nature communications

Article

https://doi.org/10.1038/s41467-025-58506-x

Gene clusters linked to insulin resistance identified in a genome-wide study of the Taiwan Biobank population

Received: 27 September 2024

Accepted: 25 March 2025

Published online: 14 April 2025

Check for updates

Eugene Lin[1,2], Yu-Ting Yan[3], Mu-Hong Chen[4,5], Albert C. Yang[6,7], Po-Hsiu Kuo[3,8] ✉ & Shih-Jen Tsai[4,5,6] ✉

This pioneering genome-wide association study examined surrogate markers for insulin resistance (IR) in 147,880 Taiwanese individuals using data from the Taiwan Biobank. The study focused on two IR surrogate markers: the triglyceride to high-density lipoprotein cholesterol (TG:HDL-C) ratio and the TyG index (the product of fasting plasma glucose and triglycerides). We identified genome-wide significance loci within four gene clusters: *GCKR*, *MLXIPL*, *APOA5*, and *APOC1*, uncovering 197 genes associated with IR. Transcriptome-wide association analysis revealed significant associations between these clusters and TyG, primarily in adipose tissue. Gene ontology analysis highlighted pathways related to Alzheimer's disease, glucose homeostasis, insulin resistance, and lipoprotein dynamics. The study identified sex-specific genes associated with TyG. Polygenic risk score analysis linked both IR markers to gout and hyperlipidemia. Our findings elucidate the complex relationships between IR surrogate markers, genetic predisposition, and disease phenotypes in the Taiwanese population, contributing valuable insights to the field of metabolic research.

Insulin resistance (IR) is a well-documented pathophysiological condition characterized by the diminished responsiveness of cells to insulin, a hormone essential for the regulation of glucose metabolism[1]. In individuals with IR, cellular sensitivity to insulin is markedly reduced, resulting in elevated blood glucose levels and compensatory hyperinsulinemia as the pancreas attempts to overcome this resistance[2]. This metabolic dysregulation is closely associated with an increased risk of several metabolic disorders, including type 2 diabetes (T2D), obesity, dyslipidemia, and cardiovascular diseases[3,4]. The impaired insulin action in key target tissues— especially skeletal muscle, liver, and adipose tissue—plays a central role in the pathogenesis of these conditions[5]. This underscores the critical importance of IR in the development and progression of metabolic disorders, necessitating targeted therapeutic strategies to mitigate its impact[6].

The euglycemic-hyperinsulinemic clamp[7] is considered the most reliable test for evaluating IR, but its complexity limits its use in large-scale studies. Prior research has demonstrated a robust correlation between IR scores derived from the homeostasis model assessment (HOMA) approach and IR assessed by glucose clamp techniques[8]. The triglyceride to high-density lipoprotein cholesterol (TG:HDL-C) ratio is an alternative IR marker that is more cost-effective and accessible in larger population studies compared to the HOMA test[9–12]. A recent genome-wide association study (GWAS) on TG:HDL-C in European populations discovered 114 single-nucleotide polymorphisms (SNPs)

[1]Department of Genome Sciences, University of Washington, Seattle, WA, USA. [2]Graduate Institute of Biomedical Sciences, China Medical University, Taichung, Taiwan, ROC. [3]Department of Public Health & Institute of Epidemiology and Preventive Medicine, National Taiwan University, Taipei, Taiwan, ROC. [4]Department of Psychiatry, Taipei Veterans General Hospital, Taipei, Taiwan, ROC. [5]Department of Psychiatry, College of Medicine, National Yang Ming Chiao Tung University, Taipei, Taiwan, ROC. [6]Brain Research Center, National Yang Ming Chiao Tung University, Taipei, Taiwan, ROC. [7]Department of Medical Research, Taipei Veterans General Hospital, Taipei, Taiwan, ROC. [8]Department of Psychiatry, National Taiwan University Hospital, Taipei, Taiwan, ROC. ✉e-mail: phkuo@ntu.edu.tw; tsai610913@gmail.com

associated with IR[13]. However, this previous GWAS on TG:HDL-C was limited to populations of European ancestry. Expanding genetic studies to diverse populations presents a valuable opportunity to uncover insights into the complex genetic foundations of IR[14]. In addition, IR is associated with impaired fatty acid utilization; increased free fatty acid flow from adipose to nonadipose tissue leads to aberrant fat metabolism and worsens IR. Hence, the TyG index, calculated from fasting plasma glucose and triglycerides, has emerged as another simple surrogate marker for IR[15,16]. The TyG index has high sensitivity for recognizing IR among apparently healthy subjects, compared with the HOMA-IR index[16]. As of the time of this writing and to our current knowledge, GWASs focusing on TyG have not been undertaken.

In this study, we conducted the GWAS on IR surrogate markers, including TyG, TG:HDL-C, and the logarithmically transformed TG:HDL-C ratio (log(TG:HDL-C)), in the Taiwanese population utilizing the Taiwan Biobank. We also carried out comparisons of heritability and genetic correlations among these markers. Due to the strong correlation between these IR surrogate markers and the lack of GWASs on TyG—an emerging and robust marker of IR—our subsequent analyses were centered on TyG. Additionally, fine mapping and polygenic risk score (PRS) analysis were performed for TyG. Moreover, we explored the relationships between disease phenotypes and IR surrogate markers. Furthermore, because sex differences in IR have been suggested[17], we undertook sex-stratified and sex-differentiated analyses of TyG in both female and male cohorts, along with a transcriptome-wide association study (TWAS) of TyG in the whole cohort. Finally, pathway analysis was accomplished on genes associated with TyG to reveal additional clues to the causes and consequences of IR-related phenotypes.

Results

Taiwan Biobank study cohort

Supplementary Table 1 presents baseline characteristics of the Taiwanese population within the Taiwan Biobank, stratified by whole, female, and male cohorts. In our study utilizing the Taiwan Biobank, we analyzed a total of 136,735 individuals with measurements for TyG and 138,303 individuals for TG:HDL-C and log(TG:HDL-C) (Supplementary Table 1). The gender distribution in both samples consisted of ~35.5% male participants and 64.5% female participants.

GWASs of IR markers

We conducted GWASs on three IR surrogate markers–TyG, TG:HDL-C, and log(TG:HDL-C)–within the Taiwan Biobank sample. The GWAS test statistics were calibrated at a genome-wide level, ensuring robustness and reliability (Supplementary Fig. 1). Figure 1 presents the Manhattan plot illustrating the distribution of association p values across the genome for SNPs associated with these markers.

Supplementary Table 2 provides the number of significant SNPs identified at various significant thresholds for each marker, offering a comprehensive overview of the genetic landscape associated with IR. Furthermore, Supplementary Table 3 presents the GWAS results of the top 20 significant SNPs after clumping for each marker, highlighting the genetic variants with the strongest associations with IR.

Four gene clusters are associated with TyG

Supplementary Tables 4 and 5 provide an overview of all identified variants and the top 20 variants/genes with genome-wide significance for TyG, respectively. These variants primarily clustered within four gene regions, centered on *GCKR* (chromosome 2), *MLXIPL* (chromosome 7), *APOA5* (chromosome 11), and *APOC1* (chromosome 19). Further analysis identified additional significant genes within these clusters: 29 genes in the *GCKR* cluster, 17 in the *MLXIPL* cluster, 15 in the *APOA5* cluster, and 6 in the *APOC1* cluster (Tables 1–4 and Fig. 2).

We conducted additional GWASs for TG:HDL-C and glucose levels (Supplementary Tables 6-7). Supplementary Tables 8–9 present the top variants exhibiting genome-wide significance for TG:HDL-C and glucose levels, respectively. The top 20 variants for TG:HDL-C are identical to those for TyG, with different rankings (Table 5). *GCKR*, a major factor in T2D risk[18], is the only gene consistently associated with glucose levels, TG:HDL-C, and TyG (Table 5).

Supplementary Table 10 summarizes genome-wide significance for variants associated with insulin-related traits in previous studies. Our analysis revealed 76 genome-wide significant genes in TyG that had been previously associated with IR in GWAS studies on European and/or East Asian populations (Supplementary Table 11).

Genes for IR identified in the Taiwan Biobank

Our GWAS on TyG within the Taiwan Biobank unveiled 197 genes not previously documented in the context of IR (Supplementary Table 12). According to the NHGRI-EBI GWAS Catalog[19], these genes have been associated with TG (99 genes), HDL-C (58 genes), T2D (29 genes) separately, as well as traits known to be associated with T2D risk including Alzheimer's Disease (AD; 14 genes), fasting glucose levels (22 genes), and body mass index (49 genes).

The top 20 genes, previously unreported in the literature, exhibiting genome-wide significance for TyG (Table 6) are primarily clustered within four gene regions: *GCKR*, *MLXIPL*, *APOA5*, and *APOC1*. Supplementary Fig. 2 displays a Manhattan plot of these top 20 TyG genes, emphasizing SNPs neutral in European studies but significant in our cohort, indicating population-specific effects. Analysis of LD between top SNPs and SNPs previously unreported in the literature for TyG (Supplementary Table 13) revealed varying degrees of LD within each gene region, providing insights into potential functional relationships. For example, we observed strong LD between *MLXIPL* rs3812316 and *TBL2* rs13246490 ($r^2 = 0.835$), suggesting a possible shared genetic effect. In contrast, *APOA1* rs12718464 and *APOA5* rs651821 exhibited weak LD ($r^2 = 0.008$), indicating potentially independent genetic influences on TyG within this region.

Conditional analyses on the top 20 genes, previously unreported in the literature, associated with TyG (Supplementary Table 14) helped discern independent genetic effects within closely located loci. For instance, *APOA1* rs12718464 maintained a significant association with TyG ($P = 1.1E-13$) when conditioning on *APOA5* rs651821, suggesting an independent effect. Conversely, conditioning on *MLXIPL* rs3812316 nullified the association of *TBL2* rs13246490 ($P = 0.32$), indicating these variants likely represent the same genetic signal. These findings enhance our understanding of the complex interplay between genetic variants influencing TyG and help identify truly independent genetic associations.

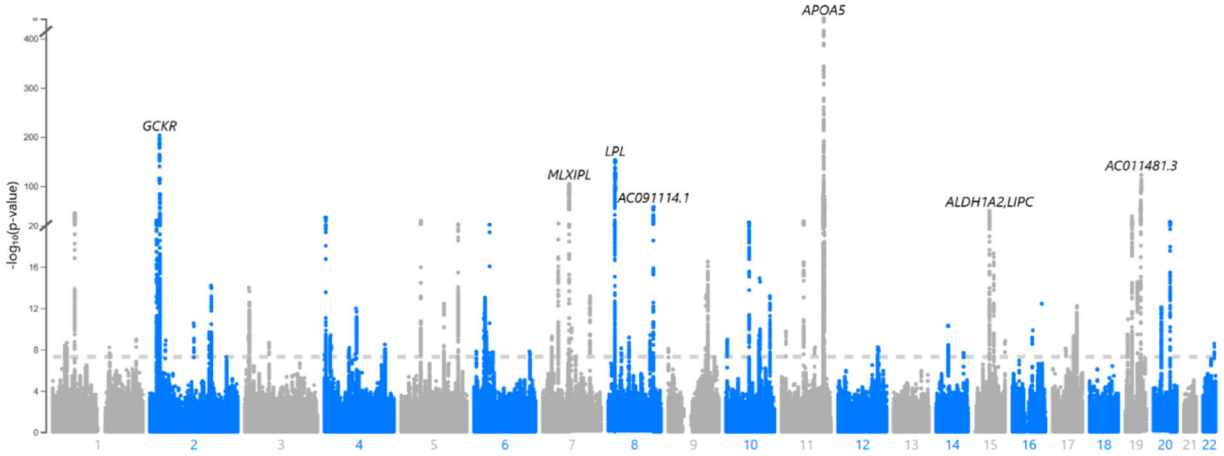
Heritability and genetic correlations for IR surrogate markers

We estimated the heritability[20] of TyG, TG:HDL-C, and log(TG:HDL-C) to be 15.5% (standard deviation (stdev) = 2.5%), 13.9% (stdev = 2.3%), and 17.3% (stdev = 2.5%), respectively (Supplementary Table 15). These values indicate the proportion of phenotypic variance attributed to genetic factors for each IR surrogate marker. Intercept values ranging from 1.07 to 1.09 suggest minimal biases due to confounding factors.

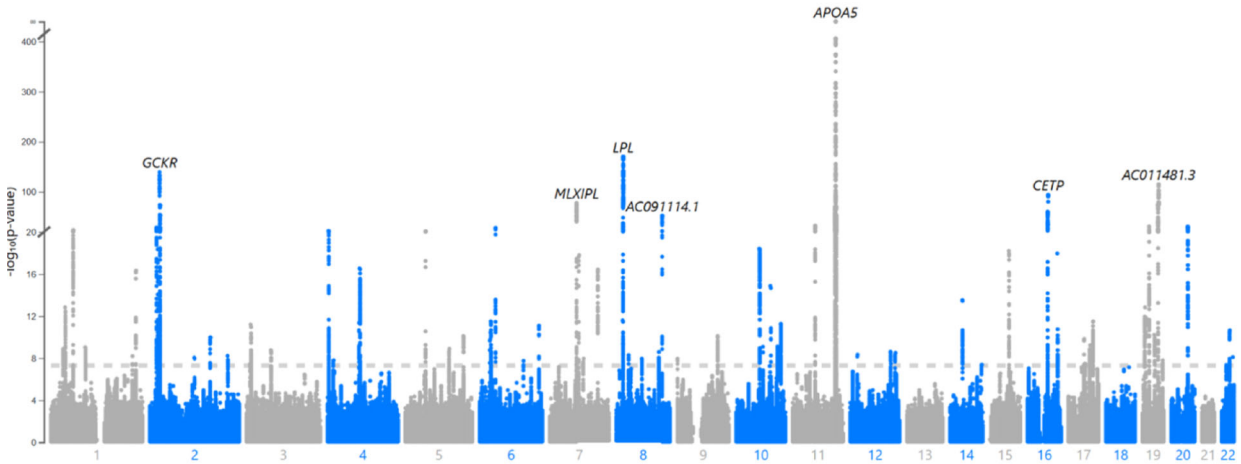
We also assessed genetic correlations to determine the extent to which genetic factors influencing TyG also affect TG:HDL-C and log(TG:HDL-C)[20]. The genetic correlations among TyG, TG:HDL-C, and log(TG:HDL-C) were notably high, ranging from 0.96 to 0.99 (Supplementary Table 15). This strong correlation underscores the close relationship among these three IR surrogate markers, highlighting their interconnectedness in the context of genetic influences on IR.

https://doi.org/10.1038/s41467-025-58506-x

(a) Manhattan plot for TyG index



(b) Manhattan plot for TG:HDL-C



(c) Manhattan plot for log(TG:HDL-C)

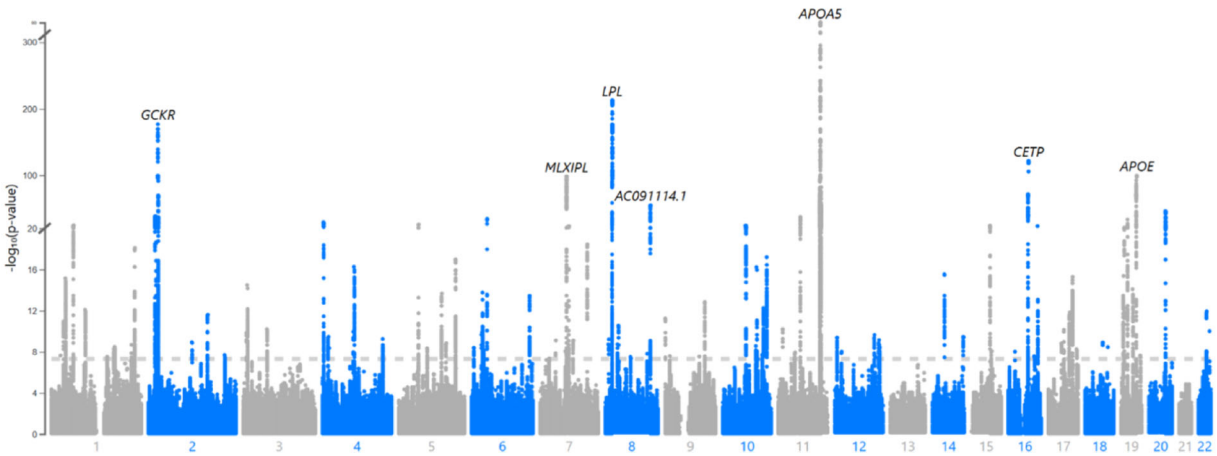


Fig. 1 | Manhattan plots. The Manhattan plots for the identified SNPs associated with the three IR surrogate markers, including **a** the TyG index, **b** TG:HDL-C ratio, and **c** log(TG:HDL-C) ratio. IR insulin resistance, TG:HDLC the triglyceride to high-density lipoprotein cholesterol ratio, TyG the product of fasting plasma glucose and triglycerides (Ln[fasting triglycerides (mg/dL) × fasting glucose (mg/dL)/2]). This GWAS analysis utilized BOLT-LMM's mixed linear models with a two-sided chi-square test. The conventional genome-wide significance threshold of $P < 5 \times 10^{-8}$ was applied. Source data is provided as a Source Data file.

Table 1 | Summary of top variants in the *GCKR* gene cluster exhibiting genome-wide significance for the TyG index, a marker of insulin resistance, identified within the Taiwan Biobank cohort

GENE	CHR	Interval (b38)	Top SNP	Pos (b38)	Top P	Novel	TWAS	GWAS catalog
DPYSL5	2	26847995..26950351	rs72804857	26,938,608	7.8E-11	no	NA	TG, HDL-C, FG, BMI, TG:HDL-C
MAPRE3	2	26970637..27027219	rs10207573	27,020,067	4.0E-09	no	1, 3	TG, HDL-C, TG:HDL-C
TMEM214	2	27032965..27041694	rs2304713	27,035,893	3.5E-10	yes	NA	NA
AGBL5	2	27050364..27070618	rs70953845	27,064,781	2.1E-10	yes	1, 5	TG
OST4[a]	2	27070472..27071654	rs4665943	27,073,017	1.8E-08	yes	7	NA
EMILIN1	2	27078615..27086403	rs2304682	27,084,901	1.8E-09	yes	2	NA
KHK	2	27086772..27100762	rs6714547	27,098,338	6.2E-11	yes	2, 8, 9, 10, 11	NA
CGREF1	2	27099353..27119128	rs6746337	27,118,195	2.6E-11	yes	5	TG
ABHD1	2	27123815..27130812	rs4665946	27,125,063	2.3E-21	yes	1, 2, 7, 8	NA
PREB	2	27130756..27134636	rs71401560	27,131,111	1.6E-15	yes	2, 9, 10	NA
PRR30	2	27136848..27139410	rs11121	27,137,019	1.3E-15	yes	NA	TG
TCF23	2	27149004..27156974	rs1275513	27,148,102	7.9E-18	yes	NA	TG
SLC5A6[a]	2	27199587..27212787	rs74684611	27,191,771	2.3E-21	yes	8	TG, HDL-C
CAD[a]	2	27217369..27243943	rs118130043	27,245,399	3.8E-13	yes	3	TG
TRIM54	2	27282429..27307435	rs189387480	27,292,538	5.9E−12	yes	NA	TG
MPV17	2	27309492..27323097	rs146502189	27,322,314	3.3E-13	yes	8	TG
GTF3C2	2	27325854..27356764	rs201232671	27,348,271	6.4E-14	yes	1	TG
PPM1G	2	27381199..27409591	rs79803862	27,399,271	4.0E-22	yes	6, 8, 9	TG, FG, BMI
NRBP1[a]	2	27427790..27442259	rs147530299	27,419,228	2.2E-23	yes	5	TG, FG, BMI
IFT172	2	27444377..27489743	rs117414910	27,455,222	1.8E-17	no	1, 10, 11	TG, TG:HDL-C
GCKR	2	27496839..27523684	rs1260326	27,508,073	4.3E-204	no	5, 11	IG, TG, T2D, FG, BMI, TG:HDL-C
SPATA31H1	2	27537386..27582722	rs1919127	27,578,626	6.9E-114	yes	1, 2, 4, 9, 11	NA
ZNF512	2	27583042..27623217	rs12989678	27,598,615	1.3E-115	yes	NA	TG, T2D
GPN1	2	27628247..27651511	rs34502053	27,631,657	7.9E-68	yes	4, 8, 10, 11	TG, BMI
SUPT7L	2	27642568..27663614	rs4666010	27,655,819	9.9E-45	yes	1, 5, 6, 9, 10, 11	NA
SLC4A1AP	2	27663889..27694969	rs13021208	27,678,861	1.6E−57	yes	6, 10	TG
MRPL33	2	27771719..27779733	rs3792252	27,773,064	2.2E-37	yes	2, 8, 9	TG, FG
RBKS	2	27781379..27890387	rs898034	27,867,953	1.5E-38	yes	8, 11	TG
BABAM2	2	27888709..28338901	rs141361525	27,896,053	8.7E-18	yes	1, 2, 11	TG, HDL-C, FG. BMI
TAD	**CHR**	**Interval (b38)**	--	--	--	--	--	--
pancreas	2	25920000..27320000	--	--	--	--	--	--
pancreas	2	27680000..28520000	--	--	--	--	--	--
hippocampus	2	25360000..27360000	--	--	--	--	--	--
hippocampus	2	27680000..28480000	--	--	--	--	--	--
liver	2	25920000..27320000	--	--	--	--	--	--
liver	2	27360000..28520000	--	--	--	--	--	--

Tissues in TWAS: 1 = Adipose Subcutaneous; 2 = Adipose Visceral Omentum; 3 = Adrenal Gland; 4 = Brain Cortex; 5 = Brain Hippocampus; 6 = Brain Hypothalamus; 7 = Liver; 8 = Muscle Skeletal; 9 = Pancreas; 10 = Pituitary; 11 = Thyroid. Previous GWASs reported in the NHGRI-EBI GWAS Catalog have identified various phenotypes that reach genome-wide significance ($P < 5 \times 10^{-8}$). Coordinates for TADs were retrieved from the following source (https://3dgenome.fsm.northwestern.edu/publications.html).

AD Alzheimer's disease, *b38* Genome Research Consortium human build 38, *BMI* body mass index, *CHR* chromosome, *FG* fasting glucose, *HDL-C* high-density lipoprotein cholesterol, *IR* insulin resistance, *NA* not available, *Pos* position, *T2D* type 2 diabetes, *TAD* topologically associating domains, *TG* triglycerides, *TG:HDLC* the triglyceride to high-density lipoprotein cholesterol ratio, *TWAS* transcriptome-wide association analysis.

[a]Nearby gene. -- not applicable. This GWAS analysis utilized BOLT-LMM's mixed linear models with a two-sided chi-square test. The conventional genome-wide significance threshold of $P < 5 \times 10^{-8}$ was applied.

Fine mapping analysis for TyG

We employed fine mapping analysis to identify candidate causal variants linked to TyG[21], consolidating overlapping loci into 11 distinct genomic regions of interest (Supplementary Table 16 and Supplementary Fig. 3). This approach enhanced analytical precision and prioritized areas for further investigation. For instance, within the chromosomal region 43378777–46429300 base pair on chromosome 19, five credible sets were identified with purity values ranging from 0.92 to 1.0 (Supplementary Table 16), suggesting a high degree of independence among the SNPs. Coverage values for these sets ranged from 0.95 to 1.0, indicating a high probability that the true causal variant for IR is represented within the credible sets. Certain SNPs exhibited PIP values exceeding 0.9, signifying a heightened probability of these variants being linked to IR (Supplementary Fig. 3). This indicates a robust likelihood that these specific genetic variants may play a significant role in IR manifestation or susceptibility, based on evidence from the Bayesian variable selection approach.

https://doi.org/10.1038/s41467-025-58506-x

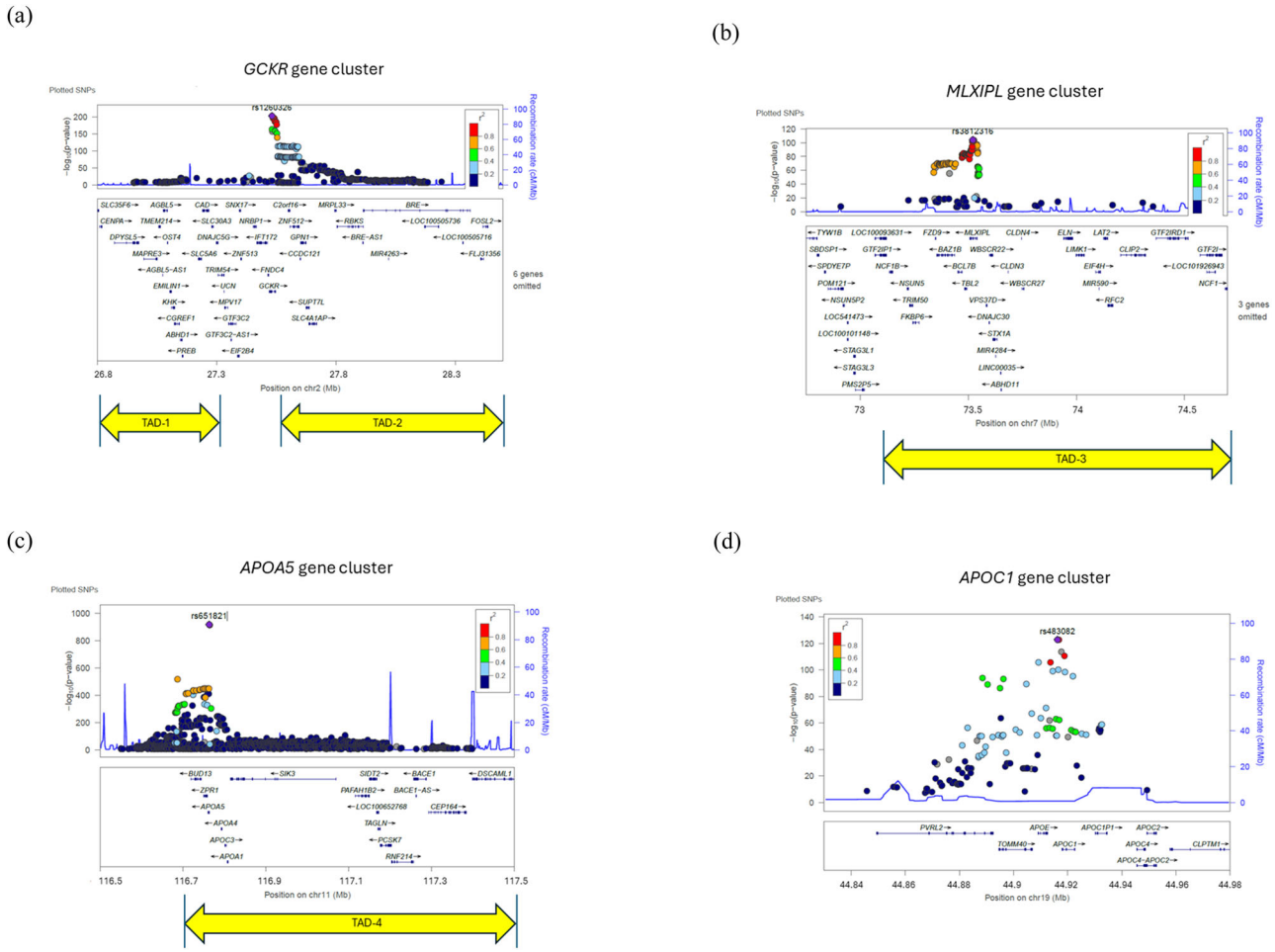


Fig. 2 | Locus zoom plots of gene clusters associated with TyG index in the Taiwan Biobank. The locus zoom plot for the **a** *GCKR*, **b** *MLXIPL*, **c** *APOA5*, and **d** *APOC1* gene clusters concerning the TyG index in the Taiwan Biobank illustrates single nucleotide polymorphisms (SNPs) by their chromosomal positions and their association with TyG (−log10 P). The SNPs are color-coded to indicate their linkage disequilibrium with the top SNPs **a** rs1260326; **b** rs3812316; **c** rs651821; and **d** rs483082. Additionally, estimated recombination rates are depicted in cyan, derived from Asian subjects in the 1000 Genomes Project. The plot was generated using LocusZoom. This GWAS analysis utilized BOLT-LMM's mixed linear models with a two-sided chi-square test. The conventional genome-wide significance threshold of $P < 5 \times 10^{-8}$ was applied. The topologically associating domain (TAD) TAD-1 spans from 25,920,000 to 27,320,000 (GRCh38) in pancreatic tissue. The TAD-2 spans from 27,680,000 to 28,520,000 (GRCh38) in pancreatic tissue. The TAD-3 spans from 73,160,000 to 74,760,000 (GRCh38) in pancreatic tissue. The TAD−4 spans from 116,760,000 to 119,360,000 (GRCh38) in pancreatic tissue. Notably, these TADs cover a significant portion of the *GCKR*, *MLXIPL*, and *APOA5* gene clusters. Source data are provided as a Source Data file.

PRS analysis for TyG

We conducted PRS analysis to quantify the cumulative impact of multiple genetic variants on TyG[22]. The analysis used 854,050 SNPs to calculate the PRS for TyG (Supplementary Table 17 and Supplementary Fig. 4). The PRS model demonstrated substantial explanatory power, with 23.80% of the variance in TyG attributed to genetic factors. The incremental predictive value (R^2–R^2 Null) achieved by integrating genetic information into the PRS model was 10.57% compared to a model without genetic predictors. The R^2 and R^2–R^2 Null values for TyG surpassed those reported in diverse phenotypes from prior investigations (e.g., triglyceride[23] in Supplementary Table 17 and Supplementary Fig. 5), underscoring the robustness and effectiveness of the PRS model in elucidating the genetic underpinnings of TyG in the context of IR.

Association of disease phenotypes with IR surrogate markers

This study investigated the relationship between 27 disease phenotypes and IR surrogate markers (TyG, TG:HDL) and their PRS within the Taiwan Biobank (Table 7). Statistical significance was determined using the Bonferroni correction threshold of $p = 0.00185$. TyG exhibited associations with 20 different diseases, including coronary heart disease and mood disorders ($P < 0.0001$). TG:HDL showed associations with 14 diseases, such as gout and hyperlipidemia ($P < 0.0001$). PRS of TyG was linked to gout ($P = 0.0016$) and hyperlipidemia ($P < 0.0001$). PRS of TG:HDL was also associated with gout ($P = 0.001$) and hyperlipidemia ($P < 0.0001$). These findings highlight the relationships between IR surrogate markers, their PRS, and various disease phenotypes within the Taiwan Biobank.

Sex-stratified and sex-differentiated analyses of TyG

Sex-specific genetic differences have been found previously in the context of IR, and exploring sex-specific genetic effects may reveal a substantial number of important genes, previously unreported in the literature[17]. We conducted sex-stratified analyses, revealing 5787 significant SNPs in females (Supplementary Table 18) and 2817 SNPs in males (Supplementary Table 19).

Sex-differentiated analyses identified 915 SNPs with statistically heterogeneous effects between sexes (heterogeneity p value < 0.05) among those with genome-wide significance (Supplementary

Table 2 | Summary of top variants in the *MLXIPL* gene cluster exhibiting genome-wide significance for the TyG index, a marker of insulin resistance, identified within the Taiwan Biobank cohort

GENE	CHR	Interval (b38)	Top SNP	Pos (b38)	Top P	Novel	TWAS	GWAS catalog
POM121	7	72879357..72951459	rs117135057	72,938,316	1.8E-08	yes	NA	TG, BMI
NSUN5	7	73302516..73308826	rs3750074	73,308,628	1.1E-17	yes	NA	NA
FKBP6[a]	7	73328161..73358625	rs186391487	73,384,030	9.2E-18	yes	1, 2	TG
FZD9	7	73433778..73436120	rs1178947	73,435,848	5.8E-69	yes	9	TG, HDL-C
BAZ1B	7	73440406..73522293	rs111837003	73,493,129	1.2E-71	yes	1, 4, 6, 10, 11	TG, HDL-C
BCL7B	7	73536356..73557690	rs13244614	73,559,524	9.0E-84	yes	1, 2, 8, 9	HDL-C
TBL2	7	73567537..73578579	rs13246490	73,578,020	2.6E-84	yes	NA	TG, HDL-C, BMI
MLXIPL	7	73593202..73647907	rs3812316	73,606,007	8.5E-105	no	1, 2, 3, 9, 10, 11	TG, HDL-C, BMI, TG:HDL-C
VPS37D[a]	7	73665347..73672110	rs371515311	73,660,465	6.1E-16	yes	1	TG, HDL-C
DNAJC30	7	73680918..73683453	rs3828973	73,682,169	1.3E-09	yes	NA	TG
ABHD11[a]	7	73736094..73738802	rs10226706	73,751,424	3.6E-08	yes	5	TG
CLDN3	7	73768997..73770270	rs6460054	73,770,204	3.0E-08	yes	NA	TG
TMEM270[a]	7	73860848..73865890	rs73144858	73,922,565	2.7E-13	yes	7	NA
ELN[a]	7	74028173..74069907	rs188072556	73,999,294	1.3E-15	yes	10	NA
RFC2	7	74231502..74254399	rs148741166	74,252,973	1.2E-10	yes	NA	TG
CLIP2	7	74289407..74405935	rs151126504	74,392,493	1.0E-13	yes	NA	NA
GTF2IRD1[a]	7	74453906..74602605	rs145049525	74,440,602	9.0E-09	no	NA	TG, BMI, TG:HDL-C
TAD	**CHR**	**Interval (b38)**	--	--	--	--	--	--
pancreas	7	73160000..74760000	--	--	--	--	--	--
hippocampus	7	73280000..74760000	--	--	--	--	--	--
liver	7	73280000..74760000	--	--	--	--	--	--

Tissues in TWAS: 1 = Adipose Subcutaneous; 2 = Adipose Visceral Omentum; 3 = Adrenal Gland; 4 = Brain Cortex; 5 = Brain Hippocampus; 6 = Brain Hypothalamus; 7 = Liver; 8 = Muscle Skeletal; 9 = Pancreas; 10 = Pituitary; 11 = Thyroid. Previous GWASs reported in the NHGRI-EBI GWAS Catalog have identified various phenotypes that reach genome-wide significance ($P < 5 \times 10^{-8}$). Coordinates for TADs were retrieved from the following source (http://3dgenome.fsm.northwestern.edu/publications.html).

AD Alzheimer's disease, *b38* Genome Research Consortium human build 38, *BMI* body mass index; *CHR* chromosome, *FG* fasting glucose, *HDL-C* high-density lipoprotein cholesterol, *IR* insulin resistance, *NA* not available, *Pos* position, *T2D* type 2 diabetes, *TAD* topologically associating domains, *TG* triglycerides, *TG:HDLC* the triglyceride to high-density lipoprotein cholesterol ratio, *TWAS* transcriptome-wide association analysis.

[a]Nearby gene. -- not applicable. This GWAS analysis utilized BOLT-LMM's mixed linear models with a two-sided chi-square test. The conventional genome-wide significance threshold of $P < 5 \times 10^{-8}$ was applied.

Table 20). For instance, the SNP rs7412 in *APOE* showed a heterogeneity *p* value of 2.37E-4 between sexes, with a *p* value of 2.29E-33 (effect size = 0.0796) for males and a *p* value of 4.43E-24 (effect size = 0.0494) for females. We identified 42 genes with sex-specific effects in females and 25 in males (Supplementary Table 21). Notably, female-specific genes were found in the *MLXIPL* cluster (*FZD9* and *BAZ1B*) and *APOA5* cluster (*APOA4*, *APOA1*, *PAFAH1B2*, *SIDT2*, *TAGLN*, and P*CSK7*). Male-specific genes such as *MRPL33* (*GCKR* cluster), *BACE1*, *CEP164*, *DSCAML1* (*APOA5* cluster), and *APOC1P1* (*APOC1* cluster) were identified. The top loci with stronger female-specific effects were mapped to *KLF14*, *APOA4*, *NID2*, and *TNFAIP8*. Conversely, loci mapping to *GMIP*, *TRIB1*, and *PBX4* showed stronger male-specific effects. Specifically, the locus rs1364422 (*KLF14*) demonstrated the strongest female-specific effect (heterogeneity *p* value = 1.51E-05). The locus rs190712692 (*APOC1*) showed the strongest male-specific effect (heterogeneity *p* value = 1.24E-05). However, some *APOC1* loci also exhibited female-specific effects.

In our parent-offspring cohort, we detected parent-of-origin effects at the *KLF14* locus (rs1364422), with 0 paternal and 1173 maternal transmissions of the minor allele in 1173 heterozygous offspring, indicating strong maternal inheritance (*p* value < 2.2E-16; Supplementary Table 22).

TWAS of TyG in the whole cohort

We used a TWAS approach[24] to identify genes with expression differences associated with TyG. Supplementary Tables 23–27 present genes exhibiting significant associations with TyG in TWAS at genome-wide significance across chromosomes 2, 7, 11, 19, and in aggregate. In the *GCKR* gene cluster, 23 of 29 genes showed significant associations in both TWAS and GWAS, predominantly in adipose-subcutaneous, muscle-skeletal, and thyroid tissues (Table 1). The *MLXIPL* gene cluster revealed 9 of 17 genes with significant associations in both analyses, primarily in adipose-subcutaneous tissue (Table 2). In the *APOA5* gene cluster, 13 of 15 genes demonstrated significant associations, mainly in adipose-visceral-omentum tissue (Table 3). The *APOC1* gene cluster showed 4 of 6 genes with significant associations in both TWAS and GWAS (Table 4).

Pathway analysis of TyG

Pathway analysis was employed to elucidate key pathways influenced by TyG and gene-metabolic interactions[25]. The analysis included genome-wide significant genes, sex-specific genes, and genes from previous studies (Supplementary Tables 28–31). A summary of GO analysis results for four gene groups is presented in Table 8 and Supplementary Table 32.

Genome-wide significant genes are involved in AD pathogenesis, glucose homeostasis, insulin resistance and signaling, neurodegenerative pathways, Wnt signaling cascade, and plasma lipoprotein dynamics (Table 8). These enriched GO terms and pathways demonstrate substantial overlap with those associated with genes from previous studies (Table 8), suggesting a consistent functional profile across multiple investigations.

Female-specific genes exhibited similar enrichment to genome-wide significant genes (Table 8). The congruence between female-specific and genome-wide significant genes implies a potentially prominent role for these pathways in female-specific disease mechanisms. Conversely, male-specific genes revealed a distinct functional profile (Table 8) and lacked several pathways enriched in female-specific genes

Table 3 | Summary of top variants in the *APOA5* gene cluster exhibiting genome-wide significance for the TyG index, a marker of insulin resistance, identified within the Taiwan Biobank cohort

GENE	CHR	Interval (b38)	Top SNP	Pos (b38)	Top P	Novel	TWAS	GWAS catalog
BUD13	11	116748173..116772987	rs6589565	116,769,521	7.3E-445	yes	8	TG, HDL-C
ZPR1	11	116773799..116788023	rs10750096	116,786,072	3.4E-451	no	NA	TG, HDL-C, TG:HDL-C
APOA5	11	116789367..116792420	rs651821	116,791,863	1.4E-919	no	9	TG, HDL-C, TG:HDL-C
APOA4	11	116820700..116823304	rs5104	116,821,618	1.1E-235	no	8, 11	TG, HDL-C, TG:HDL-C
APOC3	11	116829907..116833072	rs5128	116,832,924	2.8E-152	yes	6	TG, HDL-C
APOA1	11	116835751..116837622	rs12718464	116,836,685	3.1E-39	yes	2, 6, 7, 10	TG, HDL-C, BMI
SIK3	11	116843402..117098428	rs6589574	116,859,922	1.2E-112	no	1, 2, 5, 6, 8, 10	TG, HDL-C, BMI, TG:HDL-C
PAFAH1B2	11	117144287..117178173	rs7925256	117,174,619	8.0E-75	yes	5, 9	TG, HDL-C, BMI
SIDT2	11	117178743..117197442	rs6589603	117,185,123	1.1E-59	yes	1, 2, 3, 4, 6, 9, 10, 11	TG, HDL-C
TAGLN	11	117199294..117207465	rs12970	117,203,393	6.1E-40	yes	4, 8, 9	TG, HDL-C
PCSK7	11	117204337..117232073	rs236911	117,214,554	5.3E-56	no	1, 2, 7, 11	TG, HDL-C, BMI, TG:HDL-C
RNF214	11	117232671..117286454	rs200694867	117,244,658	6.3E-28	yes	NA	TG, HDL-C
BACE1	11	117285698..117316256	rs3741288	117,315,499	4.2E−12	yes	2	TG, HDL-C
CEP164	11	117321778..117413266	rs11216364	117,355,903	1.2E-34	no	5, 10	TG, HDL-C, BMI, TG:HDL-C
DSCAML1[a]	11	117427772..117817514	rs7115491	117,424,434	8.6E-17	no	NA	TG, HDL-C, TG:HDL-C
TAD	**CHR**	**Interval (b38)**	--	--	--	--	--	--
pancreas	11	116760000..119360000	--	--	--	--	--	--
hippocampus	11	116760000..117200000	--	--	--	--	--	--
hippocampus	11	117200000..118960000	--	--	--	--	--	--
liver	11	116720000..117200000	--	--	--	--	--	--
liver	11	117360000..118120000	--	--	--	--	--	--

Tissues in TWAS: 1 = Adipose Subcutaneous; 2 = Adipose Visceral Omentum; 3 = Adrenal Gland; 4 = Brain Cortex; 5 =Brain Hippocampus; 6 = Brain Hypothalamus; 7 = Liver; 8 = Muscle Skeletal; 9 = Pancreas; 10 = Pituitary; 11 = Thyroid. Previous GWASs reported in the NHGRI-EBI GWAS Catalog have identified various phenotypes that reach genome-wide significance ($P < 5 \times 10^{-8}$). Coordinates for TADs were retrieved from the following source (http://3dgenome.fsm.northwestern.edu/publications.html).

AD Alzheimer's disease, *b38* Genome Research Consortium human build 38, *BMI* body mass index, *CHR* chromosome, *FG* fasting glucose, *HDL-C* high-density lipoprotein cholesterol, *IR* insulin resistance, *NA* not available, *Pos* position, *T2D* type 2 diabetes, *TAD* topologically associating domains, *TG* triglycerides, *TG:HDLC* the triglyceride to high-density lipoprotein cholesterol ratio, *TWAS* transcriptome-wide association analysis.

[a]Nearby gene. -- not applicable. This GWAS analysis utilized BOLT-LMM's mixed linear models with a two-sided chi-square test. The conventional genome-wide significance threshold of $P < 5 \times 10^{-8}$ was applied.

(e.g., estrogen signaling pathway). The male-specific genes showed enrichment in the tight junction pathway, which was absent in the female-specific gene set. This sexual dimorphism in pathway enrichment suggests potential differences in disease mechanisms between sexes.

Discussion

A recent European GWAS pinpointed 114 genetic loci that play a role in TG:HDL-C[13]. However, despite the large number of identified loci, they do not fully account for the heritability of IR, suggesting the existence of additional undiscovered genetic variants. Because the majority of GWAS studies on IR have focused predominantly on European populations, genetic variants that contribute to variations in traits across different populations are likely to be discovered by expanding genetic investigations to diverse populations[14,26]. Indeed, this inaugural GWAS of IR markers in the Taiwan Biobank has provided a wealth of loci, previously unreported in the literature, enhancing our understanding of the genetic underpinnings of IR-related diseases.

In our GWAS of IR markers, we identified genome-wide significant associations with loci in four distinct genomic regions: *GCKR*, *MLXIPL*, *APOA5*, and *APOC1*. SNPs with minimal impact in European studies are significant here (Supplementary Fig. 2), indicating potential unique genetic or environmental influences in Asians, warranting further investigation. The enrichment of IR-associated loci over such large regions might be attributed to the unique population structure and haplotype patterns present in the Taiwanese population. However, we found that these clusters are included within haploblocks identified in broader population studies[14,27–31]. For instance, our analysis of the East Asian population from the 1000 Genomes Project revealed peaks in recombination rates at the boundaries of the *APOA5* gene cluster (Fig. 2). Furthermore, although they have not been highlighted in previous studies, similar clusters over the same genetic regions have also appeared in studies within the European, African, Korean, South Asian, and diverse populations[14,27–31]. For example, a recent study identified these four gene clusters in a diverse cohort, associating them with blood glucose, HDL-C, and TG levels (Supplementary Table 33)[28]. These clustered associations, therefore, may reflect low recombination rates within the regions that lead to a high probability of co-inheritance driven by strong IR-related loci, such as *GCKR*[32,33].

Another explanation for genetic association to extend across such broad genomic regions is speculative but bears some mention here. We noted that the clustered high-association regions were mostly contained within the same or closely adjacent topologically associated domains (TADs), three-dimensional chromatin structures that enclose adjacent genes that are subject to some degree of co-regulation and co-expression in vivo (Tables 1–4 and Fig. 2). Regulatory elements within TAD structures can influence the disease-related expression of multiple neighboring genes[34,35], and especially considering the TWAS results indicating the common mis-expression for these clustered genes in adipocytes (Tables 1–4), we speculate that the extended association clusters we identified could reflect the influence of noncoding variants that alter regulatory relationships and extend across the TADs. In contrast to the other three clusters,

Table 4 | Summary of top variants in the *APOC1* gene cluster exhibiting genome-wide significance for the TyG index, a marker of insulin resistance, identified within the Taiwan Biobank cohort

GENE	CHR	Interval (b38)	Top SNP	Pos (b38)	Top P	Novel	TWAS	GWAS catalog
NECTIN2	19	44846297..44889223	rs283811	44,885,243	1.6E-94	yes	4	TG, HDL-C, T2D, AD, BMI
TOMM40	19	44891254..44903689	rs157582	44,892,962	7.5E-94	yes	NA	TG, HDL-C, AD, BMI
APOE	19	44905796..44909393	rs440446	44,905,910	1.7E-106	yes	1	TG, HDL-C, T2D, AD, BMI
APOC1	19	44914325..44919346	rs5117	44,915,533	2.8E-111	no	8	TG, HDL-C, AD, BMI, TG:HDL-C
APOC1P1	19	44926803..44931386	rs7259004	44,929,300	2.0E–59	no	NA	TG, HDL-C, AD, BMI, TG:HDL-C
APOC2, APOC4-APOC2	19	44946051..44949565	rs2288911	44,946,027	2.8E-10	yes	3, 10	TG, HDL-C, AD
TAD	**CHR**	**Interval (b38)**	--	--	--	--	--	--
pancreas	19	43760000..44560000	--	--	--	--	--	--
pancreas	19	45440000..47760000	--	--	--	--	--	--
hippocampus	19	43520000..44720000	--	--	--	--	--	--
hippocampus	19	45640000..47720000	--	--	--	--	--	--
liver	19	42080000..44560000	--	--	--	--	--	--
liver	19	45000000..47720000	--	--	--	--	--	--

Tissues in TWAS: 1 = Adipose Subcutaneous; 2 = Adipose Visceral Omentum; 3 = Adrenal Gland; 4 = Brain Cortex; 5 = Brain Hippocampus; 6 = Brain Hypothalamus; 7 = Liver; 8 = Muscle Skeletal; 9 = Pancreas; 10 = Pituitary; 11 = Thyroid. Previous GWASs reported in the NHGRI-EBI GWAS Catalog have identified various phenotypes that reach genome-wide significance ($P < 5 \times 10^{-8}$). Coordinates for TADs were retrieved from the following source (http://3dgenome.fsm.northwestern.edu/publications.html).

AD Alzheimer's disease, *b38* Genome Research Consortium human build 38, *BMI* body mass index, *CHR* chromosome, *FG* fasting glucose, *HDL-C* high-density lipoprotein cholesterol, *IR* insulin resistance, *NA* not available, *Pos* position, *T2D* type 2 diabetes, *TAD* topologically associating domains, *TG* triglycerides, *TG:HDLC* the triglyceride to high-density lipoprotein cholesterol ratio, *TWAS* transcriptome-wide association analysis.

the *APOC1* cluster is not enclosed within an established TAD (Fig. 2); however, the genes in this cluster are known to interact with the same long-distance enhancers[36,37], suggesting that enhancer variation could contribute to a cluster-wide association with IR in this region as well.

Our sex-differentiated analyses of TyG identified 40 female-specific and 25 male-specific genes. Notably, *KLF14* exhibited the most pronounced sex-specific effect in females, consistent with recent findings in European populations[13]. *KLF14*-associated SNPs have consistently shown sex-specific associations with metabolic traits, including T2D, waist-to-hip ratio, TG, HDL-C, and low-density lipoprotein cholesterol[38–41], and corroborating these human studies, *KLF14* deficiency in adipocytes leads to increased adiposity in female but not male mice[42]. The observed stronger associations in females are hypothesized to stem from the modulation of *KLF14* expression rather than from hormonal influences[43,44]. We also identified several female-specific genes located within the *APOA5* gene cluster, aligning with previous research[45–48]. These findings underscore sex-dependent mechanisms in lipid regulation and metabolism, exemplifying the importance of considering sex-specific effects in genetic studies of metabolic traits.

The imprinting analysis of *KLF14* not only corroborates its known role in sex-specific genetic influence but also highlights the utility of large biobank datasets with familial information for dissecting complex genetic mechanisms[49]. Our findings suggest that the imprinting status of *KLF14* could influence its biological roles, potentially affecting metabolic pathways or disease risk in a manner dependent on the sex of the parent from whom the allele is inherited[44,50]. Future studies could expand this approach to other genes with suspected imprinting effects, furthering our understanding of how genetic and epigenetic factors interplay in human health and disease.

Comparing gene sets from previous studies and the current genome-wide significant genes revealed substantial concordance in enriched pathways, particularly those implicated in insulin resistance/signaling, lipid metabolism, and glucose homeostasis, highlighting the critical role of these pathways in the etiology of IR. Interesting, GO analysis of the four distinct gene groups (genome-wide significant, female-specific, male-specific, and previously identified genes) also consistently demonstrated significant involvement in neurodegenerative processes; pathways associated with AD, including those related to nervous system development, are prominently represented across all groups. This shared enrichment suggests a potential mechanistic link underlying the high association between IR and neurodegenerative disorders[51–53].

Our pathway analysis revealed some potential differences in male and female groups. For example, the GO analysis of female-specific genes revealed a distinct enrichment in pathways associated with estrogen signaling; this pathway's prominence in female-specific genes is consistent with its proposed role in modulating disease risk or progression in women[54]. Conversely, the male-specific gene set exhibits a unique enrichment in the tight junction pathway, which is crucial for maintaining cellular barriers[55]. This pathway is absent in the female-specific gene set, indicating a potential sex-based difference in cellular mechanisms. Furthermore, insulin signaling pathway genes were more highly represented in the female-specific genes, pointing to potentially divergent metabolic regulation between sexes[56,57]. The identification of distinct molecular pathways in male and female-specific gene sets underscores the potential for developing sex-specific therapeutic approaches[58,59]. Sex-specific genetic factors and disease mechanisms interplay complexly[17], suggesting that sex-tailored interventions may improve treatment efficacy[60]. Further research is warranted to elucidate the functional implications of these sex-specific pathway enrichments and their impact on disease risk and treatment strategies[61,62].

The absence of medication data limits our understanding of drug effects on insulin resistance markers. Our sensitivity analysis, excluding self-reported diabetes mellitus (DM) cases (Supplementary Fig. 6 and Supplementary Table 34), shows minimal change in genetic associations for the TyG index, suggesting robustness despite potential DM treatment effects. Nonetheless, this does not substitute for medication records, emphasizing the need for future studies to

https://doi.org/10.1038/s41467-025-58506-x

Table 5 | Comparison summary of top variants exhibiting genome-wide significance for the TyG index, TG:HDL-C index, and glucose levels, identified within the Taiwan Biobank (sorted by gene)

TyG				TG:HDL-C				Glucose			
CHR	**GENE**	**SNP**	***P***	**Chr**	**GENE**	**SNP**	***P***	**Chr**	**GENE**	**SNP**	***P***
11	*APOA1*[a]	rs10750098	4.3E-150	11	*APOA1*[a]	rs10750098	9E-99	2	*ABCB11*	rs139014876	1.8E-209
11	*APOA4*	rs5104	1.1E-235	11	*APOA4*	rs5104	2E-191	7	*AGMO*[a]	rs6947830	5.2E-59
19	*APOC1*	rs5117	2.8E-111	19	*APOC1*	rs5117	2E-101	7	*CAMK2B*	rs732360	3.6E-140
11	*APOC3*	rs5128	2.8E-152	11	*APOC3*	rs5128	2E-100	6	*CDKAL1*	rs34499031	3.0E-51
19	*APOE*	rs440446	1.7E-106	19	*APOE*	rs429358	8E-112	9	*CDKN2B*[a]	rs10811661	5.1E-68
7	*BAZ1B*	rs111837003	1.2E-71	7	*BAZ1B*	rs111837003	7.3E-55	20	*FOXA2*	rs3833331	3.6E-92
7	*BCL7B*	rs7793710	1.1E-83	7	*BCL7B*	rs7793710	3.6E-62	2	*G6PC2*	rs2232326	1.9E-235
2	*C2orf16*	rs1919127	6.9E-114	2	*C2orf16*	rs1919127	6.3E-73	7	*GCK*	rs2908289	2.9E-161
7	*FZD9*	rs1178947	5.8E-69	16	*CETP*	rs12720926	3.3E-61	2	*GCKR*	rs1260326	6.8E-60
2	*GCKR*	rs1260326	4.3E-204	7	*FZD9*	rs1178947	6.6E-54	9	*KANK1*	rs16922302	2.6E-51
2	*GPN1*	rs34502053	7.9E-68	2	*GCKR*	rs1260326	7E-140	20	*LINC00261*	rs6113722	4.1E-89
8	*LPL*	rs75278536	2.0E-153	8	*LPL*	rs12679834	6E-171	20	*LOC101929685*	rs6048211	3.6E-88
7	*MLXIPL*	rs3812316	8.5E-105	7	*MLXIPL*	rs3812316	4.3E-78	11	*MTNR1B*	rs10830963	5.5E-144
19	*NECTIN2*	rs283811	1.6E-94	19	*NECTIN2*	rs283811	5.7E-97	7	*MYL7*	rs741033	3.9E-58
11	*PAFAH1B2*	rs7925256	8.0E-75	11	*PAFAH1B2*	rs117592676	3.8E-68	2	*NOSTRIN*	rs143166986	1.2E-90
11	*SIK3*	rs6589574	1.2E-112	11	*SIK3*	rs199986368	3.4E-71	7	*POLD2*	rs3087370	2.5E-57
8	*SLC18A1*[a]	rs9644568	2.8E-90	8	*SLC18A1*[a]	rs9644568	2E-106	2	*SIX3*[a]	rs12712928	1.4E−126
7	*TBL2*	rs13246490	2.6E-84	7	*TBL2*	rs12540011	1.1E-62	8	*SLC30A8*	rs13266634	4.3E-95
19	*TOMM40*	rs157582	7.5E-94	19	*TOMM40*	rs10119	1E-103	2	*SPC25*	rs493816	2.2E-102
2	*ZNF512*	rs12989678	1.3E-115	2	*ZNF512*	rs12989678	1.1E-74	7	*YKT6*	rs2971667	8.1E-159

[a]Nearby gene. This GWAS analysis utilized BOLT-LMM's mixed linear models with a two-sided chi-square test. The conventional genome-wide significance threshold of $P < 5 \times 10^{-8}$ was applied.

Table 6 | Top 20 novel genes exhibiting genome-wide significance for the TyG index, a marker of insulin resistance, identified within the Taiwan Biobank cohort

CHR	Rank	Gene	Top SNP	Top P	Interval (b38)		GWAS catalog
2	4	*SPATA31H1*	rs1919127	6.90E-114	27,576,521	27,582,722	NA
2	3	*ZNF512*	rs12989678	1.30E-115	27,582,968	27,623,215	T2D, TG
2	13	*GPN1*	rs34502053	7.90E-68	27,628,647	27,650,846	TG, FG, BMI
2	17	*SUPT7L*	rs4666010	9.90E-45	27,650,809	27,663,840	NA
2	15	*SLC4A1AP*	rs13021208	1.60E−57	27,663,470	27,694,980	AD, TG
2	19	*RBKS*	rs898034	1.50E-38	27,781,379	27,890,387	TG
2	20	*MRPL33*	rs3792252	2.20E-37	27,771,719	27,779,733	TG, FG
7	12	*FZD9*	rs1178947	5.80E-69	73,433,778	73,436,120	TG, HDL-C
7	11	*BAZ1B*	rs111837003	1.20E-71	73,440,406	73,522,293	TG, HDL-C
7	9	*BCL7B*	rs7793710	1.10E-83	73,536,352	73,557,735	HDL-C
7	8	*TBL2*	rs13246490	2.60E-84	73,568,945	73,578,683	TG, HDL-C, BMI
11	1	*BUD13*	rs6589565	7.3E-445	116,748,173	116,772,987	TG, HDL-C
11	2	*APOC3*	rs5128	2.80E-152	116,829,907	116,833,071	TG, HDL-C
11	18	*APOA1*	rs12718464	3.10E-39	116,835,752	116,837,622	TG, HDL-C, BMI
11	10	*PAFAH1B2*	rs7925256	8.00E-75	117,144,283	117,178,173	TG, HDL-C, BMI
11	14	*SIDT2*	rs6589603	1.10E−59	117,178,743	117,197,442	TG, HDL-C
11	16	*TAGLN*	rs588534	3.40E−51	117,199,294	117,207,465	TG, HDL-C
19	6	*NECTIN2*	rs283811	1.60E-94	44,846,135	44,889,228	AD, T2D, TG, HDL-C, BMI
19	7	*TOMM40*	rs157582	7.50E-94	44,891,219	44,903,689	AD, TG, HDL-C, BMI
19	5	*APOE*	rs440446	1.70E-106	44,905,796	44,909,393	AD, T2D, TG, HDL-C, BMI

AD Alzheimer's disease, *b38* Genome Research Consortium human build 38, *BMI* body mass index, *CHR* chromosome, *FG* fasting glucose, *HDL-C* high-density lipoprotein cholesterol, *IR* insulin resistance, *NA* not available, *Pos* position, *T2D* type 2 diabetes, *TAD* topologically associating domains, *TG* triglycerides.
Previous GWASs reported in the NHGRI-EBI GWAS Catalog have identified various phenotypes that reach genome-wide significance ($P < 5 \times 10^{-8}$). This GWAS analysis utilized BOLT-LMM's mixed linear models with a two-sided chi-square test. The conventional genome-wide significance threshold of $P < 5 \times 10^{-8}$ was applied.

include such data to explore medication's influence on genetic associations with metabolic traits.

Methods

Taiwan Biobank

The study cohort comprised 147,880 Taiwanese subjects from the Taiwan Biobank[63–69]. Detailed inclusion and exclusion criteria[26] are presented in Supplementary Methods. Ethical approval for the study was obtained from the Institutional Review Board of Taipei Veterans General Hospital (approval number: 2023-04-007CC#1). All participants provided informed consent in accordance with established guidelines and regulations.

Genotyping and imputation

Stringent quality control measures were applied to ensure reliable SNP analysis[66,70]. SNPs were excluded if they deviated from Hardy-Weinberg equilibrium ($P < 1 \times 10^{-6}$), exhibited a genotyping call rate below 95%, or a minor allele frequency less than 1%. From an initial set of 686,370 directly genotyped and 15,851,039 imputed SNPs, 7,604,854 SNPs remained after quality control and were used for functional prediction.

Statistical analysis

We conducted a GWAS to investigate the genetic underpinnings of three IR surrogate markers: TyG[16], TG:HDLC[13], and log(TG:HDLC). The TyG index was derived using the formula Ln[fasting triglycerides (mg/dL)×fasting glucose (mg/dL)/2]. The GWAS analysis employed mixed linear models in BOLT-LMM (version 2.4.1)[71], adjusting for sex and the top 10 principal components (PCs) to account for potential confounding factors. Manhattan and quantile-quantile (Q-Q) plots were generated utilizing the R package 'qqman'. The study used a genome-wide significance threshold of $P < 5 \times 10^{-8}$.

To identify independent genetic variants, we employed the clumping method in PLINK[70], which accounts for linkage disequilibrium (LD) patterns in the genome. We performed fine mapping using susieR (v.0.12.35)[21], which includes two metrics: Posterior Inclusion Probability (PIP) and Residual Sum of Squares Posterior Inclusion Probability (RSS PIP). We estimated heritability and genetic correlations using LD Score Regression (LDSC) v.1.0.1[20]. We constructed PRS analysis using PRS-CS (v.1.1.0)[22]. Detailed methodologies for clumping, fine mapping, LDSC statistics, and PRS models are available in Supplementary Methods.

Sex-stratified and sex-differentiated analyses

Sex-stratified analyses used mixed linear models in BOLT-LMM (version 2.4.1)[71], separately for male ($n = 48,189$) and female ($n = 88,564$) subjects, adjusting for sex and the top 10 PCs. Sex-differentiated analysis[72] utilized GWAMA[73] with the "-sex" option to identify sex-specific allelic effects and test for heterogeneity between sexes.

Transcriptome-wide association studies

We conducted TWAS to investigate the relationship between gene expression and TyG. Using FUSION[24], we combined gene expression measurements with GWAS summary statistics. We obtained pre-computed expression reference weights from GTEx(v7) for 11 human tissues, including various adipose, brain, and organ tissues. FUSION was then used to calculate TWAS *P*-values, identifying genes with significant associations between their expression levels and TyG susceptibility.

Pathway analysis

We constructed protein-protein interaction (PPI) networks utilizing NDEx-The Network Data Exchange[74] with significant genes associated with TyG. The Human Integrated Protein–Protein Interaction Reference database[75] was employed for PPI analysis within NDEx. We then conducted pathway analysis on the genes within the PPI network using ClueGO[25], a Cytoscape[76] plugin. ClueGO leverages various ontology source databases to identify gene ontology (GO) terms and

https://doi.org/10.1038/s41467-025-58506-x

Table 7 | The association of disease phenotypes with the TyG index, TG:HDL ratio, PRS of the TyG index, and PRS of the TG:HDL ratio

		TyG index			TG:HDL-C			PRS of TyG index			PRS of TG:HDL-C		
Disease	**N (%)**	**Case**	**Control**	***P***	**Case**	**Control**	***P***	**Case**	**Control**	***P***	**Case**	**Control**	***P***
Osteoporosis	1635 (4.01)	8.43 (0.52)	8.38 (0.56)	**0.0002**	2.11 (1.52)	2.22 (1.68)	0.0058	−0.04 (1.01)	0.00 (1.00)	0.1199	−0.03 (0.99)	0.00 (1.00)	0.1499
Arthritis	1973 (4.84)	8.47 (0.52)	8.38 (0.56)	**<0.0001**	2.23 (1.54)	2.22 (1.69)	0.7792	−0.01 (0.99)	0.00 (1.00)	0.6785	−0.01 (0.97)	0.00 (1.00)	0.5976
Gout	1458 (3.58)	8.76 (0.53)	8.37 (0.55)	**<0.0001**	3.46 (2.23)	2.17 (1.64)	**<0.0001**	0.08 (0.98)	−0.00 (1.00)	**0.0016**	0.08 (0.98)	−0.00 (1.00)	**0.001**
Asthma	1519 (3.73)	8.38 (0.55)	8.38 (0.56)	0.9077	2.22 (1.66)	2.22 (1.68)	0.899	−0.00 (0.99)	0.00 (1.00)	0.9353	−0.02 (0.98)	0.00 (1.00)	0.4724
Emphysema bronchitis	499 (1.22)	8.43 (0.53)	8.38 (0.56)	0.0595	2.27 (1.76)	2.22 (1.68)	0.469	−0.03 (0.99)	0.00 (1.00)	0.4625	−0.04 (0.96)	0.00 (1.00)	0.383
Coronary heart disease	4087 (10.03)	8.41 (0.54)	8.38 (0.56)	**<0.0001**	2.23 (1.64)	2.21 (1.68)	0.5341	0.00 (0.99)	−0.00 (1.00)	0.9633	−0.02 (1.00)	0.00 (1.00)	0.2719
Valve heart disease	1827 (4.48)	8.31 (0.53)	8.38 (0.56)	**<0.0001**	1.98 (1.51)	2.23 (1.69)	**<0.0001**	−0.01 (1.01)	0.00 (1.00)	0.6366	−0.03 (1.01)	0.00 (1.00)	0.2179
Coronary Artery Disease	524 (1.29)	8.66 (0.52)	8.38 (0.56)	**<0.0001**	2.89 (1.91)	2.21 (1.67)	**<0.0001**	0.13 (0.97)	−0.00 (1.00)	0.0035	0.09 (0.96)	−0.00 (1.00)	0.0454
Arrhythmia	1836 (4.50)	8.43 (0.53)	8.38 (0.56)	**0.0001**	2.21 (1.54)	2.22 (1.68)	0.9548	−0.02 (0.99)	0.00 (1.00)	0.323	−0.04 (1.00)	0.00 (1.00)	0.106
Cardiomyopathy	334 (0.82)	8.62 (0.55)	8.38 (0.56)	**<0.0001**	2.84 (2.03)	2.21 (1.67)	**<0.0001**	−0.06 (1.04)	0.00 (1.00)	0.2711	−0.05 (1.03)	0.00 (1.00)	0.3761
Hyperlipidemia	2891 (7.09)	8.73 (0.54)	8.35 (0.55)	**<0.0001**	3.06 (2.08)	2.15 (1.63)	**<0.0001**	0.24 (1.04)	−0.02 (0.99)	**<0.0001**	0.21 (1.05)	−0.02 (0.99)	**<0.0001**
Hypertension	4684 (11.49)	8.68 (0.51)	8.34 (0.55)	**<0.0001**	2.92 (1.90)	2.12 (1.63)	**<0.0001**	0.04 (1.01)	−0.00 (1.00)	0.0058	0.03 (1.00)	−0.00 (1.00)	0.0194
Apoplexia	271 (0.66)	8.53 (0.55)	8.38 (0.56)	**<0.0001**	2.64 (1.85)	2.21 (1.68)	**0.0002**	−0.03 (1.07)	0.00 (1.00)	0.6692	−0.01 (1.06)	0.00 (1.00)	0.8817
Diabetes	1683 (4.13)	8.83 (0.52)	8.36 (0.55)	**<0.0001**	3.03 (1.93)	2.18 (1.66)	**<0.0001**	0.06 (0.99)	−0.00 (1.00)	0.0137	0.05 (1.00)	−0.00 (1.00)	0.0424
Gastrointestinal	9500 (23.31)	8.42 (0.55)	8.37 (0.56)	**<0.0001**	2.26 (1.69)	2.20 (1.67)	**0.0015**	0.00 (0.99)	−0.00 (1.00)	0.7086	−0.01 (1.00)	0.00 (1.00)	0.427
Peptic ulcer	5859 (14.38)	8.43 (0.54)	8.37 (0.56)	**<0.0001**	2.27 (1.68)	2.21 (1.68)	0.004	0.01 (1.00)	−0.00 (1.00)	0.4851	−0.00 (1.00)	0.00 (1.00)	0.422
Gastroesophageal reflux	5674 (13.92)	8.41 (0.55)	8.37 (0.56)	**<0.0001**	2.25 (1.67)	2.21 (1.68)	0.117	−0.01 (1.00)	0.00 (1.00)	0.2397	−0.02 (1.01)	0.00 (1.00)	0.1207
Irritable bowel syndrome	1034 (2.54)	8.40 (0.56)	8.38 (0.56)	0.2199	2.25 (1.69)	2.22 (1.68)	0.5202	−0.04 (0.98)	0.00 (1.00)	0.1835	−0.03 (1.00)	0.00 (1.00)	0.283
Mental disease	1810 (4.44)	8.42 (0.56)	8.38 (0.56)	**0.0006**	2.29 (1.71)	2.21 (1.68)	0.067	0.01 (1.01)	−0.00 (1.00)	0.5375	0.02 (1.01)	−0.00 (1.00)	0.4359
Depression	1460 (3.58)	8.42 (0.56)	8.38 (0.56)	0.0025	2.27 (1.73)	2.21 (1.68)	0.1848	−0.01 (1.01)	0.00 (1.00)	0.7788	0.00 (1.01)	−0.00 (1.00)	0.9474
Manic depression	277 (0.68)	8.50 (0.55)	8.38 (0.56)	**0.0004**	2.47 (1.76)	2.21 (1.68)	0.011	0.17 (1.01)	−0.00 (1.00)	0.004	0.11 (1.01)	−0.00(1.00)	0.0676
Nervous system disease	1360 (3.34)	8.35 (0.56)	8.38 (0.56)	0.0785	2.11 (1.65)	2.22 (1.68)	0.017	−0.02 (1.01)	0.00 (1.00)	0.3977	−0.03 (1.00)	0.00 (1.00)	0.3298
Hemicrania	1174 (2.88)	8.34 (0.56)	8.38 (0.56)	0.0115	2.06 (1.61)	2.22 (1.68)	**0.0009**	−0.02 (1.02)	0.00 (1.00)	0.3922	−0.03 (1.00)	0.00 (1.00)	0.2331
Stone	4172 (0.24)	8.53 (0.54)	8.36 (0.56)	**<0.0001**	2.58 (1.81)	2.17 (1.66)	**<0.0001**	0.00 (0.99)	−0.00 (1.00)	0.9289	−0.00 (1.00)	0.00 (1.00)	0.8771
Liver gall stone	1871 (4.59)	8.5 (0.53)	8.37 (0.56)	**<0.0001**	2.46 (1.7)	2.2 (1.68)	**<0.0001**	0.02 (0.99)	−0.00 (1.00)	0.476	0.00 (0.99)	−0.00 (1.00)	0.8524
Kidney stone	2490 (6.11)	8.55 (0.55)	8.37 (0.56)	**<0.0001**	2.69 (1.88)	2.19 (1.66)	**<0.0001**	−0.00 (0.99)	0.00 (1.00)	0.9761	0.00 (1.00)	−0.00 (1.00)	0.9619
Vertigo	2524 (6.19)	8.39 (0.53)	8.38 (0.56)	0.2181	2.12 (1.53)	2.22 (1.69)	**0.001**	0.03 (0.99)	−0.00 (1.00)	0.1313	0.02 (0.99)	−0.00 (1.00)	0.2213

PRS polygenic risk score, *TG:HDL-C* the ratio of triglycerides to high-density lipoprotein cholesterol *TyG index* the triglyceride-glucose index. The TyG index was derived using the formula Ln[fasting triglycerides (mg/dL) × fasting glucose (mg/dL)/2]. Significant *p* values, denoted in bold, indicate statistical significance after applying the Bonferroni correction threshold of $p = 0.00185$ (obtained by dividing the conventional significance level of 0.05 by the number of tests conducted, which in this case is 27; 0.05/27 = 0.00185). Polygenic risk scores were generated with PRS-CS using Bayesian regression; subsequent two-sided linear regression tested associations.

Table 8 | The summary of the pathway analysis results for the whole cohort, sex-specific cohorts, and genes from previous studies

Cohort	Whole	Female	Male	Previous
Adherens junction	5.31E-13	2.59E-02	6.39E-04	4.65E-14
Alzheimer's disease	2.24E-26	1.70E-03	2.16E-02	3.25E-23
Cell cycle	1.80E-39	9.35E-06	1.18E-10	3.64E-37
Estrogen signaling pathway	6.79E-10	2.09E-11	-	8.21E-10
Glucose metabolism	3.95E-05	4.39E-02	-	2.31E-03
Insulin resistance	4.76E-10	-	-	8.07E-16
Insulin signaling pathway	8.09E-14	-	-	7.07E-22
Insulin secretion	-	1.92E-02	-	-
Neurodegeneration	1.91E-25	2.64E-06	7.79E-03	3.41E-23
Notch signaling	3.99E-03	9.98E-03	1.91E-02	1.50E-05
Nervous system development	2.49E-29	8.72E-11	7.29E-05	2.41E-43
Plasma lipoprotein assembly, remodeling, and clearance	4.64E-06	3.59E-02	-	1.90E-03
TGF-beta signaling pathway	4.92E-24	8.54E-09	3.84E-05	8.47E-24
Tight junction	2.36E-07	-	4.02E-03	4.78E-11
Wnt signaling pathway	1.30E-06	5.13E-06	6.74E-04	2.02E-08

pathway networks, elucidating the functional significance of identified genes within specific biological processes and pathways[25].

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The summary statistics data generated in this study are available in the Supplementary Information file. The Taiwan Biobank genetic data are subject to controlled access due to privacy policy requirements. Access can be obtained by submitting a formal request to the Taiwan Biobank at biobank@gate.sinica.edu.tw, with responses typically provided within 30 days, subject to approval. Approved users must adhere to data use agreements restricting secondary distribution and non-research use. Source data are provided with this paper.

References

1. Petersen, M. C. & Shulman, G. I. Mechanisms of insulin action and insulin resistance. *Physiol. Rev.* **98**, 2133–2223 (2018).
2. Thomas, D. D., Corkey, B. E., Istfan, N. W. & Apovian, C. M. Hyperinsulinemia: an early indicator of metabolic dysfunction. *J. Endocr. Soc.* **3**, 1727–1747 (2019).
3. Gluvic, Z. et al. Link between metabolic syndrome and insulin resistance. *Curr. Vasc. Pharmacol.* **15**, 30–39 (2017).
4. Yaribeygi, H., Farrokhi, F. R., Butler, A. E. & Sahebkar, A. Insulin resistance: review of the underlying molecular mechanisms. *J. Cell. Physiol.* **234**, 8152–8161 (2019).
5. White, M. F. & Kahn, C. R. Insulin action at a molecular level–100 years of progress. *Mol. Metab.* **52**, 101304 (2021).
6. Lee, S.-H., Park, S.-Y. & Choi, C. S. Insulin resistance: from mechanisms to therapeutic strategies. *Diabetes Metab. J.* **46**, 15 (2022).
7. DeFronzo, R. A., Tobin, J. D. & Andres, R. Glucose clamp technique: a method for quantifying insulin secretion and resistance. *Am. J. Physiol. Endocrinol. Metab.* **237**, E214 (1979).
8. Matthews, D. R. et al. Homeostasis model assessment: insulin resistance and β-cell function from fasting plasma glucose and insulin concentrations in man. *Diabetologia* **28**, 412–419 (1985).
9. Chiang, J.-K., Lai, N.-S., Chang, J.-K. & Koo, M. Predicting insulin resistance using the triglyceride-to-high-density lipoprotein cholesterol ratio in Taiwanese adults. *Cardiovasc. Diabetol.* **10**, 1–6 (2011).
10. Gong, R. et al. Associations between TG/HDL ratio and insulin resistance in the US population: a cross-sectional study. *Endocr. Connect.* **10**, 1502–1512 (2021).
11. Nur Zati Iwani, A. K. et al. TG: HDL-C ratio is a good marker to identify children affected by obesity with increased cardiometabolic risk and insulin resistance. *Int. J. Endocrinol.* **2019**, 8586167 (2019).
12. McLaughlin, T. et al. Use of metabolic markers to identify overweight individuals who are insulin resistant. *Ann. Intern. Med.* **139**, 802–809 (2003).
13. Oliveri, A. et al. Comprehensive genetic study of the insulin resistance marker TG: HDL-C in the UK Biobank. *Nat. Genet.* **56**, 212–221 (2024).
14. Wojcik, G. L. et al. Genetic analyses of diverse populations improves discovery for complex traits. *Nature* **570**, 514–518 (2019).
15. Guerrero-Romero, F. et al. The product of triglycerides and glucose, a simple measure of insulin sensitivity. Comparison with the euglycemic-hyperinsulinemic clamp. *J. Clin. Endocrinol. Metab.* **95**, 3347–3351 (2010).
16. Simental-Mendía, L. E., Rodríguez-Morán, M. & Guerrero-Romero, F. The product of fasting glucose and triglycerides as surrogate for identifying insulin resistance in apparently healthy subjects. *Metab. Syndr. Relat. Disord.* **6**, 299–304 (2008).
17. Ober, C., Loisel, D. A. & Gilad, Y. Sex-specific genetic architecture of human disease. *Nat. Rev. Genet.* **9**, 911–922 (2008).
18. Spracklen, C. N. et al. Identification of type 2 diabetes loci in 433,540 East Asian individuals. *Nature* **582**, 240–245 (2020).
19. Buniello, A. et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.* **47**, D1005–D1012 (2018).
20. Bulik-Sullivan, B. K. et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
21. Wang, G., Sarkar, A., Carbonetto, P. & Stephens, M. A simple new approach to variable selection in regression, with application to genetic fine mapping. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **82**, 1273–1300 (2020).
22. Ge, T., Chen, C.-Y., Ni, Y., Feng, Y.-C. A. & Smoller, J. W. Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nat. Commun.* **10**, 1776 (2019).
23. Graham, B. E., Plotkin, B., Muglia, L., Moore, J. H. & Williams, S. M. Estimating prevalence of human traits among populations from polygenic risk scores. *Hum. Genom.* **15**, 1–16 (2021).
24. Gusev, A. et al. Integrative approaches for large-scale transcriptome-wide association studies. *Nat. Genet.* **48**, 245–252 (2016).
25. Bindea, G. et al. ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics* **25**, 1091–1093 (2009).
26. Lin, E. et al. Genome-wide association study in the Taiwan Biobank identifies four novel genes for human height: NABP2, RA SA2, RNF41, and SLC39A5. *Hum. Mol. Genet.* **30**, 2362–2369 (2021).
27. Lamina, C. et al. A genome-wide association meta-analysis on apolipoprotein A-IV concentrations. *Hum. Mol. Genet.* **25**, 3635–3646 (2016).
28. Sinnott-Armstrong, N. et al. Genetics of 35 blood and urine biomarkers in the UK Biobank. *Nat. Genet.* **53**, 185–194 (2021).
29. Koskeridis, F. et al. Pleiotropic genetic architecture and novel loci for C-reactive protein levels. *Nat. Commun.* **13**, 6939 (2022).
30. Lee, S.-B. et al. Dyslipidaemia—genotype interactions with nutrient intake and cerebro-cardiovascular disease. *Biomedicines* **10**, 1615 (2022).

31. Hoffmann, T. J. et al. A large electronic-health-record-based genome-wide study of serum lipids. *Nat. Genet.* **50**, 401–413 (2018).
32. Dupuis, J. et al. New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nat. Genet.* **42**, 105–116 (2010).
33. Park, S. Association of polygenic risk scores for insulin resistance risk and their interaction with a plant-based diet, especially fruits, vitamin C, and flavonoid intake, in Asian adults. *Nutrition* **111**, 112007 (2023).
34. da Costa-Nunes, J. A. & Noordermeer, D. TADs: dynamic structures to create stable regulatory functions. *Curr. Opin. Struct. Biol.* **81**, 102622 (2023).
35. van Mierlo, G., Pushkarev, O., Kribelbauer, J. F. & Deplancke, B. Chromatin modules and their implication in genomic organization and gene regulation. *Trends Genet.* **39**, 140–153 (2023).
36. Trusca, V. G. et al. Macrophage-specific up-regulation of apolipoprotein E gene expression by STAT1 is achieved via long range genomic interactions. *J. Biol. Chem.* **286**, 13891–13904 (2011).
37. Zannis, V. I., Kan, H.-Y., Kritis, A., Zanni, E. & Kardassis, D. Transcriptional regulation of the human apolipoprotein genes. *Front. Biosci.* **6**, d456–d504 (2001).
38. Mahajan, A. et al. Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. *Nat. Genet.* **50**, 1505–1513 (2018).
39. Pulit, S. L. et al. Meta-analysis of genome-wide association studies for body fat distribution in 694,649 individuals of European ancestry. *Hum. Mol. Genet.* **28**, 166–174 (2019).
40. Shungin, D. et al. New genetic loci link adipose and insulin biology to body fat distribution. *Nature* **518**, 187–196 (2015).
41. Teslovich, T. M. et al. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* **466**, 707–713 (2010).
42. Yang, Q. et al. Adipocyte-specific modulation of KLF14 expression in mice leads to sex-dependent impacts on adiposity and lipid metabolism. *Diabetes* **71**, 677–693 (2022).
43. Small, K. S. et al. Regulatory variants at KLF14 influence type 2 diabetes risk via a female-specific effect on adipocyte size and body composition. *Nat. Genet.* **50**, 572–580 (2018).
44. Yang, Q. & Civelek, M. Transcription factor KLF14 and metabolic syndrome. *Front. Cardiovasc. Med.* **7**, 91 (2020).
45. Angelov, A., Connelly, P. J., Delles, C. & Kararigas, G. Sex-biased and sex hormone-dependent regulation of apolipoprotein A1. *Curr. Opin. Physiol.* **33**, 100654 (2023).
46. Bai, W. et al. Functional polymorphisms of the APOA1/C3/A4/A5-ZPR1-BUD13 gene cluster are associated with dyslipidemia in a sex-specific pattern. *PeerJ* **6**, e6175 (2019).
47. Lee, G., Jeon, H. K. & Yoo, H. Y. Sex-related differences in single nucleotide polymorphisms associated with dyslipidemia in a Korean population. *Lipids Health Dis.* **21**, 124 (2022).
48. Wang, Y. E. et al. SNPs in apolipoproteins contribute to sex-dependent differences in blood lipids before and after a high-fat dietary challenge in healthy US adults. *BMC Nutr.* **8**, 95 (2022).
49. Hofmeister, R. J. et al. Parent-of-Origin inference for Biobanks. *Nat. Commun.* **13**, 6668 (2022).
50. So-Youn 2 Richards Hannah B 7 the GIANT Consortium MCSKSHÅKGENACTGKATUS, Investigators M, Consortium D. et al. Identification of an imprinted master trans regulator at the KLF14 locus related to multiple metabolic phenotypes. *Nat. Genet.* **43**, 561–564 (2011).
51. Dierssen, M. & Barone, E. Brain insulin resistance in neurodevelopmental and neurodegenerative disorders: mind the gap!. Front. Neurosci. **15**, 730378 (2021).
52. Kshirsagar, V., Thingore, C. & Juvekar, A. Insulin resistance: a connecting link between Alzheimer's disease and metabolic disorder. *Metab. Brain Dis.* **36**, 67–83 (2021).
53. Ristow, M. Neurodegenerative disorders associated with diabetes mellitus. *J. Mol. Med.* **82**, 510–529 (2004).
54. Tecalco-Cruz, A. C., López-Canovas, L. & Azuara-Liceaga, E. Estrogen signaling via estrogen receptor alpha and its implications for neurodegeneration associated with Alzheimer's disease in aging women. *Metab. Brain Dis.* **38**, 783–793 (2023).
55. Harhaj, N. S. & Antonetti, D. A. Regulation of tight junctions and loss of barrier function in pathophysiology. *Int. J. Biochem. Cell Biol.* **36**, 1206–1237 (2004).
56. Shingleton, A. W. & Vea, I. M. Sex-specific regulation of development, growth and metabolism. Semin. Cell Dev. Biol. **138**, 117–127 (2023).
57. Tramunt, B. et al. Sex differences in metabolic regulation and diabetes susceptibility. *Diabetologia* **63**, 453–461 (2020).
58. Sadeghi, I., Rodríguez-Fernández, B., Navarro, A., Gispert, J. D. & Vilor-Tejedor, N. Understanding sex-specific molecular mechanisms in preclinical and Alzheimer's disease through brain proteomics and transcriptomics. *Alzheimer's Dement.* **19**, e076194 (2023).
59. Silveira, P. P., Pokhvisneva, I., Howard, D. M. & Meaney, M. J. A sex-specific genome-wide association study of depression phenotypes in UK Biobank. *Mol. Psychiatry* **28**, 2469–2479 (2023).
60. Bernstein, S. R., Kelleher, C. & Khalil, R. A. Gender-based research underscores sex differences in biological processes, clinical disorders and pharmacological interventions. *Biochem. Pharmacol.* **215**, 115737 (2023).
61. Garratt, M. Why do sexes differ in lifespan extension? Sex-specific pathways of aging and underlying mechanisms for dimorphic responses. *Nutr. Healthy Aging* **5**, 247–259 (2020).
62. Bourquard, T. et al. Functional variants identify sex-specific genes and pathways in Alzheimer's disease. *Nat. Commun.* **14**, 2765 (2023).
63. Chen, C. H. et al. Population structure of Han Chinese in the modern Taiwanese population based on 10,000 participants in the Taiwan Biobank project. *Hum. Mol. Genet.* **25**, 5321–5331 (2016).
64. Fan, C. T., Lin, J. C. & Lee, C. H. Taiwan Biobank: a project aiming to aid Taiwan's transition into a biomedical island. *Pharmacogenomics* **9**, 235–246 (2008).
65. Hou, S.-J. et al. An association study in the Taiwan Biobank reveals RORA as a novel locus for sleep duration in the Taiwanese Population. *Sleep Med.* **73**, 70–75 (2020).
66. Lin, E. et al. Association and interaction of APOA5, BUD13, CETP, LIPA, and health-related behavior with metabolic syndrome in a Taiwanese population. *Sci. Rep.* **6**, 36830 (2016).
67. Lin, E. et al. Effects of circadian clock genes and health-related behavior on metabolic syndrome in a Taiwanese population: evidence from association and interaction analysis. *PLoS ONE* **12**, e0173861 (2017).
68. Lin, E. et al. The ADAMTS9 gene is associated with cognitive aging in the elderly in a Taiwanese population. *PLoS ONE* **12**, e0172440 (2017).
69. Lin, E. et al. The rs1277306 variant of the REST gene confers susceptibility to cognitive aging in an elderly Taiwanese population. *Dement Geriatr. Cogn. Disord.* **43**, 119–127 (2017).
70. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet* **81**, 559–575 (2007).
71. Loh, P.-R. et al. Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat. Genet.* **47**, 284–290 (2015).
72. Magi, R., Lindgren, C. M. & Morris, A. P. Meta-analysis of sex-specific genome-wide association studies. *Genet. Epidemiol.* **34**, 846–853 (2010).
73. Mägi, R. & Morris, A. P. GWAMA: software for genome-wide association meta-analysis. *BMC Bioinforma.* **11**, 1–6 (2010).
74. Pratt, D. et al. NDEx, the network data exchange. *Cell Syst.* **1**, 302–305 (2015).

75. Alanis-Lobato, G., Andrade-Navarro, M. A. & Schaefer, M. H. HIPPIE v2. 0: enhancing meaningfulness and reliability of protein–protein interaction networks. *Nucleic Acids Res.* **45**, D408–D414 (2016).
76. Shannon, P. et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).

Acknowledgements

We thank Lisa Stubbs of Pacific Northwest Research Institute for her valuable suggestions and thorough proofreading of our manuscript. This work was supported in part by grants from the National Science and Technology Council (Taiwan) (grant NSTC112-2314-B-075-013-MY3, NSTC112-2321-B-A49-021, NSTC112-2321-B-A49-013; SJT) and from the Taipei Veterans General Hospital (grant VTA112-V1-8-1, V112C-008; SJT). This study was supported by the Population Health Research Center from the Featured Areas Research Center Program within the framework of the Higher Education Sprout Project by the Ministry of Education in Taiwan (grant number NTU-112L9004).

Author contributions

E.L. had a role in study conception and design, analysis and interpretation of data, draft manuscript. Y.T.Y. had a role in analysis and interpretation of data. P.H.K. had a role in study conception and design. M.H.C. and A.C.Y. had a role in acquisition of data. S.J.T. had a role in study conception and design. The authors read and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-025-58506-x.

Correspondence and requests for materials should be addressed to Po-Hsiu Kuo or Shih-Jen Tsai.

Peer review information *Nature Communications* thanks Yi-Cheng Chang and the other anonymous reviewer(s) for their contribution to the peer review of this work. A peer review file is available.