Functional Insights into Hypothyroidism Etiology through Complementary Genetic Association Methods

Roei Zucker
Hebrew University of Jerusalem
Michael Kovalerchik
Hebrew University of Jerusalem
Amos Stern
Hebrew University of Jerusalem
Hadasa Kaufman
Hebrew University of Jerusalem
Michal Linial
michall@cc.huji.ac.il

Hebrew University of Jerusalem

Research Article

Keywords: UK-Biobank, GWAS, Hashimoto's thyroiditis, Open Targets, Genotyping, PWAS, Congenital hypothyroidism, FinnGen

Posted Date: November 29th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-3658051/v1

Additional Declarations: No competing interests reported.

Abstract

Background

Hypothyroidism is a common endocrine disorder that increases with age. The disease manifests itself when the thyroid gland fails to produce sufficient thyroid hormones. The disorder includes cases of congenital hypothyroidism (CH) due to thyroid development abnormalities. However, the majority of cases in the developed world derive from dysregulation of the hormonal feedback of the pituitary gland and the destruction of the thyroid gland by autoantibodies. In this study, we sought to identify hypothyroidism causal genes by applying a diverse collection of genome association studies to large populations.

Methods

The study used the UK-Biobank (UKB) database to report 13,687 cases of European ancestry and approximately 260,000 controls. To identify the associated variants, GWAS and coding-GWAS protocols were used. In addition, applying the complementary genetic association methods of PWAS (proteome-based) and TWAS (transcriptome-based) revealed hypothyroidism-associated genes. The prevalence among the affected population was 7.5% and 2.0% for the female and male groups, respectively. We further developed a risk prediction model through sex stratification.

Results

Comparing GWAS summary statistics revealed the CH developmental program. The gene-based PWAS method identified 77 statistically significant genes. Most of these genes are located within the Chr6 MHC locus and are enriched with autoimmunity-related genes. Comparing GWAS and TWAS revealed multiple facets of the etiology of hypothyroidism. Most notably, thyroid developmental programs and dysregulation of hormone secretion capacity in the thyroid. Despite a 3.6-fold higher prevalence in females relative to males for hypothyroidism, using a permutation approach, we found no sex-dependent genetic effect, with 98% of the associated genes being identical between the sexes. The prediction of the polygenic risk score (PRS) for hypothyroidism is mostly derived from the female affected group.

Conclusions

This study highlights the importance of synthesizing complementary genome-wide association methods for this complex disease. We conclude that the integration of established association methods can improve interpretability and clinical utility.

Background

Hypothyroidism is a disorder of the endocrine system in which the thyroid gland does not produce enough hormones or when the thyroid hormones act inadequately in target tissues (Almandoz and Gharib 2012). Hypothyroidism (EFO: 0004705) and its extreme condition, myxedema (EFO: 1001055), are signified by impairment in the function of the thyroid (Chaker et al. 2022). The thyroid gland is crucial to the metabolism of all tissues and the early development of the central nervous system (CNS) (Patel et al. 2011). While over 10% of the world's population exhibits some level of iodine deficiency that may lead to hypothyroidism, it does not apply to the developed world (Taylor et al. 2018). In the USA, the prevalence of hypothyroidism has been shown to steadily increase over the last two decades, reaching 14.4% (clinical and preclinical) (Wyne et al. 2022). Subclinical hypothyroidism accounts for 4–8% of the population (Fatourechi 2009). It is estimated

that 1 in 8 people will develop a functional deficiency of the thyroid in their lifetime, with a 3-4-fold higher likelihood of females relative to males (Chung 2014; Luboshitzky et al. 1996; Unnikrishnan et al. 2013; Uzunlulu et al. 2007).

Primary hypothyroidism is defined by a failure of the thyroid gland itself. In contrast, secondary and tertiary hypothyroidism are caused by dysfunction of the pituitary and hypothalamus glands, respectively (Chaker et al. 2022). The diagnosis of hypothyroidism is determined by free and bound thyroid hormones in the blood, the level of TSH, and the composition of autoantibodies to thyroid markers (Evered et al. 1973). Specifically, autoantibodies against thyroid-specific antigens (e.g., TSHR, TG, and TPO) were found in most patients (Brown 2013). The majority of these cases can be assigned to hypothyroidism with an autoimmunity component (e.g., Hashimoto's thyroiditis and autoimmune hypothyroidism, Ord's thyroiditis) (Eriksson et al. 2012). Importantly, hypothyroidism is linked to a higher incidence of other organ-specific autoimmune diseases (Brown 2013; Matzaraki et al. 2017). Hyperthyroidism occurs in children in the form of autoimmune thyroiditis (AIT) (Brown 2013). AIT reflects some unknown defects in immunoregulation, which translate into injury to thyroid tissue, which in turn activates apoptotic cell death and thyroiditis. The genetic basis for AIT is unknown, but it is likely to combine genetics (estimated to account for 70% of cases) and environmental factors that interact with predisposed genetics.

Hyperthyroidism may also be congenital, where the incidence rate is 1 in every 2000–4000 live births. Congenital hypothyroidism (CH) is a developmental abnormality affecting the hypothalamic-pituitary-thyroid (HPT) axis (Persani et al. 2018). Primary CH, which is associated with a missing or underdeveloped thyroid (dysgenesis), is the most common neonatal disease and accounts for most CH. Most cases of CH occur sporadically and are frequently associated with an increase in neonatal malformations, which can result in further complications (Léger et al. 2014). Unfortunately, the genetics of thyroid dysgenesis are resolved in only 5% of cases (Wassner 2020). A systematic CH screen in Japanese (Narumi et al. 2010) and Czech (Al Taji et al. 2007) individuals confirmed the challenge of identifying causal mutations. While the most pathogenic variants of the TSH receptor (TSHR) are nonsyndromic, mutated Gsα (GNAS1) and PDE8B, which are components of TSHR signaling, are linked with syndromic disease (Persani et al. 2010). Candidate genes that potentially disrupt thyroid gland formation have been linked to other rare monogenic diseases (reviewed in (Stoupa et al. 2021 32)). Mutations in numerous thyroid transcription factors (TITF-1, TITF-2, PAX-8, FOXE1, GLIS3) are mostly syndromic (Kostopoulou et al. 2021). Additionally, mutated genes that act in the biosynthesis and cell biology of thyroid hormones (Panicker 2011) may cause dysfunction of thyroid hormone synthesis and secretion (dyshormonogenesis). Among these genes are thyroid peroxidase (TPO), thyroglobulin (TG), sodium iodide symporter (NIS), pendrin (PDS), thyroid oxidase 2 (THOX2), and iodotyrosine deiodinase (IYD). The iodothyronine transporter (MCT8), which is expressed in the thyroid gland membrane, was also shown to drive hypothyroidism, which is coupled to neurological deficits (Park and Chatterjee 2005). Dyshormonogenetic cases are often recessively inherited (Makretskaya et al. 2018). Interestingly, the occurrence of mutated CH causal genes differs substantially across populations (Sun et al. 2018).

In this study, we analyzed the genetic signatures among people diagnosed with ICD-10 E03 ("other hypothyroidism") (Edwards et al. 2013). We asked whether the genetic effects of hypothyroidism and myxedema are associated with sex. To this end, we applied several association methods, most notably the proteome-wide association study (PWAS) method, which detects gene-phenotype associations through the effect of variants on protein function (Brandes et al. 2020). By comparing results from the PWAS (Brandes et al. 2020), TWAS (Luningham et al. 2020), classical GWAS, and coding GWAS, we shed light on the complex etiologies of hypothyroidism with or without an immunological basis. We conclude that the integration of established association methods and partitioning the population by sex can improve interpretability and clinical utility.

Methods

UKB processing

The UK Biobank (UKB) is a population-based database with detailed medical, genotyping, and lifestyle information covering ~ 500k people aged 40–69 across the UK who were recruited from 2006–2010. The analyses herein were based on the 2019 UKB release. We restricted the analysis to European origin (codes 1, 1001, 1002, and 1003, respectively; and ethnic background, data field 21000). We applied the classification according to genetic ancestry (Genetic ethnic group, data field 22006). We further removed genetic relatives by randomly keeping only one representative of each kinship group.

Hypothyroidism is indexed by ICD-10 code E03. The analysis includes individuals who have any diagnosis within the main or secondary codes (UKB data fields 41202 and 41204, respectively) and the summary diagnosis code 41270. The latter covers the distinct diagnosis codes a participant has recorded across all their hospital inpatient records, in either the primary or secondary position. These fields cover ICD-10 from E03.0 to E03.9 (total 29,478 participants), with 98.5% of them marked as "unspecified hypothyroidism", "other specified hypothyroidism" (E03.8, 0.7%), "hypothyroidism due to medicaments and other exogenous substances" (0.3%, E03.2), and CH (0.3%, E03.0-E03.1). We included "other hypothyroidism" and excluded "iodine-deficiency-related hypothyroidism" (E00-E02) and postprocedural hypothyroidism (E89.0). This set of E03 with genotyping data includes 2,557 males and 11,094 females.

All GWAS and coding GWAS analyses

We used UKB processing as described above to perform GWAS and coding GWAS. UKB released genotyped data for all participants. In the genotyping data, there are ~ 820k preselected genetic variations (from the genotyping data of the UKB Axiome Array). Based on the UKB imputation protocol, the number of variants was expanded to 97,013,422. For the imputed variants, we calculated the probabilistic expectation for the alternative alleles (Brandes et al. 2020). For all GWAS, we considered variants with an MAF threshold > 0.001 and a Hardy-Weinberg equilibrium (HWE) exact test p value > 1e-6. The SNPs retained for the analysis were those with a 100% call rate (using the geno option). For the E03 GWAS, we had 12,435 cases and 257,948 controls.

The coding GWAS analysis includes the human proteome according to UniProt-SwissProt (labeled “reviewed”). Due to unambiguous mapping to RefSeq gene names, we cover 18,053 protein-coding genes of the ~ 20k proteins that are listed for the human proteome. The missed genes are instances where the mapping of genes to proteins is ambiguous (e.g., genes with multiple protein versions). In addition, no mapping is available to nongenuine RefSeq genes such as endogenous retroviruses (e.g., ERVK) or the constant and variable regions of immunoglobulins (e.g., IGLV and IGKV). The 2,119 UniProtKB-reviewed gene names that are not included in the PWAS analysis are listed in Additional file 1: **Table S1**. Altogether, from the overall ~ 97 M variants, we tested all 639,323 coding and splicing variants located within 18,053 protein-coding genes. For all GWAS and coding GWAS included 172 covariates that include sex (binary), year of birth (numeric), 40 principal components (PCs) that capture the ancestry stratification (numeric), the UKB genotyping batch (one-hot-encoding, 105 categories), and the UKB assessment centers associated with each sample (binary, 25 categories).

GWAS summary statistics

We used the Open Targets Genetics (OTG) platform to select current knowledge and GWAS results on hypothyroidism (Ghoussaini et al. 2021). The OTG (release date: 6/2023) unifies multiple sources of evidence for an inclusive list of 2007 genes, each ranked by an OT global score (range 0–1.0). Among these genes, 702 genes are supported by genetic-associated (GA) scores based on large-scale independent GWAS summary statistics (Carvalho-Silva et al. 2019). We extracted the list of associated variants from OTG for comparative analysis from six large-scale GWAS for associated phenotypes. The cohorts from these GWAS ranged from 250k to 580k controls and 15k to 33k cases. The number of independent loci in each GWAS is 115 (Neale v2, 2018), 45 (UKB Saige 2018), 65 (Sakaue et al. 2021), 160 (Donertas et al. 2021), 133 (Kichaev et al. 2019) and 67 (FinnGen, Freeze 6, 2022). Note that the study of Neale’s lab (2018, V2) focused on European ancestry with 361,141 participants, among whom 17,574 were labeled with "hypothyroidism/myxedema (self-reported, noncancer)", while other GWAS from UKB were performed across multiple ethnic groups (Donertas et al.

2021). Other datasets from the OTG platform include "permanent congenital hypothyroidism" with 53 associated genes, 35 of which have a GA score > 0.5. This phenotype is a merger of Orphanet: 442 (23 genes) and EFO 0016408. Most of the associated genes were derived from ClinVar and Orphanet (Sharo et al. 2023).

PWAS functional effect score per gene of the human proteome

The PWAS methodology (Brandes et al. 2019) assumes that causal variants in coding regions affect phenotypes by altering the biochemical functions of the encoded protein of a gene. In summary, the functional impact rating at the molecular level (FIRM) from the pretrained machine-learning (ML) model is then used to estimate the extent of the damage caused to each protein in the entire proteome (Brandes et al. 2019). FIRM performance was reported and validated for the pathological variants in ClinVar, reaching an AUC of 90% and accuracy of 82.7% (Brandes et al. 2019). The predicted effect score of a variant is a number between 0 (complete loss of function, LoF) and 1 (no functional effect, synonymous variant). PWAS explicitly treated in-frame indels (Brandes et al. 2020). We seek a calibrated score for the overall protein damage at an individual level. Thus, per-variant damage predictions are aggregated at the gene level according to recessive, dominant and hybrid gene heritability modes. On average, there are 35.4 nonsense and missense mutations per gene that are considered for the gene-based effect score. PWAS results are based on the same set of variants as used for the coding GWAS, i.e., 639,323 variants located within 18,053 protein-coding genes and 172 covariates.

Validation scheme

An independent cohort of FinnGen was used for the validation of genes identified as statistically significant by the PWAS method (for ICD10: E03). The analysis is based on a recent version of FinnGen with ~ 350,000 individuals with no overlap with UKB participants (Kurki et al. 2023). For validation, we used the 146 independent loci reported in Freeze 7.0 (Fz7; spring 2021). Fr7 lists of credible genes but includes loci that failed to identify creditable genes (33 loci). Note that for other loci, several genes have been reported. Altogether, there were 102 genes for E4_HYTHYNAS (hypothyroidism, other/unspecified) from 38,554 cases and 263,704 controls. Additional validation was performed for FinnGen Freeze 8.0 (Fz8, spring 2002). We tested the related phenotypes of E4_HYTHY_AI_STRICT (Fz8: 45,320 cases) and E4_HYTHYNAS (Fz8: 52,828 cases). We compared the results for PheWAS Section 4 (i.e., endocrine, nutritional and metabolic diseases, E4) analyzed by the Ristey R9 querying system (Kurki et al. 2023).

Statistical tests

Sex-specific statistics

We applied a procedure for assessing the possibility that the discovery of significant sex-specific PWAS results is due to the difference in the UKB cohort sizes (2,557 males and 11,094 females). We applied a random sampling of the sex-specific ICD-10 E03 cases following matching of the group sizes (i.e., by male group). We performed 100 PWAS runs (with repetitions) and recorded the number of occurrences (up to 100 times) for males and females. We questioned whether genes were differentiated between sex-specific populations directly from the scores of the individual gene-level risk scores. To this end, we performed a gene-based permutation analysis using the Python library mlxtend v0.21.0. The analysis performed an association test between the risk score of each person and the label. Specifically, we test the per-gene statistics of the results by running 100, 500, 1000, 5000, 10,000, 15,000, and 30,000 permutations. We applied the Spearman rank correlation test to test the hypothesis that genes identified for male and female groups originated from identical probabilities.

Effect size statistics

To determine the effect size of a gene on hypothyroidism, we applied a measure of Cohen's d values. Cohen's d, also known as standardized mean difference, measures the difference between two means divided by a standard deviation

(SD) for the data. In this study, Cohen's d is the (normalized) difference in mean gene effect scores between cases and controls (calculated independently for both dominant and recessive effect scores). For GWAS, the variant association and effect size were calculated by PLINK 2.0 default logistic regression, which produces the z score to specify the effect size and its directionality. Note that in GWAS, a positive z score indicates a positive correlation between hypothyroidism and the number of alternative alleles, thereby indicating a risk variant. In PWAS, positive values indicate a positive correlation with the gene effect scores, whose higher values mean less functional damage. Thus, negative values are indicative of protective variants in GWAS versus risk genes in PWAS.

PRS calculation: To calculate PRS, we used the PRSice-2 protocol (Choi and O'Reilly 2019). Predictive PRS models for coding GWAS, and all GWAS were based on a standard partition of 80:20 for the training and test sets. For all GWAS, we used ~ 0.5 M common variants (MAF > 1e-03, a p value for HWE test larger than > 1e − 06 and 100% call rate using geno option). In addition, we applied covariates of sex, age, UKB assessment centers, and genotype measurement batch. We performed predictive PRS for hypothyroidism E03 by the liability scale R^2 and the AUC Roc (i.e., the area under the receiver operating characteristic curve) (Choi and O'Reilly 2019) for both sexes, male and female groups. While the R^2 assesses the amount of explained variation in the regression models, the AUC Roc evaluates the ability of the set of used variants to discriminate between the classes (E03 vs controls).

Bioinformatics tools

For gene connectivity and protein—protein interaction (PPI) maps, we applied STRING at a high PPI connectivity score (Szklarczyk et al. 2021). For functional enrichment of GO annotation and KEGG pathways, we applied the Gene2Func function of FUMA-GWAS using default parameters and a set of genes as input (Watanabe et al. 2017). All values are reported by their adjusted p values, using the human proteome as background.

Resource and availability

FIRM model and prediction of variant-centric effect score (https://github.com/nadavbra/firm*).*

Exclusion and inclusion rules per outcomes and phenotypes from FinnGen are found in https://risteys.finngen.fi/endpoints/.

A gene-based permutation analysis using python ML library mlxtend v0.21.0; in https://github.com/rasbt/mlxtend/.

Results

Comparative GWAS results for hypothyroidism

Large-scale GWAS have been performed on several cohorts for hypothyroidism (see Methods). A comparative study compiling six of the largest studies is shown in Fig. 1. The comparison is performed with a GWAS of “Hypothyroidism/Myxoedema (noncancer, self-reported)” from Neale v2, 2018 that covers non-Finnish Europeans with ~ 17.5k cases and ~ 345k controls from UKB. This study reports on 115 significant (p value < 5e-8) variants. Each of the leading variants is reported along with its most likely associated genes.

We identified 21 intersecting lists for all 6 GWAS (Additional file 1: **Table S2**). Note that the individual GWAS may include overlapping participants. While accurate mapping of variants to genes is inconclusive, we observed significant functional connectivity among these overlapping associated genes (STRING PPI enrichment p value of 7.4e-07; Fig. 1B). For example, the TPO and FOXE1 genes (Fig. 1B, blue cluster) are involved in thyroid hormone production and secretion. Specifically, TPO is a key enzyme in thyroid peroxidase that acts in the iodination of tyrosine residues in thyroglobulin and thyroid hormones, while FOXE1 is implicated in thyroid gland morphogenesis (Table 1**).** The clusters in Fig. 1B list genes active in the regulation of T-cell receptors (yellow), thyroid hormone production (blue), transcriptional regulation (red), and

chromatin modifiers (green). We observed that the genes associated with hypothyroidism partition genes by their cellular properties in relation to immunity, DNA-binding proteins, and numerous enzymes (Fig. 1C).

Table 1 summarizes the variants (based on the overlap of six large-scale GWAS) along with the most likely associated genes. Most variants are common, with allele frequencies (AFs) ranging from 0.17 to 0.89. Note that for many of the variants, linkage disequilibrium (LD) identifies a large number of genes within the same haplotype block. In these cases, no conclusive assignment to a particular gene is possible without fine mapping. In fact, only 3 of the 21 lead variants are associated with a definitive gene (Table 1).

Table 1
The intersection variants (total 21) from 6 large-scale hypothyroidism GWAS results

	SNP rsID	Band	Variant	Closest gene		AF[a]	# LD genes[b]
1	rs78765971	1p13	1_107819547_GAC_G	VAV3		0.09	3
2	rs484959	1p13	1_109823461_T_C	GSTM3		0.50	23
3	rs2476601	1p13	1_113834946_A_G	PTPN22		0.89	15
4	rs11675342	2p25	2_1403856_C_T	TPO		0.44	3
5	rs1534430	2p24	2_12504610_C_T	TRIB2		0.41	1
6	rs2111485	2q24	2_162254026_A_G	FAP		0.62	6
7	rs7582694	2q32	2_191105394_C_G	STAT4		0.78	7
8	rs76897057	3q28	3_188407079_TA_T	LPP		0.48	1
9	rs34046593	4p15	4_26109971_G_A	RBPJ		0.31	6
10	rs546532456	4q31	4_148724495_C_CTT	PGR		0.19	1
11	rs2445610	8q24	8_127184843_A_G	POU5F1B		0.35	2
12	rs2123340	9p21	9_21589042_G_A	IFNE		0.65	19
13	rs7850258	9q22	9_97786731_A_G	FOXE1		0.66	14
14	rs71508903	10q21	10_62020112_C_T	ARID5B		0.18	4
15	rs736374	11p13	11_35245397_G_A	CD44		0.37	7
16	rs4409785	11q21	11_95578258_T_C	SESN3		0.17	5
17	rs3184504	12q24	12_111446804_T_C	SH2B3		0.52	17
18	rs61759532	17p13	17_7337072_C_T	ACAP1		0.23	59
19	rs10424978	19p13	19_4837545_C_A	TICAM1	0.59		21
20	rs1454294221	22q12	22_30125266_CCAG_C	LIF	0.48		23
21	rs229540	22q12	22_37195250_T_G	C1QTNF6	0.42		25

[a]AF, Allele frequency for non-Finnish European population. [b]# LD genes, the number of associated mapped genes resulting from the variant to gene (V2G) OTG protocol.

The listed shared variants are quite stable and remain valid in view of additional large-scale GWAS. For example, the addition of GWAS for autoimmune thyroid disease with 755k participants from Iceland (93 associated variants)

(Saevarsdottir et al. 2020) had only a minor influence on the overall number of intersected variants (19 of 21 listed variants). Under the assumption of accurate mapping of variants to genes (Table 1), the results expose the genetic signal of CH. Specifically, TPO was reported as causal for thyroid dyshormonogenesis 2 (OMIM 274500). In the Chinese population (Wang et al. 2017), abnormal expression of FOXE1 was linked to CH-based thyroid dysgenesis (OMIM 218700). Similarly, polymorphisms in the listed genes VAV3, SH2B3, FOXE1 and PTPN22 were identified in the 23andMe database to be associated not only with hypothyroidism but also with other autoimmune diseases (Eriksson et al. 2012). We conclude that the shared GWAS results identified pleiotropic effects of genes involved in autoimmunity and gene developmental alterations that underlie CH.

Coding GWAS highlights the abundance of genes in the MHC extended locus

We overcame the difficulty of variant-to-gene mapping by performing coding GWAS using ~ 640k coding variants. Figure 2 shows the results of the analysis for ~ 18k coding genes (Additional file 1: **Table S3**). We report 2813 variants with a relaxed p value of < 1e-02 and 149 and 61 variants by setting the significant thresholds for p values of 1e-08 and 1e-16, respectively. Importantly, 91% and 95% of the variants at these thresholds were associated with the extended region of MHC in chromosome 6 (Chr6: 25 M to 34 M), respectively (Fig. 2A). For the significant thresholds (p value < 1e-08), the directionality of the variants as reducing or increasing the risk is balanced (Fig. 2B). The 61 most significant variants are associated with 19 unique genes from the MHC locus and only 3 genes from other chromosomal locations.

We conclude that the coding gene view is driven by the signature within the gene-dense immunological region of the MHC locus.

Gene-based analysis using PWAS

The limitation of GWAS interpretability is due to the difficulty of assigning variants to their associated genes. To overcome these limitations, we applied PWAS, which exclusively focuses on alterations in the coding gene and assesses the impact of damaging variants on its biochemical function (Brandes et al. 2020). Based on the UKB cohort for ICD-10 E03, we identified 77 statistically significant PWAS genes (FDR-q-value < 0.05). We analyzed significant genes based on their risk directionality (Fig. 3). Among the top-range genes (FDR q-value, <1e-07; 26 genes, Fig. 3A**)**, genes with increased risk for hypothyroidism (colored red) dominate. As expected from other complex diseases, most genes have a rather limited effect size (calculated by Cohen's d values). There are six genes that have Cohen's d values >|0.06| and p values < 1.0e-16 (Fig. 3B**).** Among these genes, five genes are associated with elevated risk, and SH2B3 is a strong protective gene. A large effect size is associated with GPR174, G protein-coupled receptor 174, a ChrX gene that plays a role in autoimmunity pathogenesis (Napier et al. 2015). PWAS also model genes according to their inheritance modes. While for 53%, compelling evidence suggests dominant inheritance, 12% of the genes show clear recessive inheritance (Additional file 2: **Fig. S1**).

Autoimmunity-associated genes are enriched in PWAS results for hypothyroidism

Inspecting all identified genes revealed that many of these genes are associated with the immune system, specifically with autoimmunity. We asked whether the identified genes may shed light on the underlying mechanisms for hypothyroidism. To this end, we reconstructed a connectivity map among the 77 PWAS genes as represented by STRING (Szklarczyk et al. 2021) (Fig. 4).

We found that the connectivity map is statistically significant and of high confidence (p value of 3.68e-10; with a PPI STRING score > 0.9). The network (21 nodes) is mostly associated with cellular immunity, including antigen presentation,

processing, and T-cell regulation. Moreover, 36 of 77 genes (47%) are located at the Chr6p22.1-p21.32 locus that specifies the MHC locus (hypergeometric distribution test, p value 7.3e-57).

Relationships between MHC variants involved in autoimmunity determine other aspects of immunity, such as responses to infectious diseases and inflammation. Strong connectivity was observed (Fig. 4A), and the extreme enrichment in coding genes identified within the MHC locus (Fig. 4B, > 60-fold higher than expected) strongly argues for the dominant genetic signal that combines hypothyroidism and autoimmune complex diseases.

Validated hypothyroidism PWAS significant genes

To test the relevance of PWAS-identified genes, we sought an independent cohort that could validate the gene discovery. To this end, we investigated the Finnish Biobank (FinnGen) as an independent validation resource for hypothyroidism. Table 2 lists nine genes (DCLRE1B, CTLA4, TLR3, HLA-DPB1, TRMO, PCSK7, SH2B3, THOC5, and C1QTN) that were validated from the FinnGen data and shared with the PWAS discovery. Notably, PWAS only refers to coding variants, while like any standard GWAS, FinnGen identifies mostly noncoding variants (Table 2). Recall that there is no overlap between UKB and FinnGen participants (see Methods). Table 2 also shows genes associated with “Hypothyroidism (congenital or acquired) (38.6k cases, 263.7k controls; 122 genes, phenotype a), “Hypothyroidism, strict autoimmune” (33.4k cases, 227.4k controls, 105 genes, phenotype b), and a more general term of “Disorders of the thyroid gland” (45.5k cases, 263.7k controls, 80 genes, phenotype c). Further validation is based on the independent cohort of 23&me (Eriksson et al. 2012). Several of the replicated genes were supported by fine-mapping (TLR3, Additional file 2: **Fig. S2).**

Table 2
Validation of the PWAS gene by FinnGen Fz7 for Hypothyroidism phenotypes

Symbol	[a]FinnGen (a,b,c)	p value	rsID	Chr:M	Gene effect	[b]AID	Specific publication	More evidence	[c]Risk
C1QTNF6	a,b,c,	8.7E-19	rs229541	22:37 M	Intergenic	+	(Frederiksen et al. 2013)	Fine-map 23&me	I
CTLA4	a,b,c,	1.1E-44	rs3087243	2:203 M	missense	+	(Braun et al. 1998)	Fine-map 23&me	I
DCLRE1B	a	2.0E-10	rs12127377	1:113 M	intron	+	(Ban et al. 2010)	Japan	D
HLA-DPB1	a,b,c	4.8E-31	rs9277535	6:32 M	downstream	+	(Huang and Jap 2015)	Taiwan	D
PCSK7	a,b,c	8.5E-11	rs76169968	11:117 M	intron	-	Aging (Donertas et al. 2021)		D
SH2B3	a,b,c	3.3E-55	rs7310615	12:111 M	intron	+	(Auburger et al. 2014)	23andMe	D
THOC5	a	2.5E-06	rs8140060	22:29 M	intron	-			D
TLR3	a,b,c	2.6E-10	rs3775291	4:186 M	missense	+	(Caturegli et al. 2007)	Fine-map	D
TRMO	a,b,c	6.1E-06	rs8140060	9:97 M	intron	-			D

The OT platform provides a knowledge-based resource that converts the association of genes to diseases by including rich biological knowledge from multiple sources (e.g., literature, animal models, pathways, and drugs). Altogether, more

than 700 genes were scored for genetic association (GA score ranges 0–1.0, see Methods; Additional file 1: **Table S5)**. The overlap of the GA-listed genes and PWAS results is significant (p value 6.38e-25; 10.5-fold enrichment). An even higher enrichment was found for genes selected with a higher GA score (total top 222 genes with score > 0.3, p value 5.28e-13, 14.7-fold enrichment; Fig. 5A). Inspecting the functions of these 222 genes revealed many enzymes, membranous receptors, secreted proteins, and genes that act in the development, synthesis, and secretion of thyroid hormones. Surprisingly, 58% of the PWAS genes were not identified by the GWAS results reported by OTG. Interestingly, none of the top-ranked genes for "hypothyroidism" according to the OT global score (25 genes, threshold > 0.5; Additional file 1: **Table S5)** were identified as significant genes by PWAS. We conclude that the alternative association methodologies expose different aspects of hypothyroidism. We confirmed an extreme overlap with 19 of these top 25 genes as a CH-exclusive gene set, defined as "permanent congenital hypothyroidism" (Orphanet: 442; EFO: 0016408). Figure 5B shows that the genes associated with permanent CH are functionally linked (STRING PPI enrichment, p value < 1e-16). Validation of the CH causal genes was confirmed by independent studies analyzing patients from Korea (Jung et al. 2020) and China (Wang et al. 2020) (colored blue, Fig. 5B).

As the vast majority of associated variants occur in noncoding regions (Edwards et al. 2013), we tested the genetic signal that resulted from the transcriptome-wide association study (TWAS) (Luningham et al. 2020). Figure 5C emphasizes the overlap of association studies for hypothyroidism that relied on UKB entries of European origin: the PWAS (77 genes), the GWAS gene subset (GA score > 0.5; 136 genes), and the TWAS significant expression-trait associations (110 genes). We show that the overlap between PWAS and OT and TWAS to OT is limited to six genes. The overlapping genes are enriched with genes of cellular immunity. On the other hand, the overlap between the TWAS and OT lists is most substantial, with 33 genes that account for 25–30% of the gene sets. Genes that are shared belong to transcription, signaling, trafficking, immunity, and more. Only C1QTNF6 (C1q and TNF-related 6) was shared by all three orthogonal association studies. The shared variant was also identified by the shared variant from multiple GWAS (Table 1). C1QTNF6 is known to carry two coding mutations, rs229527 (22:37,185,445:C,A) and rs229526 (22:37,185,382:G,C), that are associated with hypothyroidism-related phenotypes. This gene was identified within a locus that is associated with a large number of thyroid-related pathologies (Additional file 2: **Fig. S3**).

In this study, we have not explicitly studied rare variants from whole genome UKB. However, based on UKB exome sequencing (Backman et al. 2021), models were developed and reported on the relationships between thousands of traits and diseases by analyzing rare variants from 269,171 UKB participants of European ancestry (Wang et al. 2021). The results reported for hypothyroidism (indexed as union#E03#Other hypothyroidism) and filtered by p value < 1e-10 include 320 variants that are mapped to 121 genes (Additional file 1: **Table S6**). Similar to our observation, most variants are associated with the extended MHC locus of Chr 6 (located at 25 M to 34 M, 87%). In terms of gene overlap, 34 (of the 77 PWAS genes) were supported by exome sequencing analysis (Wang et al. 2021). These results support the involvement of the immune-rich MHC genetic locus in the risk for hypothyroidism. Notably, only a small fraction of the genes located in the core segment of MHC (Chr 6: 29.5 M to 33.1 M, total 308 genes) are listed as PWAS-associated genes or identified by exome analysis, including rare variants (37 and 53 genes, respectively; Additional file 2: **Fig. S4**).

Gene-based association studies by sex

The higher prevalence of ICD-10 E03 in females relative to males raised the question of whether hypothyroidism is signified by sex-dependent genetics. To this end, we applied the PWAS gene-aggregative approach separately for males and females (Fig. 6). The Venn diagram emphasizes that females displayed 63 significant genes, while only nine were reported for males. Out of the female significant genes, 7 genes (ARID3C, ATL2, HDGF, IRAK1, LRRC47, OR13A1, and SLC15A2) did not appear to be significant in the full cohort combining both sexes (Fig. 6A). On the other hand, the genes identified in males are either shared by females (7 genes) or included in PWAS results from both sexes, with only one male-exclusive exception (THAP6, THAP domain containing 6; Fig. 6A). The Spearman rank correlation for the shared

genes in females and males confirmed the high correlation and similarity in gene effects between the sexes (r = 0.61, p value = 0.14).

We performed permutation tests over the score of each gene to assess whether any of the identified genes directly displayed significant differences in their assignment to females or males following 100 to 30,000 such tests (see Methods). Importantly, at 500 permutations, the results were stable, confirming the lack of sex-dependent genetic effects (Fig. 6B**).** Only two genes (IRAK1 and GPR174) consistently and significantly failed the permutation tests (p value of 3.3e-05). Both genes are located at Chromosome X. IRAK1 (Interleukin-1 receptor-associated kinase 1) was only significant in females **(**Fig. 6A**).** IRAK1 is fundamental in initiating the innate immune response against pathogens and participates in the antiviral state. GPR174 impacts autoimmune Addison's disease and Graves' disease (GD), and its female preponderance was confirmed (Napier et al. 2015). We conclude that all other identified genes do not support a sex difference for hypothyroidism.

The E03 prevalence in the analyzed UKB was 7.45% and 2.04% for females and males, respectively. As the group sizes of females and males are markedly different, we also performed 100 independent runs for E03 by artificially balancing the number of males and females. To this end, we randomly sampled from the female group the male-matched group size. Altogether, there were 217 significant genes that appeared across these runs. As many as 70% should be considered noise, as the gene was identified only by a single PWAS run (out of 100, Additional file 1: **Table S7**). Genes that were detected in > 20% of the runs showed no observed difference between the sexes (Additional file 2: **Fig. S5**). We further conclude that the reported difference in discovery rate for males and females is mostly attributed to group size differences.

PRS reflects the cumulative effect of the genetic variants. This allows for predicting the E03 genetic predisposition. We performed PRS for the weighted sum of allele dosages multiplied by their corresponding effect sizes. Figure 6C (all GWAS) and Fig. 6D (coding GWAS) show the distribution for the entire population and, according to partition by sex. Figure 6E shows the prediction power of the PRS as calculated by the R^2 and AUC-Roc for the test set (2492 cases and 51,630 controls). We conclude that the majority of the PRS predictive power is captured by variants within the coding regions. Moreover, the separation of the population by sex validated that most genetic signals for the calculated Roc-AUC are captured within the female group.

Discussion

In this study, we sought to identify the genetic basis of hypothyroidism (ICD-10, E03) in the adult population of the UKB. An interest in thyroid function in adults and especially in the elderly relies on the increasing links between thyroid status and cognitive function, cardiovascular diseases, healthy aging, and longevity (Aggarwal and Razvi 2013). It is imperative to identify people at higher risk and tune clinical treatment to avoid negative impacts on quality of life (Hegedus et al. 2022). Hypothyroidism occurs when there are low levels of thyroid hormones and is typically treated with synthetic thyroid hormone replacement (Persani and Bonomi 2017). Primary hypothyroidism can be roughly divided into congenital or acquired diseases. Several diseases are linked to hypothyroidism through the destruction of the thyroid by the immune system. These include Hashimoto's thyroiditis and Graves’ disease (GD) (Umar et al. 2010). In children, AIT is a complex immune disorder in which each gene has only a small effect. Less common are cases of secondary or tertiary hypothyroidism caused by insufficient stimulation of the thyroid gland. Other forms of hypothyroidism may be associated with organ resistance to thyroid hormone (Persani et al. 2010), with subclinical hypothyroidism being linked to older females (Biondi et al. 2019; Dunn and Turner 2016). Our understanding of the environmental factors that contribute to disease development is limited, and risk factors may include hormones (e.g., estrogen), stress, smoking, and dietary iodine consumption. Cases of subclinical hypothyroidism during and after pregnancy are not discussed further (Casey et al. 2017).

In this study, we exhaustively compared different association study methods and protocols. In a routine GWAS, variants are statistically tested under an additive model with a case–control setting (Korte and Farlow 2013). However, PWAS also detects nonadditive effects and allows the aggregated effect of variants that may occur at different locations within the same gene (i.e., a recessive model). Although only a small fraction of the genes identified by PWAS have been identified with a clear recessive signal (Additional file 1: Table S4 and Fig. S1), such inheritance modes have been mostly overlooked by GWAS approaches (Brandes et al. 2020; Tam et al. 2019). The vast majority of the associated variants in GWAS occur in the noncoding regions of the genome, and thus, the relevance of a SNP function to the studied phenotype is mostly lacking (discussed in (Edwards et al. 2013)). Along with the effort to capture gene causality, TWAS was developed to identify variants that affect the regulation of gene expression using tissue-based expression data and Bayesian considerations (e.g., eQTL). Significant variants from TWAS are located in regions that may alter the expression levels of transcripts, with cis or trans regulation modes (Luningham et al. 2020; Maddirevula et al. 2020). As expected, TWAS is a complementary approach exposed quite different results from GWAS findings (Fig. 5). Much of the overlap in gene function between TWAS and GWAS is linked to thyroid development and hormone secretion (e.g., PAX8, FOXE1, STAT4, TG). Interestingly, both methodologies display only a minimal signal of immunity (Khan et al. 2021).

The genetic basis for congenital hypothyroidism (CH) was exposed from population studies (Amberger et al. 2019), where dysregulation of transcription factors (TFs) characterizes individuals with CH (Grasberger and Refetoff 2011). A comprehensive screening of CH in family pedigrees from China identified DUOX2 as the most frequently mutated gene (Sun et al. 2018). CH, combined with neonatal diabetes mellitus, is caused by mutations in the TF GLIS3 gene (Fu et al. 2018). Other TFs, such as NR1D1 and PAX8, which are exclusively expressed in thyroid cell types, were identified by GWAS and TWAS. TFs such as FOXE1 and STAT4 were consistently identified by all large-scale GWAS. These genes act during embryogenesis to establish the pituitary, hypothalamus, and thyroid axes (Table 1). Thyroid signaling genes were highly represented in TWAS. For example, PDE8B is expressed primarily in the thyroid to execute the TSH effects. Combining results from multiple classical GWAS identified TFs that have strong links to thyroid function (e.g., NKX2-1; NK2 homeobox 1). It is likely that the strong effect size of rare variants dominated their discovery (see OMIM lists of 34 genes for CH, Additional file 1: Table S9). None of the top-scoring OT genes were identified by PWAS. We attribute this discrepancy in gene findings to the relatively small effect sizes of the most common variants in coding genes. Recall that GWAS identifies strong functional elements that are ignored by PWAS. For example, a cluster of variants on chromosome 9, including rs10759927, rs7850258 and rs7030280, was significantly identified by classical GWAS for hypothyroidism (with a p value ranging between 1e-82 and 2e-100). These variants occur within the introns of PTCSC2, a noncoding RNA (ncRNA) that is expressed exclusively in the thyroid. Furthermore, accumulated data propose that CH is not restricted to monogenic dominant mutations. Biallelic effects were suggested based on the sporadic occurrences of CH within families (Sun et al. 2018). Additionally, evidence of parents with clinical manifestations of thyroid (functional or morphological) supports the notion of predisposition with recessive inheritance (Léger et al. 2002). We argue that applying a recessive model in PWAS and including a comprehensive analysis of rare variants (Wang et al. 2021) can expose the overlooked genetics of CH.

The strongest signature of explainable hyperthyroidism is caused by dysregulation of the immune system. This signature was revealed by applying PWAS and, to a lesser extent, coding GWAS. Testing the genetic variations associated with thyroid autoimmunity identified a strong interaction with pathways driving the immune response (Khan et al. 2021; Luo et al. 2021). The findings agree with a clinical investigation that explains the mechanisms underlying thyroid autoimmunity. Hashimoto's thyroiditis is the most common form of hypothyroidism. The autoimmune facet of hypothyroidism is characterized by the infiltration of T lymphocytes into the thyroid gland and autoantibodies against thyroid-specific genes (e.g., thyroid peroxidase, thyroglobulin, and TSH receptor). AIT-associated genes were also identified in thyroid autoimmune Graves' disease (GD). Among these shared genes are HLA class II (HLA-DR), protein tyrosine phosphatase 22 (PTPN22), and cytotoxic T lymphocyte antigen 4 (CTLA4) (Jacobson and Tomer 2007). The genetic signals of autoimmune thyroiditis are shared with other immune-mediated diseases, such as T1D, celiac disease, rheumatoid

arthritis (RA), systemic lupus erythematosus (SLE), and psoriasis. PWAS exposes many immune-related genes that carry coding variants (e.g., PTPN22, SH2B3, and genes in the class 1 MHC region). Thus, it can help to assess the risk prediction for autoimmunity that overlaps with hypothyroidism (Eriksson et al. 2012). It is important to note that such enrichment in the MHC locus is specific to E03. We argue that an immune-related signature is a genuine contributing signal for hypothyroidism with evidence for gene-specific interpretability. Note that for a large collection of tested complex phenotypes (Brandes et al. 2020, 2021; Zucker and Linial 2022), using PWAS or coding GWAS did not identify gene associations within the Chr 6 MHC locus.

The gene-based analysis performed in this study raised the question of whether genetic effects are distinguished by sex. For hypothyroidism, a ratio of > 3.6:1 for females and males was reported in the UKB. This bias in prevalence was also validated in the Finnish population, which reported 37,942 affected females and 9,616 males. Several clinical features distinguish E03 by sex. For example, the average diagnosis age is 50.0 and 58.3 years for females and males, respectively. Despite the differences in sex prevalence (Graham et al. 2019). We recently showed that UKB hypertension and phenotypes of blood pressure that are more prevalent in males than females carry a strong signal by sex, where most of the genetic signal is attributed to females (Zucker and Linial 2022). For example, significant sex-dependent effects were enriched for neuronal development and immune-related genes in cases of schizophrenia, major depression disease (Blokland et al. 2022), autism (Lu and Cantor 2012) and Alzheimer's disease (Gamache et al. 2020). In this study, we showed that the sex stratification of hypothyroidism provided no support for genuine genetic differences between the sexes, despite the large gap in prevalence. This is in agreement with most human traits and diseases that do not support a mechanistic difference between the sexes (Traglia et al. 2017). With the continuous increase in statistical power due to cohort sizes, more cases of sex-dependent genetics have been revealed (Kostyunina and McLoughlin 2021; Pirastu et al. 2021).

We conclude that comparison results that rely on capturing different aspects of the genetic signal allowed us to reveal the complex etiology of hypothyroidism, covering recessive signals, CH, and acquired chronic damage to thyroid functionality. The discovery benefited from the use of complementary association studies, which can be generalized to other polygenic diseases with unknown etiologies. It is likely that complex diseases for which the age of diagnosis, sex prevalence and diseases that are subjected to (wrong) alternative diagnoses could benefit from integrating genetic association schemes. Moreover, using coding GWAS and PWAS, we illustrate the clinical benefits of gene-based genetics. The results from coding genes are more interpretable and direct, which can benefit unexplored therapeutic targets.

Abbreviations

AIT
autoimmune thyroiditis
AUC
Aria under the curve
CH
Congenital hypothyroidism
GA
Genetic association
GD
Graves' disease
GWAS
Genome wide association study
HPT
hypothalamic-pituitary-thyroid

MAF
Minor allele frequencies
MHC
Major histocompatibility complex
ncRNA
non coding RNA
OMIM
Online Mendelian Inheritance in Man
OTP
Open targets genetics
PPI
Protein-protein interaction
PRS
polygenic risk score
PWAS
Proteome wide association study
SNP
Single nucleotide polymorphism
TWAS
Transcriptome wide association study
UKB
UK Biobank
WGS
Whole genome sequencing

Declarations

Ethical committee approval

The study was approved by the University Committee for the Use of Human Subjects in Research Approval number 12072022 (July 2023). This study uses the UK-Biobank (UKB) application ID 26664 (Linial lab).

Funding:

This study was supported by the ISF grant 2753/20 on Sex dependent genetics and a grant for large scale data analysis from CIDR # 3035000440.

Author Contribution

R.Z. led the genetic associations, and comparative analyses M.K. and A.S. performed the clinical analytical aspects. H.K. performed the PRS study. M.L. conceptualization, mentoring, manuscript writing, and illustrations. All co-authors read the manuscript.

Acknowledgments

We thank Dr. Nadav Brandes for developing the UKB parser. We thank Dr. Guy Kelman for his help in the implementation of PWAS. We thank Linial's lab for useful discussions. We appreciate the constant support of the system team of the School of Computer Science at Hebrew University.

Data and code availability

PWAS is available through a command-line interface as part of an open-source project (MIT license) at Brandes N. pwas. Github. 2020. https://github.com/nadavbra/pwas. *Accessed 11 Apr 2020*.

References

1. Aggarwal N, Razvi S (2013) Thyroid and aging or the aging thyroid? An evidence-based analysis of the literature. J Thyroid Res 2013: 481287. doi: 10.1155/2013/481287
2. Al Taji E, Biebermann H, Limanova Z, Hnikova O, Zikmund J, Dame C, Gruters A, Lebl J, Krude H (2007) Screening for mutations in transcription factors in a Czech cohort of 170 patients with congenital and early-onset hypothyroidism: identification of a novel PAX8 mutation in dominantly inherited early-onset non-autoimmune hypothyroidism. Eur J Endocrinol 156: 521–9. doi: 10.1530/EJE-06-0709
3. Almandoz JP, Gharib H (2012) Hypothyroidism: etiology, diagnosis, and management. Med Clin North Am 96: 203–21. doi: 10.1016/j.mcna.2012.01.005
4. Amberger JS, Bocchini CA, Scott AF, Hamosh A (2019) OMIM. org: leveraging knowledge across phenotype–gene relationships. Nucleic acids research 47: D1038-D1043.
5. Auburger G, Gispert S, Lahut S, Ömür Ö, Damrath E, Heck M, Başak N (2014) 12q24 locus association with type 1 diabetes: SH2B3 or ATXN2? World journal of diabetes 5: 316.
6. Backman JD, Li AH, Marcketta A, Sun D, Mbatchou J, Kessler MD, Benner C, Liu D, Locke AE, Balasubramanian S (2021) Exome sequencing and analysis of 454,787 UK Biobank participants. Nature 599: 628–634.
7. Ban Y, Tozaki T, Taniyama M, Nakano Y, Ban Y, Ban Y, Hirano T (2010) Association of the protein tyrosine phosphatase nonreceptor 22 haplotypes with autoimmune thyroid disease in the Japanese population. Thyroid 20: 893–9. doi: 10.1089/thy.2010.0104
8. Biondi B, Cappola AR, Cooper DS (2019) Subclinical Hypothyroidism: A Review. JAMA 322: 153–160. doi: 10.1001/jama.2019.9052
9. Blokland GAM, Grove J, Chen CY, Cotsapas C, Tobet S, Handa R, Schizophrenia Working Group of the Psychiatric Genomics C, St Clair D, Lencz T, Mowry BJ, Periyasamy S, Cairns MJ, Tooney PA, Wu JQ, Kelly B, Kirov G, Sullivan PF, Corvin A, Riley BP, Esko T, Milani L, Jonsson EG, Palotie A, Ehrenreich H, Begemann M, Steixner-Kumar A, Sham PC, Iwata N, Weinberger DR, Gejman PV, Sanders AR, Buxbaum JD, Rujescu D, Giegling I, Konte B, Hartmann AM, Bramon E, Murray RM, Pato MT, Lee J, Melle I, Molden E, Ophoff RA, McQuillin A, Bass NJ, Adolfsson R, Malhotra AK, Bipolar Disorder Working Group of the Psychiatric Genomics C, Martin NG, Fullerton JM, Mitchell PB, Schofield PR, Forstner AJ, Degenhardt F, Schaupp S, Comes AL, Kogevinas M, Guzman-Parra J, Reif A, Streit F, Sirignano L, Cichon S, Grigoroiu-Serbanescu M, Hauser J, Lissowska J, Mayoral F, Muller-Myhsok B, Swiatkowska B, Schulze TG, Nothen MM, Rietschel M, Kelsoe J, Leboyer M, Jamain S, Etain B, Bellivier F, Vincent JB, Alda M, O'Donovan C, Cervantes P, Biernacka JM, Frye M, McElroy SL, Scott LJ, Stahl EA, Landen M, Hamshere ML, Smeland OB, Djurovic S, Vaaler AE, Andreassen OA, Major Depressive Disorder Working Group of the Psychiatric Genomics C, Baune BT, Air T, Preisig M, Uher R, Levinson DF, Weissman MM, Potash JB, Shi J, et al. (2022) Sex-Dependent Shared and Nonshared Genetic Architecture Across Mood and Psychotic Disorders. Biol Psychiatry 91: 102–117. doi: 10.1016/j.biopsych.2021.02.972

10. Brandes N, Linial N, Linial M (2019) Quantifying gene selection in cancer through protein functional alteration bias. Nucleic Acids Res 47: 6642–6655. doi: 10.1093/nar/gkz546
11. Brandes N, Linial N, Linial M (2020) PWAS: proteome-wide association study-linking genes and phenotypes by functional variation in proteins. Genome Biol 21: 173. doi: 10.1186/s13059-020-02089-x
12. Brandes N, Linial N, Linial M (2021) Genetic association studies of alterations in protein function expose recessive effects on cancer predisposition. Sci Rep 11: 14901. doi: 10.1038/s41598-021-94252-y
13. Braun J, Donner H, Siegmund T, Walfish P, Usadel K, Badenhoop K (1998) CTLA-4 promoter variants in patients with Graves' disease and Hashimoto's thyroiditis. Tissue antigens 51: 563–566.
14. Brown RS (2013) Autoimmune thyroiditis in childhood. J Clin Res Pediatr Endocrinol 5 Suppl 1: 45 – 9. doi: 10.4274/jcrpe.855
15. Carvalho-Silva D, Pierleoni A, Pignatelli M, Ong C, Fumis L, Karamanis N, Carmona M, Faulconbridge A, Hercules A, McAuley E (2019) Open Targets Platform: new developments and updates two years on. Nucleic acids research 47: D1056-D1065.
16. Casey BM, Thom EA, Peaceman AM, Varner MW, Sorokin Y, Hirtz DG, Reddy UM, Wapner RJ, Thorp Jr JM, Saade G (2017) Treatment of subclinical hypothyroidism or hypothyroxinemia in pregnancy. New England Journal of Medicine 376: 815–825.
17. Caturegli P, Kimura H, Rocchi R, Rose NR (2007) Autoimmune thyroid diseases. Current opinion in rheumatology 19: 44–48.
18. Chaker L, Razvi S, Bensenor IM, Azizi F, Pearce EN, Peeters RP (2022) Hypothyroidism (Primer). Nature Reviews: Disease Primers 8.
19. Choi SW, O'Reilly PF (2019) PRSice-2: Polygenic Risk Score software for biobank-scale data. Gigascience 8: giz082.
20. Chung HR (2014) Iodine and thyroid function. Annals of pediatric endocrinology & metabolism 19: 8.
21. Donertas HM, Fabian DK, Valenzuela MF, Partridge L, Thornton JM (2021) Common genetic associations between age-related diseases. Nat Aging 1: 400–412. doi: 10.1038/s43587-021-00051-5
22. Dunn D, Turner C (2016) Hypothyroidism in women. Nursing for women's health 20: 93–98.
23. Edwards SL, Beesley J, French JD, Dunning AM (2013) Beyond GWASs: illuminating the dark road from association to function. Am J Hum Genet 93: 779–97. doi: 10.1016/j.ajhg.2013.10.012
24. Eriksson N, Tung JY, Kiefer AK, Hinds DA, Francke U, Mountain JL, Do CB (2012) Novel associations for hypothyroidism include known autoimmune risk loci. PLoS One 7: e34442. doi: 10.1371/journal.pone.0034442
25. Evered DC, Ormston BJ, Smith PA, Hall R, Bird T (1973) Grades of hypothyroidism. Br Med J 1: 657–62. doi: 10.1136/bmj.1.5854.657
26. Fatourechi V Subclinical hypothyroidism: an update for primary care physicians Mayo Clinic Proceedings 2009. Elsevier, pp 65–71
27. Frederiksen BN, Steck AK, Kroehl M, Lamb MM, Wong R, Rewers M, Norris JM (2013) Evidence of stage- and age-related heterogeneity of non-HLA SNPs and risk of islet autoimmunity and type 1 diabetes: the diabetes autoimmunity study in the young. Clin Dev Immunol 2013: 417657. doi: 10.1155/2013/417657
28. Fu C, Luo S, Long X, Li Y, She S, Hu X, Mo M, Wang Z, Chen Y, He C (2018) Mutation screening of the GLIS3 gene in a cohort of 592 Chinese patients with congenital hypothyroidism. Clinica Chimica Acta 476: 38–43.
29. Gamache J, Yun Y, Chiba-Falek O (2020) Sex-dependent effect of APOE on Alzheimer's disease and other age-related neurodegenerative disorders. Dis Model Mech 13. doi: 10.1242/dmm.045211
30. Ghoussaini M, Mountjoy E, Carmona M, Peat G, Schmidt EM, Hercules A, Fumis L, Miranda A, Carvalho-Silva D, Buniello A (2021) Open Targets Genetics: systematic identification of trait-associated genes using large-scale genetics and functional genomics. Nucleic acids research 49: D1311-D1320.

31. Graham SE, Nielsen JB, Zawistowski M, Zhou W, Fritsche LG, Gabrielsen ME, Skogholt AH, Surakka I, Hornsby WE, Fermin D (2019) Sex-specific and pleiotropic effects underlying kidney function identified from GWAS meta-analysis. Nature communications 10: 1–9.
32. Grasberger H, Refetoff S (2011) Genetic causes of congenital hypothyroidism due to dyshormonogenesis. Current opinion in pediatrics 23: 421.
33. Hegedus L, Bianco AC, Jonklaas J, Pearce SH, Weetman AP, Perros P (2022) Primary hypothyroidism and quality of life. Nat Rev Endocrinol 18: 230–242. doi: 10.1038/s41574-021-00625-8
34. Huang C-J, Jap T-S (2015) A systematic review of genetic studies of thyroid disorders in Taiwan. Journal of the Chinese Medical Association 78: 145–153.
35. Jacobson EM, Tomer Y (2007) The genetic basis of thyroid autoimmunity. Thyroid 17: 949–961.
36. Jung SY, Lee J, Lee DH (2020) Persistent goiter with congenital hypothyroidism due to mutation in DUOXA2 gene. Ann Pediatr Endocrinol Metab 25: 57–62. doi: 10.6065/apem.2020.25.1.57
37. Khan Z, Hammer C, Carroll J, Di Nucci F, Acosta SL, Maiya V, Bhangale T, Hunkapiller J, Mellman I, Albert ML, McCarthy MI, Chandler GS (2021) Genetic variation associated with thyroid autoimmunity shapes the systemic immune response to PD-1 checkpoint blockade. Nat Commun 12: 3355. doi: 10.1038/s41467-021-23661-4
38. Kichaev G, Bhatia G, Loh PR, Gazal S, Burch K, Freund MK, Schoech A, Pasaniuc B, Price AL (2019) Leveraging Polygenic Functional Enrichment to Improve GWAS Power. Am J Hum Genet 104: 65–75. doi: 10.1016/j.ajhg.2018.11.008
39. Kollati Y, Akella RRD, Naushad SM, Borkar D, Thalla M, Nagalingam S, Lingappa L, Patel RK, Reddy GB, Dirisala VR (2020) Newborn screening and single nucleotide variation profiling of TSHR, TPO, TG and DUOX2 candidate genes for congenital hypothyroidism. Mol Biol Rep 47: 7467–7475. doi: 10.1007/s11033-020-05803-x
40. Korte A, Farlow A (2013) The advantages and limitations of trait analysis with GWAS: a review. Plant methods 9: 1–9.
41. Kostopoulou E, Miliordos K, Spiliotis B (2021) Genetics of primary congenital hypothyroidism—a review. Hormones 20: 225–236.
42. Kostyunina DS, McLoughlin P (2021) Sex dimorphism in pulmonary hypertension: the role of the sex chromosomes. Antioxidants 10: 779.
43. Kurki MI, Karjalainen J, Palta P, Sipila TP, Kristiansson K, Donner KM, Reeve MP, Laivuori H, Aavikko M, Kaunisto MA, Loukola A, Lahtela E, Mattsson H, Laiho P, Della Briotta Parolo P, Lehisto AA, Kanai M, Mars N, Ramo J, Kiiskinen T, Heyne HO, Veerapen K, Rueger S, Lemmela S, Zhou W, Ruotsalainen S, Parn K, Hiekkalinna T, Koskelainen S, Paajanen T, Llorens V, Gracia-Tabuenca J, Siirtola H, Reis K, Elnahas AG, Sun B, Foley CN, Aalto-Setala K, Alasoo K, Arvas M, Auro K, Biswas S, Bizaki-Vallaskangas A, Carpen O, Chen CY, Dada OA, Ding Z, Ehm MG, Eklund K, Farkkila M, Finucane H, Ganna A, Ghazal A, Graham RR, Green EM, Hakanen A, Hautalahti M, Hedman AK, Hiltunen M, Hinttala R, Hovatta I, Hu X, Huertas-Vazquez A, Huilaja L, Hunkapiller J, Jacob H, Jensen JN, Joensuu H, John S, Julkunen V, Jung M, Junttila J, Kaarniranta K, Kahonen M, Kajanne R, Kallio L, Kalviainen R, Kaprio J, FinnGen, Kerimov N, Kettunen J, Kilpelainen E, Kilpi T, Klinger K, Kosma VM, Kuopio T, Kurra V, Laisk T, Laukkanen J, Lawless N, Liu A, Longerich S, Magi R, Makela J, Makitie A, Malarstig A, Mannermaa A, Maranville J, Matakidou A, Meretoja T, et al. (2023) FinnGen provides genetic insights from a well-phenotyped isolated population. Nature 613: 508–518. doi: 10.1038/s41586-022-05473-8
44. Léger J, Marinovic D, Garel C, Bonaïti-Pellié C, Polak M, Czernichow P (2002) Thyroid developmental anomalies in first degree relatives of children with congenital hypothyroidism. The Journal of Clinical Endocrinology & Metabolism 87: 575–580.
45. Léger J, Olivieri A, Donaldson M, Torresani T, Krude H, Van Vliet G, Polak M, Butler G (2014) European Society for Paediatric Endocrinology consensus guidelines on screening, diagnosis, and management of congenital hypothyroidism. Hormone Research in Paediatrics 81: 80–103.

46. Lu AT, Cantor R (2012) Allowing for sex differences increases power in a GWAS of multiplex Autism families. Molecular psychiatry 17: 215–222.
47. Luboshitzky R, Oberman A, Kaufman N, Reichman N, Flatau E (1996) Prevalence of cognitive dysfunction and hypothyroidism in an elderly community population. Israel journal of medical sciences 32: 60–65.
48. Luningham JM, Chen J, Tang S, De Jager PL, Bennett DA, Buchman AS, Yang J (2020) Bayesian Genome-wide TWAS Method to Leverage both cis-and trans-eQTL Information through Summary Statistics. The American Journal of Human Genetics 107: 714–726.
49. Luo J, Martucci VL, Quandt Z, Groha S, Murray MH, Lovly CM, Rizvi H, Egger JV, Plodkowski AJ, Abu-Akeel M, Schulze I, Merghoub T, Cardenas E, Huntsman S, Li M, Hu D, Gubens MA, Gusev A, Aldrich MC, Hellmann MD, Ziv E (2021) Immunotherapy-Mediated Thyroid Dysfunction: Genetic Risk and Impact on Outcomes with PD-1 Blockade in Non-Small Cell Lung Cancer. Clin Cancer Res 27: 5131–5140. doi: 10.1158/1078-0432.CCR-21-0921
50. Maddirevula S, Kuwahara H, Ewida N, Shamseldin HE, Patel N, Alzahrani F, AlSheddi T, AlObeid E, Alenazi M, Alsaif HS, Alqahtani M, AlAli M, Al Ali H, Helaby R, Ibrahim N, Abdulwahab F, Hashem M, Hanna N, Monies D, Derar N, Alsagheir A, Alhashem A, Alsaleem B, Alhebbi H, Wali S, Umarov R, Gao X, Alkuraya FS (2020) Analysis of transcript-deleterious variants in Mendelian disorders: implications for RNA-based diagnostics. Genome Biol 21: 145. doi: 10.1186/s13059-020-02053-9
51. Makretskaya N, Bezlepkina O, Kolodkina A, Kiyaev A, Vasilyev EV, Petrov V, Kalinenkova S, Malievsky O, Dedov II, Tiulpakov A (2018) High frequency of mutations in'dyshormonogenesis genes' in severe congenital hypothyroidism. PLoS One 13: e0204323.
52. Matzaraki V, Kumar V, Wijmenga C, Zhernakova A (2017) The MHC locus and genetic susceptibility to autoimmune and infectious diseases. Genome Biol 18: 76. doi: 10.1186/s13059-017-1207-1
53. Napier C, Mitchell AL, Gan E, Wilson I, Pearce SH (2015) Role of the X-linked gene GPR174 in autoimmune Addison's disease. J Clin Endocrinol Metab 100: E187-90. doi: 10.1210/jc.2014-2694
54. Narumi S, Muroya K, Asakura Y, Adachi M, Hasegawa T (2010) Transcription factor mutations and congenital hypothyroidism: systematic genetic screening of a population-based cohort of Japanese patients. The Journal of Clinical Endocrinology & Metabolism 95: 1981–1985.
55. Panicker V (2011) Genetics of thyroid function and disease. Clin Biochem Rev 32: 165–75.
56. Park SM, Chatterjee VK (2005) Genetics of congenital hypothyroidism. J Med Genet 42: 379–89. doi: 10.1136/jmg.2004.024158
57. Patel J, Landers K, Li H, Mortimer R, Richard K (2011) Thyroid hormones and fetal neurological development. Journal of Endocrinology 209: 1.
58. Persani L, Bonomi M (2017) The multiple genetic causes of central hypothyroidism. Best Practice & Research Clinical Endocrinology & Metabolism 31: 255–263.
59. Persani L, Calebiro D, Cordella D, Weber G, Gelmini G, Libri D, de Filippis T, Bonomi M (2010) Genetics and phenomics of hypothyroidism due to TSH resistance. Mol Cell Endocrinol 322: 72–82. doi: 10.1016/j.mce.2010.01.008
60. Persani L, Rurale G, de Filippis T, Galazzi E, Muzza M, Fugazzola L (2018) Genetics and management of congenital hypothyroidism. Best Pract Res Clin Endocrinol Metab 32: 387–396. doi: 10.1016/j.beem.2018.05.002
61. Pirastu N, Cordioli M, Nandakumar P, Mignogna G, Abdellaoui A, Hollis B, Kanai M, Rajagopal VM, Parolo PDB, Baya N, Carey CE, Karjalainen J, Als TD, Van der Zee MD, Day FR, Ong KK, FinnGen S, andMe Research T, i PC, Morisaki T, de Geus E, Bellocco R, Okada Y, Borglum AD, Joshi P, Auton A, Hinds D, Neale BM, Walters RK, Nivard MG, Perry JRB, Ganna A (2021) Genetic analyses identify widespread sex-differential participation bias. Nat Genet 53: 663–671. doi: 10.1038/s41588-021-00846-7
62. Saevarsdottir S, Olafsdottir TA, Ivarsdottir EV, Halldorsson GH, Gunnarsdottir K, Sigurdsson A, Johannesson A, Sigurdsson JK, Juliusdottir T, Lund SH, Arnthorsson AO, Styrmisdottir EL, Gudmundsson J, Grondal GM, Steinsson K,

Alfredsson L, Askling J, Benediktsson R, Bjarnason R, Geirsson AJ, Gudbjornsson B, Gudjonsson H, Hjaltason H, Hreidarsson AB, Klareskog L, Kockum I, Kristjansdottir H, Love TJ, Ludviksson BR, Olsson T, Onundarson PT, Orvar KB, Padyukov L, Sigurgeirsson B, Tragante V, Bjarnadottir K, Rafnar T, Masson G, Sulem P, Gudbjartsson DF, Melsted P, Thorleifsson G, Norddahl GL, Thorsteinsdottir U, Jonsdottir I, Stefansson K (2020) FLT3 stop mutation increases FLT3 ligand level and risk of autoimmune thyroid disease. Nature 584: 619–623. doi: 10.1038/s41586-020-2436-0

63. Sakaue S, Kanai M, Tanigawa Y, Karjalainen J, Kurki M, Koshiba S, Narita A, Konuma T, Yamamoto K, Akiyama M, Ishigaki K, Suzuki A, Suzuki K, Obara W, Yamaji K, Takahashi K, Asai S, Takahashi Y, Suzuki T, Shinozaki N, Yamaguchi H, Minami S, Murayama S, Yoshimori K, Nagayama S, Obata D, Higashiyama M, Masumoto A, Koretsune Y, FinnGen, Ito K, Terao C, Yamauchi T, Komuro I, Kadowaki T, Tamiya G, Yamamoto M, Nakamura Y, Kubo M, Murakami Y, Yamamoto K, Kamatani Y, Palotie A, Rivas MA, Daly MJ, Matsuda K, Okada Y (2021) A cross-population atlas of genetic associations for 220 human phenotypes. Nat Genet 53: 1415–1424. doi: 10.1038/s41588-021-00931-x
64. Sharo AG, Zou Y, Adhikari AN, Brenner SE (2023) ClinVar and HGMD genomic variant classification accuracy has improved over time, as measured by implied disease burden. Genome Med 15: 51. doi: 10.1186/s13073-023-01199-y
65. Stoupa A, Kariyawasam D, Muzza M, de Filippis T, Fugazzola L, Polak M, Persani L, Carre A (2021) New genetics in congenital hypothyroidism. Endocrine 71: 696–705. doi: 10.1007/s12020-021-02646-9
66. Sun F, Zhang J-X, Yang C-Y, Gao G-Q, Zhu W-B, Han B, Zhang L-L, Wan Y-Y, Ye X-P, Ma Y-R (2018) The genetic characteristics of congenital hypothyroidism in China by comprehensive screening of 21 candidate genes. European journal of endocrinology 178: 623–633.
67. Szklarczyk D, Gable AL, Nastou KC, Lyon D, Kirsch R, Pyysalo S, Doncheva NT, Legeay M, Fang T, Bork P (2021) The STRING database in 2021: customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement sets. Nucleic acids research 49: D605-D612.
68. Tam V, Patel N, Turcotte M, Bosse Y, Pare G, Meyre D (2019) Benefits and limitations of genome-wide association studies. Nat Rev Genet 20: 467–484. doi: 10.1038/s41576-019-0127-1
69. Taylor PN, Albrecht D, Scholz A, Gutierrez-Buey G, Lazarus JH, Dayan CM, Okosieme OE (2018) Global epidemiology of hyperthyroidism and hypothyroidism. Nat Rev Endocrinol 14: 301–316. doi: 10.1038/nrendo.2018.18
70. Traglia M, Bseiso D, Gusev A, Adviento B, Park DS, Mefford JA, Zaitlen N, Weiss LA (2017) Genetic Mechanisms Leading to Sex Differences Across Common Diseases and Anthropometric Traits. Genetics 205: 979–992. doi: 10.1534/genetics.116.193623
71. Umar H, Muallima N, Adam J, Sanusi H (2010) Hashimoto's thyroiditis following Graves' disease. Acta Med Indones 42: 31–5.
72. Unnikrishnan AG, Kalra S, Sahay RK, Bantwal G, John M, Tewari N (2013) Prevalence of hypothyroidism in adults: An epidemiological study in eight cities of India. Indian journal of endocrinology and metabolism 17: 647.
73. Uzunlulu M, Yorulmaz E, Oguz A (2007) Prevalence of subclinical hypothyroidism in patients with metabolic syndrome. Endocr J 54: 71–6. doi: 10.1507/endocrj.k06-124
74. Wang F, Liu C, Jia X, Liu X, Xu Y, Yan S, Jia X, Huang Z, Liu S, Gu M (2017) Next-generation sequencing of NKX2.1, FOXE1, PAX8, NKX2.5, and TSHR in 100 Chinese patients with congenital hypothyroidism and athyreosis. Clin Chim Acta 470: 36–41. doi: 10.1016/j.cca.2017.04.020
75. Wang H, Kong X, Pei Y, Cui X, Zhu Y, He Z, Wang Y, Zhang L, Zhuo L, Chen C, Yan X (2020) Mutation spectrum analysis of 29 causative genes in 43 Chinese patients with congenital hypothyroidism. Mol Med Rep 22: 297–309. doi: 10.3892/mmr.2020.11078
76. Wang Q, Dhindsa RS, Carss K, Harper AR, Nag A, Tachmazidou I, Vitsios D, Deevi SVV, Mackay A, Muthas D, Huhn M, Monkley S, Olsson H, AstraZeneca Genomics I, Wasilewski S, Smith KR, March R, Platt A, Haefliger C, Petrovski S

(2021) Rare variant contribution to human disease in 281,104 UK Biobank exomes. Nature 597: 527–532. doi: 10.1038/s41586-021-03855-y

77. Wassner AJ (2020) Unraveling the Genetics of Congenital Hypothyroidism: Challenges and Opportunities. J Clin Endocrinol Metab 105. doi: 10.1210/clinem/dgaa454
78. Watanabe K, Taskesen E, van Bochoven A, Posthuma D (2017) Functional mapping and annotation of genetic associations with FUMA. Nat Commun 8: 1826. doi: 10.1038/s41467-017-01261-5
79. Wyne KL, Nair L, Schneiderman CP, Pinsky B, Antunez Flores O, Guo D, Barger B, Tessnow AH (2022) Hypothyroidism Prevalence in the United States: A Retrospective Study Combining National Health and Nutrition Examination Survey and Claims Data, 2009–2019. J Endocr Soc 7: bvac172. doi: 10.1210/jendso/bvac172
80. Zucker R, Linial M (2022) Recessive and sex-dependent genetic effects in primary hypertension. medRxiv: 2022.05.31.22275828.

Figures

Figure 1

Summary of independent loci identified from major GWAS results as compiled in the OTG portal. ***(A)*** *The number of participants in each study and the number of hypothyroidism cases are indicated by N (all) and n (cases). There are 21 variants that are shared by all six studies (colored red). The chromosomal position is shown (bottom, light blue).* ***(B)***

STRING analysis of the 21 mapped associated genes resulted in a network of 13 genes (interaction score >0.4). The nodes are colored by PPI clusters. Evidence of connectivity between the clusters is indicated by dashed lines. ***(C)*** *Connectivity of the 21 associated genes (****Table 1****) and their major functional annotations.*

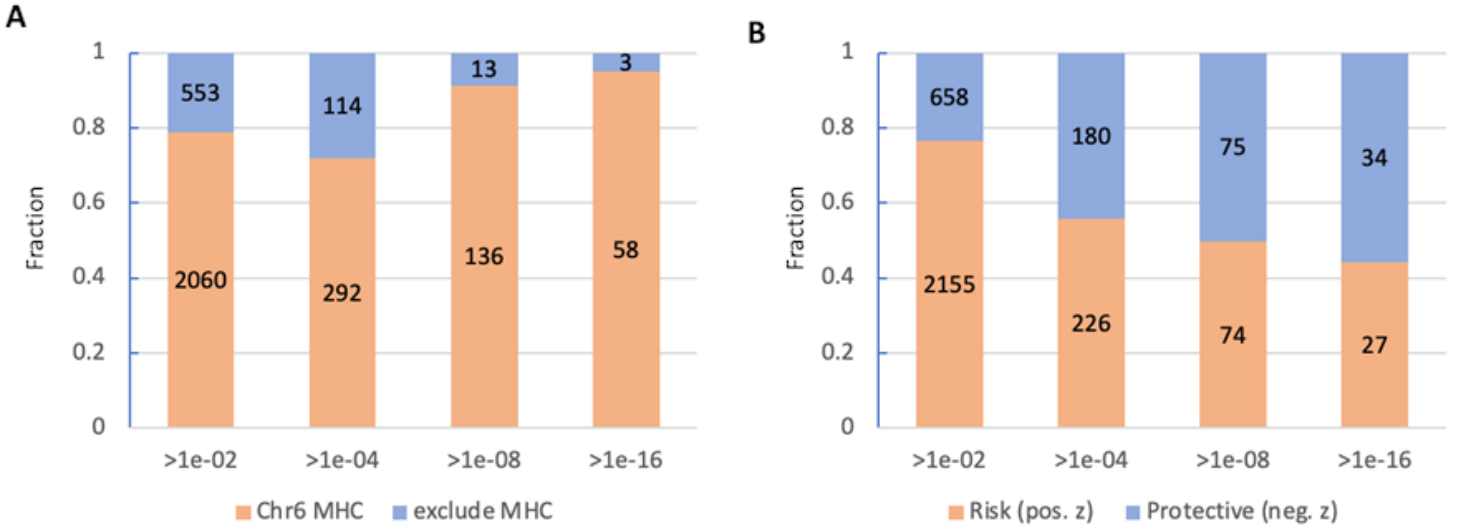

Figure 2

Partition of the significant coding GWAS variants at different thresholds. ***(A)*** *Position of the variants in the Chr6 MHC locus and in other locations. We consider the MHC locus to span between positions 25 M and 40 M on Chr6.* ***(B)*** *Partition according to the trend of variants that are protective or increase the risk for hypothyroidism.*

A

Symbol	Uniprot	Name	q-value
SH2B3	Q9UQQ2	SH2B adaptor protein 3	3.60e-40
C6orf15	Q6UXA7	chromosome 6 open reading frame 15	3.06e-37
CTLA4	P16410	cytotoxic T-lymphocyte associated protein 4	3.25e-30
DCLRE1B	Q9H816	DNA cross-link repair 1B	6.58e-30
TCF19	Q9Y242	transcription factor 19	4.11e-21
C6orf47	O95873	chromosome 6 open reading frame 47	8.03e-21
HSPA1L	P34931	heat shock protein family A (Hsp70) member 1 like	1.21e-16
AGER	Q15109	advanced glycosylation end-product specific receptor	9.52e-16
PSORS1C2	Q9UIG4	psoriasis susceptibility 1 candidate 2	5.84e-14
LST1	O00453	leukocyte specific transcript 1	2.59e-13
PSORS1C1	Q9UIG5	psoriasis susceptibility 1 candidate 1	5.11e-13
TRMO	Q9BU70	tRNA methyltransferase O	7.73e-11
PPP1R18	Q6NYC8	protein phosphatase 1 regulatory subunit 18	2.88e-10
HLA-DPA1	P20036	major histocompatibility complex, class II, DP alpha 1	4.76e-10
C4A	P0C0L4	complement C4A (Rodgers blood group)	6.84e-10
TSBP1	Q5SRN2	testis expressed basic protein 1	9.75e-10
HLA-DRB1	P01911	major histocompatibility complex, class II, DR beta 1	4.20e-9
MUC21	Q5SSG8	mucin 21, cell surface associated	5.96e-9
HLA-DRB5	Q30154	major histocompatibility complex, class II, DR beta 5	8.88e-9
PPT2	Q9UMR5	palmitoyl-protein thioesterase 2	9.46e-9
NOTCH4	Q99466	notch receptor 4	6.20e-8
CLECL1	Q8IZS7	C-type lectin like 1	2.97e-7
SFTA2	Q6UW10	surfactant associated 2	3.14e-7
RPP21	Q9H633	ribonuclease P/MRP subunit p21	3.45e-7
HLA-DQA2	P01906	major histocompatibility complex, class II, DQ alpha 2	3.96e-7
SH2D2A	Q9NP31	SH2 domain containing 2A	5.44e-7

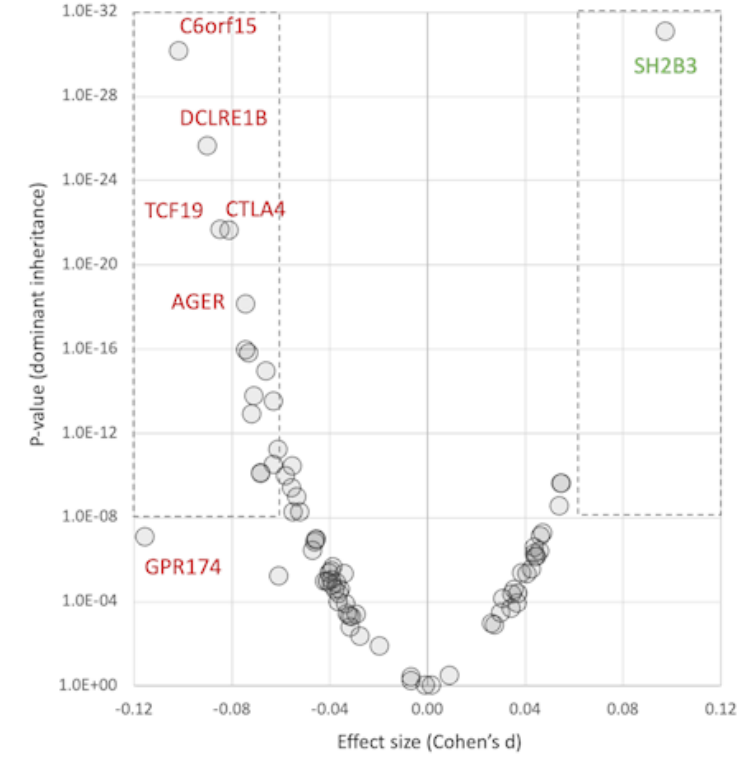

Figure 3

Associated genes from PWAS results. ***(A)*** *Statistically significant genes from PWAS for ICD-10 E03 with q-value <1e-07 (total 26 genes). Genes with an increased and decreased risk are colored purple/red and blue, respectively.* ***(B)*** *Effect size (Cohen's d) for PWAS results for the dominant model. The genes within the dashed frames are associated with Cohen's d >|0.06|. Positive (green font) and negative (red font) Cohen's d values are associated with reduced and increased risk, respectively. Additional file 1:* ***Table S4*** *lists all genes and their statistics.*

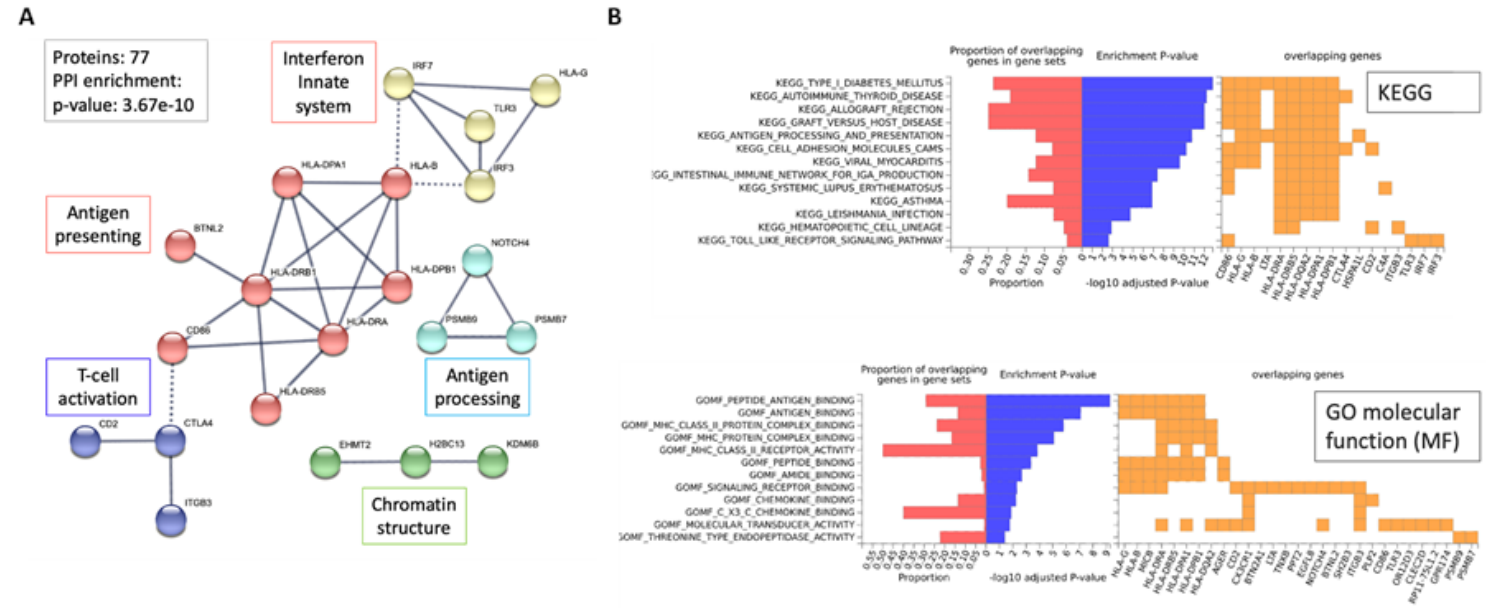

Figure 4

Network relationship and functional enrichment of PWAS results (77 genes). ***(A)*** *The STRING network represents the genes connected at an interaction score >0.9. Dashed lines mark the connections between clusters. The unified function for each cluster is colored and annotated (e.g., antigen processing).* ***(B)*** *Enrichment analysis using the FUMA-GWAS Gene2Func protocol. In red, the fraction of genes in the gene set; blue, the adjusted p value; orange, the overlapping genes for each term. The top 13 KEGG pathways and bottom, the GO_MF annotations. Note the enrichment of MHC genes (HLA-DPA1, HLA-DRB1, HLA-DRB5, HLA-B, HLA-DPA1, HLA-G) in KEGG and GO-MF analyses.*

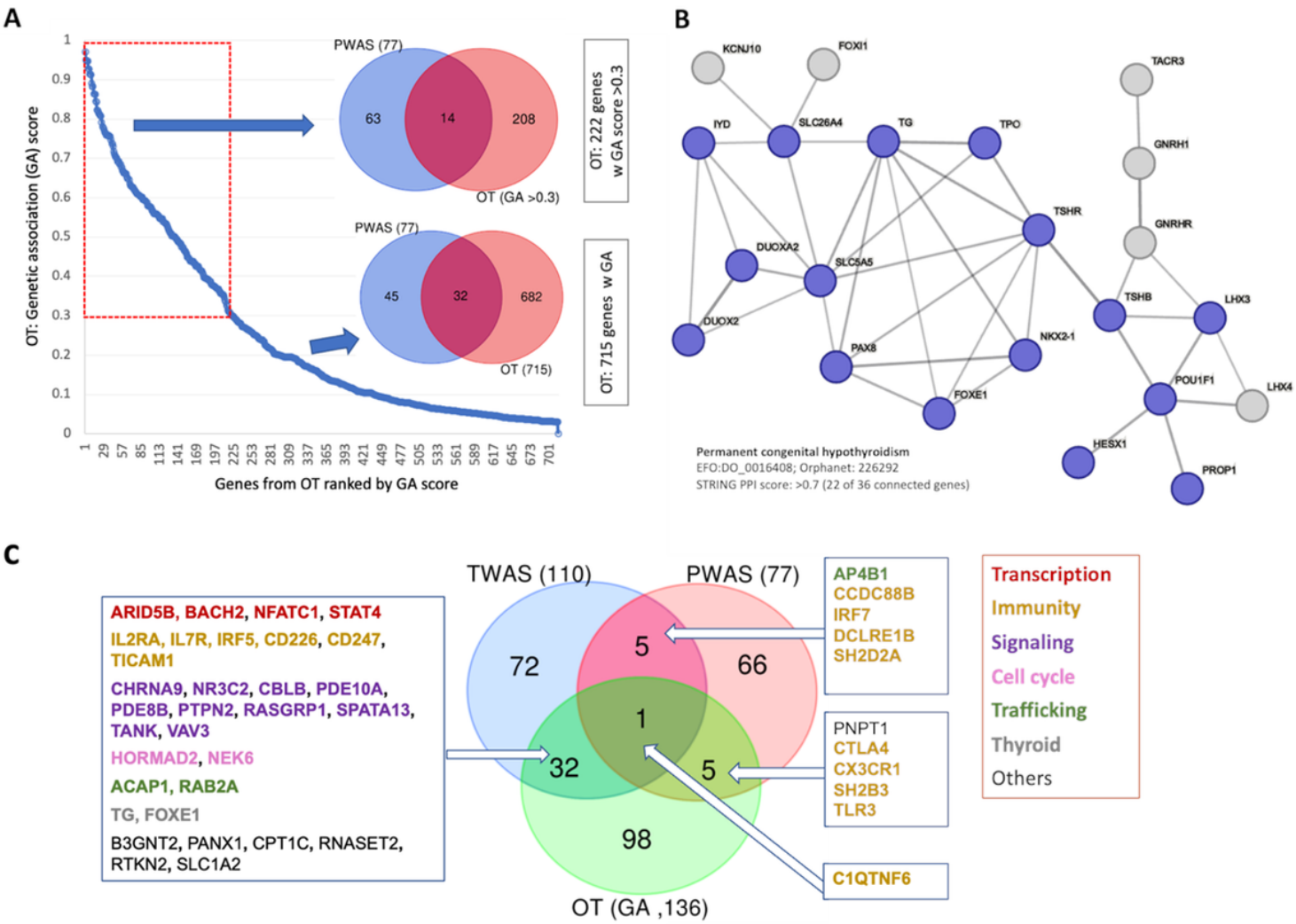

Figure 5

Genetic association with GWAS compiled by OTG. ***(A)*** *Ranked genes by their genetic association (by GA score, total of 715 genes). The overlap of 77 PWAS genes and 222 OT genes with GA scores >0.3 for all 715 genes.* ***(B)*** *A network relation of genes that are ranked by the OT global score >0.5 for the phenotype of permanent CH (total 36, 22 are connected, STRING PPI confidence score >0.7). The nodes are colored according to the match with the findings of CH causal genes from independent cohorts from India (Kollati et al. 2020) and China (Wang et al. 2020).* ***(C)*** *Venn diagram of major association resources: PWAS (77 genes), GWAS (OT, by genetic association score >0.5 (136 genes), and TWAS (transcription-based association study) for hypothyroidism/myxedema (self-reported) with 45 loci and 110 genes. The subsets of overlapping genes are color-coded according to their main functional annotations.*

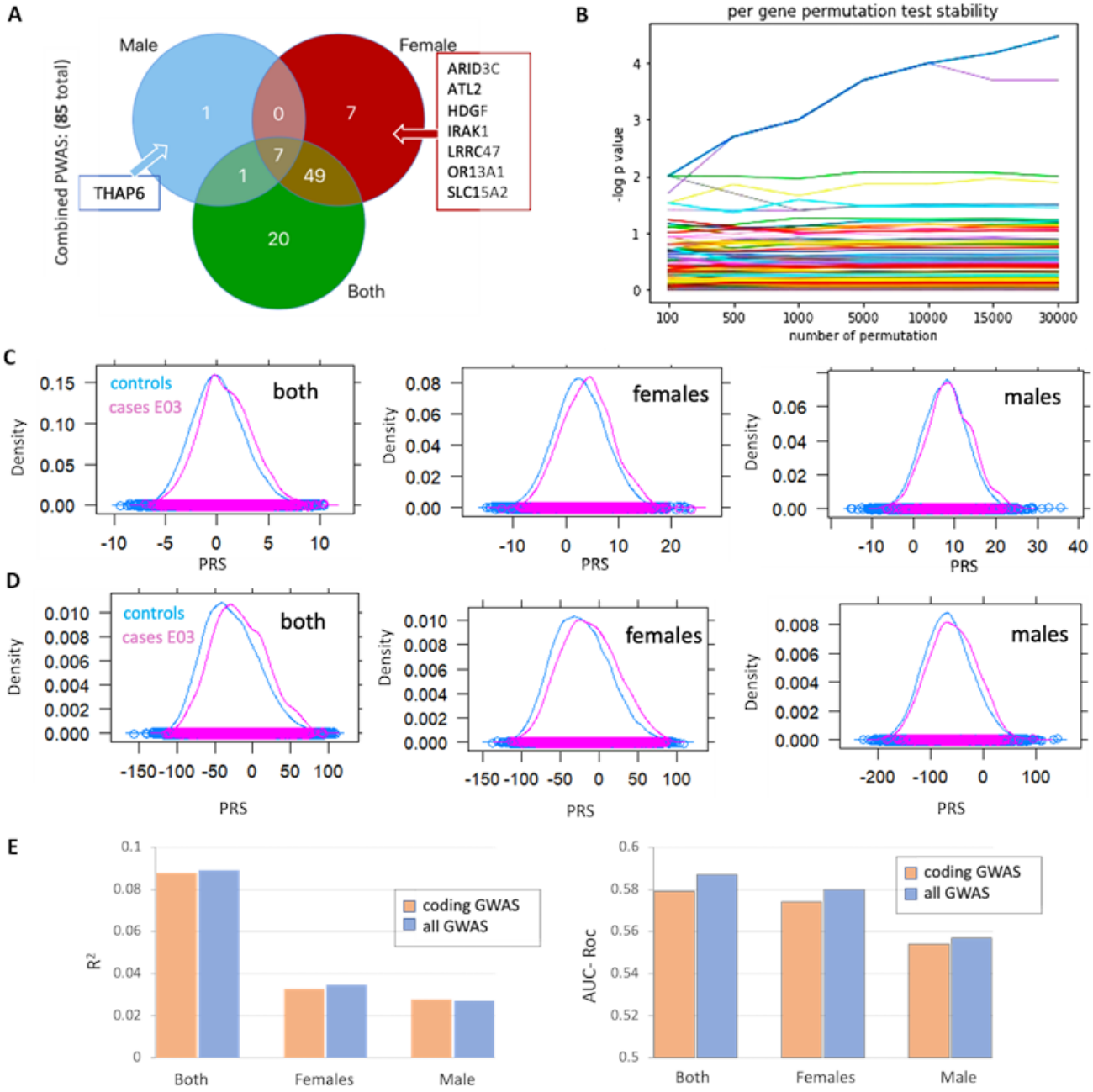

Figure 6

*Gene-based association analysis by sex **(A)** Venn diagram for the number of PWAS-identified genes that are significant (q-value <0.05) by sex. Female- and male-specific genes are listed. **(B)** The results of the permutation test between each gene found to be significant in PWAS for male and female populations. The y-axis is the received p value for each number of permutations (x-axis). Each gene is represented by a colored line. Distribution of a polygenic risk score (PRS) among individuals with and without E03 diagnosis, marked as cases (pink) and controls (blue). PRS scores were calculated for all-GWAS **(C)** and coding-GWAS **(D)** for the entire cohorts (both), females and males. **(E)** PRS prediction by the coefficient of determination (R^2, left) AUC-Roc (right) for coding GWAS (orange) and all GWAS (blue) for the entire cohort (both) and by sex. Coding GWAS variants partitioned by sex are listed in Additional file 1: **Table S8**.*

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- HGHypothyroidismE03Ver14SuppFigS1S5.docx