increased risks of heart disease, abnormal glucose, multiple myeloma, and osteopenia. Genetically predicted low levels of LIF in plasma correlated with decreased risks of hypothyroidism, Sicca syndrome, type 1 diabetes, and dementias, yet they elevated the risks of premature beats, chronic pharyngitis, and hyponatremia. Plasma TNF11 showed a negative correlation with hypothyroidism, sarcoidosis, Sicca syndrome, osteoarthrosis, and osteoporosis, but a positive correlation with duodenal ulcer, depression, and peripheral vascular disease. Elevated levels of IL7RA in plasma were linked to hypothyroidism, non-Hodgkin's lymphoma, and asthma, whereas decreased levels were associated with osteopenia, hyperpotassemia, hyperlipidemia, and hypertension. Lower levels of plasma JUND were associated with hypothyroidism and psoriatic arthropathy, while higher levels were linked to dyschromia, emphysema, obesity, and hyperlipidemia. All these protein-disease associations should be interpreted with caution as they only showed suggestive significance due to the unsatisfaction of Bonferroni-corrected *p*-value ($p = 6.38 \times 10^{-5}$; 0.05/784).

Discussion

In this proteome-wide MR study, we systematically identified plasma proteins causally associated with AITD. Eleven candidate proteins, including LIF, IL7RA, CD226, TNF11, and JUND, demonstrated robust evidence of causality across two independent datasets after rigorous sensitivity and colocalization analyses. Functional enrichment highlighted the involvement of these proteins in immune regulatory pathways, underscoring the central role of immune dysregulation in AITD pathogenesis. Notably, elevated genetically predicted levels of LIF and IL7RA were associated with an increased risk of AITD, while higher levels of CD226, TNF11, and JUND showed protective effects. These findings provide novel insights into the molecular mechanisms underlying AITD and suggest potential biomarkers and therapeutic targets, particularly among proteins with strong colocalization evidence.

In recent years, proteome-wide MR analyses have significantly contributed to advancing our understanding of disease mechanisms, biomarker discovery, and therapeutic development. For example, MR studies identifying interleukin-6 receptor (IL6R) and proprotein convertase subtilisin/kexin type 9 (PCSK9) as causal proteins have provided foundational evidence for the development of therapeutic agents targeting these molecules in cardiovascular diseases, such as tocilizumab and PCSK9 inhibitors, respectively[38,39]. These findings illustrate the translational potential of proteome-wide MR approaches in bridging genetic epidemiology and clinical

innovation. Incorporating similar strategies into the field of AITD could facilitate the discovery of novel biomarkers for early diagnosis and the identification of specific molecular targets for precision therapy, ultimately improving disease management and patient outcomes.

LIF, a member of the interleukin-6 cytokine family, is a multifunctional cytokine with significant impacts on various biological processes, including immune responses, inflammation, hematopoiesis, and embryonic development. LIF exerts its effects by binding to its specific receptor, activating JAK/STAT and other signaling pathways, thereby modulating the functions of immune cells. This action influences the differentiation and function of T cells, particularly in regulating the balance between T cell subsets, such as helper T cells and regulatory T cells, which is crucial in the development of autoimmune diseases. Additionally, LIF may affect the establishment and maintenance of immune tolerance, a vital aspect in autoimmune disorders[40,41]. Specifically, in the pathogenesis of AITD, LIF plays a significant role. It is produced within the thyroid gland by lymphocytes and thyroid follicular cells themselves, enhancing inflammatory responses by activating T and B cells, promoting antibody production and tissue damage[42]. Our MR study also confirmed a causal relationship between genetically predicted higher levels of LIF in plasma and increased AITD risk. Phe-MR suggested that LIF was not only associated with AITD but also positively correlated with the risks of Sicca syndrome and type 1 diabetes, further supporting the causality between LIF and autoimmune diseases. Sufficient evidence supported a strong colocalization between LIF and AITD, revealing potential therapeutic targets.

IL7RA, a cell surface protein predominantly expressed on lymphocytes such as T cells and B cells, activates a cascade of signal transduction pathways upon binding with its ligand IL-7. These pathways are crucial for the development and maintenance of lymphocyte survival, the formation and sustenance of memory T cells, and immune tolerance[43]. In rheumatoid arthritis, IL7RA impacts T cell function by promoting Th1 differentiation and IFN-γ secretion, enhancing T cell survival and proliferation, and facilitating T cell polarization. IL7RA also influences myeloid cells in rheumatoid arthritis, increasing their expression on monocytes or macrophages and affecting their role in inflammation and osteoclastogenesis[44]. Studies have also demonstrated that variations in IL7RA increase the risks of multiple sclerosis and neuromyelitis optica[45]. However, the mechanisms and relevance of IL7RA in AITD have not yet been explored. Utilizing MR analysis, we have revealed a positive causal relationship between genetically predicted plasma IL7RA levels and AITD risk, and a moderate colocalization effect between IL7RA and AITD. These findings hold implications for future drug research targeting AITD.

CD226 is a member of the immunoglobulin superfamily. It is highly expressed on the surface of NK cells and $CD8^+$ T cells in humans, functioning as an adhesion molecule and is important in promoting the migration, activation, proliferation, differentiation, and function of $CD8^+$ T cells[46]. CD226 plays a significant role in autoimmune diseases by influencing the function of regulatory T cells (Tregs) and their interaction with other immune cells. The expression of CD226 is negatively correlated with the suppressive function of $Foxp3^+$ Tregs. It competes with the co-inhibitory molecule T-cell immunoreceptor with Ig and ITIM domains (TIGIT) for the same ligand, impacting the regulation of immune responses[47]. Polymorphism in CD226 is linked to various autoimmune diseases including systemic lupus erythematosus, systemic sclerosis, type 1 diabetes and rheumatoid arthritis[48–51]. However, the relationship between CD226 and AITD remains unclear. Through MR analysis, we discovered that genetically predicted lower levels of CD226 in plasma were causally associated with an increased risk of AITD, aligning with previous findings on the relationship between CD226 and autoimmune diseases. Furthermore, Phe-MR also indicated a negative correlation between plasma CD226 and inflammatory bowel diseases, asthma, and osteoarthrosis. Additionally, colocalization analysis revealed a strong colocalization relationship between CD226 and AITD, suggesting a potentially therapeutic target for AITD treatment.

TNF11 plays a significant role in the immune system. It is involved in the positive regulation of T cell activation and immune response, functioning as a key regulator in immune system signaling. The interaction with its receptor RANK on T cells and dendritic cells underlines its importance in lymphoid organ development and the control of T cell-dependent immune responses. This involvement makes it crucial in immune system functioning and the body's defense mechanisms[52]. TNF11 has been linked to autoimmune diseases, particularly those affecting the musculoskeletal system. In systemic juvenile idiopathic arthritis, a study found lower levels of TNF11 in active disease compared to inactive disease[53], which was consistent with the causal relationship between genetically predicted levels of TNF11 in plasma and AITD. Phe-MR analysis suggested that plasma TNF11 had negative causal associations with sarcoidosis, Sicca syndrome, osteoarthrosis, and osteoporosis, underscoring the crucial role of TNF11 in autoimmune diseases. Colocalization analysis confirmed that the association between TNF11 and AITD was driven by linkage disequilibrium, thereby positioning TNF11 as a potential therapeutic target for AITD.

JUND, a member of the Jun family of transcription factors, plays a significant role in immunity, particularly in the context of Th17 cell differentiation and function. JUND's binding pattern in Th17 cells overlaps significantly with other key transcription factors like BATF and IRF4, highlighting its role in the activation and regulation of genes essential for Th17 cell identity and function[54]. In a bioinformatic analysis integrating bulk and single-cell RNA-sequencing, it was found that a specific gene module in $CD16^+$ monocytes, which included JUND, was significantly different in systemic sclerosis patients compared to healthy controls, suggesting a potential role in the pathogenesis of the disease[55]. However, there are few studies reporting on the relationship between JUND and AITD. Through MR analysis, we found that a decrease in genetically predicted plasma JUND was causally associated with an increased risk of AITD. Colocalization analysis suggested that JUND shared the same causal variant with AITD. Consequently, whether JUND can serve as a therapeutic target for AITD and the underlying mechanisms with AITD warrant further investigation.

Recent studies utilizing similar proteome-wide MR approaches have investigated the genetic and proteomic landscapes of thyroid cancer and hyperthyroidism. For instance, a proteome-wide MR analysis of thyroid cancer identified 26 circulating proteins with putative causal effects, with enrichment in amino acid and organic acid metabolism pathways[56]. These findings suggest that metabolic dysregulation plays a central role

in thyroid cancer pathogenesis. In contrast, another study focusing on hyperthyroidism highlighted the role of immune cell phenotypes, particularly CD25 on naive-mature B cells and CD8 + NKT cells, with thymol sulfate identified as a key metabolic mediator influencing immune cell evolution and disease risk[57]. Compared to these findings, our study emphasizes the contribution of immune regulatory proteins such as LIF, IL7RA, and CD226 to the development of AITD, underscoring the pivotal role of immune dysregulation in autoimmune thyroid diseases. Although certain immune-related genes, such as TNFAIP3 and TIGIT, have been implicated across different thyroid disorders, the distinct sets of causal proteins identified in AITD highlight its unique autoimmune pathogenesis compared to the metabolic and cellular mechanisms underpinning thyroid cancer and hyperthyroidism. This comparative insight strengthens the understanding that different thyroid diseases, while anatomically related, are driven by divergent molecular mechanisms.

One strength of this study is the use of MR and colocalization analyses, utilizing cis-genetic instruments to estimate the causal effects of plasma proteins on AITD, thereby minimizing biases caused by confounding and reverse causation. Additionally, we employed two large-scale summary-level GWAS datasets, enhancing the reliability of our results. However, understanding the limitations of this study aids in interpreting the results correctly. First, the analysis was limited to European ancestry, restricting the generalizability of our findings to other populations. Second, the proteomic data included only British and Icelandic individuals, whose genetic backgrounds may differ from other European populations, potentially introducing bias in the MR analysis. Third, although the majority of patients in the two independent AITD datasets were primarily autoimmune hypothyroidism, the AITD GWAS meta-analysis dataset still included a small number of Graves' disease and Hashimoto's thyroiditis patients, which might bias the results. Fourth, we employed stringent Bonferroni correction and required significant causal associations and colocalization effects in both the discovery and replication cohorts. This might lead to neglecting some subtle causal relationships. Fifth, to reduce horizontal pleiotropy, we used cis-SNPs as IVs, potentially ignoring the impact of trans-regulation on the proteins, which might be crucial for some protein-coding genes. Sixth, in this study, some plasma proteins lacked suitable genetic instruments for MR analysis, possibly overlooking some meaningful proteins. Seventh, the sample overlap between the exposure and outcome may bias the causal effect to some extent. Last but not least, these causal associations were solely based on MR analysis and thus require further validation through population-based studies.

In conclusion, this proteome-wide MR study identified 11 plasma proteins with robust causal associations to AITD, among which LIF, IL7RA, CD226, TNF11, and JUND emerged as key candidates supported by colocalization evidence. These proteins are predominantly involved in immune regulatory processes, such as promoting T cell differentiation, sustaining immune tolerance, and modulating cytokine signaling, thereby highlighting immune dysregulation as a central mechanism in AITD pathogenesis. Our findings provide novel insights into the molecular underpinnings of AITD and suggest promising directions for biomarker development and therapeutic targeting. Nevertheless, further experimental and clinical investigations are warranted to validate these causal associations and to explore the feasibility of targeting these proteins for early diagnosis and precision therapy in AITD.

Data availability

The datasets used and provided here are available for download from various digital archives. Both the repository(s) and accession number(s) are listed in the article/supplementary materials.

Received: 20 September 2024; Accepted: 29 May 2025
Published online: 06 June 2025

References

1. Mammen, J. S. R. & Cappola, A. R. Autoimmune thyroid disease in women. *JAMA* **325**, 2392–2393. https://doi.org/10.1001/jama.2020.22196 (2021).
2. Antonelli, A., Ferrari, S. M., Corrado, A., Di Domenicantonio, A. & Fallahi, P. Autoimmune thyroid disorders. *Autoimmun. Rev.* **14**, 174–180. https://doi.org/10.1016/j.autrev.2014.10.016 (2015).
3. Hwangbo, Y. & Park, Y. J. Genome-wide association studies of autoimmune thyroid diseases, thyroid function, and thyroid cancer. *Endocrinol. Metab. (Seoul)* **33**, 175–184. https://doi.org/10.3803/EnM.2018.33.2.175 (2018).
4. McLeod, D. S. & Cooper, D. S. The incidence and prevalence of thyroid autoimmunity. *Endocrine* **42**, 252–265. https://doi.org/10.1007/s12020-012-9703-2 (2012).
5. Caturegli, P., De Remigis, A. & Rose, N. R. Hashimoto thyroiditis: Clinical and diagnostic criteria. *Autoimmun. Rev.* **13**, 391–397. https://doi.org/10.1016/j.autrev.2014.01.007 (2014).
6. Chiovato, L., Magri, F. & Carle, A. Hypothyroidism in context: Where we've been and where we're going. *Adv. Ther.* **36**, 47–58. https://doi.org/10.1007/s12325-019-01080-8 (2019).
7. Pisetsky, D. S. Pathogenesis of autoimmune disease. *Nat. Rev. Nephrol.* **19**, 509–524. https://doi.org/10.1038/s41581-023-00720-1 (2023).
8. Effraimidis, G. & Wiersinga, W. M. Mechanisms in endocrinology: Autoimmune thyroid disease: old and new players. *Eur. J. Endocrinol.* **170**, R241-252. https://doi.org/10.1530/EJE-14-0047 (2014).
9. Vargas-Uricoechea, H. Molecular mechanisms in autoimmune thyroid disease. *Cells* **12**, 918. https://doi.org/10.3390/cells12060918 (2023).
10. Stan, M. N. et al. Randomized controlled trial of rituximab in patients with Graves' orbitopathy. *J. Clin. Endocrinol. Metab.* **100**, 432–441. https://doi.org/10.1210/jc.2014-2572 (2015).
11. Smith, T. J. et al. Teprotumumab for thyroid-associated ophthalmopathy. *N. Engl. J. Med.* **376**, 1748–1761. https://doi.org/10.1056/NEJMoa1614949 (2017).
12. Davey Smith, G. & Hemani, G. Mendelian randomization: Genetic anchors for causal inference in epidemiological studies. *Hum. Mol. Genet.* **23**, r89-98. https://doi.org/10.1093/hmg/ddu328 (2014).
13. Sun, B. B. et al. Plasma proteomic associations with genetics and health in the UK Biobank. *Nature* **622**, 329–338. https://doi.org/10.1038/s41586-023-06592-6 (2023).

14. Ferkingstad, E. et al. Large-scale integration of the plasma proteome with genetics and disease. *Nat. Genet.* **53**, 1712–1721. https://doi.org/10.1038/s41588-021-00978-w (2021).
15. Finan, C. et al. The druggable genome and support for target identification and validation in drug development. *Sci. Transl. Med.* **9**, 383. https://doi.org/10.1126/scitranslmed.aag1166 (2017).
16. Chen, J. et al. Therapeutic targets for inflammatory bowel disease: Proteome-wide Mendelian randomization and colocalization analyses. *EBioMedicine* **89**, 104494. https://doi.org/10.1016/j.ebiom.2023.104494 (2023).
17. Yuan, S. et al. Plasma proteins and onset of type 2 diabetes and diabetic complications: Proteome-wide Mendelian randomization and colocalization analyses. *Cell Rep. Med.* **4**, 101174. https://doi.org/10.1016/j.xcrm.2023.101174 (2023).
18. Zheng, J. et al. Phenome-wide Mendelian randomization mapping the influence of the plasma proteome on complex diseases. *Nat. Genet.* **52**, 1122–1131. https://doi.org/10.1038/s41588-020-0682-6 (2020).
19. Saevarsdottir, S. et al. 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 (2020).
20. Kurki, M. I. et al. FinnGen provides genetic insights from a well-phenotyped isolated population. *Nature* **613**, 508–518. https://doi.org/10.1038/s41586-022-05473-8 (2023).
21. Burgess, S., Davies, N. M. & Thompson, S. G. Bias due to participant overlap in two-sample Mendelian randomization. *Genet. Epidemiol.* **40**, 597–608. https://doi.org/10.1002/gepi.21998 (2016).
22. Bowden, J. et al. Improving the visualization, interpretation and analysis of two-sample summary data Mendelian randomization via the radial plot and radial regression. *Int. J. Epidemiol.* **47**, 1264–1278. https://doi.org/10.1093/ije/dyy101 (2018).
23. Zhu, Z. et al. 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 (2016).
24. Greco, M. F., Minelli, C., Sheehan, N. A. & Thompson, J. R. Detecting pleiotropy in Mendelian randomisation studies with summary data and a continuous outcome. *Stat. Med.* **34**, 2926–2940. https://doi.org/10.1002/sim.6522 (2015).
25. Bowden, J., Davey Smith, G. & Burgess, S. Mendelian randomization with invalid instruments: Effect estimation and bias detection through Egger regression. *Int. J. Epidemiol.* **44**, 512–525. https://doi.org/10.1093/ije/dyv080 (2015).
26. Giambartolomei, C. et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet.* **10**, e1004383. https://doi.org/10.1371/journal.pgen.1004383 (2014).
27. Wallace, C. A more accurate method for colocalisation analysis allowing for multiple causal variants. *PLoS Genet.* **17**, e1009440. https://doi.org/10.1371/journal.pgen.1009440 (2021).
28. Bowden, J., Davey Smith, G., Haycock, P. C. & Burgess, S. 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 (2016).
29. Warde-Farley, D. et al. The GeneMANIA prediction server: Biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res.* **38**, W214-220. https://doi.org/10.1093/nar/gkq537 (2010).
30. Knox, C. et al. DrugBank 6.0: The DrugBank knowledgebase for 2024. *Nucleic Acids Res.* https://doi.org/10.1093/nar/gkad976 (2023).
31. Mendez, D. et al. ChEMBL: Towards direct deposition of bioassay data. *Nucleic Acids Res.* **47**, D930–D940. https://doi.org/10.1093/nar/gky1075 (2019).
32. Denny, J. C. et al. Systematic comparison of phenome-wide association study of electronic medical record data and genome-wide association study data. *Nat. Biotechnol.* **31**, 1102–1110. https://doi.org/10.1038/nbt.2749 (2013).
33. Gagliano Taliun, S. A. et al. Exploring and visualizing large-scale genetic associations by using PheWeb. *Nat. Genet.* **52**, 550–552. https://doi.org/10.1038/s41588-020-0622-5 (2020).
34. Zhou, W. et al. 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 (2018).
35. Hemani, G. et al. The MR-base platform supports systematic causal inference across the human phenome. *Elife* https://doi.org/10.7554/eLife.34408 (2018).
36. Wu, T. et al. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation (Camb.)* **2**, 100141. https://doi.org/10.1016/j.xinn.2021.100141 (2021).
37. Purcell, S. et al. PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575. https://doi.org/10.1086/519795 (2007).
38. Collaboration, I. R. G. C. E. R. F. et al. Interleukin-6 receptor pathways in coronary heart disease: a collaborative meta-analysis of 82 studies. *Lancet* **379**, 1205–1213, https://doi.org/10.1016/S0140-6736(11)61931-4 (2012).
39. Ference, B. A., Majeed, F., Penumetcha, R., Flack, J. M. & Brook, R. D. Effect of naturally random allocation to lower low-density lipoprotein cholesterol on the risk of coronary heart disease mediated by polymorphisms in NPC1L1, HMGCR, or both: a 2 x 2 factorial Mendelian randomization study. *J. Am. Coll. Cardiol.* **65**, 1552–1561. https://doi.org/10.1016/j.jacc.2015.02.020 (2015).
40. O'Shea, J. J., Ma, A. & Lipsky, P. Cytokines and autoimmunity. *Nat. Rev. Immunol.* **2**, 37–45. https://doi.org/10.1038/nri702 (2002).
41. Metcalfe, S. M. LIF in the regulation of T-cell fate and as a potential therapeutic. *Genes Immun.* **12**, 157–168. https://doi.org/10.1038/gene.2011.9 (2011).
42. Ajjan, R. A. & Weetman, A. P. Cytokines in thyroid autoimmunity. *Autoimmunity* **36**, 351–359. https://doi.org/10.1080/08916930310001603046 (2003).
43. Chen, D., Tang, T. X., Deng, H., Yang, X. P. & Tang, Z. H. Interleukin-7 biology and its effects on immune cells: Mediator of generation, differentiation, survival, and homeostasis. *Front. Immunol.* **12**, 747324. https://doi.org/10.3389/fimmu.2021.747324 (2021).
44. Meyer, A., Parmar, P. J. & Shahrara, S. Significance of IL-7 and IL-7R in RA and autoimmunity. *Autoimmun. Rev.* **21**, 103120. https://doi.org/10.1016/j.autrev.2022.103120 (2022).
45. Kim, J. Y. et al. Association analysis of IL7R polymorphisms with inflammatory demyelinating diseases. *Mol. Med. Rep.* **9**, 737–743. https://doi.org/10.3892/mmr.2013.1863 (2014).
46. Huang, Z., Qi, G., Miller, J. S. & Zheng, S. G. CD226: An emerging role in immunologic diseases. *Front. Cell Dev. Biol.* **8**, 564. https://doi.org/10.3389/fcell.2020.00564 (2020).
47. Lozano, E., Dominguez-Villar, M., Kuchroo, V. & Hafler, D. A. The TIGIT/CD226 axis regulates human T cell function. *J Immunol* **188**, 3869–3875. https://doi.org/10.4049/jimmunol.1103627 (2012).
48. Elhai, M. et al. Targeting CD226/DNAX accessory molecule-1 (DNAM-1) in collagen-induced arthritis mouse models. *J. Inflamm. (Lond.)* **12**, 9. https://doi.org/10.1186/s12950-015-0056-5 (2015).
49. Du, Y. et al. Association of the CD226 single nucleotide polymorphism with systemic lupus erythematosus in the Chinese Han population. *Tissue Antigens* **77**, 65–67. https://doi.org/10.1111/j.1399-0039.2010.01568.x (2011).
50. Avouac, J. et al. Critical role of the adhesion receptor DNAX accessory molecule-1 (DNAM-1) in the development of inflammation-driven dermal fibrosis in a mouse model of systemic sclerosis. *Ann. Rheum. Dis.* **72**, 1089–1098. https://doi.org/10.1136/annrheumdis-2012-201759 (2013).
51. Mattana, T. C. et al. CD226 rs763361 is associated with the susceptibility to type 1 diabetes and greater frequency of GAD65 autoantibody in a Brazilian cohort. *Mediat. Inflamm.* **2014**, 694948. https://doi.org/10.1155/2014/694948 (2014).
52. Luan, X. et al. Crystal structure of human RANKL complexed with its decoy receptor osteoprotegerin. *J. Immunol.* **189**, 245–252. https://doi.org/10.4049/jimmunol.1103387 (2012).
53. Qu, H. et al. Immunoprofiling of active and inactive systemic juvenile idiopathic arthritis reveals distinct biomarkers: A single-center study. *Pediatr. Rheumatol. Online J.* **19**, 173. https://doi.org/10.1186/s12969-021-00660-9 (2021).

54. Carr, T. M., Wheaton, J. D., Houtz, G. M. & Ciofani, M. JunB promotes Th17 cell identity and restrains alternative CD4(+) T-cell programs during inflammation. *Nat. Commun.* **8**, 301. https://doi.org/10.1038/s41467-017-00380-3 (2017).
55. Kobayashi, S. et al. Integrated bulk and single-cell RNA-sequencing identified disease-relevant monocytes and a gene network module underlying systemic sclerosis. *J. Autoimmun.* **116**, 102547. https://doi.org/10.1016/j.jaut.2020.102547 (2021).
56. Fan, Q. et al. Assessment of circulating proteins in thyroid cancer: Proteome-wide Mendelian randomization and colocalization analysis. *iScience* **27**, 109961. https://doi.org/10.1016/j.isci.2024.109961 (2024).
57. Li, Y., Song, X., Huang, Y., Zhou, S. & Zhong, L. Genetic associations of plasma metabolites with immune cells in hyperthyroidism revealed by Mendelian randomization and GWAS-sc-eQTLs xQTLbiolinks analysis. *Sci. Rep.* **15**, 1377. https://doi.org/10.1038/s41598-025-85664-1 (2025).

Acknowledgements

We extend our sincere gratitude to the UK Biobank Pharma Proteomics Project (UKB-PPP), deCODE genetics, and the FinnGen database for their invaluable contributions to this study. We also wish to acknowledge the tireless efforts of all the staff and researchers involved in these projects. Additionally, our heartfelt thanks go to all the patients who participated in these studies, whose involvement has been crucial in enhancing our understanding of autoimmune thyroid diseases. Their willingness to contribute to medical science is deeply appreciated.

Author contributions

Conceptualization, Y.L., W.Z., Y.C. and Q.Z.; methodology, Y.L., W.Z., Y.C. and Q.Z.; software, Y.L., Q.K. and Y.Z.; validation, P.Y., S.W., C.L., Y.Z. and Y.L.; formal analysis, Y.L., W.Z. and Y.C.; investigation, Q.K., Y.Z., P.Y., S.W. and C.L.; resources, C.L., Y.Z., S.W., P.Y. and Q.K.; data curation, Q.K., Y.Z., P.Y., S.W. and C.L.; writing—original draft preparation, Y.L., W.Z., Y.C, Q.K., Y.Z., P.Y., S.W., C.L., Y.Z. and Q.Z.; writing—review and editing, Y.L., W.Z., Y.C, Q.K., Y.Z., P.Y., S.W., C.L., Y.Z. and Q.Z.; visualization, W.Z., Y.C., Q.K., Y.Z. and P.Y.; supervision, C.L.; project administration, Y.Z. All authors have read and agreed to the published version of the manuscript.

Funding

This research received no external funding.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-04902-8.

Correspondence and requests for materials should be addressed to C.L., Y.Z. or Q.Z.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.