It is made available under a [CC-BY 4.0 International license](http://creativecommons.org/licenses/by/4.0/) .

Title

Genome-wide association study of thyroid-stimulating hormone highlights new genes, pathways and associations with thyroid disease susceptibility and age-of-onset

Authors
Alexander T Williams*†¹, Jing Chen⁺¹, Kayesha Coley¹, Chiara Batini^{1,2}, Abril Izquierdo^{1,2}, Richard Alexander T Williams*†*, Jing Chen†*, Kayesha Coley*, Chiara Batini**, Abril Izquierdo**, Richard
Packer^{1,2}, David J Shepherd¹, Robert C Free^{1,2}, Edward J Hollox³, Nigel J Brunskill⁴, Ioanna Ntalla¹,
Nicola Ree Packer*'*, David J Shepherd*, Robert C Free*'*, Edward J Hollox*, Nigel J Brunskill*, Ioanna Ntalla*
Nicola Reeve¹, Christopher E Brightling⁵, Laura Venn¹, Emma Adams¹, Catherine Bee¹, Susan Wa
Manish Pareek⁵, $\frac{1}{2}$ Packer^{1,2}, David J Shepherd¹, Robert C Free^{1,2}, Edward J Hollox³, Nigel J Brunskill⁴, Ioanna Ntalla¹, ,
||
! Nicola Reeve*, Christopher E Brightling", Laura Venn*, Emma Adams*, Catherine Bee*, Susan Wallace*
Manish Pareek⁵, Anna L Hansell¹, William Hennah^{6,7,8}, Balasubramanya S Rao⁹, Frank Dudbridge¹,
Louise V Wain^{1,2} $\frac{1}{2}$ Christopher E Brightling", $\frac{r}{\sqrt{2}}$ \overline{a} Manish Pareek^s, *I*
Louise V Wain^{1,2},
<mark>Affiliations</mark> , Anna L Hansell', William Hennah^{o,,,o}, Balasubramanya S Rao³
², Nick Shrine¹, Martin D Tobin‡^{1,2}, Catherine John*‡^{1,2}.
S , Frank Dudbridge⁺
|
| \overline{a} Louise V Wain*'*, Nick Shrine*, Martin D Tobin‡*'*, Catherine John*‡*'*.
Affiliations
1 Department of Population Health Sciences, University of Leicester

-
- 2. National Institute for Health Research, Leicester Respiratory Biomedical Research Centre,
University of Leicester, Leicester, UK
- 3. Department of Genetics and Genome Biology, University of Leicester, Leicester, UK University of Leicester, Leicester, Lei
Department of Cardiovascular Scienc
Department of Respiratory Sciences,
-
- 4. Department of Cardiovascular Sciences University of Leicester, Leicester, UK
5. Department of Respiratory Sciences, University of Leicester, Leicester, UK
6. Orion Pharma, Espoo, Finland 5. Department of Respiratory Sciences, University of Leicester, Leicester, UK
6. Orion Pharma, Espoo, Finland
7. Neuroscience Center, HiLIFE, University of Helsinki, Helsinki, Finland
-
- 6. Orion Pharma, Espoo, Finland
7. Neuroscience Center, HiLIFE, University of Helsinki, Helsinki, Finland
8. Institute for Molecular Medicine FIMM, HiLIFE, University of Helsinki, Helsi 7. Neuroscience Center, HiLIFE, L
8. Institute for Molecular Medici
9. Orion Pharma, Nottingham, UI
- 7. Neuroscience Center, Hille, Center, Processor, Processor, Hillie, University of Helsinki
9. Orion Pharma, Nottingham, UK
6. Processor Marma, Helsinki, Finland and Pharma, Nottingham, UK
-

8. Institute for Molecular Medicine Financial Medicine Financial Actions, Finland, Finland, Finland, Finland, Finland
8. Octresponding author statement eral Corresponding author stater
* = corresponding authors Corresponding author statement

-
- * = corresponding authors

Corresponding author email: atw20@leicester.ac.uk (Alexander T Williams), cj153@leicester.ac.uk
(Catherine John) Corresponding author email: <u>attaces of corresponding</u>
|Catherine John)
Abstract

(Catherine Thyroid bormone)
Thyroid bormone Abstract

Thyroid hormones play a critical role in regulation of multiple physiological functions and thyroid
dysfunction is associated with substantial morbidity. Electronic health records were used to
undertake the largest genomedy statements assessed with substantial morthingly substanting formulating hormone (TSH) undertake the largest genome-wide association study of thyroid-stimulating hormone (TSH) with a total sample size of 247,107. We iden undertake the largest genome-mile association study of thyroid-stimulating hormone (TSH) is with a total sample size of 247,107. We identified 158 novel signals, more than doubling the number
of known associations with TSH of known associations with TSH, and implicating 112 putative causal genes, of which 78 were not
previously implicated. For the first time, we demonstrate that a polygenic score for TSH was
associated with TSH levels in all previously implicated. For the first time, we demonstrate that a polygenic score for TSH was
associated with TSH levels in all ancestries in UK Biobank, and strongly predicted age of onset of
hypothyroidism and hyperthyroi previously implicated. For the first time, we demonstrate that a polygenic score for the first time
associated with TSH levels in all ancestries in UK Biobank, and strongly predicted age of onse
hypothyroidism and hyperthy associated with TSH levels in all ancestries in all ancestry participants. We developed pathway
specific genetic risk scores for TSH levels and used these in phenome-wide association studies to
identify potential consequen potential utility of genetic associations to inform future therapeutics and risk prediction for thyro specific general interaction of the second the second product in phenome-mixed these indings demonstrate is
potential utility of genetic associations to inform future therapeutics and risk prediction for thyro
diseases. potential utility of genetic associations to inform future therapeutics and risk prediction for thyroid
diseases. potential utility of genetic associations to informations to information for the risk predictions of the risk predictions of the risk prediction for the risk prediction for the risk prediction for the risk prediction for t

It is made available under a CC-BY 4.0 International license.

Introduction
Thyroid hormones are essential for energy metabolism and act on almost all cells. Thyroid
dysfunction is associated with secondary cardiovascular, mental health, ophthalmic and other
disease(1). Hypothyroidism disease(1). Hypothyroidism has a high prevalence(2) and is most commonly due to autoimmune 0.2%-1.3%, is most commonly due to autoimmune (Graves) disease or toxic nodular goitre.(1)
Ageing, diet (including iodine deficiency), smoking status, genetic susceptibility, ethnicity, and (Hashimon) the set of the total monormal (Graves) disease or toxic nodular goitre.(1)
Ageing, diet (including iodine deficiency), smoking status, genetic susceptibility, ethnicity, and
endocrine disruptors are risk factors Ageing, diet (including iodine deficiency), smoking status, genetic susceptibility, ethnicity, and
endocrine disruptors are risk factors for thyroid diseases; defining genetic variants, genes, prot
and pathways associated endocrine disruptors are risk factors for thyroid diseases; defining genetic variants, genes, prot
and pathways associated with hypothyroidism and hyperthyroidism will inform a deeper
understanding of the mechanisms of thy and pathways associated with hypothyroidism and hyperthyroidism will inform a deeper
understanding of the mechanisms of thyroid disease and inform prevention and treatment
strategies. and pathways are stated with hyperhyroiding with hyperhyroidism and hyperthyroidism and hyperthyroidism will inform a deeper
strategies.

understanding of the mechanisms of thyroid disease and inform prevention and treatments
strategies.
Genome-wide association studies (GWAS) of quantitative traits have been particularly powe ی
Genome-w
successful i Successful in identifying new drug targets(3-5). Most genetic associations for thyroid disorders were
discovered in GWAS of thyroid-stimulating hormone (TSH) levels, a sensitive marker of thyroid
function which is suppress discovered in GWAS of thyroid-stimulating hormone (TSH) levels, a sensitive marker of thyroid
function which is suppressed when levels of thyroxine (T₄) and triiodothyronine (T₃) are high and
elevated when T₄ and T function which is suppressed when levels of thyroxine (T₄) and triiodothyronine (T₃) are high an
elevated when T₄ and T₃ levels are low. The largest GWAS to date, including 119,715 participan
brought the number of elevated when T₄ and T₃ levels are low. The largest GWAS to date, including 119,715 participants,
brought the number of known genetic signals for TSH to 99, highlighting associations with
hypothyroidism, hyperthyroidis brought the number of known genetic signals for TSH to 99, highlighting associations with
hypothyroidism, hyperthyroidism and thyroid cancer(6).

hypothyroidism, hyperthyroidism and thyroid cancer(6).
Electronic health records (EHR) are increasingly utilised in genomic studies(7, 8). In the UK, hyperthyroidism, hyperthyroidism and thyroid cancer(e).
Electronic health records (EHR) are increasingly utilised in
care EHR have been recorded prospectively for more tha Electronic health records (EHR) are increasingly alternatingly and transfer (7, 9). In the UK, primary
care EHR have been recorded prospectively for more than 25 years. TSH is frequently measured in
primary care because th primary care because thyroid disease may present with non-specific symptoms or be asymptomation.
Through harnessing such TSH measures, our study included 247,107 participants, more than
doubling the size of the largest stu primary care because any present with present with non-specific symptoms or because than
Through harnessing such TSH measures, our study included 247,107 participants, more than
doubling the size of the largest study to da Through harmesting such terms in each process, increasing the number of genetic signals for TS
19 to 260. Using these 260 signals we then (i) tested the association between TSH-associated
variants and disease; (ii) fine-ma 99 to 260. Using these 260 signals we then (i) tested the association between TSH-associated
variants and disease; (ii) fine-mapped signals through annotation-informed credible sets; (iii) a
a consensus-based framework to variants and disease; (ii) fine-mapped signals through annotation-informed credible sets; (iii) applied
a consensus-based framework to systematically investigate and identify putative causal genes,
integrating eight locusvariants and disease; (ii) fine-disperse; (iii) fine-disperse; (iii) disease; and identify putative causal genes,
integrating eight locus-based or similarity-based criteria; (iv) developed and applied a polygenic
score (PG integrating eight locus-based or similarity-based criteria; (iv) developed and applied a polygenic
score (PGS) for TSH to show associations with susceptibility and age-of-onset of thyroid disease
applied phenome-wide assoc integrating eight locus-based or similarity-based criteria; (iv) developed and applicate perspectively.
Score (PGS) for TSH to show association studies (PheWAS) to individual variants, the PGS, and molecu
pathway-specific score (PGS) for TSH to show an account in the sum of the WAS) for individual variants, the PGS, and molecular
pathway-specific genetic risk scores (GRS). Through evidence from the above, we aimed to define
putative causal ...
pathway-specific genetic risk scores (GRS). Through evidence from the above, we aimed to define
putative causal genes, and provide new insights into the mechanistic pathways underlying thyroid
disorders and their relat putative causal genes, and provide new insights into the mechanistic pathways underlying thyroid
disorders and their relationship to other long-term conditions to inform relevant drug therapies. putative causal genes, and provide new insights into the mechanism pathways underlying thyroid
disorders and their relationship to other long-term conditions to inform relevant drug therapies.
Results

Results

In UK Biobank and the EXCEED study we undertook GWAS with TSH levels, using an inverse normal In UK Biomation and adjusting for age, genotyping array, sex and the first 10 principal components of the extra
Incestry (Online Methods). Across the two studies, we analysed 53,518,359 genetic variants in
I 27.392 Europea ancestry (Online Methods). Across the two studies, we analysed 53,518,359 genetic variants in
127,392 European ancestry (EUR) participants and subsequently meta-analysed these GWAS
summary statistics with those from the in ancestry (Online Methods). Across the two studies, we analysed 33,318,339 genetic variants in
127,392 European ancestry (EUR) participants and subsequently meta-analysed these GWAS
summary statistics with those from the in 1281, Allemary statistics with those from the independent European-ancestry populations represer
the analysis by Zhou et al, (6) bringing the total sample size to 247,107 participants. the analysis by Zhou et al, (6) bringing the total sample size to 247, 107 participants.

TSH association with 260 variants

Using annotation informed fine-mapping (Online Methods), and a genome-wide significance threshold of P<5×10⁸, we identified 260 independent signals for TSH at 156 unique genomic loci, of
which 158 signals at 78 genomic loci are new. 212 of the 260 independent sentinels (118 of the 158 threshold of P<5×10°, we identified 260 independent signals for TSH at 156 unique genomic loci, of
which 158 signals at 78 genomic loci are new. 212 of the 260 independent sentinels (118 of the 158
novel signals) showed co movel signals) showed consistent direction of effect and P<0.01 in at least two of the three
2 novel signals) showed consistent direction of effect and P \geq

It is made available under a [CC-BY 4.0 International license](http://creativecommons.org/licenses/by/4.0/) .

contributing datasets in addition to reaching P<5×10⁻⁸ in the meta-analysis (**Supplementary Table 1**).
The SNP heritability of TSH was 16.6% (95% CI: 12.5%, 20.7%); together the 260 signals explain
22.6% of the TSH SNP h 22.6% of the TSH SNP heritability (Online Methods). The median number of variants per 95%
credible set was 3, and 167 (64%) of credible sets had a putative causal variant with a posterior
probability of association >50% 22.6% of the TSH SNP heritability (Online Methods). The median number of variants per 95%
credible set was 3, and 167 (64%) of credible sets had a putative causal variant with a posteric
probability of association >50%. probability of association >50%.
Identification of putative causal genes and causal variants

probability of association >50%.
Identification of putative causal genes and causal variants
To better understand the functional relevance of our signals, we undertook comprehensive variantexpression quantitative trait loci (eOTL) or (iv) protein quantitative trait loci (pOTL) within the to gene mapping by integrating evidence from eight methods: (i) included gene to the senting variant; (ii) the gene with the highest polygenic priority score (PoPS)(9); identification of (iii)
expression quantitative trait expression quantitative trait loci (eQTL) or (iv) protein quantitative trait loci (pQTL) within the redible sets; (v) proximity to a gene for a thyroid-associated Mendelian disease (±500kb); (v
annotation-informed credible expedible sets; (v) proximity to a gene for a thyroid-associated Mendelian disease (±500kb); (vi) credible set containing a missense/deleterious/damaging variant with a missense experience training a missense/deleterious/d annotation-informed credible set containing a missense/deleterious/damaging variant with a
posterior probability of association >50%; (vii) identification of a rare variant (±500kb of a TSH
sentinel) association with hyno posterior probability of association >50%; (vii) identification of a rare variant (±500kb of a TSH
sentinel) association with hypo- or hyperthyroidism using whole-exome(10) and whole-genome posterinel) association with hypo- or hyperthyroidism using whole-exome(10) and whole-genom
sequencing(11) resources; and (viii) proximity to a human ortholog of a mouse knockout gene
thyroid-related phenotyne (+500kb) sequencing(11) resources; and (viii) proximity to a human ortholog of a mouse knockout gene w
thyroid-related phenotype (±500kb). thyroid-related phenotype (±500kb).
We identified 112 putative causal genes satisfying ≥2 criteria, of which 30 were supported by ≥3

therm
We identified 112 putative causal ger
criteria (**Figure 1, Supplementary Tal** We identified 212 putative causal genes satisfying 22 criteria, of which 3 were supported by 22
criteria (Figure 1, Supplementary Table 2), and compared these to 67 genes implicated typically
single criterion in previous s single criterion in previous studies (**Supplementary Table 3)** (6, 12).
Of the 112 putative causal genes supported by ≥2 criteria, 78 genes have not been previously

single criterion in previous studies (Supplementary Table 3)(6, 12).
Of the 112 putative causal genes supported by \geq criteria, 78 genes
implicated in TSH level. The remaining genes were supported by add implicated in TSH level. The remaining genes were supported by additional criteria compared
the original reports, among which were 15 genes supported by additional signals (Supplemen
Table 2). Among the 30 genes supporte the original reports, among which were 15 genes supported by additional signals (**Supplementary**
T**able 2**). Among the 30 genes supported by ≥3 criteria, 18 were not previously implicated in TSH
levels (ADCY6, ANXA5, BCAS the original reports, among which were 15 genes supported by additional signals (Supplementary
Table 2). Among the 30 genes supported by ≥3 criteria, 18 were not previously implicated in TSH
levels (ADCY6, ANXA5, BCAS3, B Table 2). Among the 30 genes supported by 23 criteria, 18 were not previously implicated in TSH
levels (ADCY6, ANXA5, BCAS3, BNC2, CADM1, HMGA2, KIAA1217, KRT18, PDE4D, PHC2, PTEN,
SDCCAG8, SGK1, SMOC2, SPPL3, SULF1, TRIM2 levels (ADC*TO, ANXA3, BCA33, BNC2, CADM1, HMGA2, KIAA1217, KRT18, PDL4D, PHC2, PTEN,*
SDCCAG8, SGK1, SMOC2, SPPL3, SULF1, TRIM2, TSHZ3, novel genes shown in bold) and 12 were
previously reported genes supported by additio SDCCAG8, SGK1, SMOC2, SFT 13, SOLT1, TRIM2, TSH23, HOVET genes shown in bold) and 12 were
previously reported genes supported by additional criteria compared with the original reports,
among which were 6 genes also support .
among which were 6 genes also supported by additional novel signals (TG, TSHR, GLIS3, IGFBP5
PTPRS, SPATA13). The 30 genes supported by ≥3 criteria include genes involved in transcription
regulation (**BNC2, HMGA2, PHC2,** among which were 6 genes also supported by additional novel signals (TO, TSHR, G2155, IGT BF5,
PTPRS, SPATA13). The 30 genes supported by ≥3 criteria include genes involved in transcriptiona
regulation (BNC2, HMGA2, PHC2, PTPRS, SPATA13). The 30 genes supported by 23 criteria include genes involved in transcriptional
Pregulation (BNC2, HMGA2, PHC2, TSHZ3, GLIS3), production, signalling or response to thyroid
hormones (TG, TPO, TSHR, PDE4D) regulation (DNC2, TMOA2, FTC2, TSH23, GEB31), production, signaling or response to thyroid
hormones (TG, TPO, TSHR, **PDE4D**) or non-thyroid hormones (**ADCY6**, INSR, NR3C2), regulatior
thyroid-relevant pathways (**HMGA2**, IG hormones (TG, TPO, TSHR, P**DL4D**) or non-thyroid hormones (ADCTO, HSR, NR3C2₎, regulation of
thyroid-relevant pathways (**HMGA2**, IGFBP5), neuronal protection and neuropathies (**ADCY6**,
TRIM2), angiogenesis (**SDCCAG8**). thyroid-relevant pathways (HMGA2, IGFBF5), neuronal protection and neuropathies (ADCY6,
TRIM2), angiogenesis (SDCCAG8).
and ciliogenesis (SDCCAG8). and ciliogenesis (*SDCCAG8).*
To supplement understanding of the biological pathways and clinical phenotypes influenced by TSH-

associated variants, we first tested associations with circulating free T_4 levels, hypothyroidism,
hyperthyroidism (Figure 1, Supplementary Table 4), thyroid cancer and other thyroid disease. Using DeepPheWAS(7), we then undertook PheWAS of selected individual variants which mapped to hyperthyroidism (**Figure 1, Supplementary Table 4**), thyroid cancer and other thyroid disease. I
DeepPheWAS(7), we then undertook PheWAS of selected individual variants which mapped to
putative causal genes implicated by ≥ ny perthyroidism (Figure 1, Supplementary Table 4), thyroid cancer and other thyroid disease. Osing
DeepPheWAS(7), we then undertook PheWAS of selected individual variants which mapped to
putative causal genes implicated b putative causal genes implicated by \geq 3 criteria or by a single putative causal missense variant (posterior probability >50%).

putative causal genes implication by ≥3 criteria or by a single putative causal implication (posterior probability >50%).
TSH signals implicating putative causal genes with ≥2 variant-to-gene mapping criteria show v (posterior probability) = 50%).
TSH signals implicating putati
patterns of association with h patterns of association with hypothyroidism, hyperthyroidism, thyroid cancer and other thyroid
diseases. Among these are sentinel variants associated with hypothyroidism but not patterns of associated with hypothyroidism but not allocated with seasociated with hypothyroidism but not and o
hyperthyroidism (implicating AKT1, IGFBP5, INSR, GLIS3, SPATA13, **CADM1, BCAS3**, SASH1, **PDE4**
VEGFC, **SPPL3,** hyperthyroidism (implicating AKT1, IGFBP5, INSR, GLIS3, SPATA13, **CADM1, BCAS3**, S
VEGFC, **SPPL3, BNC2**, PDE10A, PDE8B, **NR3C2**, VAV3, SOX9, B4GALNT3, **CGA**, C9orf92 hyperthyroidism (implicating AKT1, IGFBF5, INSR, GEBS5, STATA13, CADM1, BCA33, SASH1, I DE4D,
VEGFC, **SPPL3**, BNC2, PDE10A, PDE8B, NR3C2, VAV3, SOX9, B4GALNT3, CGA, C9orf92, NEK6, NSF,
3 VEGFC, SPPL3, BNC2, PDE10A, PDE8B, NR3C2, VAV3, SOX9, B4GALNT3, CGA, C9orf92, NEK6, NSF,

It is made available under a [CC-BY 4.0 International license](http://creativecommons.org/licenses/by/4.0/) .

CCBE1, GNG7, TPPP, WNT4, SNX8, C1orf116, RBM47, KCTD5, PPP2R1B, PTPRJ, OCLN, C9orf156,
GATA3, WWTR1, MAL2, ZBTB17, ARNT, CDC16), sentinels associated with hyperthyroidism but not GATA3, WWTR1, MAL2, 2DTD17, ARNT, CDC10, SERRIFIERS associated with hyperthyroidism but not
hypothyroidism (implicating TSHR, NRG1, **SDCCAG8, HMGA2**, APOH, BMP2, SMAD6), and sentinels
(implicating TPO, PTEN, CAPZB, VEGFA, hypothyroidism (implicating *TSHR, NRG1, SDCCAG8, HMOA2, APOH, BMP2, SMAD0*), and sentifiers
associated with both hypothyroidism and hyperthyroidism with an opposite direction of effect
(implicating *TPO, PTEN, CAPZB,* (implicating TPO, PTEN, CAPZB, VEGFA, NFIA, MBIP, HLA-C, SLC25A37, SLC25A37, NKX2-1, NKX2-3,
FOXA2). However, tolerated SH2B3 misssense variant, rs3184504 (allele T), associated with (implicating TPO, PTEN, CAPZD, VEGPA, NTIA, MDIP, HEA-C, SEC23A37, SEC23A37, NKX2-1, NKX2-3,
FOXA2). However, tolerated SH2B3 misssense variant, rs3184504 (allele T), associated with
increased TSH, was associated with incr FOXA2). However, tolerated SH2B3 misssense variant, rs3184304 (allele T), associated with
increased TSH, was associated with increased risk of both hypothyroidism and hyperthyroid
in our PheWAS with increased risk of other in our PheWAS with increased risk of other autoimmune disorders and pleiotropic associations with
many traits (**Supplementary Tables 4 and 5**). Furthermore, the TSH-increasing allele of *SGK1* intronic
SNP rs1743963, were in our Pherma in International Pheme Line Carettinian and the Pherma procedulation in the many traits (Supplementary Tables 4 and 5). Furthermore, the TSH-increasing allele of *SGK1* intronic
SNP rs1743963, were associated many traits (Supplementary Tables 4 and 5). Furthermore, the TSH-increasing allele of SONI intronic
SNP rs1743963, were associated with decreased risk of both hypothyroidism and hyperthyroidism
and in our PheWAS, with incr and in our PheWAS, with increased calcium levels. SGK1, encoding serum glucocorticoid regulated
kinase 1, is involved in regulation of ion channels and stress response. kinase 1, is involved in regulation of ion channels and stress response.
Among relevant clinical phenotypes in the PheWAS is "secondary hypothyroidism", defined by

PheCode 244.1 which encompasses ICD codes for hypothyroidism due surgery, ablation or
medicaments used in treating hyperthyroidism. This code therefore represents consequences of Among relevant clinical phenotypes in the Phenotypes in the Phenotypes in the PheCode 244.1 which encompasses ICD codes for hypothyroidism due surgery, ablation or medicaments used in treating hyperthyroidism. This code th medicaments used in treating hyperthyroidism. This code therefore represents consequent
The Code 24 treated hyperthyroidism (and is unrelated to central hypothyroidism, which was sometime
previously referred to as secondar treated hyperthyroidism (and is unrelated to central hypothyroidism, which was sometimes
previously referred to as secondary hypothyroidism). We excluded these codes from our definitio
of hypothyroidism (Figure 1). When we previously referred to as secondary hypothyroidism). We excluded these codes from our definition
of hypothyroidism (Figure 1). When we explored single SNP associations with treated previously referred to assect on the secondary of hypothyroidism (Figure 1). When we explored single SNP associations with treated
hyperthyroidism we found consistent directions of association with hyperthyroidism clinical of hypothyroidism (Figure 1). When we explored single SNP associations with treated
hyperthyroidism we found consistent directions of association with hyperthyroidism c
and instances of associations, implicating genes *IGF* and instances of associations, implicating genes *IGFBP5, CADM1, SOX9, BMP2, TGFB2* and **SYT13**,
which were not detected by only studying hyperthyroidism codes in UK Biobank, consistent with
earlier onset hyperthyroidism which were not detected by only studying hyperthyroidism codes in UK Biobank, consistent with

Of the TSH sentinel variants implicating putative causal genes, a minority were associated with free T₄ (TPO, IGFBP5, INSR, GLIS3, NRG1, PDE10A, PDE8B, CAPZB, VEGFA, NFIA, MBIP, **HLA-C**, NEK6,
CERS6, CCBE1, GNG7, PTPRJ, KANK1, C9orf156, NKX2-1, NKX2-3, GATA3), and sentinels that were T₄ (TPO, IGFBP5, INSR, GLIS3, NRG1, PDE10A, PDE8B, CAPZB, VEGFA, NFIA, MBIP, **HLA-C**, NEK6,
CERS6, CCBE1, GNG7, PTPRJ, KANK1, C9orf156, NKX2-1, NKX2-3, GATA3), and sentinels that were
not associated with T₄ included se T₄ (TPO, IGFBP5, INSR, OLISS, NNO1, PDL10A, PDL0B, CAPZB, VLOPA, NFIA, NIBIP, **HLA-C**, NENO,
CERS6, CCBE1, GNG7, PTPRJ, KANK1, C9orf156, NKX2-1, NKX2-3, GATA3), and sentinels that we
not associated with T₄ included s CERSO, CCBE1, GNO7, PTPRJ, RANK1, CSUJ130, NKX2-1, NKX2-5, OATA3), and sentinels that were
not associated with T₄ included sentinels associated with hypothyroidism or hyperthyroidism (*TG*,
AKT1, TSHR, SPATA13, **CADM1, B** not associated with T4 included sentifiers associated with hypothyroidism or hyperthyroidism (TG,
AKT1, TSHR, SPATA13, **CADM1, BCAS3**, SASH1, **PTEN, PDE4D,** VEGFC, **SGK1, SPPL3, SDCCAG8,**
KDR, FOXA2, WWTR1, TGFB2, MAL2, AKT1, TSHR, SPATA13, **CADM1, BCAS3**, SASH1, **PTEN, PDE4D,** VEGFC, **SGK1, SPPL3, SDCCAG8,**
HMGA2, NR3C2, VAV3, **CGA**, SH2B3, **GNG7,** APOH, TPPP, WNT4, **SNX8**, BMP2, **C1orf116, KCTD5**, TSH levels is more sensitive approach to detecting genetic associations relevant to thyroid disease than the study of T_4 levels.

TSH ISH IS MOTTO DISTRIBUTE APPLICATION CONTROL SERVERTHEN TO A MOTHT THAT THE MOTHT.
Genes implicated by a single putative causal missense variant that was deleterious included SPATA Genes implicated by a single
ADCY6 and APOH. Apolipor Genes implicated by a single putative causal missense variant that was deletenous included SPATA6,
ADCY6 and APOH. Apolipoprotein H is involved in lipoprotein metabolism, coagulation and
haemostasis. The G allele of the AP ADCTO and APOH. Aponpoprotein H13 involved in lipoprotein inclabolism, coagulation and
haemostasis. The G allele of the APOH missense deleterious variant, rs1801690 (minor allel
frequency [MAF] 5.7% in EUR), was associated maemostasis. The G allele of the APOH missense deleterious variant, rs1801690 (minor allele
frequency [MAF] 5.7% in EUR), was associated with reduced TSH, increased risk of hyperthyre
increased risk of congenital anomalies frequency [MAF] 5.7% in EUR), was associated with reduced to by mensure with reduced risk of congenital anomalies of endocrine glands and thyrotoxicosis in the UK Biobank
DeepPheWAS analysis, as well as increased aspartate DeepPheWAS analysis, as well as increased aspartate aminotransferase (AST) and alanine aminotransferase (AST) levels, increased height, reduced triglycerides, reduced carotid intima media
thickness and, in FinnGen(13), reduced deep venous thrombosis risk. **C9orf156** (encoding TRNA
Methyltransferase O) was im thickness and, in FinnGen(13), reduced deep venous thrombosis risk. C9orf156 (encoding TRNA thickness and, in FinnGen(13), reduced deep venous thrombosis risk. C90/3130 (encoding TRNA
Methyltransferase O) was implicated by a single putative causal tolerated missense variant,
rs2282192 (T allele, frequency 28.8% i rs2282192 (T allele, frequency 28.8% in EUR), associated with increased TSH, increased
hypothyroidism risk and decreased risk of nontoxic multinodular goitre and thyroid can
lower mean corpuscular volume and HbA1c. rs 22822192 (T allele, frequency 28.8%) in EUR), associated with increased to thy, increased
hypothyroidism risk and decreased risk of nontoxic multinodular goitre and thyroid can
lower mean corpuscular volume and HbA1c.

hypothyroidism risk and decreased risk and decreased risk of going and any processes in the cancer mean corpuscular volume and HbA1c.
Novel TSH signals associated with thyroid diseases also implicated relatively understudi lower mean corpuscular volume and having
Novel TSH signals associated with thyroid dis
causal genes, such as SPPL3 and SDCCAG8. S. Novel The System association with thy side also incrementatively understanding putatively also increases causa
In T cell receptor signalling, regulation of calcineurin-NFAT signalling and protein dephosphorylation
In T cel causargenes, such as SPPL3 and SDCCAG8. SPPL3 encodes Signal Peptide Peptidase Like 3 involved
in T cell receptor signalling, regulation of calcineurin-NFAT signalling and protein dephosphorylation
4 in T cell receptor signalling, regulation of calcineurin-NFAT signalling and protein dephosphorylation.

The **SPPL3** intronic variant rs2393717 G allele (frequency 47.3% in EUR) associated with increased
TSH was associated with reduced hypothyroidism risk, increased tyrosine (a thyroid hormone
precursor), as well as decreased precursor), as well as decreased C-reactive protein, increased insulin-like growth factor 1 (IGF-1),
reduced height, whole body fat-free mass and reduced sex hormone binding globulin (especially i
males), decreased gamma g preduced height, whole body fat-free mass and reduced sex hormone binding globulin (especially i
males), decreased gamma glutamyltransferase (GGT), increased alkaline phosphatase, reduced
platelet count and eosinophils, in reduced height, where body fat-heighted and reduced tenthermally globulin (especially in
males), decreased gamma glutamyltransferase (GGT), increased alkaline phosphatase, reduced
platelet count and eosinophils, increased males), decreasing animalgeman, promotions (CCT), increased and with lipid composition traits. Mutation
SDCCAG8, encoding the sonic hedgehog (SHH) signalling and ciliogenesis regulator, SDCCAG8, ca
Bardet-Biedl Syndrome 16 **SDCCAG8**, encoding the sonic hedgehog (SHH) signalling and ciliogenesis regulator, SDCCAG8, cause
Bardet-Biedl Syndrome 16 (BBS16). Hypothyroidism and hyperthyroidism have been observed in
commoner forms of Bardet-Biedl S SDCCAG8, encoding the sonic hedgehog (SHH) signalling and chlogenesis regulator, SDCCAG8, cause
Bardet-Biedl Syndrome 16 (BBS16). Hypothyroidism and hyperthyroidism have been observed in
commoner forms of Bardet-Biedl Synd commoner forms of Bardet-Biedl Syndrome(14). The TSH increasing allele, C (frequency 53.8% in
EUR), of **SDCCAG8** intronic variant rs10926981 was associated with reduced risk of thyrotoxicosis
and thyroid cancer, as well as EUR), of **SDCCAG8** intronic variant rs10926981 was associated with reduced risk of thyrotoxicosis
and thyroid cancer, as well as reduced creatinine and increased eGFR. EUR), of SDCCAG8 infronic variant rs10926981 was associated with reduced risk or thyrotoxicosis
and thyroid cancer, as well as reduced creatinine and increased eGFR.
As smoking is known to influence thyroid function, we te

and thyroid cancer, as well as reduced creations with the latent creation.
As smoking is known to influence thyroid function, we tested whether
attenuated sentinel SNP associations with TSH in UK Biobank. Adjustm attenuated sentinel SNP associations with TSH in UK Biobank. Adjustments for current smoking of pack-years resulted in betas <50% of unadjusted betas for eight SNPs, including variants implicat
DIO2, TSHR, C9orf156, DET1 a attenuated sentinel SNP associations with TSH in UK Biobank. Adjustments for current smoking or pack-years results in betas for eight states in a light state, years in generating variants in pressing block,
Druggable targets DIO2, TSHR, C9orf156, DET1 and DZANK1.
Druggable targets
For the 112 genes supported by ≥2 criteria, we surveyed gene-drug interactions using the Drug Gene

Druggable targets

Interaction Database (DGIDB). The protein products of these genes include targets for treatments to
stimulate (thyrotropin [TSH]) or suppress (methimazole, targeting thyroid peroxidase, TPO) thyroid stimulate (thyrotropin [TSH]) or suppress (methimazole, targeting thyroid peroxidase, TPO) thyroid
function, and drugs to treat thyroid cancer (e.g. the KDR inhibitor, vandetanib) as well as PDE4
inhibitors and AKT inhibit stimulate (thyrotropin [TSH]) or suppress (methimazole) angeling thyrotropic matter) in equipment
function, and drugs to treat thyroid cancer (e.g. the KDR inhibitor, vandetanib) as well as PDE4
inhibitors and AKT inhibito inhibitors and AKT inhibitors utilised in immunoinflammatory conditions and cancers
(Supplementary Table 7). infibitors and AKT in immunoinflammators in immunoles in immunos and cancers in immunoinflammators in immunoin
in immunoinflammatory conditions and cancers and cancers and cancers and cancers and cancers and cancers and c

Pathway analysis

Employing ConsensusPathDB (15) , we tested biological pathways enrichment for the 112 putative causal genes supported by ≥ 2 criteria, highlighting signal transduction, particularly G protein (Reactome) and cAMP (KEGG) signalling, and the overlapping phosphodiesterases in neuronal
function pathway (Wikipathways, including novel genes *PDE4D, PDE7A, PDE4B, ADCY6*). The
thyroxine production (Wikipathways) pathway function pathway (Wikipathways, including novel genes PDE4D, PDE7A, PDE4B, ADCY6). The glycoprotein hormones subunit alpha, which is common to TSH, chorionic gonadotropin (CG), thyroxine production (Wikipathways) pathway included novel gene COA, encoding anterior pitulitary
glycoprotein hormones subunit alpha, which is common to TSH, chorionic gonadotropin (CG),
luteinizing hormone (LH), and foll gract in the integration of the stimulating hormone (FSH). New pathways of interest in
VEGF hypoxia and angiogenesis (Biocarta), including ARNT, BDKRB2, and KDR alongside *VEGF,*
AKT1, as well as opioid signalling, and pla VEGF hypoxia and angiogenesis (Biocarta), including ARNT, BDKRB2, and KDR alongside VEGFA and
AKT1, as well as opioid signalling, and platelet activation (Supplementary Table 8). VEGF hypoxia and anglogenesis (Biocarta), including ARNT, BBKRB2, and KBR alongside VEGTA and
AKT1, as well as opioid signalling, and platelet activation (Supplementary Table 8).
Phenome-wide associations of pathway-specif

Phenome-wide associations of pathway-specific TSH genetic risk scores

AKT1, as well as opioid signalling, and platelet activation (Supplementary Table 8).
Phenome-wide associations of pathway-specific TSH genetic risk
We hypothesised that partitioning a TSH genetic risk score (GRS) into path according to the biological pathway(s) that each variant influences could inform understandi
mechanisms underlying TSH and thyroid disease, and possible consequences of pathway
perturbation. Informed by the above prioritis mechanisms underlying TSH and thyroid disease, and possible consequences of pathway
perturbation. Informed by the above prioritisation of putative causal genes and classification of
these genes by pathway, we undertook Phe perturbation. Informed by the above prioritisation of putative causal genes and classificat
these genes by pathway, we undertook PheWAS for TSH-weighted GRSs partitioned by ea
enriched pathways (FDR<5x10⁻⁴) after droppi these genes by pathway, we undertook PheWAS for TSH-weighted GRSs partitioned by each of 2
enriched pathways (FDR<5x10⁻⁴) after dropping redundant pathways (GRS correlation r^2 <0.7). enriched pathways (FDR<5x10⁻⁴) after dropping redundant pathways (GRS correlation r^2 <0.7).
We highlight examples of pathway-specific GRSs showing differing patterns of associations with

enriched pathways (FDR<5x10") after dropping redundant pathways (GRS correlation r°<0.7).
We highlight examples of pathway-specific GRSs showing differing patterns of associations wit
thyroid and non-thyroid diseases (**Sup** thyroid and non-thyroid diseases (**Supplementary Table 9**). The GRS for higher TSH specific to th
cAMP signalling pathway (KEGG, including novel genes **GNAS, PDE4D, PDE4B, ADCY6, CGA**) was
specific to increased risk of hyp thyroid and non-thyroid diseases (Supplementary Table 5). The GRS for higher TSH specific to the
cAMP signalling pathway (KEGG, including novel genes **GNAS, PDE4D, PDE4B, ADCY6, CGA**) was
(Supplementary Figure 2a). A GRS f cAMP signalling pathway (KEGG, including novel genes **GNAS, I DE4D, I DE4D, ADCT0, CGA**) was
specific to increased risk of hypothyroidism; no associations (FDR<1%) with other traits were sho
(**Supplementary Figure 2a**). A (**Supplementary Figure 2a**). A GRS for higher TSH specific to the activin receptor-like kinase (ALK) in
cardiac myocytes pathway (Biocarta, including novel genes **SMAD6, BMP4**) showed associations with (Supplementary Figure 2a). A GRS for higher TSH specific to the activin receptor-like kinase (ALK) in cardiac myocytes pathway (Biocarta, including novel genes **SMAD6**, **BMP4**) showed associations wit cardiac myocytes pathway (Biocarta, including novel genes SMAD6, BMP4) showed associations with

It is made available under a [CC-BY 4.0 International license](http://creativecommons.org/licenses/by/4.0/) .

reduced risk of nontoxic nodular and multinodular goitre, and simple goitre, as well as raised heel
bone mineral density, standing height and whole-body fat-free mass and reduced FEV₁/FVC (**Figure**
2a). GRSs specific t bone mineral density, standing height and whole-body fat-free mass and reduced FEV1/FVC (Figure
2a). GRSs specific to several pathways showed association to PheCode 244.1 capturing
consequences of treated hyperthyroidism: 2a). GRSs specific to several pathways showed association to PheCode 244.1 capturing
consequences of treated hyperthyroidism: the activin receptor-like kinase (ALK) in card
pathway, pathways in cancer (KEGG), factors and p pathway, pathways in cancer (KEGG), factors and pathways affecting insulin-like growth factor
(IGF1)-Akt signalling (Wikipathways), myometrial relaxation and contraction pathways
(Wikipathways), FGFR3 signalling in chondro pathway, pathways in cancer (KEGG), factors and pathways and contraction pathways
(Kikipathways), FGFR3 signalling in chondrocyte proliferation and terminal differentiation
(Wikipathways). The GRS for higher TSH specific t (Wikipathways), FGFR3 signalling in chondrocyte proliferation and terminal differentiat
(Wikipathways). The GRS for higher TSH specific to the platelet activation, signalling an
pathway (Reactome, including novel genes **GN** (Wikipathways), FEFR3 spanning in chorocyte proliferation and terminal anteremented
(Wikipathways). The GRS for higher TSH specific to the platelet activation, signalling and ag
hypothyroidism risk, reduced nontoxic nodula (Wikipathway). The GRS for higher TSH specific to the platelet activation, signalling and aggregation
pathway (Reactome, including novel genes **GNG7, ANXA5, PRKC2**) was associated with increased
hypothyroidism risk, reduce pathway (Reactome, including novel genes **GNO7, ANXA5, PRKCZ**) was associated with increased
hypothyroidism risk, reduced nontoxic nodular goitre and simple goitre risk, raised urate, creatinii
and reduced eGFR, reduced se my permy consumed in the measure in stating going and points and, these made years in the properties and reduced eGFR, reduced sex hormone binding globulin and testosterone, increased waist
circumference, handgrip strength circumference, handgrip strength and whole body fat-free mass, osteochondropathies, increa
triglyceride levels, and with lipid composition traits (**Figure 2b**).
Data triglyceride levels, and with lipid composition traits (Figure 2b).

Polygenic score associations

We constructed a polygenic score (PGS) for TSH using the summary statistics of approximately 1.12
million SNPs from our discovery GWAS (meta-analysed from UK Biobank, EXCEED and results from We construct a polygenic score (PGS) for Terms and give contributing the summary statistics of approximation
Thou et al, with total sample size of 247,107 European-ancestry individuals) as the training dataset
(Online Meth million SNPs from our discovery General and personal controlled in the controlled from the strain of the control
Zhou et al, with total sample size of 247,107 European-ancestries).
Conline Methods, Figure 3: PGS performanc

(Online Methods, Figure 3: PGS performance across ancestries)
The TSH PGS showed distinct patterns of associations with relevant thyroid and non-thyroid (Online Methods, Figure 3: POS performance across ancestries).
The TSH PGS showed distinct patterns of associations with relevar
phenotypes in our PheWAS (Supplementary Figure 3). Thyroid-re phenotypes in our PheWAS (**Supplementary Figure 3**). Thyroid-relevant PGS associations included
increased risk of hypothyroidism, lower risk of non-toxic (multi)nodular goitre, thyrotoxicosis,
Graves' disease, and thyroid phenotypes in our FiewAS (Supplementary Figure 3). Thyroid Televant P as associations included
increased risk of hypothyroidism, lower risk of non-toxic (multi)nodular goitre, thyrotoxicosis,
Graves' disease, and thyroid c Graves' disease, and thyroid cancer, and reduced tyrosine. Other PGS associations included
increased FEV₁/FVC, lower risk of chronic obstructive pulmonary disease (COPD), pneumonia,
disease, common cancers and multi-site disease, common cancers and multi-site chronic pain, lower arterial stiffness, increased creatinine
and urate, increased alkaline phosphatase and aspartate aminotransferase, increased eosinophils,
decreased sex hormone-bin and urate, increased alkaline phosphatase and aspartate aminotransferase, increased eosinophils,
decreased sex hormone-binding globulin, testosterone and IGF-1, decreased glucose
(Supplementary Table 10) as well as altered and urate, increased alta unionism prospinsors and uspartate and IGF-1, decreased glucose
(Supplementary Table 10) as well as altered lipid levels and composition. We found little or no
attenuation of these PGS association (Supplementary Table 10) as well as altered lipid levels and composition. We found li
attenuation of these PGS associations after adjustment for whether the individuals ha
smoked. attenuation of these PGS associations after adjustment for whether the individuals had ever
smoked.

We then tested PGS associations across ancestries. Strong associations were shown with TSH levels in all ancestry groups tested (African, AFR; Admixed American, AMR; Central/South Asian, CSA; East
Asian, EAS; Middle Eastern, MID; **Figure 3, Supplementary Table 11**). The TSH PGS was strongly in all ancestry groups tested (African, AFR; Admixed American, AMR; Central/South Asian, CSA; East
Asian, EAS; Middle Eastern, MID; F**igure 3, Supplementary Table 11**). The TSH PGS was strongly
associated with free T₄ le in all ancestry groups tested (African, African, American, African, Asian, EAS; Middle Eastern, MID; Figure 3, Supplementary Table 11). The TSH PGS was strongly
associated with free T₄ levels in E ancestry individuals (P Asian, EAS, Middle Eastern, MID, Figure 3, Supplementary Table 11). The TSHT 05 was strongly
associated with free T₄ levels in E ancestry individuals (P=5.93x10⁻⁵⁸), nominally associated with 1
T₄ in the next larges T_4 in the next largest ancestry group, CSA (P=0.0010, 1307 participants), and showed a consistent direction of effect in other ancestry groups (**Supplementary Table 11**).

The next largest ancestry, group, Carry Territon and participants), and showed a consistent and showed direction of effect in other ancestry groups (Supplementary Table 11).
To inform understanding of the relevance of the direction of effect in other ancestry groups (Supplementary Table 11).
To inform understanding of the relevance of the TSH PGS for disease, w
disease susceptibility risk in all UK Biobank ancestry subgroups with at l disease susceptibility risk in all UK Biobank ancestry subgroups with at least 100 cases. In Euro
ancestry UK Biobank participants the TSH PGS was associated with risk of hypothyroidism
(P<1x10⁻³⁰⁰), hyperthyroidism (P=4 ancestry UK Biobank participants the TSH PGS was associated with risk of hypothyroidism
(P<1x10⁻³⁰⁰), hyperthyroidism (P=4.79x10⁻¹⁶⁹), thyroid cancer (P=7.64x10⁻¹⁰) and other thyroid diseas.
(P=1.55x10⁻⁴⁴, **Supplem** $(P< 1x10^{-300})$, hyperthyroidism $(P=4.79x10^{-169})$, thyroid cancer $(P=7.64x10^{-10})$ and other thyr
 $(P=1.55x10^{-44})$, **Supplementary Table 12**. In other ancestry groups, a consistent direction of
association was shown with eac (P<1x10 ³⁰⁰), hyperthyroidism (P=4.79x10 ³⁰³), thyroid cancer (P=7.64x10 ³⁰) and other thyroid disease
(P=1.55x10⁻⁴⁴, **Supplementary Table 12**. In other ancestry groups, a consistent direction of
association was sh (P=1.55x10⁻⁺⁺, **Supplementary Table 12**. In other ancestry groups, a consistent direction of
association was shown with each of these traits, the largest case numbers being seen for
hypothyroidism in CSA (720 cases, P=4. hypothyroidism in CSA (720 cases, $P=4.68 \times 10^{-17}$, Supplementary Table 12).

To further understand the clinical relevance of the polygenic score we examined risk of thyroid hypothyroidism in CSA (720 cases, P=4.68x10 ^{- ,}, **Supplementary Table 12**).
To further understand the clinical relevance of the polygenic score we exan
disease per decile of the PGS in European ancestry individuals. Indi To further understand the clinical relevance of the polygenic score in a manufacture of the polygenic disease
3.53-fold higher odds of hypothyroidism compared with those in the lowest decile, whilst those
3.53-fold higher 3.53-fold higher odds of hypothyroidism compared with those in the lowest decile, whilst those in
6 3.53-fold higher odds of hypothyroidism compared with those in the lowest decile, whilst those in

It is made available under a [CC-BY 4.0 International license](http://creativecommons.org/licenses/by/4.0/) .

the lowest decile had 4.12-fold and 2.15-fold higher odds of hyperthyroidism and thyroid cancer,
respectively, and 2.98-fold higher odds of other thyroid disease compared with those in the highest
decile (Figure 4). Given decile (**Figure 4**). Given questions about how best to deploy and repeat testing for thyroid disease in
asymptomatic patients and in patients with non-specific symptoms, we explored whether
membership of a high or low risk decile (Figure 4). Given questions about how best to deploy and repeat testing for thyroid disease in
asymptomatic patients and in patients with non-specific symptoms, we explored whether
membership of a high or low risk d asymptomatic patients and in patients with non-specific symptoms, we explored with
membership of a high or low risk decile for TSH PGS was associated with differences in age
of hypothyroidism or hyperthyroidism (Online Met of hypothyroidism or hyperthyroidism (**Online Methods**). Between individuals with median, highest
and lowest deciles of the TSH PGS, clear differences were seen in age of onset of hypothyroidism
(P=1.0x10⁻³⁰⁰, Fi**gure 5a** of hypothyroidism of hyperthyroidism (Online Methods). Between marviduals with median, highest
and lowest deciles of the TSH PGS, clear differences were seen in age of onset of hypothyroidism
(P=1.0x10⁻³⁰⁰, **Figure 5a**) (P=1.0x10³⁰⁰, Figure 5a) and hyperthyroidism (P=5.3x10⁻⁶⁵, Figure 5b). For example, a 5% prevaler
of hypothyroidism was reached by age 51.1 years in the highest TSH PGS decile versus by 74.7 yea
in the lowest TSH PGS d of hypothyroidism was reached by age 51.1 years in the highest TSH PGS decile versus by 74.7 years in the lowest TSH PGS decile. Similarly a 1% prevalence of hyperthyroidism was reached by age 47.3 years in the lowest PGS decile compared with 71.2 years in the highest TSH PGS decile.
DiSCUSSION

years in the lowest PG decide compared with 71.2 years in the highest TSH TO 2008.
Discussion
The large sample size of our study, achieved through utilising quality-controlled TSH me **Discussion**
The large sample size of our study, achieved through utilising quality-controlled TSH measures from
UK primary health care records, increased the yield of TSH genetic signals by over 2.5-fold, to 260.
Through UK primary health care records, increased the yield of TSH genetic signals by over 2.5-fold, to 260.
Through the most comprehensive initiative to identify putative causal variants and genes for TSH
levels, we defined 112 h develop pathway-specific GRS for TSH levels and to use these in PheWAS, through our new
DeepPheWAS platform(7), to investigate the potential consequences of intervening in relevant levels) are defined 112 mg. Confidence generally process to many transmising the first study to
develop pathway-specific GRS for TSH levels and to use these in PheWAS, through our new
DeepPheWAS platform(7), to investigate DeepPheWAS platform(7), to investigate the potential consequences of intervening in relev
pathways. It is also the first to develop and use a polygenic score to predict age of onset of
hypothyroidism and hyperthyroidism, s pathways. It is also the first to develop and use a polygenic score to predict age of onset of
hypothyroidism and hyperthyroidism, showing marked differences in ages of onset of these
conditions according to PGS deciles. hypothyroidism and hyperthyroidism, showing marked differences in ages of onset of these
conditions according to PGS deciles.

onditions according to PGS deciles.
We implicate novel putative causal variants and genes, which alongside those previously rep Example according to Polymer.
We implicate novel putative causal v
12, 16), provide a more complete pic 12, 16), provide a more complete picture of relevant pathways and putative mechanisms. Pathways
we highlight include signal transduction and cAMP signalling, as well as pathways not confidently
implicated previously such a we highlight include signal transduction and cAMP signalling, as well as pathways not confidently
implicated previously such as VEGF hypoxia and angiogenesis, AKT signalling and platelet activation.
Our findings are consis implicated previously such as VEGF hypoxia and angiogenesis, AKT signalling and platelet activatic
Our findings are consistent with signalling or response to thyroid or non-thyroid hormones (includ
IGF signalling), neurona implicated previously such as VEGF hypothematically gapterity and algorithy and platelet activation.
Our findings are consistent with signalling or response to thyroid or non-thyroid hormones (including
IGF signalling), ne Our findings are consistent with signalling or response to thyroid or non-thyroid hormones (including

diseases.
Pleiotropic effects of aggregated TSH-associated variants have been previously shown through meenses.
Pleiotropi
PheWAS. PheWAS. Partitioning TSH-associated variants by pathway provides a more nuanced understanding
of the consequences of pathway perturbation on thyroid and non-thyroid disorders. We show Phenelet wave and support the Wash-association on thyroid and non-thyroid disorders. We show
contrasting patterns of phenotype association – for example highly specific associations for
hypothyroidism (cAMP signalling) ver contrasting patterns of phenotype association – for example highly specific associations for
hypothyroidism (cAMP signalling) versus associations also with body composition, renal fun
lipid traits (platelet activation path contrasting patterns of phenotype associations also with body composition, renal fun-
hypothyroidism (cAMP signalling) versus associations also with body composition, renal fun-
lipid traits (platelet activation pathways). ind traits (platelet activation pathways). As individuals may have high GRS for one or more
pathways and low GRS for other pathways(17), individuals' pathway GRS profiles may relate to
patterns of comorbidities, and could pathways and low GRS for other pathways(17), individuals' pathway GRS profiles may relate to
patterns of comorbidities, and could have implications for treatment choices in thyroid diseases.
Here we adopted a powerful stra

patterns of comorbidities, and could have implications for treatment choices in thyroid disease:
Here we adopted a powerful strategy for discovery of signals associated with thyroid diseases. ' studied TSH as a quantitative measure within reference ranges, and detect novel signals which
individually and in aggregate are associated with thyroid diseases. Not all TSH-associated variants studied TSH as a quantitative measure within reference ranges, and detect novel signals which
individually and in aggregate are associated with thyroid diseases. Not all TSH-associated variants
showed association with fre individually and in aggregate are associated with thyroid diseases. Not all TSH-associated variar
showed association with free T₄ levels, even those associated with thyroid disease, highlighting
value of TSH as a sensiti showed association with free T₄ levels, even those associated with thyroid disease, highlighting the value of TSH as a sensitive marker of thyroid disorders. In addition, thyroid function within the euthyroid range is a value of TSH as a sensitive marker of thyroid disorders. In addition, thyroid function within the
euthyroid range is associated with adverse outcomes(18) and thus TSH-associated variants that
not yet been overtly associate euthyroid range is associated with adverse outcomes(18) and thus TSH-associated variants that have
not yet been overtly associated with thyroid disease remain highly relevant. GWAS of other
quantitative traits – including not yet been overtly associated with thyroid disease remain highly relevant. GWAS of other
quantitative traits – including those on the UK Biobank biomarker panel – have highlighted a number
of targets leading to active dr not yet a discribing those on the UK Biobank biomarker panel – have highlighted and tanglet as a prototon the
of targets leading to active drug development for related diseases(3-5). However, TSH has no
of targets leading of targets leading to active drug development for related diseases(3-5). However, TSH has not yet
7 of targets leading to active drug development for related diseases(3-5). However, TSH has not yet

EXCEED and UK Biobank studies more than doubled available sample sizes.
As with other contemporary genome-wide association meta-analysis, maximising power for As with other contemporary genome-wide association meta-analysis, maxir
discovery does not leave available datasets suitable for independent replica discovery does not leave available datasets suitable for independent replication. In our stud
the 260 independent sentinels (118 of the 158 novel signals) showed consistent direction of
and P<0.01 in at least two of the th the 260 independent sentinels (118 of the 158 novel signals) showed consistent direction of effect
and P<0.01 in at least two of the three contributing datasets in addition to reaching genome-wide
significance in the metaand P<0.01 in at least two of the three contributing datasets in addition to reaching genome-wide
significance in the meta-analysis. Not all SNPs were represented in the Zhou et al dataset due to
contributing datasets usin and Person at least the Person and Person and B and Person and B contributing dataset due to
contributing datasets using less dense imputation platforms than we were able to use in UK Biobar
and EXCEED. significance in the meta-analysis. Note and structure in the Zhop estimation in the Zhou et al. Analysis contributing datasets using less dense imputation platforms than we were able to use in UK Biob and EXCEED. and EXCEED.
The PGS we developed for TSH utilised the full genome-wide association statistics to maximise

The PGS we c
predictive po predictive power. To our knowledge this is the first TSH PGS to be shown to be associated with
levels across all ethnic groups in UK Biobank. Power was more limited for testing other traits ar
diseases in the much smaller revels across all ethnic groups in UK Biobank. Power was more limited for testing other traits and
diseases in the much smaller sample sizes available in non-European ancestries in UK Biobank.
Nevertheless, we showed a con diseases in the much smaller sample sizes available in non-European ancestries in UK Biobank.
Nevertheless, we showed a consistent direction of association on T4 levels in all ancestries and
association with hypothyroidism Nevertheless, we showed a consistent direction of association on T₄ levels in all ancestries and
association with hypothyroidism in South Asian participants (P=4.68x10⁻¹⁷, 720 cases).
Understanding the genetic archite association with hypothyroidism in South Asian participants (P=4.68x10⁻¹⁷, 720 cases).
Understanding the genetic architecture of thyroid diseases within and across ancestries require
larger sample sizes in non-European a larger sample sizes in non-European ancestries, and an urgent global effort is required to include
much more diverse populations in genomic studies than has been the case to date(19). larger sample sizes in non-European ancestries, and an urgent global effort is required to include

much more diverse populations in the PGS shows a strong association with age of onset of hypothyroidism and hyperthyroid is than the case is not recomment. The Puropean ancestry individuals. Universal screening for thyroid disease is not recommended(20, 2
Thstead, case finding strategies are adapted to personal risk factors such as age, family history an
The Postin and hypert Instead, case finding strategies are adapted to personal risk factors such as age, family history and
relevant long-term conditions. Our findings raise the possibility of tailoring case finding strategies fo
thyroid diseas Instead, case manning strategies are adapted to personal risk factors such as age, family model, interesting
Individual risk factors and the medicing strategies are added to personally if genome-wide data become
Internal r relevant long-term conding-term conditions. Our possibility of tailoring cases in thyroid disease according to a PGS for thyroid disease, especially if genome-wide data become
available as part of the medical record. Furth available as part of the medical record. Further development and testing of a PGS would be required
in independent, diverse populations, and alongside risk factors already employed in case finding.
Although we used TSH as in independent, diverse populations, and alongside risk factors already employed in case finding.

thyroid diseases using electronic healthcare records in our investigating the clinical relevance of
these signals. We undertook careful curation of clinical codes used to define clinical thyroid disease Although we used the trait for signal discovery, there is positioned the clinical relevance of
these signals. We undertook careful curation of clinical codes used to define clinical thyroid dise
to test SNP associations. F these signals. We undertook careful curation of clinical codes used to define clinical thyroid disea
to test SNP associations. For example, we excluded cases of hypothyroidism resulting from
treatment of hyperthyroidism. T to test SNP associations. For example, we excluded cases of hypothyroidism resulting from
treatment of hyperthyroidism. To avoid classifying cases incorrectly as hyperthyroidism or
hypothyroidism we created a separate cate treatment of hyperthyroidism. To avoid classifying cases incorrectly as hyperthyroidism or
hypothyroidism we created a separate category "other thyroid disease" containing disorder
goitre (where the clinical code was not e treatment of hyperthyroidism. To avoid classifying classification, as hyperthyroidism.
By pothyroidism we created a separate category "other thyroid disease" containing disorde
goitre (where the clinical code was not expli goitre (where the clinical code was not explicit regarding thyroid status) and thyroiditis. In the latter group, we found that the direction of PGS associations were similar to those for hyperthyroidism,
albeit with a smal group, we found that the direction of PGS associations were similar to those for hyperthyroidism,

group, we found that the direction of the direction of the similar to the comp_{rement}, country, to those about a
In summary, we more than doubled the number of TSH-associated signals to 260, confidently albeit with a smaller effect estimate.
In summary, we more than doubled t
implicated 112 priority genes, showe implicated 112 priority genes, showed their relevance to thyroid diseases, and developed pathway-
specific genetic risk scores which show differential patterns of pleiotropy of relevance in
understanding co-morbidities and and age of onset of hypothyroidism and hyperthyroidism, of potential utility in future case finding specific general control of the matrice control patterns of pleisingly controlled in
understanding co-morbidities and treatment choices. The PGS we developed predicts risk
and age of onset of hypothyroidism and hyperthyroi and age of onset of hypothyroidism and hyperthyroidism, of potential utility in future case finding
strategies, subject to further development and appropriate evaluation. strategies, subject to further development and appropriate evaluation. strategies, subject to further development and appropriate evaluation.

It is made available under a CC-BY 4.0 International license.

Methods

Cohort details

We analysed data from UK Biobank is a cohort of approximately 500,000 individuals recruited from
across the United Kingdom. Individuals aged between 40 and 69 years were recruited from the
general population between 2006 a)
(Excross the United Kingdom. Individuals aged between 40 and 69 years were recruited from the
general population between 2006 and 2010. EXCEED recruited approximately 10,000 individuals
primarily through local general pr general population between 2006 and 2010. EXCEED recruited approximately 10,000 individual:
primarily through local general practices in Leicester City, Leicestershire and Rutland. Recruitme
started in 2013. Individuals in general population between 2012 and 2010 to the suppresentation, 2010 in the primarily through local general practices in Leicester City, Leicestershire and Rutland. Recruitmer
started in 2013. Individuals invited to contr primarily throught in throught states in the state of throught started in 2013. Individuals invited to contribute to EXCEED were aged between 40 and 69 years.
Both UK Biobank and EXCEED collected information at baseline co started in 2022. Individuals invited to contribute the 2013-2013 of the Castle Contribute to the 2013.
Both UK Biobank and EXCEED collected information at baseline concerning lifestyle and health
outcomes, as well as provi outcomes, as well as providing linkage to electronic health records and genetic data.

Phenotype

We captured all TSH results reported in the primary care data available in up to 230,000 individuals
in UK Biobank and 8500 individuals in EXCEED utilising codes from Read version 2 and Read version 3 In UK Biobank and 8500 individuals in EXCEED utilising codes from Read version 2 and Read version
(Clinical Terms Version 3 or "CTV3"). We took an individual's first non-missing TSH measurement to
minimise the effect of th (Clinical Terms Version 3 or "CTV3"). We took an individual's first non-missing TSH measurement to
minimise the effect of thyroid function-altering medications on our phenotype as an individual is
unlikely to have received ,
minimise the effect of thyroid function-altering medications on our phenotype as an individual is
unlikely to have received these medications before their first thyroid function test. Where an
individual's first non-miss unlikely to have received these medications before their first thyroid function test. Where an
individual's first non-missing TSH measurement was 0, they were excluded from the analysis unle
the individual also had a code individual's first non-missing TSH measurement was 0, they were excluded from the analysis unless
the individual also had a code in their primary care data for hyperthyroidism where a TSH
measurement of 0 may be clinically individual also had a code in their primary care data for hyperthyroidism where a TSH
measurement of 0 may be clinically feasible. In all other instances, we were unable to disentangle
true 0 measurements from those that m the individual also had a code in their primary care data for primary candidary measurement of 0 may be clinically feasible. In all other instances, we were unable to dise
true 0 measurements from those that may have arise true 0 measurements from those that may have arisen due to, for example, an individual's TSH bei
below the detectable range of the test apparatus used and, therefore, being entered into the
primary care data as 0 or some o the concentration from the control may have are to, for example, an individual concentration of below the detectable range of the test apparatus used and, therefore, being entered into the primary care data as 0 or some ot primary care data as 0 or some other value such as "<0.05" which may have later been conver
0. We excluded individuals with a TSH measurement <0.4 or >4.0 mU/L as has been done
previously(12). 0. We excluded individuals with a TSH measurement <0.4 or >4.0 mU/L as has been done previously (12). 0. We exclude in the two states with a TSH measurement of the two states with a TSH measurement of the two states of the two states previously (12).

Genome-wide association study

previous_{/(12)}.
Genome-wid
We applied an i We applied an inverse normal transformation to the residuals from linear regression of the TSH
phenotype against age (at time of measurement) and sex. This transformed phenotype was used for
genome-wide association testing individuals of European ancestry, as defined by the Pan-UK Biobank initiative (25), who were not genome-wide association in area and the array, sex and the first and the first analysed array, sex analysed in
Individuals of European ancestry, as defined by the Pan-UK Biobank initiative(25), who were not
Inore closely r individuals of European ancestry, as defined by the Pan-UK Biobank initiative(25), who were n
more closely related than third-degree relatives using a KING software relatedness coefficient
>0.0884 to indicate second-degree individuals of European ancestry, as defined by the Pan-UK Biobank initiative (25), the Pan-UK Biobank increas

20.0884 to indicate second-degree relatives or closer(26). In UK Biobank, we tested genetic variativity a mino more closely related than the degree relatives or closer(26). In UK Biobank, we tested genetic variants with a minor allele count >20 and imputation score >0.5. In EXCEED, due to the smaller sample
we tested genetic varian with a minor allele count >20 and imputation score >0.5. In EXCEED, due to the smaller sample size,
we tested genetic variants with a minor allele frequency >0.1% and imputation score >0.5.

we tested genetic variants with a minor allele frequency >0.1% and imputation score >0.5.
We derived the LD Score regression intercept using LDSC(27) to estimate inflation in our test we derived the LD Score regression intercept using LDSC(27) to estimate inflation in our te
statistics due to confounding, such as by cryptic relatedness or population stratification. W Statistics due to confounding, such as by cryptic relatedness or population stratification. We
estimated, separately, the LD Score regression intercept for the GWAS in UK Biobank and EXC
The UK Biobank test statistics were stimated, separately, the LD Score regression intercept for the GWAS in UK Biobank and EX
The UK Biobank test statistics were corrected for inflation (λ_{LDSC} = 1.05) prior to meta-analysis
EXCEED test statistics were The UK Biobank test statistics were corrected for inflation (λ_{LDSC} = 1.05) prior to meta-analysis. The EXCEED test statistics were not corrected for inflation (λ_{LDSC} = 0.98). EXCEED test statistics were not corrected for inflation $(\lambda_{LDSC} = 0.98)$.
We used METAL(28) to meta-analyse the results from the GWAS in UK Biobank and EXCEED and th

ر
We used METAL(28) to meta-analyse the results from the GWAS in I
previous largest GWAS of TSH(6). Since the EXCEED results were alig previous largest GWAS of TSH(6). Since the EXCEED results were aligned to GRCh38, we ran LiftOver
to map the results to GRCh37. Over 99.5% of the genetic variants tested in the GWAS in EXCEED
were successfully mapped to GR previous largest CMAS of NCR (6). Since the EXCELC version is angles to CMAS in EXCEED
were successfully mapped to GRCh37. Following meta-analysis, we estimated the LD Score
regression intercept once more. were successfully mapped to GRCh37. Following meta-analysis, we estimated the LD Score
regression intercept once more. were succession intercept once more. regression intercept once more.

It is made available under a [CC-BY 4.0 International license](http://creativecommons.org/licenses/by/4.0/) .

$$
\frac{\sum_{i=1}^{n} 2f_i(1-f_i)\beta_i^2}{V}
$$

where n is the number of SNPs, f_i and β_i are the frequency and effect estimate of the i^{th} variant where *n* is the number of SNPs, f_i and p_i are the frequency and effect estimate of the *i* "variant"
from our UK Biobank discovery analysis, and *V* is the phenotypic variance (always 1 as TSH was
inverse-normal tran from our ok Biobank discovery analysis, and V is the phenotypic variance (always 1 as TSH was
inverse-normal transformed).
Signal selection

Signal selection

inverse-normal transformed).
Signal selection
We selected 2 Mb loci centred on the most significant variant for all regions containing a variant
with P<5×10⁻⁸. Loci within 500 kb of each other were merged for fine mappin

with P<5×10⁻⁸. Loci within 500 kb of each other were merged for fine mapping.
PolyFun(29) and Sum of Single Effects (SuSiE)(30) was used to fine-map autosomal non-HLA loci
utilising pre-computed functional prior causal p utilising pre-computed functional prior causal probabilities based on a meta-analysis of 15 UK
Biobank traits. The functional priors are proportional to per-SNP heritabilities estimated from th
functional enrichments of 18 Biobank traits. The functional priors are proportional to per-SNP heritabilities estimated from the
functional enrichments of 187 variant annotations from the baseline-LF 2.2.UKB model (Gazal S,
2018), including those rela Biodim traits. The functional priors are proportional to per-STM in training to committed from the functional enrichments of 187 variant annotations from the baseline-LF 2.2.UKB model (Gazal S, 2018), including those relat 2018), including those relating to conservation, regulation, MAF and linkage disequilibrium, whic
were estimated using an extension of stratified-LDSC(31). Imputed genotype data from 10,000
randomly selected European indiv 2022), including these relating to conservation, regulation, may also many measure quind in the set of those re
where estimated using an extension of stratified-LDSC(31). Imputed genotype data from 10,000
randomly selected randomly selected European individuals from UK Biobank was used as an LD reference. Loci for which PolyFun and SuSiE did not identify any credible sets, as well as HLA and chromosome X loci, were fine-mapped using the Wakefield method (32) with the prior W set as 0.04 in the approximate Bayes factor formula. 95% credible sets were generated for all loci, and variants with the highest posterior inclusion probability (PIP) per credible set identified (Supplementary Table 13).

Novel signals

We searched PubMed and GWAS Catalog to identify applied studies focused on thyroid stimulating hormone and associations reaching P<5x10⁻⁸. These included the following sources: Gudmundsson et al. (33), Kwak et al. (34), Malinowski et al. (35), Medici et al. (36), Nielsen et al. (37), Popović et al. (38), Porcu *et al.* (16), Taylor *et al.* (39), Teumer *et al.* (12), and Zhou *et al.* (6). We determined whether a signal was novel if its extent of linkage disequilibrium (LD) with nearby previously reported signals w whether a signal was novel if its extent of linkage disequilibrium (LD) with nearby previously

Epidemiological associations with clinical thyroid disease

reported signals was <0.2 (R²).
Epidemiological associations with clinical thyroid disease
We tested the association between our sentinel variants and free thyroxine (T₄) or five clinica ,
ia
tv thyroid diseases: (i) hypothyroidism ; (ii) hyperthyroidism; (iii) thyroid cancer; and (iv) other (no
cancer) thyroid disease. Using the UK Biobank primary care data, we extracted free T4
measurements that co-occurred with maximise potential cases for hypo- and hyperthyroidism, we utilised data available in primary care, measurements that co-occurred with the corresponding individual's first TSH measure
maximise potential cases for hypo- and hyperthyroidism, we utilised data available in _l
secondary care and self-reported diagnoses (UK B maximise potential cases for hypo- and hyperthyroidism, we utilised data available in primary ca
secondary care and self-reported diagnoses (UK Biobank Data-Field 20002). For the remaining
clinical disease phenotypes, we d secondary care and self-reported diagnoses (UK Biobank Data-Field 20002). For the remaining
clinical disease phenotypes, we defined these using primary care alone (other thyroid disease), and
cancer register data (thyroid clinical disease phenotypes, we defined these using primary care alone (other thyroid disease), and
cancer register data (thyroid cancer). To reduce the overlap in cases for the clinical disease
phenotypes, we defined a ca cancer register data (thyroid cancer). To reduce the overlap in cases for the clinical disease
phenotypes, we defined a case by their first diagnosis of hypothyroidism, hyperthyroidism, thyroid
cancer and other thyroid di phenotypes, we defined a case by their first diagnosis of hypothyroidism, hyperthyroidism,
cancer and other thyroid diseases. The clinical codes used to define free T₄ and the clinical t
diseases are presented in (**Suppl** phenotypes, we defined a case by their first angles bourges by permain, hypermyrenesing, hyperthyroid
cancer and other thyroid diseases. The clinical codes used to define free T_4 and the clinical thyroid
diseases are p diseases are presented in (**Supplementary Table 14).** We used PLINK 2.0<mark>(24)</mark> to test the associatior
under an additive genetic model adjusted for sex, genotyping array, age at recruitment to UK
Biobank and the first 10 pr diseases are presented in (Supplementary Table 14). We used PERK 2.0(24) to test the associations
under an additive genetic model adjusted for sex, genotyping array, age at recruitment to UK
Biobank and the first 10 princi Biobank and the first 10 principal components of ancestry.
These phenotypes were further tested for association with our polygenic score (described bel

These phenotypes were further tested for association with These phenotypes were further tested for association with our polygenic score (described below).

It is made available under a [CC-BY 4.0 International license](http://creativecommons.org/licenses/by/4.0/) .

To assess the robustness of the associations between our sentinel SNPs and TSH to smoking
behaviour, we fit separate models that included current smoking status (UK Biobank data field 1239)
and pack-years of smoking (UK Bi behaviour, we fit separate models that included current smoking status (UK Biobank data field 20161). We determined that an association was
attenuated upon adjusting for smoking if the smoking-adjusted effect estimate was and pack-years of smoking (UK Biobana field at a Biobany). We determined that an associate in the stress of the
attenuated upon adjusting for smoking if the smoking-adjusted effect estimate was less than 50% is
an associat attenuation upon adjusting for smoking-adjusted effect estimate was less than 50% in the smoking-adjusted effect
Identification of putative causal genes

Identification of putative causal genes

original magnetic
Identification of
To systematically pri To systematically prioritise putative causal genes for TSH associated signals, we integrated eight
sources of evidence including: (i) the nearest gene to the sentinel variant; (ii) the gene with the
highest polygenic prior highest polygenic priority score (PoPS)(9), a method based on the assumption that causal genes
different chromosomes share similar functional characteristics; identification of (iii) expression
quantitative trait loci (eQT different chromosomes share similar functional characteristics; identification of (iii) expression
quantitative trait loci (eQTLs) or (iv) protein quantitative trait loci (pQTLs) within the credible sets; (
proximity to a quantitative trait loci (eQTLs) or (iv) protein quantitative trait loci (pQTLs) within the credible se
proximity to a gene for a thyroid-associated Mendelian disease (±500kb); (vi) an annotation-
informed credible set cont proximity to a gene for a thyroid-associated Mendelian disease (±500kb); (vi) an annotation-
informed credible set containing a missense/deleterious/damaging variant with a posterior association with hypo- or hyperthyroidism using whole-exome(10) and whole-genome(11) probability of association >50%; (vii) identification of a rare variant (\pm 500kb of a TSH sentinel) probability of association with hypo- or hyperthyroidism using whole-exome(10) and whole-genome(11)
sequencing resources; and (viii) proximity to a human ortholog of a mouse knockout gene with
thyroid-related phenotype (±5 association with hyperthyroidism using whole-exome(21) and whole-general general sequencing resources; and (viii) proximity to a human ortholog of a mouse knockout gene is
thyroid-related phenotype (±500kb). thyroid-related phenotype (±500kb).
For (i), (v), (vii) and (viii), all 260 signals were used as input; for (ii), the 257 autosomal signals were

there of the set of the set of the set of the form of the form of the set of th For (i), (vi), (vi), and (vii), and to signals were used as input; for (iii), the 257 autosomal signals were
there were two signals with the joint-highest posterior probability in their credible set, the one with
the small there were two signals with the joint-highest posterior probability in their credible set, the one with
the smallest P-value was used. the smallest P-value was used.
We catalogued previously reported genes (Supple<mark>mentary Table 3</mark>) implicated by mapping genome-

the smallest P-value was used.
We catalogued previously repo
wide significant sentinels for th wide significant sentinels for thyroid traits using eQTL colocalization (P<1x10-7)(12) or DEPICT $(FDR \le 0.01, ref. (6))$, to define whether the genes we implicated were novel.

Expression quantitative trait loci (eQTLs)

We used the SNP2GENE function implemented by $FUMA(40)$ to facilitate the eQTL analysis. FUMA contains several eQTL datasets across a broad range of tissue types. We ran SNP2GENE requesting
eQTL results from the GTEx v8 (thyroid, hypothalamus and pituitary tissues) and eQTLGen (blood, eQTL results from the GTEx v8 (thyroid, hypothalamus and pituitary tissues) and eQTLGen (blood,
cis- and trans-eQTLs) datasets. We performed approximate colocalisation between our GWAS and
eQTL signals by identifying wheth cis- and trans-eQTLs) datasets. We performed approximate colocalisation between our GWAS and
eQTL signals by identifying whether the top variant in an eQTL signal was in one of our 95% credib
sets (**Supplementary Table 15** eQTL signals by identifying whether the top variant in an eQTL signal was in one of our 95% credible sets (Supplementary Table 15). exter symmals, in an external was in an extendion set of our protein quantitative trait loci (pQTLs)

Two pQTL datasets were included in the pQTL analyses: deCODE Genetics(41), with data for 4,719
proteins measured by 4,907 aptamers, and the SCALLOP Consortium(42), including 90 cardiovascular proteins. The significance level for pQTL associations were set as in the original publications: P-
value<1.8×10⁻⁹ for deCODE Genetics(41) and P-value<5×10⁻⁸ for the SCALLOP Consortium(42). We proteins. The significance level for pQTL associations were set as in the original publications: P-
value<1.8×10⁻⁹ for deCODE Genetics(41) and P-value<5×10⁻⁸ for the SCALLOP Consortium(42). We
performed approximate col performed approximate colocalisations were set as in the original politications: The sentinel variant in a pQTL signal was in one of our 95% credible sets.
Polygenic priority score (PoPS) value<1.8×10 ° for deCODE Genetics(41) and P-value<5×10 ° for the SCALLOP Consortium(42). We
performed approximate colocalisation between our GWAS and pQTL signals by identifying whether
the sentinel variant in a pQTL sig

Polygenic priority score (PoPS)

We used a gene prioritization tool, PoPS(9), to calculate gene features enrichment based polygenic
priority score(42) to predict genes for our TSH signals. The full set of gene features used in the analysis included 57,543 total features – 40,546 derived from gene expression data, 8,718 extracted
from a protein-protein interaction network, and 8,479 based on pathway membership. In this study, .
analysis included 57,543 total features – 40,546 derived from gene expression data, 8,718 extr
from a protein-protein interaction network, and 8,479 based on pathway membership. In this :
we prioritized genes for all aut from a protein-protein interaction network, and 8,479 based on pathway membership. In this study,
we prioritized genes for all autosomal TSH signals within a 500kb (±250kb) window of the sentinel from a protein-protein-interaction network, and 8,479 based on pathway, included on protein-increment charge,
we prioritized genes for all autosomal TSH signals within a 500kb (±250kb) window of the sentinel
11 we prioritize g enes for all autosomal TSH signals within a 500kb (\pm 250kb) window of the sentinelli \pm

It is made available under a [CC-BY 4.0 International license](http://creativecommons.org/licenses/by/4.0/) .

and reported the top prioritised genes in the region. If there was no gene prioritized within a 500kb
window of the sentinel, we reported any top prioritized genes within a 1Mb window
(Supplementary Table 16). window of the senting, we reported any top prioritized genes within a 1Mb window (Supplementary Table 16).
Nearby Mendelian disease genes:

We selected rare Mendelian-disease genes from ORPHANET (https://www.orpha.net/) within ±500kb of a TSH sentinel that were associated with thyroid-related diseases. We implicated the gene
if the string "thyro" (but not "parathyro") was included in either the disease name or appeared if the string "thyro" (but not "parathyro") was included in either the disease name or appeared
frequently in human phenotype ontology (HPO) terms for that disease. We manually checked the
diseases and HPO terms identified frequently in human phenotype ontology (HPO) terms for that disease. We manually checked the diseases and HPO terms identified for relevance (Supplementary Table 17). diseases and HPO terms identified for relevance (Supplementary Table 17).
Nearby mouse knockout orthologs with thyroid related phenotype:

diseases and HPO terms identified for relevance (**Supplementary Table 17**).
Nearby mouse knockout orthologs with thyroid related phenotype:
We selected human orthologs of mouse knockout genes with thyroid related phenotype We selected human orthologies in the case with the genes with thyroid phenotype.org/) within
t500kb of a TSH sentinel. The thyroid related phenotypes included enlarged thyroid gland, abnormal
thyroid gland morphology and i the International Mouse Phenotyping constraint (<u>https://www.mousephen.org/)</u> within
±500kb of a TSH sentinel. The thyroid related phenotypes included enlarged thyroid gland, abne
thyroid gland morphology and increased/dec thyroid gland morphology and increased/decreased circulating thyroxine level (**Supplementary**
Table 18). thyroid gland morphology and increased/decreased circulating thyroxine level (Supplementary
Table 18).
Functional annotation of credible sets:

Functional annotation of credible sets:

We annotated variants in the 95% credible sets using Variant Effect Predictor (VEP) (43) . We implicated the gene if there was a variant with >50% posterior probability in the credible set that
was also either a missense variant, annotated as "deleterious" by SIFT, annotated as "damaging"
PolyPhen-2 or had a CADD was also either a missense variant, annotated as "deleterious" by SIFT, annotated as "damaging" by was also either a missense variant, and there as "deleterious" by SIFT, and there are "damaging" by
PolyPhen-2 or had a CADD PHRED score 220.
Rare variant analysis from whole exome and whole genome sequencing

Rare variant analysis from whole exome and whole genome sequencing

,
Rare variant analysis from whole e:
We performed a lookup for rare variant assod ±500kb of a TSH sentinel using the following resources: (i) single variant and gene-based exonic
associations from the AstraZeneca PheWAS Portal(10) (<u>https://azphewas.com/</u>); (ii) single variant
whole-genome associations associations from the AstraZeneca PheWAS Portal(10) (https://azphewas.com/); (ii) single variar
whole-genome associations in 150,119 UK Biobank participants(11). For all tests, we used MAF<
and P<5x10⁻⁶ (**Supplementary T** whole-genome associations in 150,119 UK Biobank participants(11). For all tests, we used MAF<1% and $P < 5x10^{-6}$ (Supplementary Table 19).

Pathway analysis

We used Consensus Path DB (15) to test for enrichment of our prioritised genes in up to 31 pathway and gene set ontology databases. Pathways with FDR<5% are reported.

Pathway-specific GRS

We selected 26 pathways that were enriched at FDR <5x10⁻⁴ for our 112 genes implicated by 2 or more lines of evidence (**Supplementary Table 8**). We created a weighted GRS (weights estimated We selected 26 pathways that were enriched at FDR <5x10 " for our 112 genes implicated by 2 or
more lines of evidence (**Supplementary Table 8**). We created a weighted GRS (weights estimated
from the TSH meta-analysis of UK more lines of evidence (Supplementary Table 8). We created a weighted GRS (weights estimated
from the TSH meta-analysis of UK Biobank and EXCEED) for each of the 26 pathways by including,
each gene in the pathway, the vari each gene in the pathway, the variant with the most significant P-value that implicates the gene in our variant-to-gene mapping (Supplementary Table 2). Each of the 26 GRS were then checked for association with up to 1939 traits in the PheWAS. association with up to 1939 traits in the PheWAS.

It is made available under a [CC-BY 4.0 International license](http://creativecommons.org/licenses/by/4.0/) .

Polygenic score (PGS)

We approach the Constraint Processor (PC) in the summary statistic and the summary statistics from
Michigan Genomic Initiative [MGI], ThyroidOmics)) as the training dataset. PRS-CS-auto is a Bayesian
approach, which automa our discovery GWAS (a meta-analysis of OK Biobank, EXCEED and results from Zhou et al (HONT,
Michigan Genomic Initiative [MGI], ThyroidOmics)) as the training dataset. PRS-CS-auto is a Baye
approach, which automatically le Michigan General Michigan General, Thyroidomics), as the training dataset the securities are provided approach, which automatically learns hyper-parameters from the training data; no validation dataset.
Is required. We tes approach, which automatically reacheshyper-parameters from the training auto, its cannot altace to
is required. We tested the association of this PGS trained from EUR ancestry group with TSH in non-
EUR ancestry groups in first 10 principal components of ancestry. We evaluated the association of the TSH PGS with EUR ancessing groups in UK Biobana, including AFR, AMR, 2007, 2007, 2007, 2008 and the association of the TSH PGS with first 10 principal components of ancestry. We evaluated the association of the TSH PGS with susceptibil first 10 principal components of ancestry. We evaluated the association of the TSH PGS with
susceptibility to hypothyroidism, hyperthyroidism, thyroid cancer, other thyroid diseases and thyroid
eye disease in ancestry grou susceptibility to hypothyroidism, hyperthyroidism, thyroid cancer, other thyroid diseases and
eye disease in ancestry groups with more than 100 cases in UK Biobank. Associations were te
using logistic regression models, ad eye disease in ancestry groups with more than 100 cases in UK Biobank. Associations were tested
using logistic regression models, adjusted for genotyping array, sex and the first ten 10 principal
components of ancestry. To using logistic regression models, adjusted for genotyping array, sex and the first ten 10 principal
components of ancestry. To further aid clinical interpretation, we divided individuals into deciles
according to their PGS components of ancestry. To further aid clinical interpretation, we divided individuals into deciles
according to their PGSs and using logistic regression, investigated disease risk associated, compar
each decile to a refer according to their PGSs and using logistic regression, investigated disease risk associated, compar
each decile to a reference decile. To evaluate the age-dependent PGS performance, we used the
Kaplan–Meier method to gener each decile to a reference decile. To evaluate the age-dependent PGS performance, we used the
Kaplan–Meier method to generate a cumulative incidence plot and a log-rank test to test for
differences between groups. Kaplan–Meier method to generate a cumulative incidence plot and a log-rank test to test for
differences between groups. Kaplan–Merences between groups.
Meifferences between groups.

differences between groups. Phenome-wide association study (PheWAS)

 $\begin{array}{c} \n\downarrow \\ \n\downarrow \n\end{array}$ To identify pleiotropic associations with a wide range of phenotypes, we used DeepPheWAS(7), a
flexible PheWAS framework which incorporates phenotypes not present in other PheWAS platforms
for: (i) sentinel variants impli missense/damaging/deleterious/phred-scaled CADD score \geq 20 that also had a posterior probability for: (Supplementary Table 2), variants in a credible set that were annotated as
thissense/damaging/deleterious/phred-scaled CADD score ≥20 that also had a posterior pro
>50% (Supplementary Table 20), and low-frequency sen Supplementary Table 2), variants in a credible set that were annotated as
missense/damaging/deleterious/phred-scaled CADD score ≥20 that also ha
>50% (Supplementary Table 20), and low-frequency sentinel variants (MAF
Table >50% (Supplementary Table 20), and low-frequency sentinel variants (MAF<1%, Supplementary
Table 1); (ii) the PGS for TSH; (iii) pathway-specific genetic risk score (GRS).
□ Table 1); (ii) the PGS for TSH; (iii) pathway-specific genetic risk score (GRS).

Druggability

To identify gene products that are the targets of drugs, we queried the Drug Gene Interaction
Database (DGIDB) (https://www.dgidb.org) for the 112 putative causal genes supported by ≥ 2 To identify gene products that are the targets of drugs, it of queried the origion and the targets of Database
Database (DGIDB) (https://www.dgidb.org) for the 112 putative causal genes supported by \geq 2
variant-to-gene Database (DGIDB) (parameter of the 112 putative carrier of the 112 putations of the 112 putations of the 112 puta
headings). $\mathsf{headings}.$

..............
Acknowledgements

 $\frac{1}{1}$ an NIHR Senior Investigator Award to **M.D. Tobin**; views expressed are those of the author(s) and
not necessarily those of the NHS, the NIHR or the Department of Health. The funders had no role in
the design of the study. an NIHR Senior Investigator Award to M.D. Tobin; views expressed are those of the author(s) and
not necessarily those of the NHS, the NIHR or the Department of Health. The funders had no role is
the design of the study.

not necessarily the State of the NHS, the NHMS are Separtment of Health. The Chapter had no role in
This research was funded in whole, or in part, by the Wellcome Trust: Wellcome Trust Investigator the design of the study.
This research was funde
Award (WT202849/Z/16 Award (WT202849/Z/16/Z, MD Tobin) and Wellcome Trust Discovery Award (WT225221/Z/22/Z, MD
Tobin and L.V.Wain). For the purpose of open access, the author has applied a CC BY public copyright
licence to any Author Accepted Tobin and L.V. Wain). For the purpose of open access, the author has applied a CC BY public copyright Interace to any Author Accepted Manuscript version arising from this submission.
L. V. Wain was supported by GSK/British Lung Foundation Chair in Respiratory Research.

licence to any Author Accepted Manuscript version aning mean this submission.
L. V. Wain was supported by GSK/British Lung Foundation Chair in Respiratory Re L. V. Wain was supported by GSK/British Lung Foundation Chair in Respiratory Research. It is made available under a [CC-BY 4.0 International license](http://creativecommons.org/licenses/by/4.0/) .

C. Batini was supported by a OKRI Innovation Fellowship at Health Data Research OK
(MR/S003762/1).
<mark>C. John</mark> held a Medical Research Council Clinical Research Training Fellowship (MR/PC

、
C. John held a Me
EXCEED is support

C. John held a Medical Research Council Clinical Research Training Fellowship (MR/P00167X/1).
EXCEED is supported by the University of Leicester, the NIHR Leicester Respiratory Biomedical
Research Centre, by Wellcome [2028 EXCEED is supported by Wellcome [202849, https://doi.org/10.35802/202849] and by Cohort Acce
fees from studies funded by the Medical Research Council (MRC), BBRSC, NIHR, the UK Space
Agency, and GSK. It was previously supp fees from studies funded by the Medical Research Council (MRC), BBRSC, NIHR, the UK Space
Agency, and GSK. It was previously supported by MRC grant G0902313. This work is supported by
BREATHE - The Health Data Research Hub Agency, and GSK. It was previously supported by MRC grant G0902313. This work is supported
BREATHE - The Health Data Research Hub for Respiratory Health [UKR_PC_19004] in partnersl
with SAIL Databank. We also thank all par Agency, and General Mac previously supported by MRC grant GOG COND (DKR_PC_19004) in partnership
BREATHE - The Health Data Research Hub for Respiratory Health [UKR_PC_19004] in partnership
with SAIL Databank. We also thank BREATHE - The Health Data Research Hub for Respiratory Health [UKR_PC_19004] in partnership
with SAIL Databank. We also thank all participants and staff who have contributed their time to the
study. study.
The UK Biobank genetic and phenotypic data were analysed under UK Biobank Application 43027.

We would like to thank the ThyroidOmics consortium, Nord-Trøndelag Health (HUNT) Study, and The UK Biobaning Biobana phenotypic and nucleanalysed under UK Biobana ppercentic 1992.
We would like to thank the ThyroidOmics consortium, Nord-Trøndelag Health (HUNT) Study, and
Michigan Genomics Initiative for depositin Michigan Genomics Initiative for depositing genome-wide summary statistics from previous studion.
In GWAS Catalog, and we are grateful to the staff who develop and maintain the GWAS Catalog.
—————————————————————————————— in GWAS Catalog, and we are grateful to the staff who develop and maintain the GWAS Catalog.

Ethical approval

The UK Biobank genetic and phenotypic data were analysed under UK Biobank Application 43027.
UK Biobank has ethical approval from the UK National Health Service (NHS) National Research Ethics The Entertaining Press and presses, pressure and pressure, pressure and pressure pressure in teach R
UK Biobank has ethical approval from the UK National Health Service (NHS) National Research Ethic
Service (11/NW/0382). E Service (11/NW/0382). EXCEED received ethical approval from the Leicester Central Research Ethics
Committee (13/EM/0226). Informed consent was obtained from all participants. Committee (13/EM/0226). Informed consent was obtained from all participants.

Data availability

Access to UK Biobank and EXCEED datasets is available to bona fide researchers upon application (in accordance with the terms of ethical approval and participant consent). Genome-wide summary statistics will be made publicly available via the EMBL-EBI GWAS Catalog.

Author Contributions

C.J. and M.D.T. supervised the study. A.T.W., J.C., N.S., M.D.T. and C.J. designed the study. A.T.W.,
J.C., K.C., C.B., A.I., R.P. and N.S. performed statistical analyses. A.T.W., J.C., K.C., C.B., A.I., R.P., J.C., K.C., C.B., A.I., R.P. and N.S. performed statistical analyses. A.T.W., J.C., K.C., C.B., A.I., R.P.,
E.J.H., W.H., B.S., F.D., L.V.W., N.S., M.D.T., C.J. analysed and/or interpreted the data. F.D. and L.V.
provided I.C., M.E., M.E., M.E., M.E., M.E., Protocology and Article Book and Article Book and Article Book and Legislation.
B.J.H., W.H., B.S., F.D., L.V.W., N.S., M.D.T., C.J. analysed and/or interpreted the data. F.D. and L
prov provided methodological and statistical advice. A.T.W., J.C., N.S., M.D.T. and C.J. wrote the
manuscript. All remaining co-authors contributed to data collection, management or analysis of the
EXCEED study. All co-authors provided methodological and statistical advices methodological and statistical advice. A.T. and consider manuscript.
EXCEED study. All co-authors critically reviewed the manuscript.
Competing interests EXCEED study. All co-authors critically reviewed the manuscript.

Competing interests

M.D.T. and L.V.W. have previously received funding from GSK for collaborative research projects outside of the submitted work. R.J.P., M.D.T., C.J. and L.V.W. have a funded research collaboration with Orion for collaborative research projects outside the submitted work.
References
1. Taylor PN. Albrecht D. Scholz A. Gutierrez-Buey G. Lazarus JH. Dava

References

Taylor PN, Albrecht D, Scholz A, Gutierrez-Buey G, Lazarus JH, Dayan CM, et al. Global epidemiology of hyperthyroidism and hypothyroidism. Nature Reviews Endocrinology.
2018;14(5):301-16.
2. Hollowell JG, Staehling NW, Flanders WD, Hannon WH, Gunter EW, Spencer CA, et al. Serum

TSH, T(4), and thyroid antibodies in the United States population (1988 to 1994): National Health 2018;14(5):301-16. 2. Hollowell Statehling Hingmann (1988), The Hing, School and Nutritian J. Statehling TSH, T(4), and thyroid antibodies in the United States population (1988 to 1994): National Health
and Nutrition Examination Survey (NHAN

TEM, T(1), and any the antibodies in the UNES III). J Clin Endocrinol Metab. 2002;87(2):489-99.
and Nutrition Examination Survey (NHANES III). J Clin Endocrinol Metab. 2002;87(2):489-99.
3. The Praveen K, Patel GC, Gurski 3. Praveen K, Patel GC, Gurski L, Ayer AH, Persaud T, Still MD, et al. ANGPTL7, a therap.
target for increased intraocular pressure and glaucoma. Commun Biol. 2022;5(1):1051.
. 3. Practice Contraction, Patel GC, Gurski L, Ayer Andy Persaud C, Gurskin C, Gurski L, Ayer Atheney C, Angels
Target for increased intraocular pressure and glaucoma. Commun Biol. 2022;5(1):1051. t is target for increased intraocular pressure and glaucoma. Communistic t 2022; t

4. Nioi P, Sigurdsson A, Thorleifsson G, Helgason H, Agustsdottir AB, Norddahl GL, et al. Variant
ASGR1 Associated with a Reduced Risk of Coronary Artery Disease. N Engl J Med.
2016;374(22):2131-41. 4.

2016;374(22):2131-41.
5. Verweij N, Haas ME, Nielsen JB, Sosina OA, Kim M, Akbari P, et al. Germlii
CIDEB and Protection against Liver Disease. New England Journal of Medicine. 20: 2016;374(22):2131-41. 5. Verweij N, Haas ME, Nielsen JB, Sosina OA, Kim M, Akbari P, et al. Germline Mutations in

6. Zhou W, Brumpton B, Kabil O, Gudmundsson J, Thorleifsson G, Weinstock J, et al. GWAS of thyroid stimulating hormone highlights pleiotropic effects and inverse association with thyroid cancer. Nat Commun. 2020;11(1):3981

7. Packer R, Williams A, Hennah W, Eisenberg M, Fawcett K, Pearson W, et al. Deep-PheWAS: a thyroid stimulating hormone highlights pleicits pleicits and inverse association with thyroid
cancer. Nat Commun. 2020;11(1):3981.
pipeline for phenotype generation and association analysis for phenome-wide association stu cancer. Nat Commun. 2020;11(1):3981. pipeline for phenotype generation and association analysis for phenome-wide association studies.
medRxiv. 2022.
8. Fabbri C, Hagenaars SP, John C, Williams AT, Shrine N, Moles L, et al. Genetic and clinical

pipedRxiv. 2022.
8. Fabbri C, Hagenaars SP, John C, Williams AT, Shrine N, Moles L, et al. Genetic and clinical
characteristics of treatment-resistant depression using primary care records in two UK cohorts. medRxiv. 2022. 8. Fabbri C, Hagenaars SP, John C, Hambri C, John C, Harle B, John Central and chinem
Characteristics of treatment-resistant depression using primary care records in two UK cohorts.
9. Weeks EM, Ulirsch JC, Cheng N

Molecular Psychiatry. 2021;26(7):3363-73.
9. Weeks EM, Ulirsch JC, Cheng NY, Trippe BL, Fine RS, Miao J, et al. Leveraging polygenic
enrichments of gene features to predict genes underlying complex traits and dis Molecular Psychiatry. 2021;26(7):3363-73. 9. Weeks Em, School, Cheng Al, Miller, Miller, Miller, Pause Prignig polygenic
2020.
10. Wang Q, Dhindsa RS, Carss K, Harper AR, Nag A, Tachmazidou I, et al. Rare variant

enrichments of gene features to predict genes underlying complex traits and diseases. medRxiv. 10.
contril
11.

contribution to human disease in 281,104 UK Biobank exomes. Nature. 2021;597(7877):52
11. Halldorsson BV, Eggertsson HP, Moore KHS, Hauswedell H, Eiriksson O, Ulfarsson N
The sequences of 150,119 genomes in the UK 11. Halldorsson BV, Eggertsson HP, Moore KHS, Hauswedell H, Eiriksson O, Ulfarsson MO, et al.
The sequences of 150,119 genomes in the UK Biobank. Nature. 2022;607(7920):732-40.
12. Teumer A, Chaker L, Groeneweg S, Li Y, Di

11. Hall documents of the UK Biobank. Nature, 2022;607(7920):732-40.
12. Teumer A, Chaker L, Groeneweg S, Li Y, Di Munno C, Barbieri C, et al. Genome-wide analyses
identify a role for SLC17A4 and AADAT in thyroid hormone r The sequences of 2007-2022; Seminon and Subsember of 150,119 (150,11920):
12. Teumer A, Chaker L, Groeneweg S, Li Y, Di Munno C, Barbieri C, et al. Genome-w
identify a role for SLC17A4 and AADAT in thyroid hormone regulati identify a role for SLC17A4 and AADAT in thyroid hormone regulation. Nat Commun. 2018;9(1):4455.
13. Kurki MI, Karjalainen J, Palta P, Sipilä TP, Kristiansson K, Donner K, et al. FinnGen: Unique
genetic insights fr

13. Kurki MI, Karjalainen J, Palta P, Sipilä TP, Kristiansson K, Donner K, et al. FinnGen: Unique
genetic insights from combining isolated population and national health register data. 2022.
14. Mujahid S, Hunt KF, Cheah Y

genetic insights from combining isolated population and national health register data. 2022.
14. Mujahid S, Hunt KF, Cheah YS, Forsythe E, Hazlehurst JM, Sparks K, et al. The Endocrine ar
Metabolic Characteristics of genetic insights from the KF, Cheah YS, Forsythe E, Hazlehurst JM, Sparks K, et al. The Endocri
Metabolic Characteristics of a Large Bardet-Biedl Syndrome Clinic Population. The Journal of
Endocrinology & Metabolism. 2018; Metabolic Characteristics of a Large Bardet-Biedl Syndrome Clinic Population. The Journal of Clinical
Endocrinology & Metabolism. 2018;103(5):1834-41.
15. Herwig R, Hardt C, Lienhard M, Kamburov A. Analyzing and in

Endocrinology & Metabolism. 2018;103(5):1834-41.
15. Herwig R, Hardt C, Lienhard M, Kamburov A. Analyzing and interpreting genome data at the
network level with ConsensusPathDB. Nature Protocols. 2016;11(10):1889-9 network level with ConsensusPathDB. Nature Protocols. 2016;11(10):1889-907.
16. Porcu E. Medici M. Pistis G. Volpato CB. Wilson SG. Cappola AR. et al. A

network level with ConsensusPathDB. Nature Protocols. 2016;11(10):1889-907.
16. Porcu E, Medici M, Pistis G, Volpato CB, Wilson SG, Cappola AR, et al. A Meta-Analysis of
Thyroid-Related Traits Reveals Novel Loci an network level with Consensus Protocols. 2019
16. Porcu E, Medici M, Pistis G, Volpato CB, Wilson SG, Cappola AR, et al. A I
Thyroid Function. PLoS Genetics. 2013;9(2):e1003266. Thyroid-Related Traits Reveals Novel Loci and Gender-Specific Differences in the Regulation of
Thyroid Function. PLoS Genetics. 2013;9(2):e1003266.
17. Shrine N, Izquierdo AG, Chen J, Packer R, Hall RJ, Guyatt AL,

Thyroid Function. PLoS Genetics. 2013;9(2):e1003266.
17. Shrine N, Izquierdo AG, Chen J, Packer R, Hall RJ, Guyatt AL, et al. Multi-ancestry genon
wide association study improves resolution of genes, pathways and ple 17. Shrine N, Izquierdo AG, Chen J, Packer R, Hall F
wide association study improves resolution of genes, p
chronic obstructive pulmonary disease. 2022.

18. Cappola AR, Arnold AM, Wulczyn K, Carlson M, Robbins J, Psaty BM. Thyroid function in the whe associated association study improves resolution of genes, pathways and pleiotropy for lung function and
18. Cappola AR, Arnold AM, Wulczyn K, Carlson M, Robbins J, Psaty BM. Thyroid function in the
euthyroid range and chronic obstructive politician, measure of the Cappola AR, Arnold AM, Wulczyn K, Cappo euthyroid range and adverse outcomes in older adults. J Clin Endocrinol Metab. 2015;100(3):1088-
96.
19. Sirugo G, Williams SM, Tishkoff SA. The Missing Diversity in Human Genetic Studies. Cell.

96.
19. Sirugo G, Williams SM, Tishkoff SA. The Missing Diversity in Human Genetic Studies. Cell.
2019;177(1):26-31.
20. Association for Clinical Biochemistry BTA. British Thyroid Foundation. UK Guidelines for t 2019;177(1):26-31.

Association for Clinical Biochemistry BTA, British Thyroid Foundation. UK Guidelines for the 2019;1777(1):26-21.
20. Association
Use of Thyroid Func
association.org/san 20. Association for Clinical Biochemistry Biochemistry Biochemistry Discoversion of Thyroid-
20. Association.org/sandbox/bta2016/uk guidelines for the use of thyroid function tests.pdf.
21. MICE. Thyroid disease: assessmen

http://www.nice.org.uk association.org/sandbox/bta2016/uk_guidelines_for_the_use_of_thyroid_function_tests.pdf
21. NICE. Thyroid disease: assessment and management 2019 [Available from:
http://www.nice.org.uk.
22. Sudlow C, Gallacher

21. Interaction, the assessment and management 2019 [Araman Press)
22. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. UK Bioban
Access Resource for Identifying the Causes of a Wide Range of Complex Dis and an Extended Control Control
Access Resource for Iden
Age. PLOS Medicine. 201 Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and
Age. PLOS Medicine. 2015;12(3):e1001779. Age. PLOS Medicine. 2015;12(3):e1001779.
 $\frac{1}{2}$ Age. PLOS Medicine. 2015;12(3):e1001779.

It is made available under a CC-BY 4.0 International license.

23. John C, Reeve NF, Free RC, Williams AT, Ntalla I, Farmaki A-E, et al. Cohort Profile: Extended
Cohort for E-health, Environment and DNA (EXCEED). International Journal of Epidemiology.
2019;48(3):678-9j.
24. Chang CC. 23.

24. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising
to the challenge of larger and richer datasets. Gigascience. 2015;4:7.
25. Pan-ancestry genetic analysis of the UK Biobank

to the challenge of larger and richer datasets. Gigascience. 2015;4:7.
25. Pan-ancestry genetic analysis of the UK Biobank 2020 [Availa
https://pan.ukbb.broadinstitute.org/. 24. Chang CC, Chang CC, Chang CC, Change of Language of larger and richer datasets. Gigascience. 2015;4:7.
25. Pan-ancestry genetic analysis of the UK Biobank 2020 [Available from:
https://pan.ukbb.broadinstitute.org/ 25. Pan-ancestry genetic analysis of the UK Biobank 2020 [Availa
<u>https://pan.ukbb.broadinstitute.org/</u>
26. Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen

25. Pan-ancestry generic analysis of the UK Biobana 2021 [Araman Promoth
26. Pan-ancestry pan-analysis of the UK Biobana 2020
126. Pan-ance in genome-wide association studies. Bioinformatics. 2010;26(22):286 man. 1988.
26. Manichaikul A, Mychaleckyj JC
inference in genome-wide association
27. Bulik-Sullivan BK, Loh PR, Finu

26. Manutematic, Mychalecky, Journal, Journal, A, 28. Mychalecky, 2010;26(22):2867-73.
27. Bulik-Sullivan BK, Loh PR, Finucane HK, Ripke S, Yang J, Patterson N, et al. LD Score
regression distinguishes confounding from pol 27. Bulik-Sullivan BK, Loh PR, Finucane HK, Ripke S, Yang J, Patterson N, et al. LD
regression distinguishes confounding from polygenicity in genome-wide association s
Genet. 2015;47(3):291-5. regression distinguishes confounding from polygenicity in genome-wide association studies
Genet. 2015;47(3):291-5.
28. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide Genet. 2015;47(3):291-5.
28. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide

association scans. Bioinformatics. 2010;26(17):2190-1.
29. Weissbrod O, Hormozdiari F, Benner C, Cui R, Ulirsch J, Gazal S, et al. Functionally informed association scans. Bioinformatics. 2010;26(17):2190-1.
29. Weissbrod O, Hormozdiari F, Benner C, Cui R, Ulirsch J, Gazal S, et al. Functionally in
fine-mapping and polygenic localization of complex trait heritabilit 29. Weissbrod O, Hormozdiari F, Benner C, Cui R, L
fine-mapping and polygenic localization of complex trai
2020;52(12):1355-63.

2020;52(12):1355-63.
30. Wang G, Sarkar A, Carbonetto P, Stephens M. A simple new approach to variable selection in 2020;52(12):1355-63.
30. Wang G, Sarkar A, Carbonetto P, Stephens M. A simple new approach to varial
regression, with application to genetic fine mapping. Journal of the Royal Statistical So 2020;52(12):1355-63. regression, with application to genetic fine mapping. Journal of the Royal Statistical Society: Series B
(Statistical Methodology). 2020;82(5):1273-300.
31. Finucane HK, Bulik-Sullivan B, Gusev A, Trynka G, Reshef

regression, with application to genetic fine mapping. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 2020;
31. Finucane HK, Bulik-Sullivan B, Gusev A, T
heritability by functional annotation using genon
Genetics. 2015;47(11):1228-35. 31. Final Muslim May 2008 Manuslin Cyllister A, A, Muslim A, 2008, A, Partitioning
Senetics. 2015;47(11):1228-35.
32. Makefield J. A Bayesian Measure of the Probability of False Discovery in Genetic

Genetics. 2015;47(11):1228-35.
32. Wakefield J. A Bayesian Measure of the Probability of False Discovery in Genetic
Epidemiology Studies. The American Journal of Human Genetics. 2007;81(2):208-27. France Breed, (11):1111-111
32. Makefield J. A Bayesian
Epidemiology Studies. The Amer
33. Gudmundsson J, Sulem

Epidemiology Studies. The American Journal of Human Genetics. 2007;81(2):208-27.
33. Gudmundsson J, Sulem P, Gudbjartsson DF, Jonasson JG, Masson G, He H, et al. I
common variants associated with low TSH levels and 33. Gudmundsson J, Sulem P, Gudbjartsson DF, Jonasson JG, Masson G, He H, et al. Discovery of
common variants associated with low TSH levels and thyroid cancer risk. Nature Genetics.
2012;44(3):319-22.
34. Kwak SH. Park YJ common variants associated with low TSH levels and thyroid cancer risk. Nature Genetics.
2012;44(3):319-22.
34. Kwak SH, Park YJ, Go MJ, Lee KE, Kim SJ, Choi HS, et al. A genome-wide association study on

common variants associated with the state with low the surface with the theorem.
2012;44(3):319-22.
thyroid function and anti-thyroid peroxidase antibodies in Koreans. Hum Mol Genet. thyroid function and anti-thyroid peroxidase antibodies in Koreans. Hum Mol Genet.
2014;23(16):4433-42.

35. Malinowski JR, Denny JC, Bielinski SJ, Basford MA, Bradford Y, Peissig PL, et al. Genetic thyroid function and any the permanic and anti-the more and molecules in the sense.
2014;23(16):4433-42.
variants associated with serum thyroid stimulating hormone (TSH) levels in European 2014;23(16):4433-42. variants associated with serum thyroid stimulating hormone (TSH) levels in European American
African Americans from the eMERGE Network. PLoS One. 2014;9(12):e111301.
36. Medici M, Van Der Deure WM, Verbiest M, Verm

African Americans from the eMERGE Network. PLoS One. 2014;9(12):e111301.
36. Medici M, Van Der Deure WM, Verbiest M, Vermeulen SH, Hansen PS, Kiemeney LA, et al. A
large-scale association analysis of 68 thyroid hor African American Americans from the emergency of the emergency of the emergency of the emerge S6. And an emerge S6. I
Harge-scale association analysis of 68 thyroid hormone pathway genes with servelevels. European Journal 36. Medici M, Van Der Deure Win, Verbiest M, Vermeusen Styrman (Symmetry) any of any
large-scale association analysis of 68 thyroid hormone pathway genes with serum TSH and FT4
levels. European Journal of Endocrinology. 20

lays scale association analysis of 19 the 19 th
levels. European Journal of Endocrinology. 2011;164(5):781-8.
association study of thyroid stimulating hormone 37. Nielsen TR, Appel EV, Svendstrup M, Ohrt JD, Dahl M, F
association study of thyroid stimulating hormone and free thyr
adolescents. PLoS One. 2017;12(3):e0174204. 33. Nielsen TR, Appel EV, Scenaring M, Scholz, Scholz, Scholz, Scheme Thatassociation study of thyroid stimulating hormone and free thyroxine in Danish children and
adolescents. PLoS One. 2017;12(3):e0174204.
38. Popović M

adolescents. PLoS One. 2017;12(3):e0174204.
adolescents. PLoS One. 2017;12(3):e0174204.
38. Popović M, Matana A, Torlak V, Boutin T, Brdar D, Gunjača I, et al. Genome-wide me
analysis identifies novel loci associated with 38. Popović M, Matana Á, Torlak V, Boutin
analysis identifies novel loci associated with fro
J Endocrinol Invest. 2019;42(10):1171-80.

39. Taylor PN, Porcu E, Chew S, Campbell PJ, Traglia M, Brown SJ, et al. Whole-genome analysis identifies novel 1013-1200.
1 Endocrinol Invest. 2019;42(10):1171-80.
39. Taylor PN, Porcu E, Chew S, Campbell PJ, Traglia M, Brown SJ, et al. Whole-genome.
sequence-based analysis of thyroid function. Nat J Endocrinol Invest. 2019;42(10):1171-80.
39. Taylor PN, Porcu E, Chew S, Campbell PJ, Traglia M, Brown SJ, et a
sequence-based analysis of thyroid function. Nat Commun. 2015;6:5681.
40. Watanabe K, Taskese

Watanabe K, Taskesen E, Van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. Nature Communications. 2017;8(1).
41. Ferkingstad F. Sulem P. Atlason BA. Sveinbiornsson G. Magnusso

Ferkingstad E, Sulem P, Atlason BA, Sveinbjornsson G, Magnusson MI, Styrmisdottir EL, et al. 41. Ferkingstad E, Sulem P, Atlason BA, Sveinbjornsson G, Magnussc
Large-scale integration of the plasma proteome with genetics and diseas
2021;53(12):1712-21. Large-scale integration of the plasma proteome with genetics and disease. Nature Genetics.
2021;53(12):1712-21. $2021;53(12):1712-21.$ $2025(12)$:1713-21.

It is made available under a [CC-BY 4.0 International license](http://creativecommons.org/licenses/by/4.0/) .

42. Folkersen L, Gustafsson S, Wang Q, Hansen DH, Hedman ÅK, Schork A, et al. Genomic and
drug target evaluation of 90 cardiovascular proteins in 30,931 individuals. Nature Metabolism.
2020;2(10):1135-48. Cunningham F, All 42.

43. Cunningham F, Allen JE, Allen J, Alvarez-Jarreta J, Amode M R, Armean Irina M, et al.
Ensembl 2022. Nucleic Acids Research. 2022;50(D1):D988-D95.
44. Ge T, Chen C-Y, Ni Y, Feng Y-CA, Smoller JW. Polygenic prediction vi Ensembl 2022. Nucleic Acids Research. 2022;50(D1):D988-D95.

Ge T, Chen C-Y, Ni Y, Feng Y-CA, Smoller JW. Polygenic prediction via Bayesian regression and 44. Ge T, Chen C-Y, Ni Y, Feng Y-CA, Smoller JW. Polygenic p
continuous shrinkage priors. Nature Communications. 2019;10(
. 44. Gentinuous shrinkage priors. Nature Communications. 2019;10(1). continuous shrinkage priors. Nature Communications. 2019;10(1).

hypothyroid
hyperthyroid h'ypothyroid
hyperthyroid
cancer edible hypothyroid õ nyperthyro iypothyroi cre ٠ě yperthy ouse ouse cancer
other hcer ত্ ਨ ㅎ 片 cance
other 동 공 $\bar{5}$ 풂 玉 $\overline{\mathbf{e}}$ ਨ ಹ nal 듕 ō a
othe $SDCCAG8(3)$ $BUB1B(2)$ $TG(5)$ $THADA(2)$ $AKT1(4)$ $TRIM2(3)$ $CCBE1(2)$ $CRTC3(2)$ $TPO(4)$ $PDE10A(2)$ $WWTR1(2)$ $PDE4B(2)$ $TSHR(4)$ $NR3C2(2)$ $GNGZ(2)$ PPP2R1B(2) $INSR(3)$ $PDE8B(2)$ $APOH(2)$ $PTPAA1(2)$ $GLIS3(3)$ $VEGFA(2)$ $TPPP(2)$ $SMAD6(2)$ $NRG1(3)$ $CAPZB(2)$ $IHH(2)$ **BDKRB2(2)** $IGFBP5(3)$ $NFIA(2)$ $WNT4(2)$ **DZANK1(2)** 10^{-20} $VAV3(2)$ $DLEU1(2)$ SPATA13(3) $ARNL(2)$ $MBIP(2)$ $SNX8(2)$ $SMOC2(3)$ $PTPRJ(2)$ 5×10^{-8} $NKX2-1(2)$ $LDLRAD4(2)$ $PCMTD1(2)$ $CADM1(3)$ 10^{-5} $BCAS3(3)$ $NKX2-3(2)$ $BMP2(2)$ $CPT1A(2)$ 10^{-2} $NSE(2)$ $C1orf116(2)$ $BNC2(3)$ $GNAS(2)$ $SOX9(2)$ $PTEN(3)$ $MAL2(2)$ $TRHR(2)$ $SULF1(3)$ $FOX42(2)$ $NRXN3(2)$ $PRKCZ(2)$ 10^{-2} $SASH1(3)$ $CGA(2)$ $TSPAN6(2)$ **FAM96A(2)** 10^{-5} $PDE4D(3)$ $HLA-C(2)$ $SPATA6(2)$ $SNX22(2)$ 5×10^{-8} $ANXA5(3)$ $GATA3(2)$ $RBM4Z(2)$ **KANK1(2)** $HMGA2(3)$ SLC25A37(2) PPARGC1A(2) $CDC16(2)$ 10^{-20} $CDKN2C(2)$ $PDEZA(2)$ FAM220A(2) **KIAA1217(3)** Yes $PTPRS(3)$ B4GALNT3(2) $$ $OCLN(2)$ $\overline{ADCY6}(3)$ $PRKX(2)$ $NEB(2)$ $BMP4(2)$ $VEGFC(3)$ $C9$ orf $92(2)$ $ETS1(2)$ $TGFB2(2)$ $PHC2(3)$ $NEK6(2)$ $RNF144B(2)$ $DIO2(2)$ $SGK1(3)$ GNG12(2) $KDR(2)$ $KCTD5(2)$ $SPPL3(3)$ $ASXL2(2)$ $B4GALNT4(2)$ **DET1(2)** $KRT18(3)$ $SH2B3(2)$ $FRS2(2)$ $SYT13(2)$ $TSHZ3(3)$ $CERS6(2)$ $C9$ orf156 (2) $ZBTB17(2)$

Figure 1: 112 genes prioritised by two or more variant-to-gene criteria. The first seven columns indicate that at least one variant implicates the corresponding gene via the evidence for that column. The remaining six columns indicate the strength of association of the most significant variant implicating the corresponding gene as causal with respect to the TSH increasing allele, such that shades of blue represent associations with the other thyroid phenotypes that have the same direction of effect as the TSH association and shades of red represent an opposite direction of effect to the TSH association.

. [CC-BY 4.0 International license](http://creativecommons.org/licenses/by/4.0/) It is made available under a

It is made available under a CC-BY 4.0 International license

It is made available under a [CC-BY 4.0 International license](http://creativecommons.org/licenses/by/4.0/) .

Figure 2: PheWAS for pathway-specific TSH-weighted GRS partitioned by: (a, top) activin receptorlike kinase (ALK) in cardiac myocytes pathway (Biocarta); (b, bottom) platelet activation, signalling and aggregation pathway (Reactome).

It is made available under a [CC-BY 4.0 International license](http://creativecommons.org/licenses/by/4.0/) .

PGS associations with TSH/Free T4 in non-EUR ancestry groups in UK Biobank

Figure 3: PGS performance across ancestries. Prediction performance of the TSH PGS across ancestry groups in UK Biobank shown as standard deviation (SD) change in TSH/free T4 per SD
increase in the PGS. The ancestry groups were as defined by the Pan-UK Biobank initiative – AFR=African ancestry, AMR=admixed American ancestry, CSA=Central/South Asian ancestry, AFR=African ancestry, AMR=admixed American ancestry, CSA=Central/South Asian ancestry, EAS=East Asian ancestry, MID=Middle Eastern ancestry. Error bars indicate 95% confidence intervals.

It is made available under a CC-BY 4.0 International license.

Figure 4: PGS performance across clinical diseases. Prediction performance of the TSH PGS for four
clinical thyroid phenotypes – (a, top left) hypothyroidism, (b, top right) hyperthyroidism, (c, bottom clinical thyroid phenotypes – (a, top left) hypothyroidism, (b, top right) hyperthyroidism, (c, bottom left) thyroid cancer, and (d, bottom right) other thyroid disease. Error bars indicate 95% confidence intervals.

Figure 5: Age-of-onset analysis. Proportion of hypothyroidism (a, left) and hyperthyroidism (b, right) cases diagnosed by age stratified into lowest (green), median (blue) and highest (red) decile for the TSH PGS. Shaded bands indicate 95% confidence intervals.

. [CC-BY 4.0 International license](http://creativecommons.org/licenses/by/4.0/) It is made available under a

It is made available under a CC-BY 4.0 International license