Supplementary Methods 1 Description of the source of outcome GWAS.

EPITHYR

EPITHYR is a collaborative consortium[1] comprising seven population-based case-control studies focused on thyroid cancer research. These studies include three conducted in Metropolitan France (CATHY, YOUNG-Thyr, and E3N), two in the South Pacific Islands (French Polynesia and New Caledonia), one in Cuba, and one in the Gomel region of Belarus, which was contaminated following the Chernobyl disaster. The CATHY study was carried out across three French administrative regions, involving patients aged 25 years and older diagnosed with DTC between 2002 and 2007, with DNA samples collected from 516 cases and 569 controls. The New Caledonia study encompassed the entire country, including DTC patients diagnosed between 1993 and 1999, with available DNA samples from 224 cases and 261 controls. The YOUNG-Thyr study focused on individuals under 35 years old, born after January 1, 1971, and diagnosed with DTC between 2002 and 2006, with DNA samples collected from 715 cases and 692 controls. The Cuban study included patients aged 17-60 years, residing in Havana, who were diagnosed and treated for DTC between 2000 and 2011, consisting of 199 cases and 193 controls. The French Polynesia study involved patients diagnosed between 1983 and 2003, younger than 56 years, and born and living in French Polynesia, including 144 cases and 231 controls. The Chernobyl study examined individuals from the most contaminated areas of Belarus, including 83 cases of papillary thyroid cancer (PTC) and 324 matched controls who were under 15 years old at the time of the Chernobyl accident. The E3N study was a prospective cohort study comprising 98,995 women in France born between 1925 and 1950, with sufficient DNA material available for 568 participants. Genotyping in EPITHYR was conducted using the Infinium OncoArray-500K BeadChip (Illumina), with quality control performed on 4,553 genotyped individuals, excluding those with a call rate below 95%, sex discordance, duplicates, and related individuals (pi-hat threshold of 0.2[2]). We filtered by ethnicity to include only individuals of European ancestry, and ultimately, the study included 1,552 incident DTC cases and 1,954 controls. SNP quality control reduced the dataset to 401,525 SNPs by removing variants with poor call rates (n=8,327), duplicates (n=814), monomorphic SNPs (n=13,832), and those not meeting Hardy-Weinberg equilibrium ($P < 10^{-7}$ in controls, $P < 10^{-12}$ in cases) or minor allele frequency (<1%) thresholds. Imputation used the 1000 Genomes Project dataset, employing SHAPEIT2 and IMPUTEv2 for European populations, resulting in 9,665,802 SNPs. Logistic regression analysis, adjusted for sex, region, and the top 10 principal components, was used to assess DTC associations.

EPIC

The EPIC cohort[3] is a large multicenter prospective study initiated in 1992 across 7 European countries (Denmark, France, Germany, Italy, Spain, Netherland, and United Kingdom) to investigate links between nutrition, lifestyle, metabolism, genetic factors, and cancer risk. It includes approximately 370,000 women and 150,000 men. Thyroid cancer cases were identified among initially cancer-free participants who donated blood and were later diagnosed with TC during follow-up. DTC cases were classified using ICD-10 codes, with papillary cases corresponding to morphologic codes 8050, 8130, 8260, 8340-8344, and 8350, while follicular cases were assigned codes 8290, 8330-8335. Controls were selected from living, cancer-free cohort members at the time of diagnosis and matched based on recruitment center, sex, age, date, time, fasting status at blood collection, and follow-up duration. Genotyping and quality control in the EPIC study followed the same protocols as EPITHYR, leading to the analysis of 9,652,732 SNPs. Logistic regression adjusted for sex, country, and the top 10 principal components was employed to measure associations, with the study including 345 incident DTC cases and 783 controls.

deCODE genetics

deCODE genetics study[4] examined genetic associations with thyroid cancer across European ancestry populations, including individuals from Iceland, the Netherlands, Spain, and the United States. All participants were of European descent. In Iceland, thyroid cancer cases were identified through the Icelandic Cancer Registry (ICD-10: C73), with genotypic data available for 1,003 individuals, predominantly women. The Dutch cohort comprised 85 non-medullary thyroid cancer cases and 4,956 cancer-free individuals, primarily women, with a mean diagnosis age of 40 years. The Columbus, USA cohort included 1,580 cases with a mean diagnosis age of 43 years and 1,628 controls with a mean age of 45 years. The Houston, USA cohort consisted of 250 PTC cases and 363 cancer-free controls, with mean ages at diagnosis of 45 and 53 years, respectively. The Spanish cohort included 83 non-medullary thyroid cancer cases and 1,612 controls, with a mean diagnosis age of 49 years. The Icelandic thyroid cancer GWAS dataset was generated using Illumina technology with an average sequencing depth of 34X, whereas genotyping for the Dutch, Spanish, and US samples was conducted using Omni-1 Quad-bead chips. Variants were excluded based on a yield below 94%, minor allele frequency under 1%, failure in Hardy-Weinberg equilibrium testing ($P < 1.0 \times 10^{-6}$), or significant genotype batch differences ($P < 1.0 \times 10^{-6}$), yielding a final dataset of 9,863,645 SNPs. Logistic regression, adjusted for study center, sex, age, and the top 10 principal components,

assessed genetic associations. Results across study groups were combined using a Mantel-Haenszel model[5]. The study included 3,001 incident DTC cases and 287,550 controls.

UK Biobank

The UK Biobank[6] is a large population-based health research resource comprising approximately 500,000 participants aged 40-69 years, recruited between 2006 and 2010 across the UK. DTC cases were identified using ICD-10 (Data field: 40006, code C730-739) and ICD-9 (Data field: 40013, code 1930-1939) codes, covering histological subtypes such as papillary, follicular, and oxyphilic cell thyroid cancer (Data field: 40011). Only first primary cancer occurrences were considered. Controls were cancer-free (except for non-melanoma skin cancer) throughout follow-up. Genotyping was conducted using the UK Biobank Axiom Array, measuring approximately 850,000 variants, with over 90 million imputed using the Haplotype Reference Consortium and UK10K + 1000 Genomes reference panels. Only European participants were analyzed using PCA. Logistic regression, adjusted for sex, age at recruitment, and the top 10 principal components, assessed genetic associations. The study included 518 DTC cases (262 incident cases) and 358,640 controls.

Italian Study

The Italian study[7] enrolled 701 consecutive DTC cases from the Department of Endocrinology at the University Hospital of Pisa. Controls (n=499) were healthy individuals recruited from blood donors and hospital workers during routine surveillance. Both cases and controls were required to be at least 18 years old, of Caucasian origin, and unrelated. Controls were excluded if they had any history of malignancy, chronic inflammatory disease, or benign thyroid disorders. Genotyping was performed using Illumina BeadChips, with genotype calling via Illumina GenomeStudio 2010. Samples with SNP call rates below 95% were excluded, and markers were filtered out if they had a genotype call rate under 95%, minor allele frequency below 5%, or violated Hardy-Weinberg equilibrium ($P<5.0 \times 10^{-5}$). Additional study details were previously published alongside the GWAS[7]. A total of 632 cases and 430 controls were included.

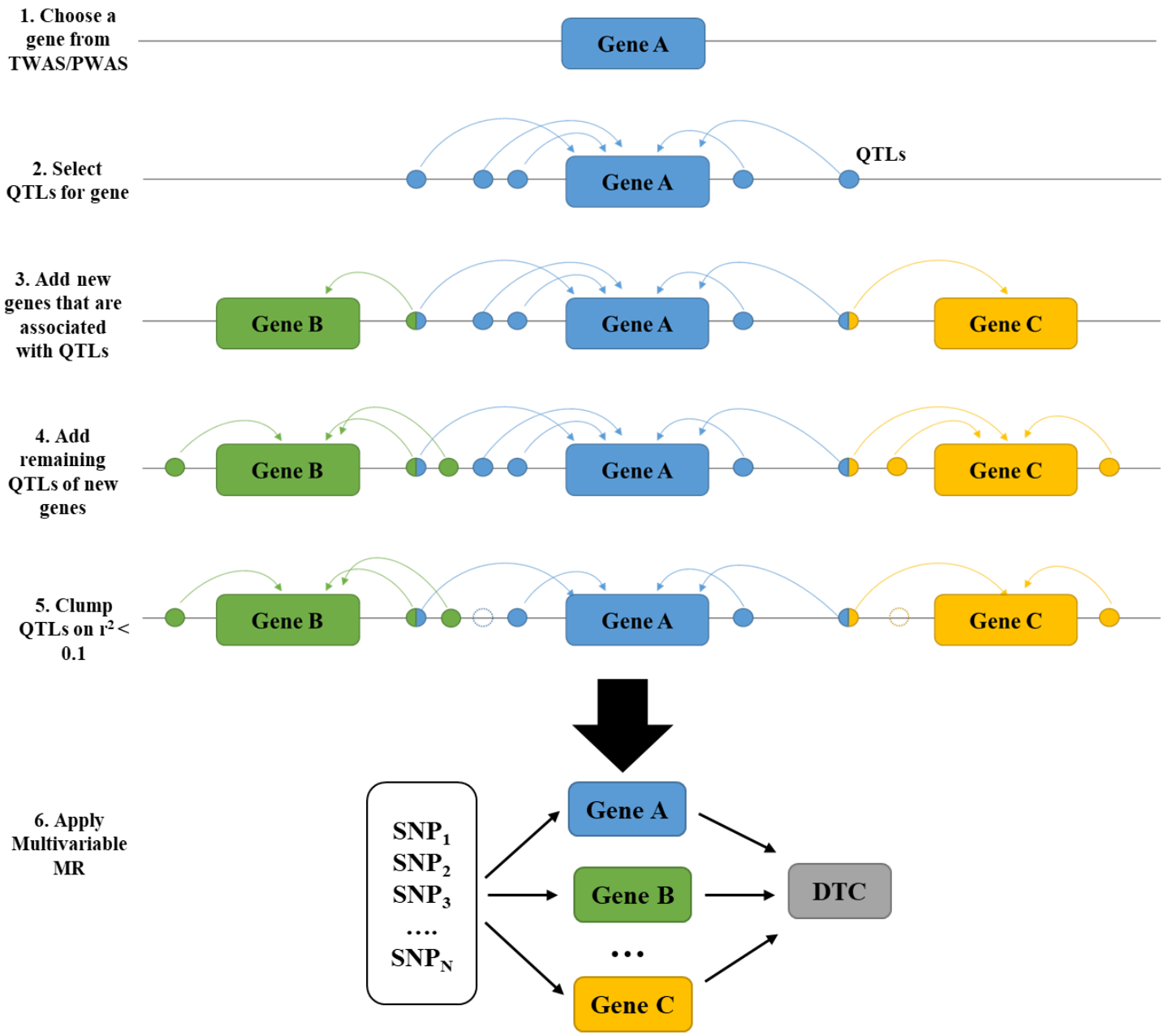
FinnGen

The FinnGen[8] R10 dataset comprises 412,181 individuals of Finnish descent, sourced from Finnish biobanks and digital health registries. Study details are available at (https://www.finngen.fi/en/access_results). Exclusion criteria included ambiguous gender, genotype missingness above 5%, extreme heterozygosity (±4 SD), and non-Finnish ancestry. SNP genotyping was conducted using Illumina and Affymetrix arrays. The dataset included 1,471 papillary and 162 follicular

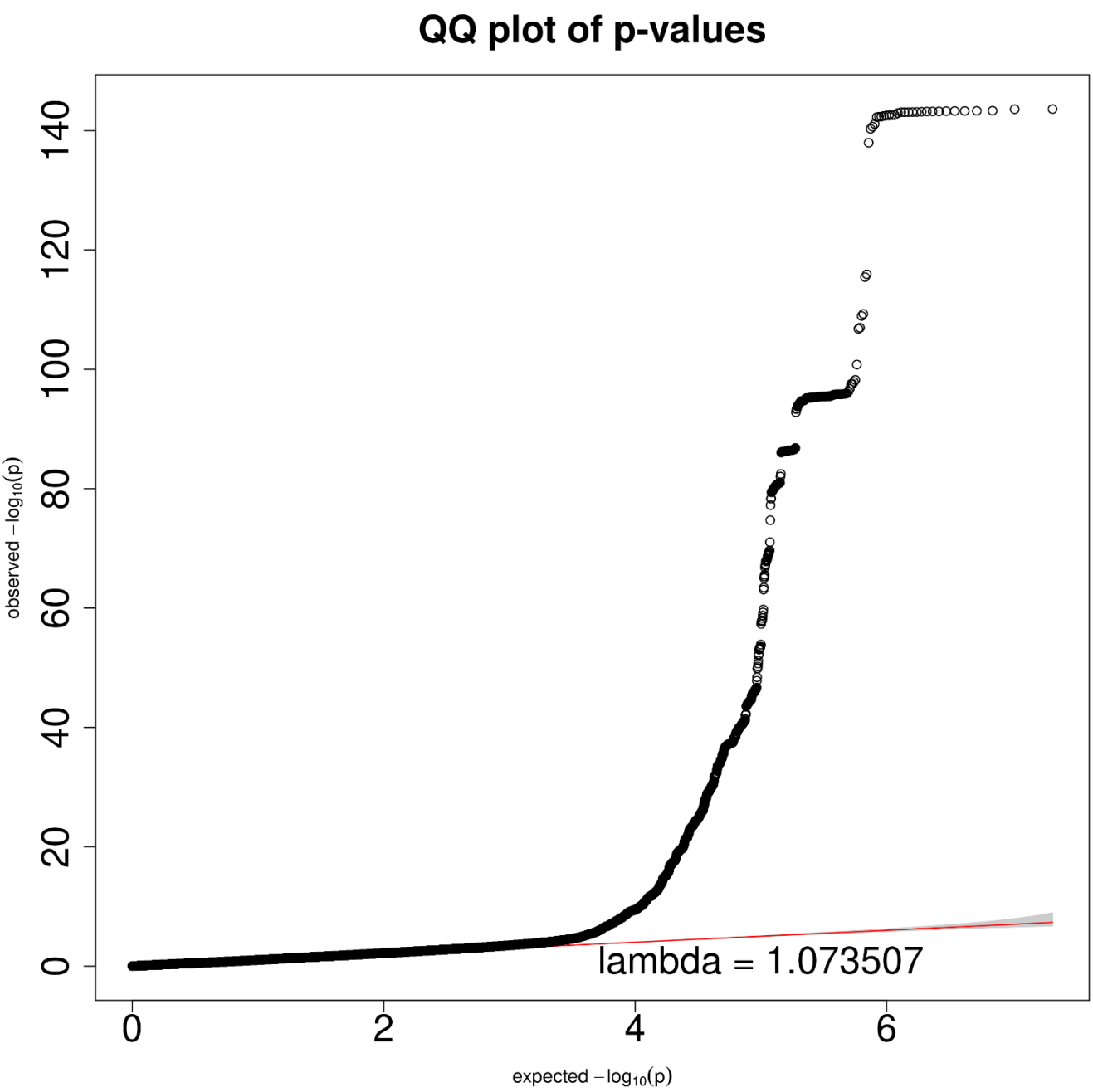
thyroid carcinoma cases, with cancer-free controls identified using ‘EXALLC’ variables that excluded other cancers. GWAS analysis applied GAIGE, a mixed model logistic regression, adjusting for sex, age, genotyping batch, and the top 10 principal components.

Supplementary Methods 2 Development of a polygenic risk score framework incorporating genome-, transcriptome-, and proteome-wide association studies for multi-omics risk prediction.

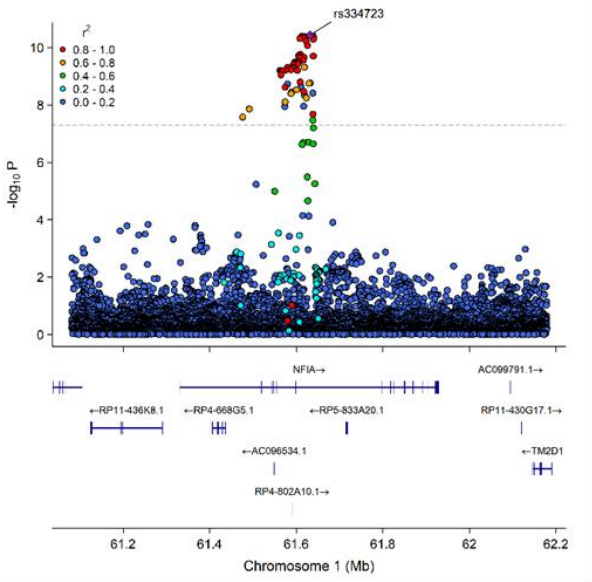
To construct the polygenic risk score, we utlized summary statistics from genome-wide association studies (GWAS), transcriptome-wide association studies (TWAS), and proteome-wide association studies (PWAS). The polygenic risk score (PRS) was constructed by first using significant single nucleotide polymorphisms (SNPs) identified in GWAS. Each SNP was weighted based on its effect size, and the weighted sum of risk alleles was computed for each individual. The formula for PRS can be expressed as $PRS_j = \sum_{j=1}^{m} \gamma_j x_{ij}$, where γ_j is the effect size of SNP *j*, x_{ij} is the risk allele count for SNP *j* in individual *i, and m* represents the total number of SNPs included in the PRS model. For the polygenic TWAS-based risk score (PTRS), we applied the PrediXcan framework to estimate genetically regulated gene expression (GReX) for each individual. The PTRS was calculated by summing the predicted gene expression levels, weighted by the effect sizes of each gene, derived from the S-PrediXcan[9] framework. Predicted gene expression levels were calculated as $GeneExp_j = \sum_{j=1}^{m} X_j \beta_j$, where $GeneExp_j$ is the predicted expression level of gene *j*, X_j is the count of effect alleles for variant *j*, and β_j is the variant's effect on gene expression derived from a regression model across *m* variants. And then PTRS were calculated as $PTRS = \sum_{j=1}^{g} GeneExp_j \times Z_j$, where Z_j represents the TWAS effect size of gene *j*, $GeneExp_j$ is the predicted expression for gene *j*, and g is the total number of genes included[10]. To integrate proteomic information, we incorporated PWAS results by considering genetically predicted protein abundances for genes identified in brain and plasma blood tissues. The polygenic PWAS-based risk score (PPRS) was constructed similarly to the PTRS approach, by weighting the predicted protein levels by the effect sizes from proteomic data. We applied the FUSION[11] framework to estimate genetically regulated protein abundance expression for each individual. The formula for PRPS is $PPRS = \sum_{j=1}^{g} ProteinExp_j \times Z_j$, where $ProteinExp_j$ represents the predicted protein abundance for gene *j*, and Z_j is its corresponding PWAS effect size across g genes. For all models, gene selection was refined using a p-value threshold corrected for multiple testing via Bonferroni correction. Genes overlapping with GWAS loci were excluded to avoid redundancy, ensuring that the final risk scores reflected distinct contributions from each omics layer. Finally, we evaluated the predictive performance of the PRS by using receiver operating characteristic (ROC) curves and area under the curve (AUC).

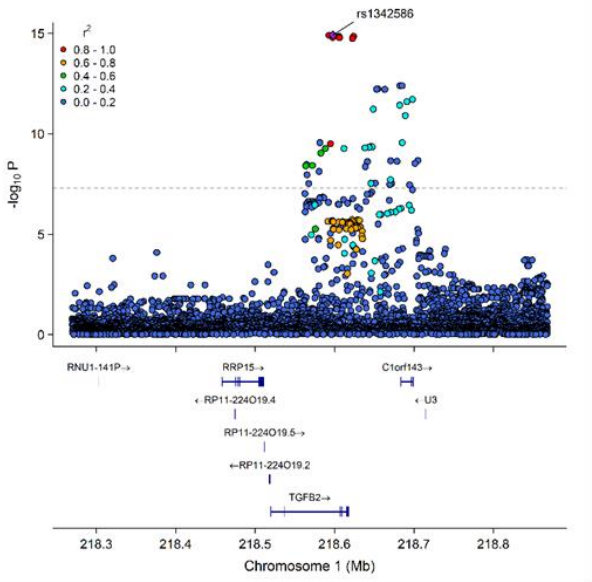


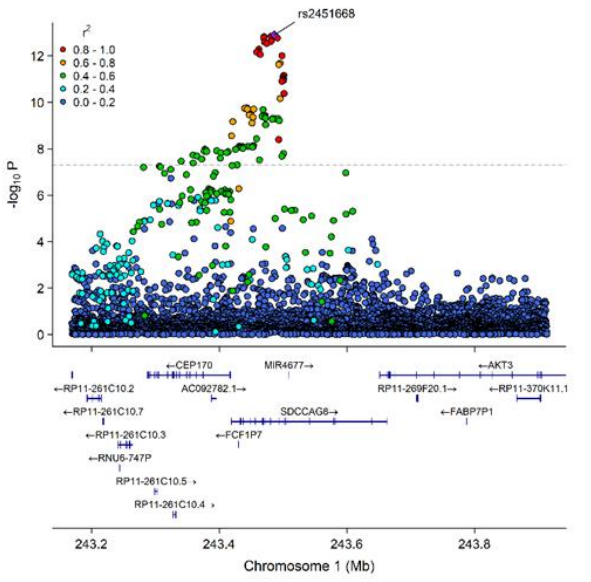
Supplementary Figure 1 A Mendelian Randomization framework utilizing genes as exposures.

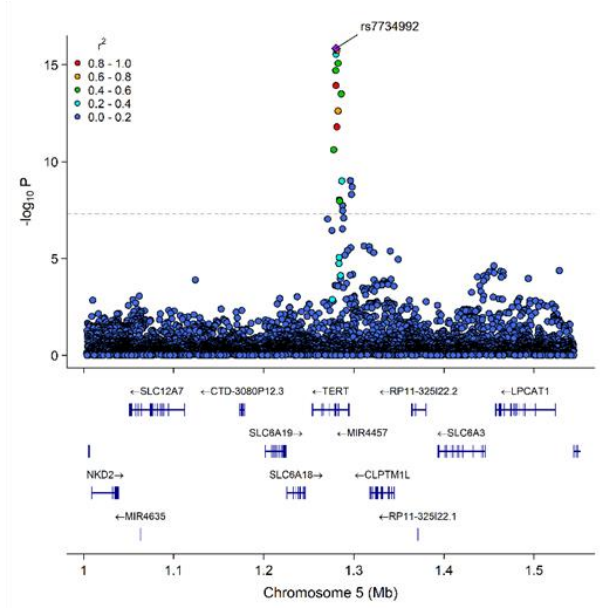


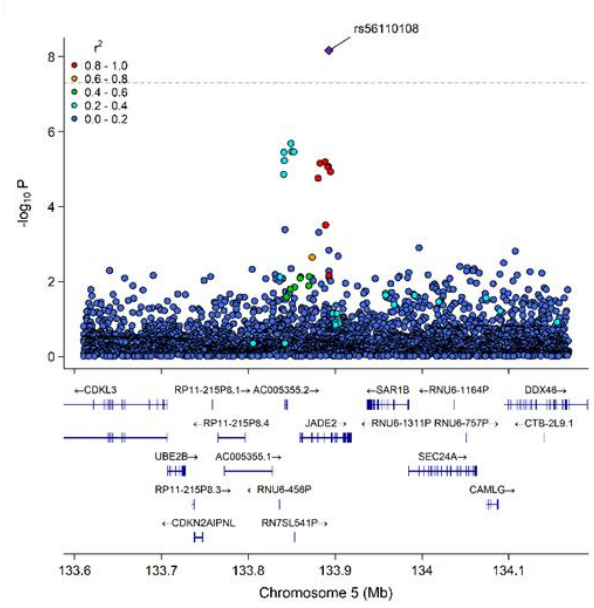
Supplementary Figure 2 Quantile-Quantile plots of DTC GWAS including 7,681 cases and 963,550 controls of European descent. The red diagonal line represents expected distribution under the assumption that there is no inflation of the chi-square statistics.

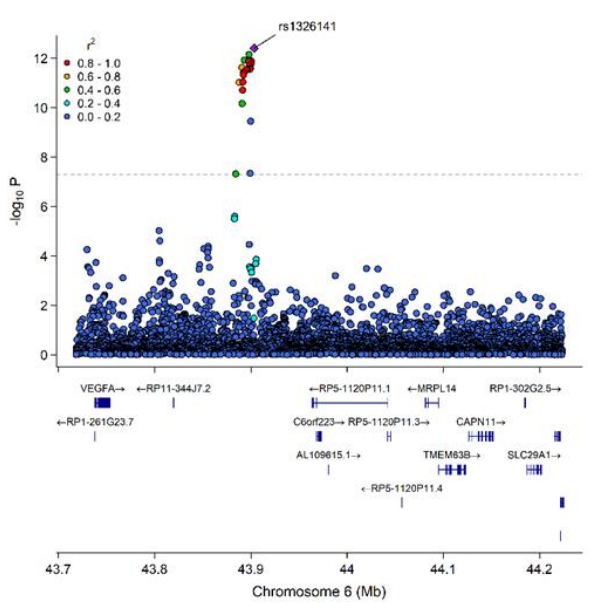

rs334723
r²
0.8 - 1.0
0.6 - 0.8
0.4 - 0.6
0.2 - 0.4
0.0 - 0.2
-log10 P
NFIA→
AC099791.1→
←RP11-436K8.1
←RP4-668G5.1
←RP5-833A20.1
RP11-430G17.1→
←AC096534.1
←TM2D1
RP4-802A10.1→
61.2
61.4
61.6
61.8
62
62.2
Chromosome 1 (Mb)

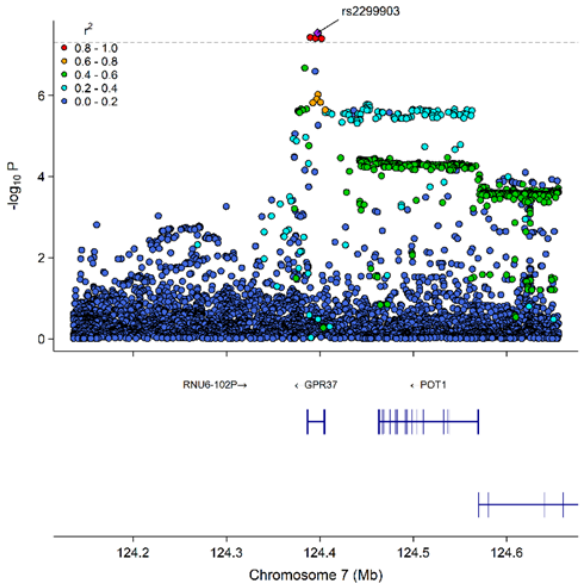

rs1342586
r²
0.8 - 1.0
0.6 - 0.8
0.4 - 0.6
0.2 - 0.4
0.0 - 0.2
-log10 P
RNU1-141P→
RRP15→
C1orf143→
← RP11-224O19.4
← U3
RP11-224O19.5→
←RP11-224O19.2
TGFB2→
218.3
218.4
218.5
218.6
218.7
218.8
Chromosome 1 (Mb)

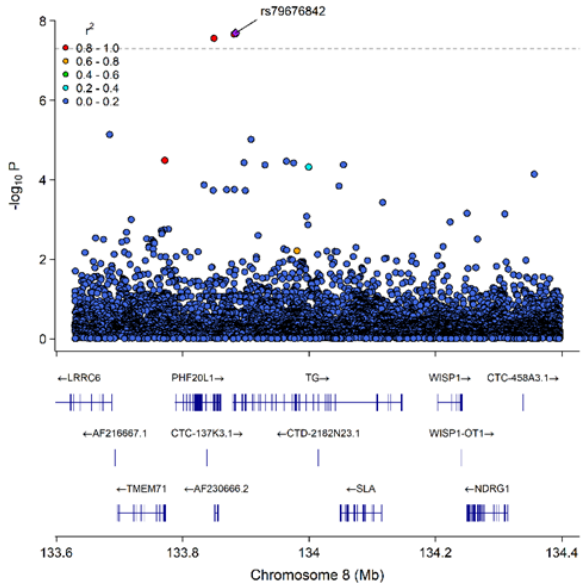

rs2451668
r²
0.8 - 1.0
0.6 - 0.8
0.4 - 0.6
0.2 - 0.4
0.0 - 0.2
-log10 P
←CEP170
MIR4677→
←AKT3
←RP11-261C10.2
AC092782.1→
RP11-269F20.1→
←RP11-370K11.1
←RP11-261C10.7
SDCCAG8→
←FABP7P1
←RP11-261C10.3
←FCF1P7
←RNU6-747P
RP11-261C10.5 ›
RP11-261C10.4 ›
243.2
243.4
243.6
243.8
Chromosome 1 (Mb)

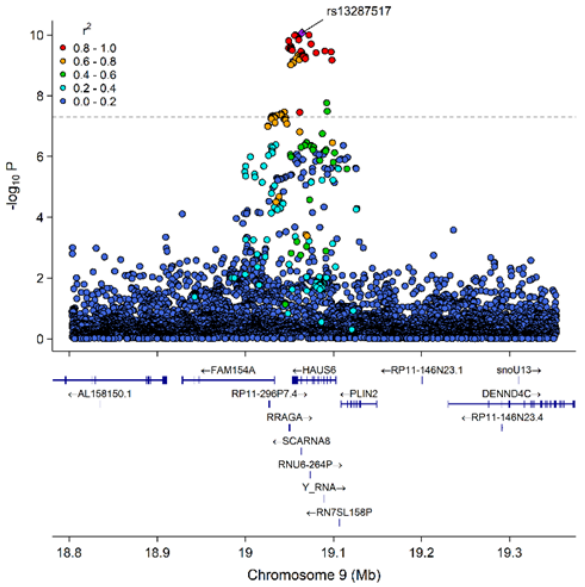

rs7734992
r²
0.8 - 1.0
0.6 - 0.8
0.4 - 0.6
0.2 - 0.4
0.0 - 0.2
-log10 P
← SLC12A7
← CTD-3080P12.3
← TERT
← RP11-325I22.2
← LPCAT1
SLC8A19→
← MIR4457
← SLC6A3
NKD2→
SLC6A18→
← CLPTM1L
← MIR4635
← RP11-325I22.1
1
1.1
1.2
1.3
1.4
1.5
Chromosome 5 (Mb)

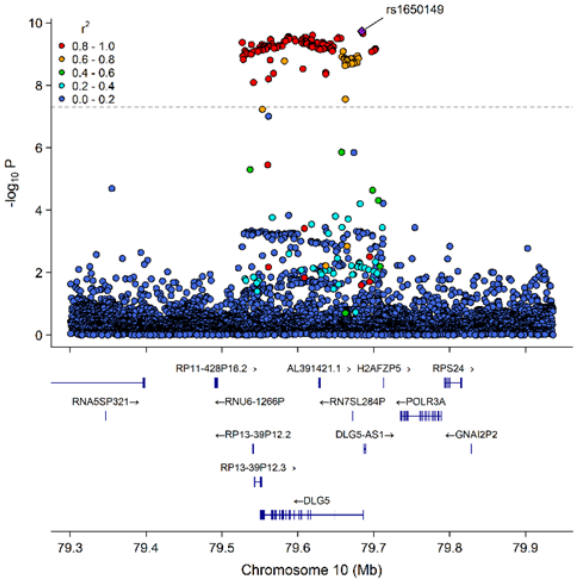

rs56110108
r²
0.8 - 1.0
0.6 - 0.8
0.4 - 0.6
0.2 - 0.4
0.0 - 0.2
-log10 P
←CDKL3
RP11-215P8.1→ AC005355.2→
←SAR1B
←RNU6-1164P
DDX46→
← RP11-215P8.4
JADE2 ›
‹ RNU6-1311P RNU6-757P ›
← CTB-2L9.1
UBE2B→
AC005355.1→
SEC24A→
RP11-215P8.3→
‹ RNU6-456P
CAMLG→
←CDKN2AIPNL
RN7SL541P→
133.6
133.7
133.8
133.9
134
134.1
Chromosome 5 (Mb)

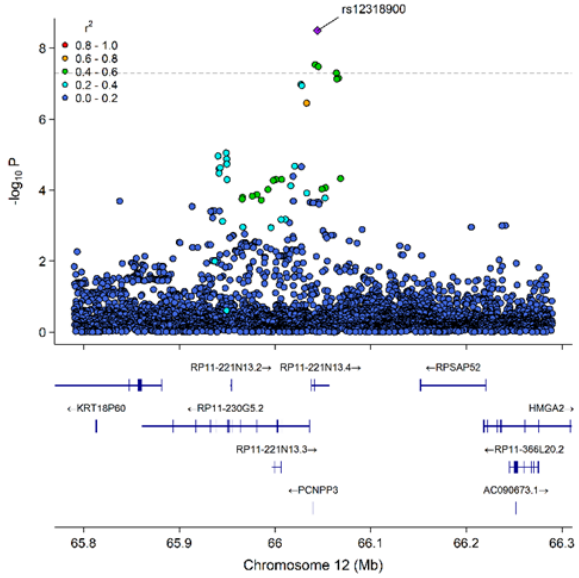

rs1326141
r²
0.8 - 1.0
0.6 - 0.8
0.4 - 0.6
0.2 - 0.4
0.0 - 0.2
-log10 P
VEGFA→
←RP11-344J7.2
←RP5-1120P11.1
←MRPL14
RP1-302G2.5→
←RP1-261G23.7
C6orf223→ RP5-1120P11.3→
CAPN11→
AL109615.1→
TMEM63B→
SLC29A1→
← RP5-1120P11.4
43.7
43.8
43.9
44
44.1
44.2
Chromosome 6 (Mb)

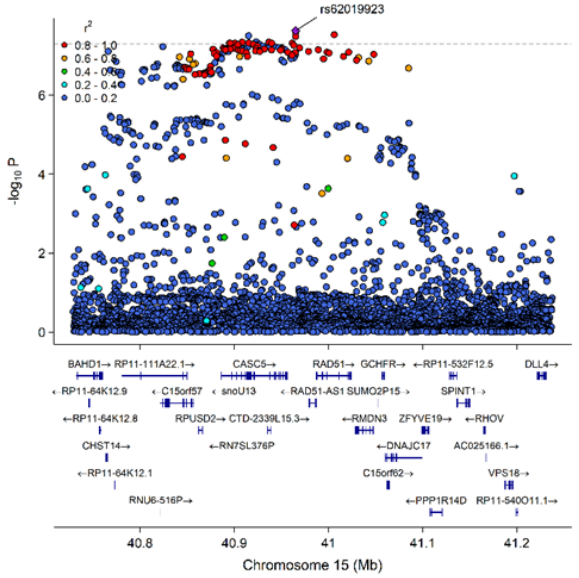

rs2299903
r²
0.8 - 1.0
0.6 - 0.8
0.4 - 0.6
0.2 - 0.4
0.0 - 0.2
-log10 P
RNU6-102P→
‹ GPR37
‹ POT1
124.2
124.3
124.4
124.5
124.6
Chromosome 7 (Mb)

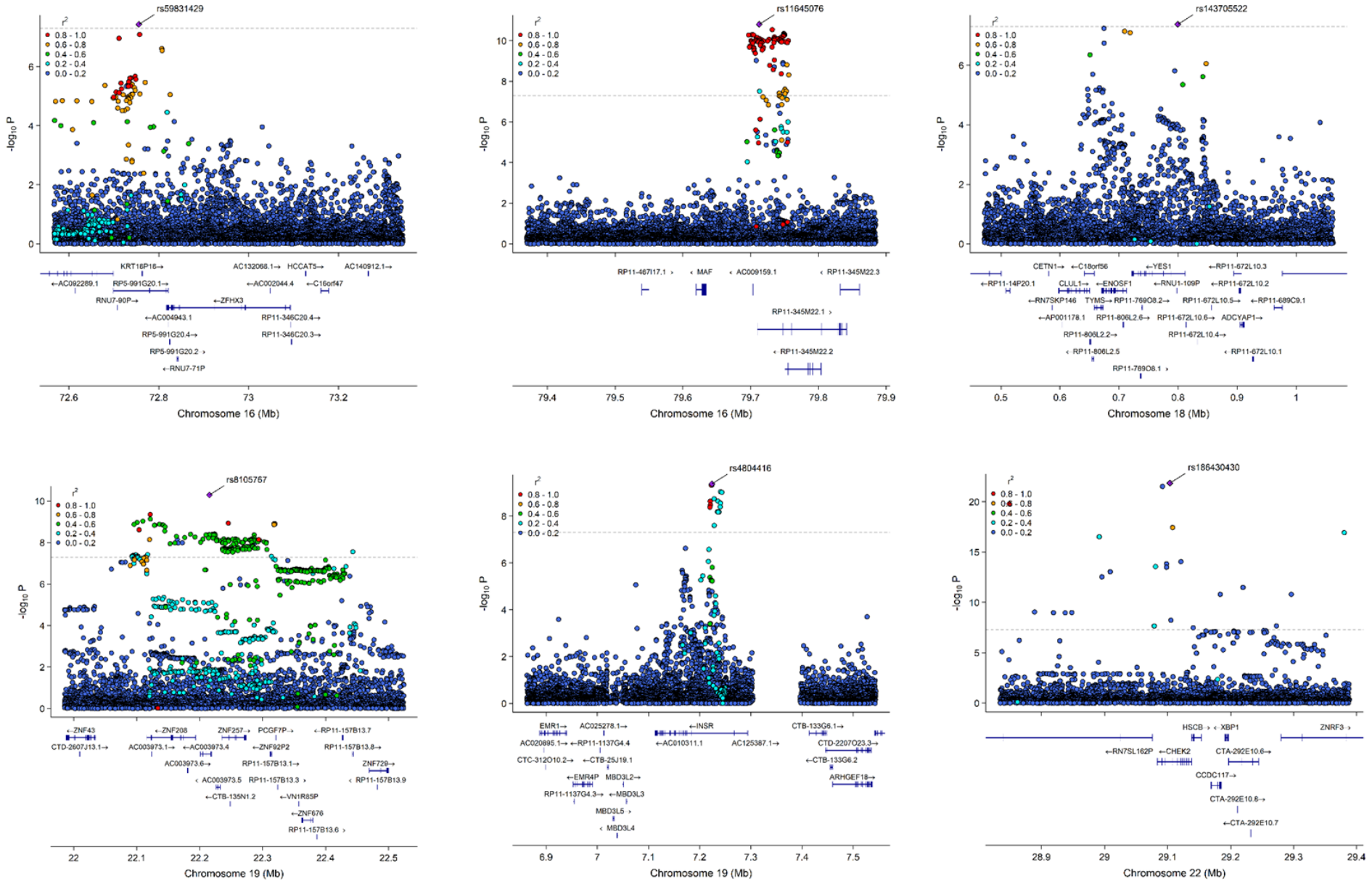

rs79676842
r²
0.8 - 1.0
0.6 - 0.8
0.4 - 0.6
0.2 - 0.4
0.0 - 0.2
-log10 P
←LRRC6
PHF20L1→
TG→
WISP1→
CTC-458A3.1→
←AF216667.1
CTC-137K3.1→
←CTD-2182N23.1
WISP1-OT1→
←TMEM71
←AF230666.2
←SLA
←NDRG1
133.6
133.8
134
134.2
134.4
Chromosome 8 (Mb)


rs13287517
r²
0.8 - 1.0
0.6 - 0.8
0.4 - 0.6
0.2 - 0.4
0.0 - 0.2
-log10 P
←FAM154A
←HAUS6
←RP11-146N23.1
snoU13→
‹ AL158150.1
RP11-296P7.4 ›
‹ PLIN2
DENND4C ›
RRAGA ›
‹ RP11-146N23.4
‹ SCARNA8
RNU6-264P ›
Y_RNA→
←RN7SL158P
18.8
18.9
19
19.1
19.2
19.3
Chromosome 9 (Mb)

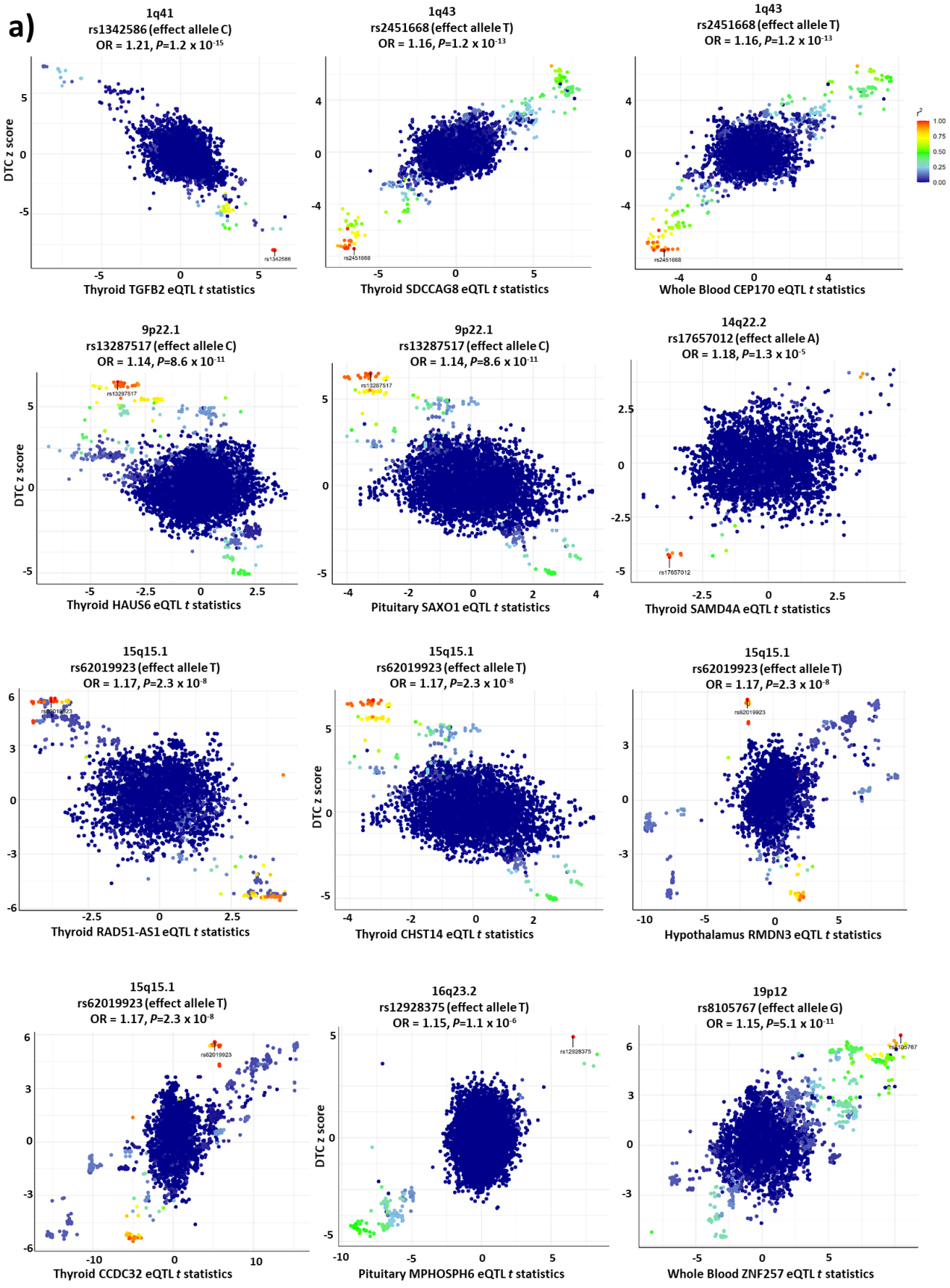

rs1650149
r²
0.8 - 1.0
0.6 - 0.8
0.4 - 0.6
0.2 - 0.4
0.0 - 0.2
-log10 P
RP11-428P16.2 ›
AL391421.1 ›
H2AFZP5 ›
RPS24 ›
RNA5SP321→
←RNU6-1266P
←RN7SL284P
←POLR3A
←RP13-39P12.2
DLG5-AS1→
←GNAI2P2
RP13-39P12.3 ›
←DLG5
79.3
79.4
79.5
79.6
79.7
79.8
79.9
Chromosome 10 (Mb)

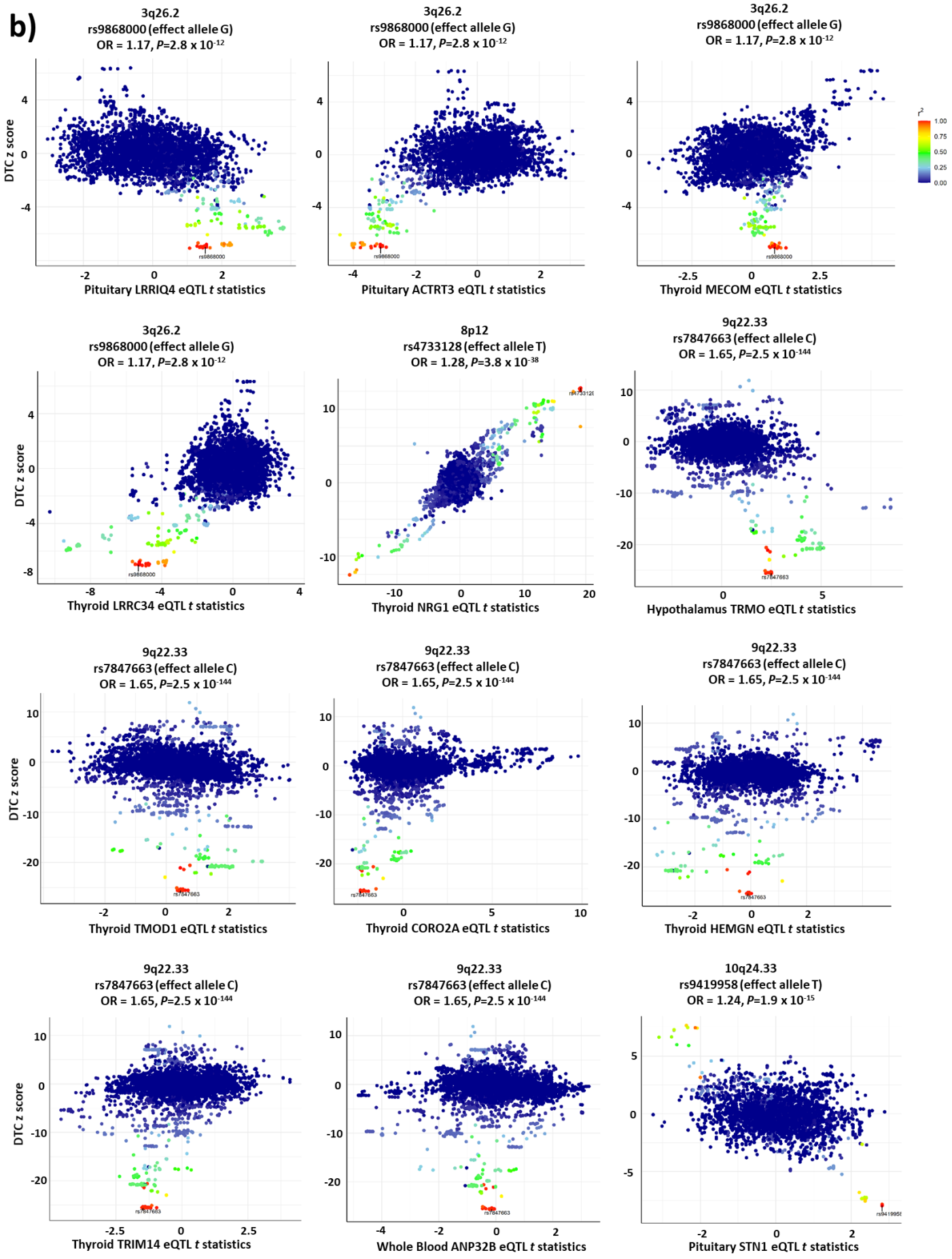

rs12318900
r²
0.8 - 1.0
0.6 - 0.8
0.4 - 0.6
0.2 - 0.4
0.0 - 0.2
-log10 P
RP11-221N13.2→
RP11-221N13.4→
←RPSAP52
‹ KRT18P60
‹ RP11-230G5.2
HMGA2 ›
RP11-221N13.3→
←RP11-366L20.2
←PCNPP3
AC090673.1→
65.8
65.9
66
66.1
66.2
66.3
Chromosome 12 (Mb)


rs62019923
r²
0.8 - 1.0
0.6 - 0.8
0.4 - 0.6
0.2 - 0.4
0.0 - 0.2
-log10 P
BAHD1→
RP11-111A22.1→
CASC5→
RAD51→
GCHFR→
←RP11-532F12.5
DLL4→
‹ RP11-64K12.9
‹ C15orf57
‹ snoU13
‹ RAD51-AS1
SUMO2P15 ›
SPINT1 ›
←RP11-64K12.8
RPUSD2→
CTD-2339L15.3→
←RMDN3
ZFYVE19→
←RHOV
CHST14→
←RN7SL378P
←DNAJC17
AC025166.1→
←RP11-64K12.1
C15orf62→
VPS18→
RNU6-516P→
←PPP1R14D
RP11-540O11.1→
40.8
40.9
41
41.1
41.2
Chromosome 15 (Mb)

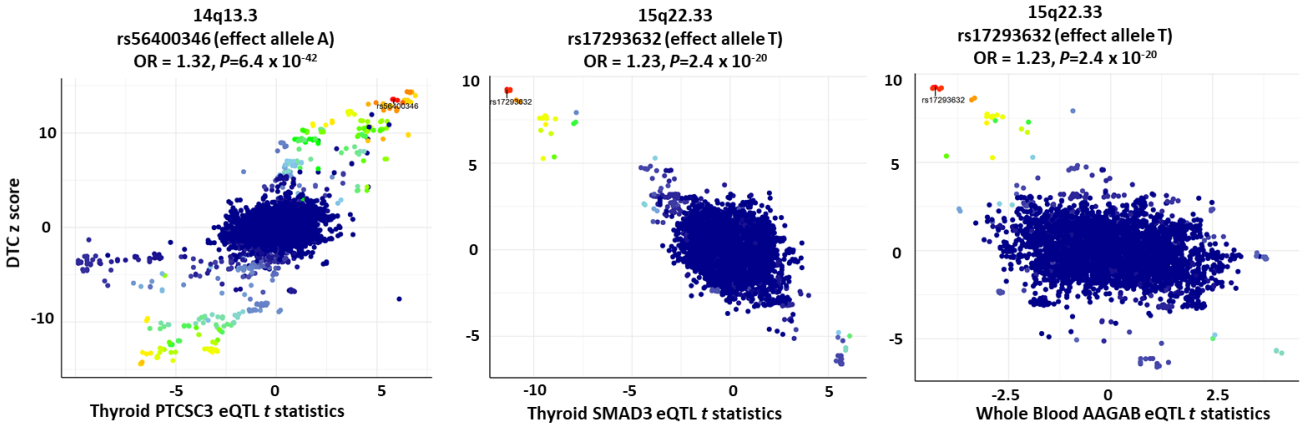


Supplementary Figure 3 Regional plots of the DTC GWAS are displayed for all significant novel loci at 1p31.3, 1q41, 1q43, 5p15.33, 5q31.1, 6p21.1, 7q31.33, 8q24.22, 9p22.1, 10q22.3, 12q14.3, 15q15.q, 16q22.2, 16q23.2, 18p11.32, 19p13.2, 19p12 and 22q12.1. The x-axis represents the position of SNPs and gene on the chromosome, while the y-axis represents the –log10 (P-value). The color of each SNP dots indicates its correlation with LD.

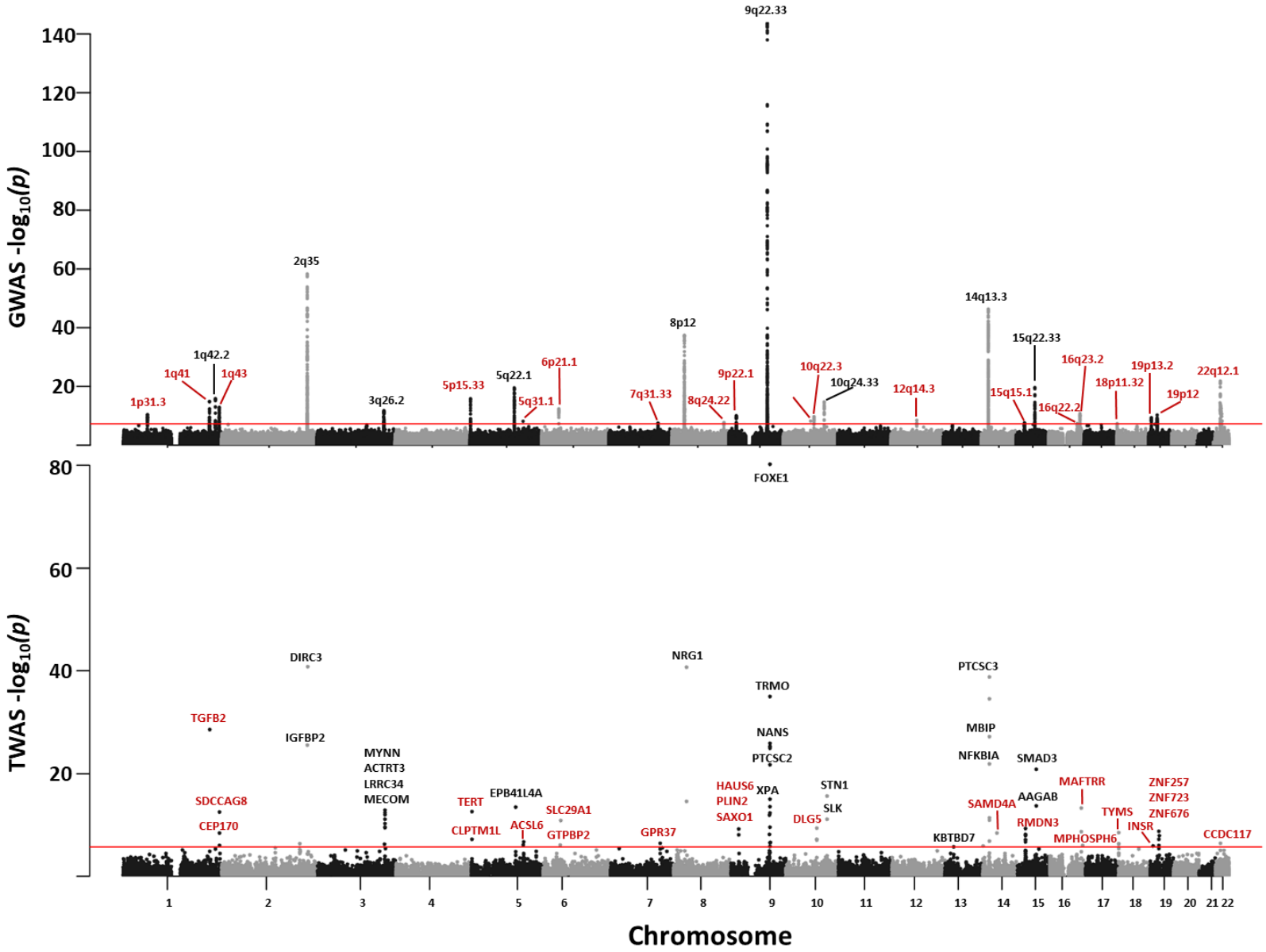

a)
1q41
rs1342586 (effect allele C)
OR = 1.21, P=1.2 x 10-15
DTC z score
Thyroid TGFB2 eQTL t statistics
1q43
rs2451668 (effect allele T)
OR = 1.16, P=1.2 x 10-13
Thyroid SDCCAG8 eQTL t statistics
1q43
rs2451668 (effect allele T)
OR = 1.16, P=1.2 x 10-13
r2
1.00
0.75
0.50
0.25
0.00
Whole Blood CEP170 eQTL t statistics
9p22.1
rs13287517 (effect allele C)
OR = 1.14, P=8.6 x 10-11
DTC z score
Thyroid HAUS6 eQTL t statistics
9p22.1
rs13287517 (effect allele C)
OR = 1.14, P=8.6 x 10-11
Pituitary SAXO1 eQTL t statistics
14q22.2
rs17657012 (effect allele A)
OR = 1.18, P=1.3 x 10-5
Thyroid SAMD4A eQTL t statistics
15q15.1
rs62019923 (effect allele T)
OR = 1.17, P=2.3 x 10-8
Thyroid RAD51-AS1 eQTL t statistics
15q15.1
rs62019923 (effect allele T)
OR = 1.17, P=2.3 x 10-8
DTC z score
Thyroid CHST14 eQTL t statistics
15q15.1
rs62019923 (effect allele T)
OR = 1.17, P=2.3 x 10-8
Hypothalamus RMDN3 eQTL t statistics
15q15.1
rs62019923 (effect allele T)
OR = 1.17, P=2.3 x 10-8
Thyroid CCDC32 eQTL t statistics
16q23.2
rs12928375 (effect allele T)
OR = 1.15, P=1.1 x 10-6
DTC z score
Pituitary MPHOSPH6 eQTL t statistics
19p12
rs8105767 (effect allele G)
OR = 1.15, P=5.1 x 10-11
Whole Blood ZNF257 eQTL t statistics


b)
3q26.2
rs9868000 (effect allele G)
OR = 1.17, P=2.8 x 10-12
DTC z score
Pituitary LRRIQ4 eQTL t statistics
3q26.2
rs9868000 (effect allele G)
OR = 1.17, P=2.8 x 10-12
Pituitary ACTRT3 eQTL t statistics
3q26.2
rs9868000 (effect allele G)
OR = 1.17, P=2.8 x 10-12
Thyroid MECOM eQTL t statistics
r2
1.00
0.75
0.50
0.25
0.00
3q26.2
rs9868000 (effect allele G)
OR = 1.17, P=2.8 x 10-12
DTC z score
Thyroid LRRC34 eQTL t statistics
8p12
rs4733128 (effect allele T)
OR = 1.28, P=3.8 x 10-38
Thyroid NRG1 eQTL t statistics
9q22.33
rs7847663 (effect allele C)
OR = 1.65, P=2.5 x 10-144
Hypothalamus TRMO eQTL t statistics
9q22.33
rs7847663 (effect allele C)
OR = 1.65, P=2.5 x 10-144
DTC z score
Thyroid TMOD1 eQTL t statistics
9q22.33
rs7847663 (effect allele C)
OR = 1.65, P=2.5 x 10-144
Thyroid CORO2A eQTL t statistics
9q22.33
rs7847663 (effect allele C)
OR = 1.65, P=2.5 x 10-144
Thyroid HEMGN eQTL t statistics
9q22.33
rs7847663 (effect allele C)
OR = 1.65, P=2.5 x 10-144
DTC z score
Thyroid TRIM14 eQTL t statistics
9q22.33
rs7847663 (effect allele C)
OR = 1.65, P=2.5 x 10-144
Whole Blood ANP32B eQTL t statistics
10q24.33
rs9419958 (effect allele T)
OR = 1.24, P=1.9 x 10-15
Pituitary STN1 eQTL t statistics

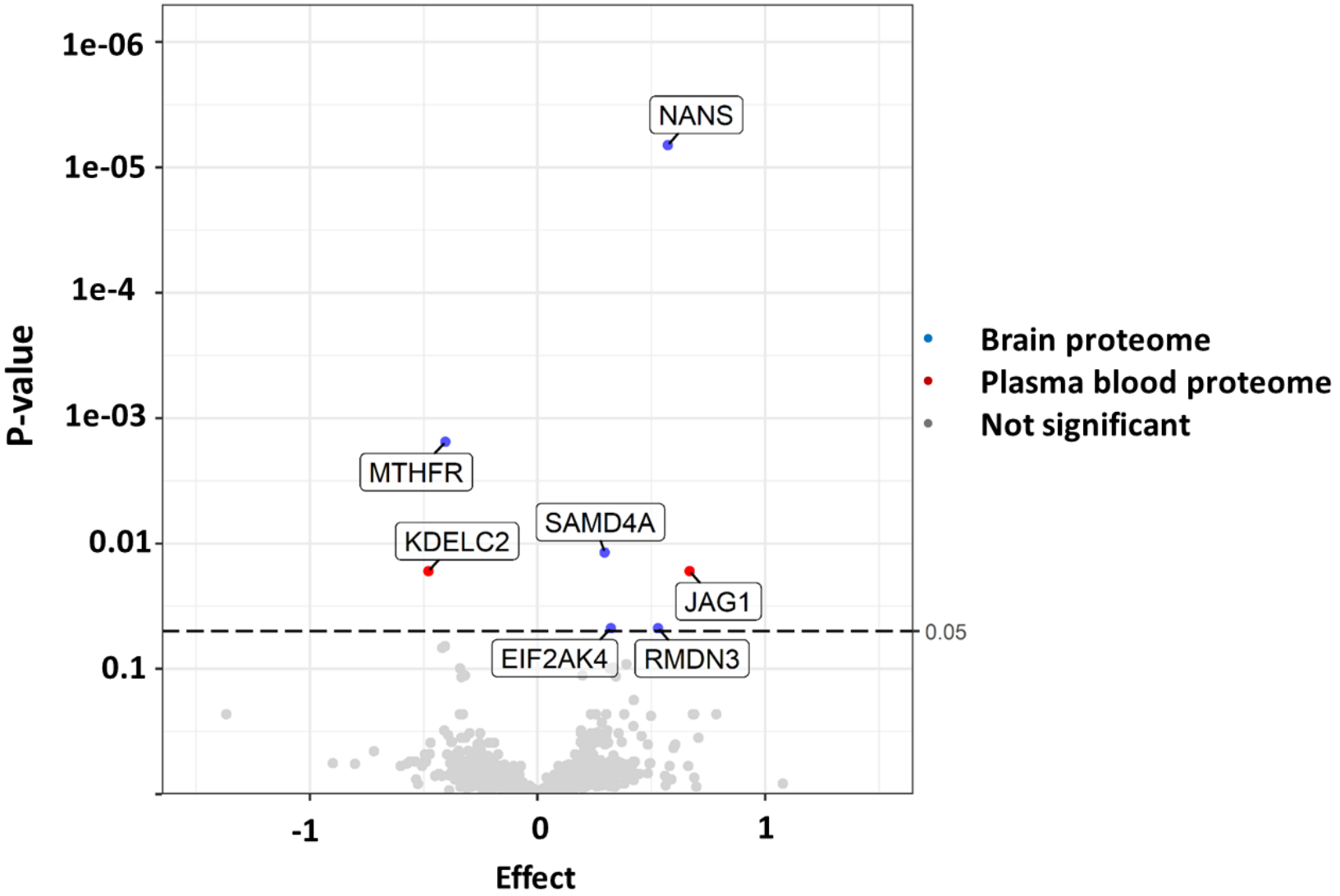
Cont.



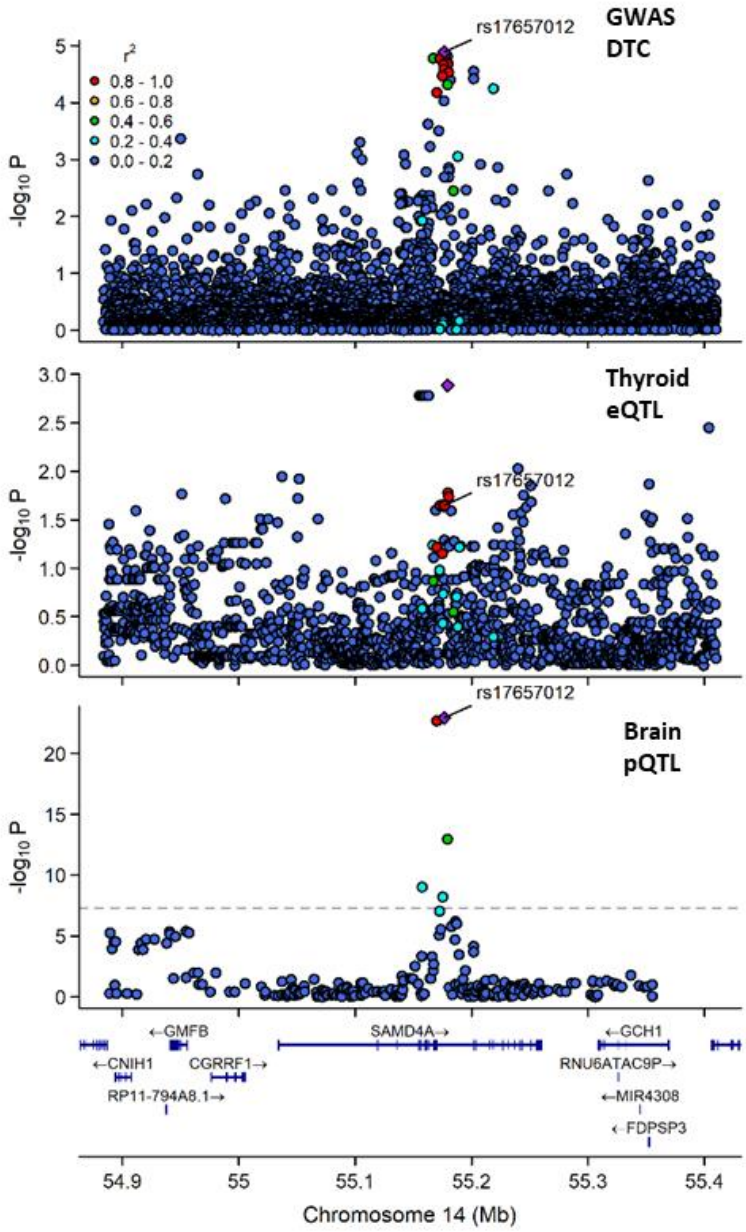
Supplementary Figure 4 Scatterplots of the **a) novel** and **b) known** susceptibility loci and identified by TWAS using thyroid-relevant tissues, showing their associations with DTC and corresponding cis-eQTLs in the tissue. The y-axis represents the association of variants with DTC, while the x-axis shows their association with gene expression levels. Each variant is color-coded according to its LD correlation.



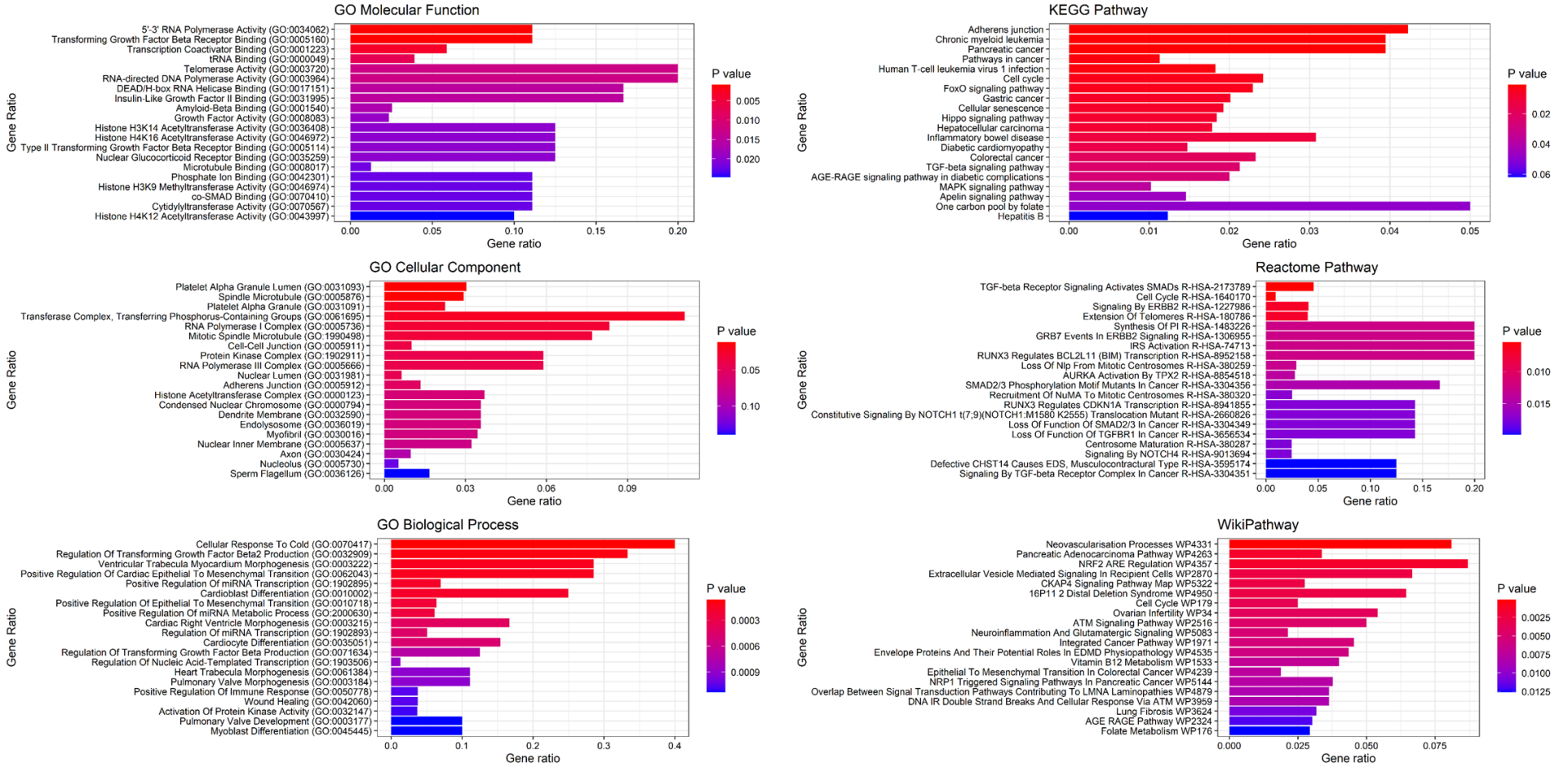
Supplementary Figure 5 The Manhattan plot displays the results of the DTC GWAS and TWAS analyses. A fixed-effect meta-analysis was performed on 20.8 million variants, including 7,681 cases and 963,550 controls of European descent. Previously identified risk loci are marked in black, while novel risk loci are highlighted in red, with the associated genes indicated in the TWAS Manhattan plot. Each dot in TWAS Manhattan plot corresponds to an association test between genetically predicted gene expression in all 49 tissues available from GTEx and differentiated thyroid cancer. The red line indicates the significant boundary for GWAS and TWAS after Bonferroni correction (5.0×10^{-8} and 2.45×10^{-6} respectively).



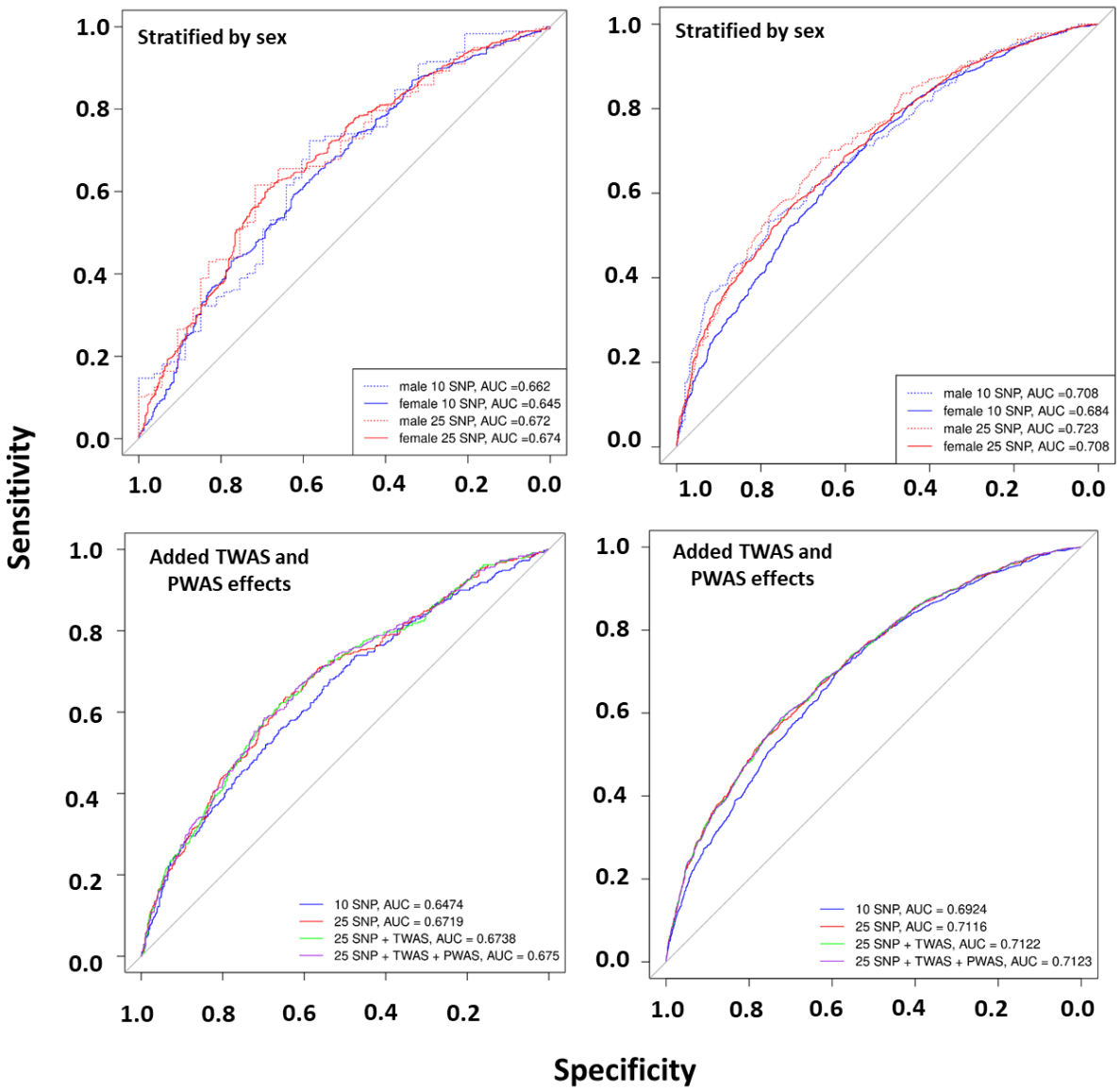
Supplementary Figure 6 Volcano plot showing results from proteome-wide association study (PWAS)



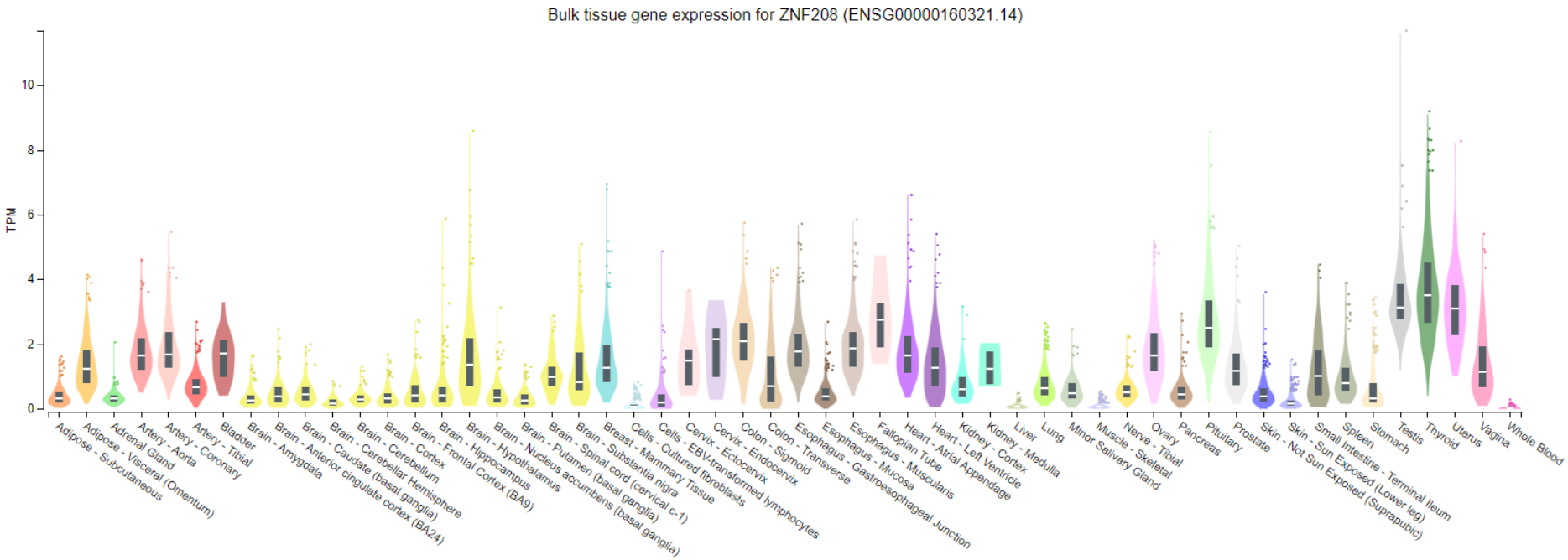
Supplementary Figure 7 A regional locus plots are shown from GWAS of differentiated thyroid cancer, gene expression (eQTL) and brain proteome levels (pQTL). *SAMD4A* genes were both identified in TWAS using thyroid-relevant tissue (thyroid tissue and brain hypothalamus respectively) as well as in brain proteome PWAS. *SAMD4A* genes displayed significant colocalization with high posterior probability (92% for eQTL and 99.1% for pQTL). SNPs are colored based on their LD with the GWAS lead variant (rs17657012).



Supplementary Figure 8 Gene ontology (GO) and Pathway Enrichment analysis from genes identified by both TWAS and PWAS



Supplementary Figure 9 A receiver operating characteristic (ROC) curve evaluating the discriminative ability of PRS models based on 25 GWAS-identified significant SNPs compared to a previously reported model using 10 SNPs in the EPIC and EPITHYR study. Both 25 and 10 SNPs models were stratified by sex. The PRS models were further enhanced by incorporating TWAS and PWAS results to assess their effect on predictive performance.



Supplementary Figure 10 The expression levels of *ZNF208* at 19q12 across various tissues from the GTEx database. Expression values are presented in TPM (Transcripts Per Million), calculated from a gene model with isoforms collapsed into a single gene. Recently GTEx has adopted TPM instead of RPKM as the unit for comparing RNA-seq samples. TPM can be converted from RPKM using the formula: TPM = RPKM / sum(RPKM) * $1.0x10^{-6}$ for each sample/column.

Reference

1. Truong, T. *et al.* Multiethnic genome-wide association study of differentiated thyroid cancer in the EPITHYR consortium. *International Journal of Cancer* **148**, 2935–2946 (2021).
2. Anderson, C. A. *et al.* Data quality control in genetic case-control association studies. *Nat Protoc* **5**, 1564–1573 (2010).
3. Riboli, E. *et al.* European Prospective Investigation into Cancer and Nutrition (EPIC): study populations and data collection. *Public Health Nutr* **5**, 1113–1124 (2002).
4. Gudmundsson, J. *et al.* A genome-wide association study yields five novel thyroid cancer risk loci. *Nat Commun* **8**, 14517 (2017).
5. Mantel, N. & Haenszel, W. Statistical aspects of the analysis of data from retrospective studies of disease. *J Natl Cancer Inst* **22**, 719–748 (1959).
6. Sudlow, C. *et al.* UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. *PLoS Med* **12**, e1001779 (2015).
7. Köhler, A. *et al.* Genome-wide association study on differentiated thyroid cancer. *J Clin Endocrinol Metab* **98**, E1674-1681 (2013).
8. Kurki, M. I. *et al.* FinnGen provides genetic insights from a well-phenotyped isolated population. *Nature* **613**, 508–518 (2023).
9. Barbeira, A. N. *et al.* Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. *Nat Commun* **9**, 1825 (2018).
10. Pain, O. *et al.* Imputed gene expression risk scores: a functionally informed component of polygenic risk. *Hum Mol Genet* **30**, 727–738 (2021).
11. Gusev, A. *et al.* Integrative approaches for large-scale transcriptome-wide association studies. *Nat Genet* **48**, 245–252 (2016).