Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

# Gene-based map of cardiometabolic diseases

**Barış Kayaalp**
Bilkent University https://orcid.org/0009-0006-6741-5966
**Kerem Çil**
Bilkent University https://orcid.org/0009-0002-8851-1667
**Onur Emre Onat**
Bezmialem Vakıf University https://orcid.org/0000-0002-7105-1572
**Tayfun Özçelik**
tozcelik@bilkent.edu.tr
Bilkent University https://orcid.org/0000-0001-5937-1082

**Research Article**

**Keywords:**

**Posted Date:** July 16th, 2026

**DOI:** https://doi.org/10.21203/rs.3.rs-10366105/v1

**License:** This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

**Additional Declarations:** The authors declare no competing interests.

## Abstract

Cardiometabolic disorders are common, interrelated conditions with substantial genetic contribution. We performed sex- and ancestry-stratified rare-variant gene-burden analyses in UK Biobank across obesity, fatty liver disease, dyslipidemia, hypertension, chronic kidney disease, and diabetes mellitus, with Turkish Variome contributing to obesity analyses, using state-of-the-art variant classification tools. This approach replicated established gene–phenotype relationships and expanded the number of causal cardiometabolic gene–phenotype associations from 368 to 534. Complementary genetic-epidemiological analyses showed that approximately one in six individuals carries risk-increasing alleles and one in nine carries protective alleles.


## Main

Cardiometabolic disorders, affect more than a quarter of the adult population[1]. These phenotypes can be viewed as a constellation of physiologically connected traits that may occur independently but can also cluster as part of a broader metabolic syndrome, historically referred to as syndrome X. This clustering reflects interrelated pathways involving appetite regulation, lipid and glucose metabolism, blood pressure control, adiposity, and kidney function[2–4].

Human genetics has yielded major causal insights into cardiometabolic biology. Classical examples include *LDLR*-mediated lipoprotein clearance in familial hypercholesterolemia[5], imprinted *SNRPN*-region Prader–Willi pathogenesis[6] or the leptin–melanocortin pathway in severe obesity[7], monogenic blood-pressure disorders affecting mineralocorticoid signaling and renal ion transport[8], and monogenic diabetes genes involved in glucose sensing and pancreatic beta-cell function[9]. Yet canonical Mendelian genes, syndromes, extreme phenotypes and family-based discoveries explain only part of inherited susceptibility.

Population-scale sequencing now enables systematic assessment of rare germline variation in adult-onset cardiometabolic disease. Unlike fully penetrant Mendelian variants, many rare alleles contributing to common cardiometabolic phenotypes are expected to show incomplete penetrance, variable expressivity, or effects that become detectable only in specific biological contexts. This creates two major challenges. First, rare-variant burden testing depends strongly on how functionally relevant variants are selected[10,11]. Second, pooled analyses may obscure associations whose effects differ by the genetic sex[12].

Sex is an especially relevant context for cardiometabolic genetics because females and males differ in disease prevalence, presentation, hormonal regulation, fat distribution, vascular biology, renal physiology, and metabolic complications[13,14]. Similarly, cross-ancestry analyses can improve discovery by capturing differences in allele frequencies, variant spectra, and genetic architecture across populations[15]. These considerations suggest that sex- and ancestry-stratified rare-variant analyses may reveal associations that are diluted or missed in conventional pooled analyses.

Here, we performed sex- and ancestry-stratified rare-variant gene-burden analyses of cardiometabolic disorders in the UK Biobank[16], with additional obesity analyses incorporating the Turkish Variome[17]. We used AAVC[18] for ACMG-based variant classification and FuncVEP[10] for missense variant prioritization, and tested high-impact, ACMG-classified, and damaging missense variant masks across multiple allele-frequency thresholds. This

strategy replicated established cardiometabolic gene–phenotype relationships and identified novel risk-increasing, protective, and sex-specific associations. Lastly, we used genetic-epidemiological analyses to quantify the population-level burden of these variants.

## Results

We conducted gene-burden meta-analyses across ancestry and genetic-sex strata in 483,311 UK Biobank whole-genome sequencing participants; 981 obesity patients from the Turkish Variome were included only in obesity analyses. Binary and continuous phenotypes spanned obesity, fatty liver disease, dyslipidemia, hypertension, chronic kidney disease, and diabetes mellitus (Supplementary Table 1).

Association testing was performed with REGENIE[19] to address both polygenic estimates from common variants and cryptic relatedness, and meta-analysis was conducted using remeta[20]. To define replicated associations, we curated known cardiometabolic disease genes from the literature, PanelApp[21], and OMIM (https://www.omim.org/), including 51 obesity genes, 48 dyslipidemia genes, 28 hypertension genes, 160 chronic kidney disease genes, and 81 diabetes genes (Supplementary Table 2). We identified 260 gene–phenotype associations reaching genome-wide significance ($p < 2.6 \times 10^{-6}$, 0.05/18,821), of which 94 (36%) replicated previously reported associations and 166 (64%) were novel (Supplementary Information: Replication rate differences). We used independent GWAS hits, previous linkage evidence and recapitulation of the phenotype in mouse models as external supporting evidence. Of the 166 novel associations, 74 (45%) had at least one line of supporting evidence (Fig. 1; Supplementary Tables 3, 4 and 5).

# Sex- and ancestry-stratification improve association power

To evaluate the contribution of multi-ancestry meta-analysis and sex-stratified testing, we compared association yields across alternative cohort definitions using the summary statistics from our gene-burden analyses. Among the 94 significant associations corresponding to previously reported findings, only 8 were identified when analysis was restricted to European-ancestry individuals without sex stratification. This number increased modestly to 10 with sex-stratified analysis within Europeans, and to 82 when all ancestry groups were included in the meta-analysis. Similarly, among the 166 novel associations, only 9 were detected without ancestry or sex stratification, compared with 28 after sex stratification and 74 after inclusion of all ancestry groups (Supplementary Table 6; Supplementary Information: Sex-specific associations).

## Identification of risk and protective genes for obesity and fatty liver disease

We focused on body mass index (BMI) and “Obesity” phecode for obesity-related phenotypes. We replicated 20 of the 51 (39%) previously reported genes, including 18 associated with increased BMI or obesity risk, such as *MC4R, PCSK1, NTRK2, DIDO1*, and *TNK2*, and 2 associated with lower BMI or protection, *GPR151* and *GIPR*. In addition, we identified 50 novel associations, 25 of which were supported by prior GWAS or linkage evidence, or by animal models (Fig. 2A and 2B; Supplementary Table 4).

Among the novel findings, AAVC-classified P/LP/VUS-H variants in *GNAZ* (G protein subunit alpha z) were identified in 70 individuals and were associated with higher BMI (β = 2.63 kg/m², p = 1.31 x $10^{-6}$, lambda = 1.03). *GNAZ* is predominantly expressed in nervous tissue and the pituitary gland[22]. Notably, distal 22q11.2 deletions encompassing *GNAZ* have been reported in severe early-onset obesity[23], and knockout mice show increased total body fat in the MGI database[24], supporting a possible role in body-weight regulation.

*ATP1A2* (ATPase Na+/K+ transporting subunit alpha 2) also emerged as a novel obesity-associated gene. Monoallelic gain-of-function variants in *ATP1A2* are known to cause familial hemiplegic migraine type 2 (OMIM: 104290), whereas biallelic loss-of-function variants cause fetal akinesia (OMIM: 619602). In our analysis, AAVC-classified P/LP/VUS-H variants in *ATP1A2* were present in 550 individuals in monoallelic form and were associated with a 0.99 kg/m² higher BMI (p = 3.63 x $10^{-7}$, lambda = 1.03). This observation is consistent with prior mouse model data showing obesity at older ages[25].

*HCRTR2* encodes hypocretin receptor 2 and is primarily expressed in the hypothalamus[22]. Damaging missense variants in *HCRTR2* were identified in 393 individuals and were associated with higher BMI (β = 1.20 kg/m², p = 8.17 x $10^{-7}$, lambda = 1.03). *HCRTR2* has also been implicated in obesity in a GWAS from the Million Veteran Program[26].

*RUVBL2* encodes RuvB-like AAA ATPase 2 and is a highly constrained gene (missense Z score = 3.52, LOEUF = 0.271)[27]. In our analysis, singleton damaging missense variants in *RUVBL2* were associated with obesity in females (OR = 5.50, p = 1.56 x $10^{-6}$, lambda = 0.96). Previous mouse studies have shown that RUVBL2-expressing neurons in the paraventricular hypothalamus exert anorexigenic effects through projections to the arcuate hypothalamus, dorsomedial hypothalamus, and parabrachial complex, and form synaptic connections with POMC and AgRP neurons in the arcuate hypothalamus. Consistent with this role, paraventricular hypothalamic knockout of *Ruvbl2* in mice resulted in hyperphagic obesity[28].

*HECTD4* (HECT domain E3 ubiquitin protein ligase 4) showed one of the broadest metabolic signatures and is primarily expressed in nervous tissue[22]. Biallelic loss-of-function variants in this gene are known to cause a neurodevelopmental disorder with seizures, spasticity, and complete or partial agenesis of the corpus callosum (OMIM: 620250). In our analysis, singleton AAVC-classified P/LP/VUS-H variants, present in 172 individuals, were associated with higher BMI (β = 2.21 kg/m², p = 3.01 x $10^{-10}$, lambda = 1.00), higher body fat percentage (β = 2.75 percentage points, p = 7.34 x $10^{-9}$, lambda = 1.00), higher glucose levels (β = 0.56 mmol/L, p = 6.49 x $10^{-7}$, lambda = 0.68), and higher HbA1c levels (β = 3.09 mmol/mol, p = 4.22 x $10^{-9}$, lambda = 0.83). Previous GWAS have also linked *HECTD4* to BMI in Korean populations[29] and to type 2 diabetes in East Asian populations[30].

Among the remaining 20 supported novel associations, 19 involved genes associated with higher BMI (*AFF3, CASTOR2, CKAP5, CPSF6, DENR, DNMT3B, DOCK1, DTNB, EI24, HIF1AN, KCNMA1, LONRF3, PCSK2, RB1CC1, TAOK2, THUMPD3, TMED2, TOX4*, and *VCF1*), whereas *COL16A1* was associated with lower BMI. Across genome-wide significant obesity associations, effect sizes generally increased with decreasing allele frequency (Fig. 2C).

In our analysis, we identified one gene associated with liver fat accumulation. Damaging missense variants in *SCYGR5* were associated with higher liver PDFF in 47 individuals (β = 4.85 percentage points, $p = 1.63 \times 10^{-6}$, lambda = 0.97) (Extended Data Fig. 1; Supplementary Table 4).

## Identification of genes associated with dyslipidemia

We used both binary dyslipidemia phenotypes and continuous lipid traits to identify genes associated with dyslipidemia. For continuous traits, we required concordant effects on multiple lipid traits. Atherogenic direction was defined as increasing LDL cholesterol, total cholesterol, triglycerides, lipoprotein(a), or apolipoprotein B while decreasing HDL cholesterol or apolipoprotein A. Protective direction was defined as the reverse pattern (Supplementary Table 7; Supplementary Information: Dyslipidemia traits).

We replicated 17 genes associated with atherogenic lipid profiles, including *LDLR*, *LPL*, and *APOA5*. In addition, we replicated 24 genes with protective lipid associations, including *APOB*, *APOC3*, and *PCSK9*, replicating 41 of the 48 (85%) known dyslipidemia-associated genes. We identified 20 novel associations, including 14 atherogenic and 6 protective signals, of which 15 were supported by at least one external line of evidence, including GWAS associations, linkage data, or animal models (Table 1; Extended Data Fig. 2; Supplementary Table 4).

Table 1
Novel dyslipidemia genes with supporting evidence

| Gene | Sex[a] | Mask[b] | N carriers | Trait | Cohort[c] | Effect | P val. | AF |
|---|---|---|---|---|---|---|---|---|
| **Protective** | | | | | | | | |
| *DNAJC13* | F | ACMG | 406 | Pure hypercholesterolemia | --+--- | OR = 0.76 | 5E-08 | 8E-03 |
| *GLTPD2* | F | ACMG | 1,017 | LDL | -+-+-- | -0.16mmol/L | 2E-09 | 2E-03 |
| | | | | Cholesterol | ---+-- | -0.2mmol/L | 4E-09 | 2E-03 |
| | | | | Apolipoprotein B | -+-+-- | -0.04g/L | 4E-07 | 2E-03 |
| *HNF1A* | F | HI | 32 | Cholesterol | ????-? | -1mmol/L | 3E-07 | 7E-05 |
| | | | | LDL | ????-? | -0.75mmol/L | 3E-07 | 7E-05 |
| | | | | Apolipoprotein B | ????-? | -0.2g/L | 1E-06 | 7E-05 |
| *SLC4A1* | B | HI | 38 | LDL | ??-?-- | -0.91mmol/L | 8E-11 | 4E-05 |
| | | | | Cholesterol | ??-?-- | -1.18mmol/L | 2E-10 | 4E-05 |
| | | | | Apolipoprotein B | ??-?-+ | -0.23g/L | 2E-09 | 4E-05 |
| *SMAD6* | B | ACMG | 1,222 | HDL | +?++++ | 0.07mmol/L | 1E-10 | 1E-03 |
| | | | | Apolipoprotein A | +?++++ | 0.05g/L | 3E-09 | 1E-03 |
| *UGT2B7* | B | ACMG | 7,221 | Cholesterol | ---+-- | -0.09mmol/L | 4E-11 | 8E-03 |
| | | | | Apolipoprotein B | ---+-+ | -0.02g/L | 4E-10 | 8E-03 |
| | | | | LDL | ---+-- | -0.06mmol/L | 4E-10 | 8E-03 |

[a], Genetic sex: B, both genetic sexes; F, female; M, male.

[b], Variant mask: ACMG, combination of ACMG-classified pathogenic, likely pathogenic, VUS-H, and damaging missense variants; HI, high-impact variants.

[c], Direction of association across genetic ancestry cohorts. “?” indicates that no individuals were observed in the corresponding cohort for the phenotype and variant mask. Cohorts are ordered as AFR, AMR, CSA, EAS, EUR, and OTH.

| Gene | Sex[a] | Mask[b] | N carriers | Trait | Cohort[c] | Effect | P val. | AF |
|---|---|---|---|---|---|---|---|---|
| | | | | Triglycerides | --+--- | -0.06mmol/L | 3E-07 | 8E-03 |
| **Atherogenic** | | | | | | | | |
| *GAS6* | B | ACMG | 262 | Apolipoprotein B | -?++++ | 0.1g/L | 5E-12 | 3E-04 |
| | | | | LDL | -?++++ | 0.33mmol/L | 2E-09 | 3E-04 |
| | | | | Cholesterol | +?++++ | 0.41mmol/L | 9E-09 | 3E-04 |
| | | | | Triglycerides | -?-+++ | 0.31mmol/L | 5E-07 | 3E-04 |
| *HRG* | F | ACMG | 14 | Hyperlipidemia | +?+++ | OR = 5.25 | 9E-07 | 3E-04 |
| *KCNN4* | B | ACMG | 365 | Pure hypercholesterolemia | +-+-++ | OR = 1.38 | 4E-07 | 2E-03 |
| *NFIC* | B | ACMG | 1,425 | Apolipoprotein A | ++---- | -0.04g/L | 7E-07 | 2E-03 |
| | | | | HDL | ++---- | -0.05mmol/L | 2E-06 | 2E-03 |
| *SMARCA4* | B | HI | 3,336 | LDL | +?+?+- | 0.07mmol/L | 2E-06 | 4E-03 |
| | | | | Apolipoprotein B | +?+?+- | 0.02g/L | 2E-06 | 4E-03 |
| *STAB1* | B | ACMG | 1,625 | Cholesterol | -++-++ | 0.18mmol/L | 2E-10 | 4E-03 |
| | | | | LDL | -++-++ | 0.12mmol/L | 1E-08 | 4E-03 |
| *TIMD4* | B | ACMG | 759 | Cholesterol | +++-++ | 0.31mmol/L | 2E-14 | 8E-04 |
| | | | | LDL | +++-++ | 0.21mmol/L | 4E-12 | 8E-04 |

[a], Genetic sex: B, both genetic sexes; F, female; M, male.

[b], Variant mask: ACMG, combination of ACMG-classified pathogenic, likely pathogenic, VUS-H, and damaging missense variants; HI, high-impact variants.

[c], Direction of association across genetic ancestry cohorts. “?” indicates that no individuals were observed in the corresponding cohort for the phenotype and variant mask. Cohorts are ordered as AFR, AMR, CSA, EAS, EUR, and OTH.

| Gene | Sex[a] | Mask[b] | N carriers | Trait | Cohort[c] | Effect | P val. | AF |
|---|---|---|---|---|---|---|---|---|
| | | | | Triglycerides | ++++++ | 0.23mmol/L | 7E-12 | 8E-04 |
| | | | | Apolipoprotein B | +-+-++ | 0.06g/L | 1E-11 | 8E-04 |
| *TOMM40* | F | ACMG | 946 | Apolipoprotein B | -?--++ | 0.06g/L | 6E-14 | 2E-03 |
| | | | | LDL | -?--++ | 0.17mmol/L | 2E-09 | 2E-03 |
| | | | | Cholesterol | -?--++ | 0.17mmol/L | 2E-06 | 2E-03 |
| *UBE2Z* | B | HI | 16 | Pure hypercholesterolemia | ????+- | OR = 7 | 1E-06 | 3E-05 |

[a], Genetic sex: B, both genetic sexes; F, female; M, male.

[b], Variant mask: ACMG, combination of ACMG-classified pathogenic, likely pathogenic, VUS-H, and damaging missense variants; HI, high-impact variants.

[c], Direction of association across genetic ancestry cohorts. "?" indicates that no individuals were observed in the corresponding cohort for the phenotype and variant mask. Cohorts are ordered as AFR, AMR, CSA, EAS, EUR, and OTH.

Among the protective associations, *GLTPD2* showed concordant effects across multiple lipid traits. *GLTPD2* encodes glycolipid transfer protein domain-containing 2 and is primarily expressed in liver tissue[22]. A combined mask of AAVC-classified P/LP/VUS-H variants and FuncVEP-damaging variants was associated with lower LDL cholesterol ($\beta = -0.16$ mmol/L, $p = 1.53 \times 10^{-9}$, lambda = 0.98), lower total cholesterol ($\beta = -0.20$ mmol/L, $p = 3.85 \times 10^{-9}$, lambda = 0.95), and lower apolipoprotein B levels ($\beta = -0.04$ g/L, $p = 3.74 \times 10^{-7}$, lambda = 0.98) in 1,017 female participants. Previous multi-ancestry GWAS studies have reported associations between the *GLTPD2* locus and both total cholesterol and LDL cholesterol levels[31].

In contrast, *STAB1* was associated with an atherogenic lipid profile. *STAB1* encodes a multifunctional scavenger receptor, and biallelic variants in this gene have been reported in benign hyperferritinemia. AAVC-classified P/LP/VUS-H and damaging missense variants in *STAB1* were associated with higher total cholesterol ($\beta = 0.18$ mmol/L, $p = 2.31 \times 10^{-10}$, lambda = 0.95) and higher LDL cholesterol levels ($\beta = 0.12$ mmol/L, $p = 1.25 \times 10^{-8}$, lambda = 0.97) in 1,625 participants. Previous GWAS studies have reported associations of *STAB1* with total cholesterol[31] and LDL cholesterol levels[29].

Among the remaining supported associations, we identified five gene–phenotype associations with favorable lipid profiles, involving *DNAJC13, HNF1A, SLC4A1, SMAD6*, and *UGT2B7*, and eight associations with atherogenic lipid profiles, involving *GAS6, HRG, KCNN4, NFIC, SMARCA4, TIMD4, TOMM40*, and *UBE2Z*.

## Identification of genes associated with hypertension and chronic kidney disease

We used both the binary phenotype of essential hypertension and the continuous traits of mean arterial pressure, systolic blood pressure, and diastolic blood pressure to identify genes involved in blood pressure regulation. We replicated 12 of 28 previously reported hypertension-associated genes, including 9 associated with increased risk or higher blood pressure values (*COL21A1, COL4A4, FES, GUCY1A1, NOS3, NPR1, PKD1, PKD2*, and *UBR3*) and 3 associated with protection (*DBH, NHERF2*, and *RGL3*).

In addition, we identified 14 new candidate genes for blood pressure regulation, of which 7 had support from previous GWAS findings and/or animal models. Damaging variants in *AHSP* and *UBE2Z* were associated with higher systolic blood pressure and mean arterial pressure, whereas high-impact variants in *ATP8A2* were associated only with higher systolic blood pressure. For binary phenotypes, AAVC-classified damaging variants in *SIPA1* (OR = 0.80, $p = 3.71 \times 10^{-10}$, lambda = 1.03) and *ECD* (OR = 0.56, $p = 1.46 \times 10^{-6}$, lambda = 1.04, males only) were associated with reduced risk (Fig. 3a; Extended Data Fig. 3; Supplementary Table 4).

Damaging missense variants in *GATA5* were associated with both binary and continuous blood pressure phenotypes, including essential hypertension (OR = 1.14, $p = 1.25 \times 10^{-6}$, lambda = 1.03), systolic blood pressure (β = 1.20 mmHg, $p = 9.59 \times 10^{-7}$, lambda = 0.87), diastolic blood pressure (β = 0.78 mmHg, $p = 5.49 \times 10^{-8}$, lambda = 0.86), and mean arterial pressure (β = 0.79 mmHg, $p = 5.67 \times 10^{-8}$, lambda = 0.85). The same damaging missense variant set for *GATA5* was also associated with increased eGFR (β = 0.81 mL/min/1.73 m², $p = 3.37 \times 10^{-8}$, lambda = 0.94). In humans, loss-of-function variants in *GATA5* are known to cause “congenital heart defects, multiple types, 5,” in both monoallelic and biallelic forms (OMIM: 617912). In mouse models, *Gata5* has been shown to be expressed in endothelial tissue, and endothelial-specific inhibition of Gata5 was reported to recapitulate a hypertensive phenotype[32].

*ABAT* encodes 4-aminobutyrate aminotransferase, a neuronal enzyme involved in the catabolism of gamma-aminobutyric acid. High-impact variants in *ABAT* were associated with lower diastolic blood pressure (β = −3.77 mmHg, $p = 2.60 \times 10^{-6}$, lambda = 0.82). Consistent with a possible role in blood pressure regulation, pharmacological inhibition of 4-aminobutyrate aminotransferase by vigabatrin delayed the development of hypertension in spontaneously hypertensive rats, potentially through increased brain GABA concentrations[33].

We identified 14 gene–phenotype associations for chronic kidney disease. Five of the 160 (3%) known CKD genes were replicated (*COL4A4, JAK2, PKD1, PKD2*, and *UMOD*). The remaining nine were novel associations, of which four had supporting evidence from previous human genetic studies or animal models (*ORAI3, SLC22A2, SLC22A7*, and *ZRANB1*) (Fig. 3b; Extended Data Fig. 3; Supplementary Table 4). *ORAI3* encodes a calcium channel component, whereas *ZRANB1* encodes a deubiquitinating enzyme; both genes show broad tissue expression.

*SLC22A7* encodes an organic anion transporter that is primarily expressed in liver and kidney. Damaging missense variants in *SLC22A7* were associated with an increased risk of chronic kidney disease (OR = 1.63, $p = 4.60 \times 10^{-7}$, lambda = 0.98) and lower eGFR (β = −4.00 mL/min/1.73 m², $p = 2.02 \times 10^{-38}$, lambda = 0.89). *SLC22A2* encodes organic cation transporter 2 and is expressed predominantly in the kidney, particularly in the

renal cortex and medulla[22]. Similarly, high-impact variants in *SLC22A2* were associated with increased chronic kidney disease risk (OR = 1.78, p = 4.48 x $10^{-8}$, lambda = 1.02) and lower eGFR (β = −6.02 mL/min/1.73 m², p = 3.42 x $10^{-68}$, lambda = 0.89). Common variants near both *SLC22A7* and *SLC22A2* have previously been associated with eGFR in a multi-ancestry GWAS[34].

## Identification of genes associated with diabetes mellitus

We used both binary phenotypes, type 1 diabetes and type 2 diabetes, and continuous glycaemic traits, glucose and HbA1c, to identify genetic associations with diabetes. To ensure consistency of effect, we included genes only when the same variant set significantly altered both glucose and HbA1c in the same direction within the same sex stratum (Supplementary Information: Simultaneous HbA1c and glucose altering associations). In this analysis, we replicated 2 protective genes, *G6PC2* and *MAP3K15*, and 14 risk genes, *ANKH, GCK, GIGYF1, HNF1A, HNF4A, IGF1R, IRS2, NEUROD1, PAM, PDX1, PTEN, SLTM, TNRC6B*, and *UBR3*, replicating 16 of the 81 (20%) known diabetes-associated genes (Fig. 4; Supplementary Tables 4 and 8).

We identified 72 novel associations for diabetes risk and none for protection. Of these, 23 had support from either previous GWAS findings or animal models. Four were identified through the type 2 diabetes phenotype alone, namely *ACVR1*, *FAM118A*, *HMGA2*, and *PSG11*, whereas the others were identified through concordant increases in both glucose and HbA1c levels. These included *APOC3, ASCL2, BMAL1, CELF4, CREBBP, DDR2, HECTD4, HSD17B6, IL10, KAT6A, MGME1, NKX2-1, SIN3A, SOS2, TACC2, UBE2D4, UBE2Z, WWTR1*, and *ZMIZ1*.

Singleton high-impact variants in *BMAL1* were associated with higher HbA1c levels (β = 20.16 mmol/mol, p = 6.65 x $10^{-9}$, lambda = 0.85) and higher glucose levels (β = 4.19 mmol/L, p = 1.51 x $10^{-10}$, lambda = 0.72) in six males. *BMAL1* encodes a core circadian clock component, and pancreatic knockout of *Bmal1* in mice has been shown to result in diabetes mellitus[35]. In addition, common non-coding variants near *BMAL1* were associated with type 2 diabetes in a multi-ancestry meta-analysis[36].

AAVC-classified P/LP/VUS-H variants in *MYCBP2* were associated with higher HbA1c levels (β = 2.69 mmol/mol, p = 4.09 x $10^{-8}$, lambda = 0.83), higher glucose levels (β = 0.55 mmol/L, p = 3.54 x $10^{-7}$, lambda = 0.68), and an increased risk of type 2 diabetes (OR = 3.07, p = 3.69 x $10^{-9}$, lambda = 0.90) in 187 individuals, although *MYCBP2* did not have prior support from GWAS or animal models.

# Sex-hormone associations

To assess whether associations prioritized in sex-specific analyses were also associated with circulating sex-hormone traits, we tested 92 novel associations against testosterone, free androgen index, sex hormone-binding globulin, and oestradiol levels. Eleven associations were also associated with at least one hormone trait, indicating that most were not captured by the measured circulating hormone phenotypes (Extended Data Fig. 4; Supplementary Table 5).

Seven hormone-associated signals involved diabetes-related phenotypes and were associated with higher testosterone levels or free androgen index, suggesting a potential contribution of androgen-related biology to a subset of sex-stratified diabetes associations [37].

## Genetic prevalence of cardiometabolic risk and protective genotypes

We calculated the genetic prevalence of P/LP genotypes in known disease genes and novel genes identified in this study across eight gnomAD genetic ancestry groups: African, Admixed American, Ashkenazi Jewish, East Asian, Finnish, Middle Eastern, non-Finnish European, and South Asian. For epidemiological reporting, we prioritized associations supported by non-singleton variant sets and classified variants in the corresponding genes using AAVC (Supplementary Table 9; Supplementary Information: Calculation of carrier frequency and genetic prevalence).

Among obesity-associated genes, the most common risk genotype was observed for *TNK2* (1 in 489 individuals), followed by *NR0B2* (1 in 506) and *MC4R* (1 in 546). The most common novel obesity-associated genotype was in *THUMPD3*, observed in 1 in 1,375 individuals, ranking as the sixth most common obesity gene overall (Fig. 5A). The most common novel dyslipidemia-associated genotype was in *ABCA6* (1 in 133), followed by *GCKR* (1 in 134) and *LDLR* (1 in 150) (Fig. 5B). The *LDLR* prevalence was largely driven by East Asian individuals, among whom it was present in 1 in 73 individuals. The most common hypertension risk gene was *COL21A1* (1 in 237), followed by *COL4A4* (1 in 293) and *WNK4* (1 in 446) (Fig. 5C). For chronic kidney disease, the highest genetic prevalence was observed for the newly identified gene *SLC22A2*, with a prevalence of 1 in 97 individuals (Fig. 5D). For diabetes, *HFE* was the most common risk gene (1 in 64), followed by *WFS1* (1 in 147) and *ABCC8* (1 in 281) (Fig. 5E).

When analyzed cumulatively, nearly 1 in 6 individuals carried a risk genotype for at least one cardiometabolic phenotype. The cumulative prevalence was 1 in 18 for dyslipidemia, 1 in 23 for diabetes mellitus, 1 in 24 for chronic kidney disease, 1 in 55 for hypertension, and 1 in 62 for obesity. In contrast, 1 in 9 individuals carried a genotype associated with a favorable cardiometabolic profile: 1 in 11 for dyslipidemia, 1 in 110 for hypertension, 1 in 144 for diabetes mellitus, and 1 in 221 for obesity (Fig. 5F).

## Discussion

We conducted sex- and ancestry-stratified gene-based association tests for six cardiometabolic phenotype groups: obesity, fatty liver disease, dyslipidemia, hypertension, chronic kidney disease, and diabetes mellitus. Across the UK Biobank analyses with Turkish Variome contributing to obesity, we replicated 94/368 (26%) previously reported gene–phenotype associations and identified 166 novel associations.

Although no ancestry-specific analysis outside the European subgroup reached genome-wide significance, multi-ancestry meta-analysis substantially increased both replication and discovery yields, with a 9.4-fold increase in replicated associations and a 5.9-fold increase in novel discoveries. This gain likely reflected larger effective sample size and improved representation of ancestry-specific allele frequencies. Consistent with this, leave-one-ancestry-out analyses showed that excluding individual ancestry groups reduced discovery yield for several phenotypes (Supplementary Table 10). For example, in the obesity analysis, excluding the Turkish Variome would have missed nine novel associations, five of which had independent supporting evidence. Sex stratification increased replicated findings (+ 15.2%) more modestly than multi-ancestry meta-analysis, but had a particularly strong effect on novel discovery, increasing novel associations from 74 to 166 (+ 124%). This pattern may reflect reduced effect heterogeneity in sex-stratified analyses, as well as the limited number of prior biobank-scale rare-variant studies that performed sex-stratified analyses (Supplementary Table 6).

Consistent with recent benchmarking of gene-level burden tests, choices in variant annotation, transcript selection, allele-frequency thresholds, and mask definition can substantially influence both discovery yield and replication across studies. This supports the use of robust and accurate variant-prioritization strategies to improve interpretability and comparability of rare-variant association results[10,11].

We also observed cross-phenotype cardiometabolic associations among several genes identified in our rare-variant burden analyses. *HECTD4* and *RB1CC1* linked obesity-related traits to glycaemic phenotypes, whereas *UBE2Z* showed broader associations across blood pressure, lipid and glycaemic traits. These pleiotropic findings are consistent with previous reports[38–42] and further suggest that a subset of rare-variant associations may influence shared cardiometabolic pathways rather than isolated disease categories.

This study expands the number of causal cardiometabolic gene–phenotype associations from 368 to 534, corresponding to a 45% increase, and shows that approximately one in six individuals carries risk-increasing alleles and one in nine carries protective alleles. These findings may provide clinically actionable signals for prevention, risk stratification, or earlier diagnosis of cardiometabolic phenotypes, which are often chronic and clinically silent in their early stages and may be modifiable through lifestyle or therapeutic intervention.

## Methods

# Study cohorts

We analyzed two cohorts in this study: the UK Biobank and the Turkish Variome. Whole-genome sequencing data for 490,640 UK Biobank participants were described previously[10,16]. We excluded individuals who were outliers for heterozygosity, had discordance between self-reported and chromosomal sex, had sex chromosome aneuploidy, or failed to meet the overall quality-control criteria for whole-genome sequencing. After sample-level quality control, the UK Biobank cohort comprised 483,311 individuals, including 262,115 females and 221,196 males. Genetic ancestry was subsequently assigned to UK Biobank participants using a reference-panel-based classifier (Supplementary Information: Genetic ancestry).

The Turkish Variome includes 3,362 individuals, and its processing pipeline was described previously[17]. From this cohort, we selected individuals with available BMI information and excluded those with sex chromosome aneuploidy, resulting in a final sample of 981 individuals, including 752 females and 229 males. BMI ranged from 15 to 96 kg/$m^2$, with a median of 45 kg/$m^2$.

# Variant processing

For UK Biobank whole-genome sequencing data, variant-level quality control was performed using DRAGEN machine-learning recalibration, and variants were excluded if read depth was below 10, missingness exceeded 10%, or the Hardy–Weinberg equilibrium p value was below $1 \times 10^{-15}$. For the Turkish Variome[17], whole-exome sequencing variants were filtered according to GATK Best Practices and excluded if read depth was below 10, missingness exceeded 10%, or the Hardy–Weinberg equilibrium p value was below $1 \times 10^{-8}$.

We classified variants using AAVC[18] and used FuncVEP-CTE[10] for missense predictions. We defined variant sets under four broad mask categories: ACMG-classified, damaging missense, high-impact, and synonymous variants used for inflation assessment. For the ACMG-classified and damaging-missense categories, we used five masks: (1) P, (2) P + LP, (3) P + LP+VUS-H, (4) Missense-D (FuncVEP-CTE > 0.5192), and (5) P + LP+VUS-H+Missense-D. For high-impact variants, we used seven masks derived from AAVC classifications: (1) PVS1-VS, (2) PVS1-S, (3) PVS1-M, (4) PVS1-P, (5) PVS1-VS + S, (6) PVS1-VS + S+M, and (7) PVS1-VS + S+M + P. PVS1-VS denotes variants predicted to undergo nonsense-mediated decay; PVS1-S denotes variants predicted to result in loss of more than 10% of the amino acid sequence or loss of a protein domain; PVS1-M denotes variants predicted to result in loss of less than 10% of the protein while preserving all annotated domains; and PVS1-P denotes start-lost variants. To assess genomic inflation and calculate lambda statistics for each phenotype, allele-frequency threshold, variant mask, and genetic-sex combination, we included non-splice-altering synonymous variants (SpliceAI score < 0.2) as a negative-control variant set. In total, this yielded 13 variant masks. We then applied three allele-frequency thresholds to each mask: ≤0.01, ≤ 0.001, and singleton. In total, 18,821 protein-coding genes had at least one variant included in at least one mask.

# Phenotype definitions

We defined both binary and continuous phenotype groups to capture complementary aspects of cardiometabolic health. We interrogated cardiometabolic disease burden under six main categories: obesity, fatty liver disease, dyslipidemia, hypertension, chronic kidney disease, and diabetes mellitus.

For binary disease phenotypes, we mapped ICD-10 diagnosis codes from UK Biobank to phecodes using the PheCodeX[43] mapping framework. We selected 15 phecodes representing the six cardiometabolic phenotype groups (Supplementary Table 1). In addition to individual phecodes, we constructed combined phenotype-group definitions by aggregating related phecodes within each category; for example, the obesity group included both “Obesity” and “Morbid Obesity”.

For continuous phenotypes, we selected 16 cardiometabolic traits. For obesity, we used body mass index and body fat percentage. For fatty liver disease, we used liver proton density fat fraction. For dyslipidemia, we used LDL cholesterol, total cholesterol, triglycerides, lipoprotein(a), apolipoprotein B, HDL cholesterol, and apolipoprotein A. For hypertension, we used systolic blood pressure, diastolic blood pressure, and calculated mean arterial pressure as: MAP = (SBP + 2 x DBP) / 3. For diabetes mellitus, we used HbA1c and glucose levels. For chronic kidney disease, we calculated estimated glomerular filtration rate (eGFR) using age, genetic sex, serum creatinine according to the 2021 CKD-EPI ancestry-free creatinine equation.

To investigate potential links between sex-stratified genetic effects and sex-hormone biology, we additionally selected total testosterone, sex hormone-binding globulin (SHBG), and oestradiol. We also derived free androgen index (FAI) using total testosterone and SHBG: FAI = 100 x (total testosterone / SHBG).

# Statistics

We performed gene-based association analyses using REGENIE v4.1[19]. In the combined-sex analyses, models were adjusted for age, age squared, sex, Townsend deprivation index, and the first 20 genetic principal

components in UK Biobank. In sex-specific analyses, association testing was performed separately in males and females, and sex was not included as a covariate. For the Turkish Obesity Cohort, Townsend deprivation index was not included because this variable was unavailable.

For binary phenotypes, association tests were performed using Firth correction to reduce small-sample bias and improve calibration for rare-variant burden tests. Continuous phenotypes were analyzed using linear regression models after phenotype preprocessing described above.

Gene-based meta-analysis was performed using remeta v0.11.1[20] across six UK Biobank genetic ancestry groups and the Turkish Variome obesity cohort from the REGENIE output files. Linkage disequilibrium matrices were constructed using a 1 Mb window around the selected variants for each cohort.

Genes were included in association testing if at least five individuals carried a qualifying variant in the corresponding burden mask in the combined dataset. Non-splice-altering synonymous variant mask was used to estimate the genomic inflation factor separately for each sex, phenotype, and allele-frequency threshold. Association results with genomic inflation factor greater than 1.15 were excluded from downstream analyses (Supplementary Table 11).

For each cardiometabolic phenotype, genome-wide significance was defined using a Bonferroni threshold based on the number of tested protein-coding genes ($2.6 \times 10^{-6}$, 0.05/18,821 genes).

# Gene prioritization

We classified genome-wide significant gene–phenotype associations as either replicated or novel based on prior evidence from external resources including OMIM, PanelApp and literature curations.

We collected independent evidence for newly identified gene–phenotype associations that reached genome-wide significance. We evaluated three sources of external support: (i) previously reported common-variant associations from the GWAS Catalog, (ii) prior linkage evidence curated in OMIM, and (iii) mouse phenotype resources from Mouse Genome Informatics (MGI) and the International Mouse Phenotyping Consortium (IMPC).

The GWAS Catalog[44] was queried in May 2026. Reported associations were considered supportive if the GWAS Catalog trait either directly matched the prioritized phenotype or represented a closely related trait within the same cardiometabolic phenotype group. Using this approach, we identified supporting GWAS evidence for 63 associations.

For linkage evidence, OMIM “Phenotype Only Entries” were downloaded as Gene Map Tables in October 2025. Cytogenetic band annotations for genome-wide significant genes were obtained from Ensembl BioMart[45] and used to match candidate gene–phenotype associations to OMIM cytogenetic band–phenotype records. Each candidate match based on cytogenetic band overlap was then manually reviewed to assess phenotypic relevance. This approach identified 5 associations with supporting linkage evidence.

Finally, we queried MGI[46] and IMPC[47] to identify mouse homologs of genes implicated in novel associations. Mouse phenotypes were considered supportive when the reported phenotype was concordant with the corresponding human cardiometabolic phenotype. Using concordant mouse phenotypes as external support, 16 associations were supported by mouse model evidence.

# Declarations


### Acknowledgement

We gratefully acknowledge İclal Büyükdevrim Özçelik and Nezahat Doğan for their invaluable contributions. This study was funded by a grant from the Leducq Foundation. This work was done under UK Biobank application 268863. The application cost for UK Biobank was covered by Global Researcher Access Fund (GRAF). The Turkish Academy of Sciences supported this work. B.K. and K.Ç. are the recipients of fellowship 2211-A National Doctorate Scholarship Program of Scientific and Technological Research Council of Türkiye (TÜBİTAK) Directorate of Science Fellowships and Grant Programmes.


### Author Contributions

B.K. and K.Ç. designed the study. B.K. performed the UK Biobank and downstream analyses. O.E.O. coordinated field recruitment and sample collection and contributed to the generation of the Turkish Variome resource. B.K., K.Ç., and T.Ö. interpreted the results. B.K. drafted the manuscript with input from K.Ç. and T.Ö. O.E.O. reviewed and edited the manuscript. B.K. and T.Ö. conceived the project, and T.Ö. supervised the study. All authors read and approved the final manuscript.

### Competing Interests

The authors declare no competing interest.

### Data availability

Gene-phenotype association results produced in this study are available at https://zenodo.org/records/21154951. All other data generated or analyzed during this study are included in the Supplementary Tables.

### Code availability

We used publicly available software together with software/methods. AAVC and FuncVEP available at https://github.com/OzcelikLab/AAVC and https://github.com/OzcelikLab/FuncVEP. Regenie and remeta available at https://rgcgithub.github.io/regenie/ and https://rgcgithub.github.io/remeta/.

# References


1. Noubiap, J. J. *et al.* Worldwide trends in metabolic syndrome from 2000 to 2023: a systematic review and modelling analysis. *Nat. Commun.* **17**, 573 (2025).
2. Cornier, M.-A. *et al.* The metabolic syndrome. *Endocr. Rev.* **29**, 777–822 (2008).

3. Lin, L. *et al.* Metabolic Syndrome-Related Kidney Injury: A Review and Update. *Front. Endocrinol. (Lausanne).* **13**, 904001 (2022).
4. Farooqi, I. S. & O'Rahilly, S. Hypothalamic obesity. *Rev. Endocr. Metab. Disord.* **27**, 617–623 (2026).
5. Lehrman, M. A. *et al.* Mutation in LDL receptor: Alu-Alu recombination deletes exons encoding transmembrane and cytoplasmic domains. *Science* **227**, 140–6 (1985).
6. Ozçelik, T. *et al.* Small nuclear ribonucleoprotein polypeptide N (SNRPN), an expressed gene in the Prader-Willi syndrome critical region. *Nat. Genet.* **2**, 265–9 (1992).
7. Zhang, Y. *et al.* Positional cloning of the mouse obese gene and its human homologue. *Nature* **372**, 425–32 (1994).
8. Lifton, R. P. *et al.* A chimaeric 11 beta-hydroxylase/aldosterone synthase gene causes glucocorticoid-remediable aldosteronism and human hypertension. *Nature* **355**, 262–5 (1992).
9. Vionnet, N. *et al.* Nonsense mutation in the glucokinase gene causes early-onset non-insulin-dependent diabetes mellitus. *Nature* **356**, 721–722 (1992).
10. Kayaalp, B. *et al.* Prediction of human missense variant effects from functional evidence. *Res. Sq.* https://doi.org/10.21203/rs.3.rs-7536763/v1 (2025) doi:10.21203/rs.3.rs-7536763/v1.
11. Nguyen, T. *et al.* Empirically determined baseline masking strategies and other considerations for gene-level burden tests. *Nat. Genet.* **58**, 1226–1236 (2026).
12. Bernabeu, E. *et al.* Sex differences in genetic architecture in the UK Biobank. *Nat. Genet.* **53**, 1283–1289 (2021).
13. Peters, S. A. E., Huxley, R. R. & Woodward, M. Diabetes as risk factor for incident coronary heart disease in women compared with men: a systematic review and meta-analysis of 64 cohorts including 858,507 individuals and 28,203 coronary events. *Diabetologia* **57**, 1542–51 (2014).
14. Yang, M.-L. *et al.* Sex-specific genetic architecture of blood pressure. *Nat. Med.* **30**, 818–828 (2024).
15. Shrine, N. *et al.* Multi-ancestry genome-wide association analyses improve resolution of genes and pathways influencing lung function and chronic obstructive pulmonary disease risk. *Nat. Genet.* **55**, 410–422 (2023).
16. UK Biobank Whole-Genome Sequencing Consortium. Whole-genome sequencing of 490,640 UK Biobank participants. *Nature* **645**, 692–701 (2025).
17. Kars, M. E. *et al.* The genetic structure of the Turkish population reveals high levels of variation and admixture. *Proceedings of the National Academy of Sciences* **118**, e2026076118 (2021).
18. İnan, R. A. *et al.* AAVC: an automated framework for high-accuracy ACMG-based variant classification. *Genetics in Medicine* 102624 (2026) doi:10.1016/j.gim.2026.102624.
19. Mbatchou, J. *et al.* Computationally efficient whole-genome regression for quantitative and binary traits. *Nat. Genet.* **53**, 1097–1103 (2021).
20. Joseph, T. A. *et al.* Computationally efficient meta-analysis of gene-based tests using summary statistics in large-scale genetic studies. *Nat. Genet.* **57**, 3193–3200 (2025).
21. Martin, A. R. *et al.* PanelApp crowdsources expert knowledge to establish consensus diagnostic gene panels. *Nat. Genet.* **51**, 1560–1565 (2019).
22. Lonsdale, J. *et al.* The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* **45**, 580–585 (2013).

23. Serra-Juhé, C. *et al.* Novel genes involved in severe early-onset obesity revealed by rare copy number and sequence variants. *PLoS Genet.* **13**, e1006657 (2017).
24. Baldarelli, R. M. *et al.* Mouse Genome Informatics: an integrated knowledgebase system for the laboratory mouse. *Genetics* **227**, (2024).
25. Kawakami, K., Onaka, T., Iwase, M., Homma, I. & Ikeda, K. Hyperphagia and obesity in Na,K-ATPase alpha2 subunit-defective mice. *Obes. Res.* **13**, 1661–71 (2005).
26. Huang, J. *et al.* Genomics and phenomics of body mass index reveals a complex disease network. *Nat. Commun.* **13**, 7973 (2022).
27. Guez, J. *et al.* Integrating 730,947 exome sequences with clinical literature improves gene discovery. Preprint at https://doi.org/10.64898/2026.03.23.26349081 (2026).
28. Xing, M. *et al.* Paraventricular hypothalamic RUVBL2 neurons suppress appetite by enhancing excitatory synaptic transmission in distinct neurocircuits. *Nat. Commun.* **15**, 8939 (2024).
29. Jee, Y. H. *et al.* Genome-wide association studies in a large Korean cohort identify quantitative trait loci for 36 traits and illuminate their genetic architectures. *Nat. Commun.* **16**, 4935 (2025).
30. Spracklen, C. N. *et al.* Identification of type 2 diabetes loci in 433,540 East Asian individuals. *Nature* **582**, 240–245 (2020).
31. Graham, S. E. *et al.* The power of genetic diversity in genome-wide association studies of lipids. *Nature* **600**, 675–679 (2021).
32. Messaoudi, S. *et al.* Endothelial Gata5 transcription factor regulates blood pressure. *Nat. Commun.* **6**, 8835 (2015).
33. Singewald, N., Pfitscher, A. & Philippu, A. Effects of gamma-vinyl GABA (vigabatrin) on blood pressure and body weight of hypertensive and normotensive rats. *Naunyn. Schmiedebergs. Arch. Pharmacol.* **345**, 181–6 (1992).
34. Sinnott-Armstrong, N. *et al.* Genetics of 35 blood and urine biomarkers in the UK Biobank. *Nat. Genet.* **53**, 185–194 (2021).
35. Marcheva, B. *et al.* Disruption of the clock components CLOCK and BMAL1 leads to hypoinsulinaemia and diabetes. *Nature* **466**, 627–31 (2010).
36. Suzuki, K. *et al.* Genetic drivers of heterogeneity in type 2 diabetes pathophysiology. *Nature* **627**, 347–357 (2024).
37. Ruth, K. S. *et al.* Using human genetics to understand the disease impacts of testosterone in men and women. *Nat. Med.* **26**, 252–258 (2020).
38. Zorn, S. *et al.* Obesity due to MC4R deficiency is associated with reduced cholesterol, triglycerides and cardiovascular disease risk. *Nat. Med.* **31**, 4180–4188 (2025).
39. Deaton, A. M. *et al.* Gene-level analysis of rare variants in 379,066 whole exome sequences identifies an association of GIGYF1 loss of function with type 2 diabetes. *Sci. Rep.* **11**, 21565 (2021).
40. Kaisinger, L. R. *et al.* Large-scale exome sequence analysis identifies sex- and age-specific determinants of obesity. *Cell genomics* **3**, 100362 (2023).
41. Banerjee, D. & Girirajan, S. Discovery of obesity genes through cross-ancestry analysis. *Nat. Commun.* **16**, 9319 (2025).

42. Zhao, Y. *et al.* Population-scale gene-based analysis of whole-genome sequencing provides insights into metabolic health. *Nat. Genet.* **57**, 2436–2444 (2025).
43. Shuey, M. M. *et al.* Next-generation phenotyping: introducing phecodeX for enhanced discovery research in medical phenomics. *Bioinformatics* **39**, (2023).
44. Cerezo, M. *et al.* The NHGRI-EBI GWAS Catalog: standards for reusability, sustainability and diversity. *Nucleic Acids Res.* **53**, D998–D1005 (2025).
45. Dyer, S. C. *et al.* Ensembl 2025. *Nucleic Acids Res.* **53**, D948–D957 (2025).
46. Baldarelli, R. M. *et al.* Mouse Genome Informatics: an integrated knowledgebase system for the laboratory mouse. *Genetics* **227**, (2024).
47. Muñoz-Fuentes, V. *et al.* The International Mouse Phenotyping Consortium (IMPC): a functional catalogue of the mammalian genome that informs conservation. *Conserv. Genet.* **19**, 995–1005 (2018).

# Figures

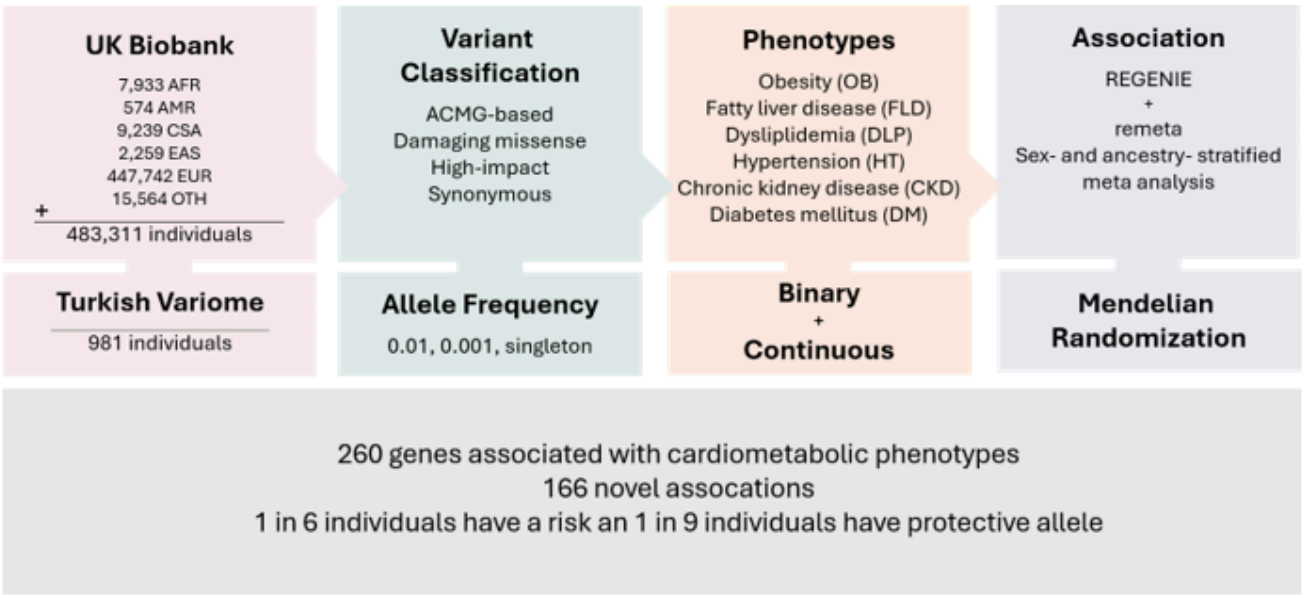
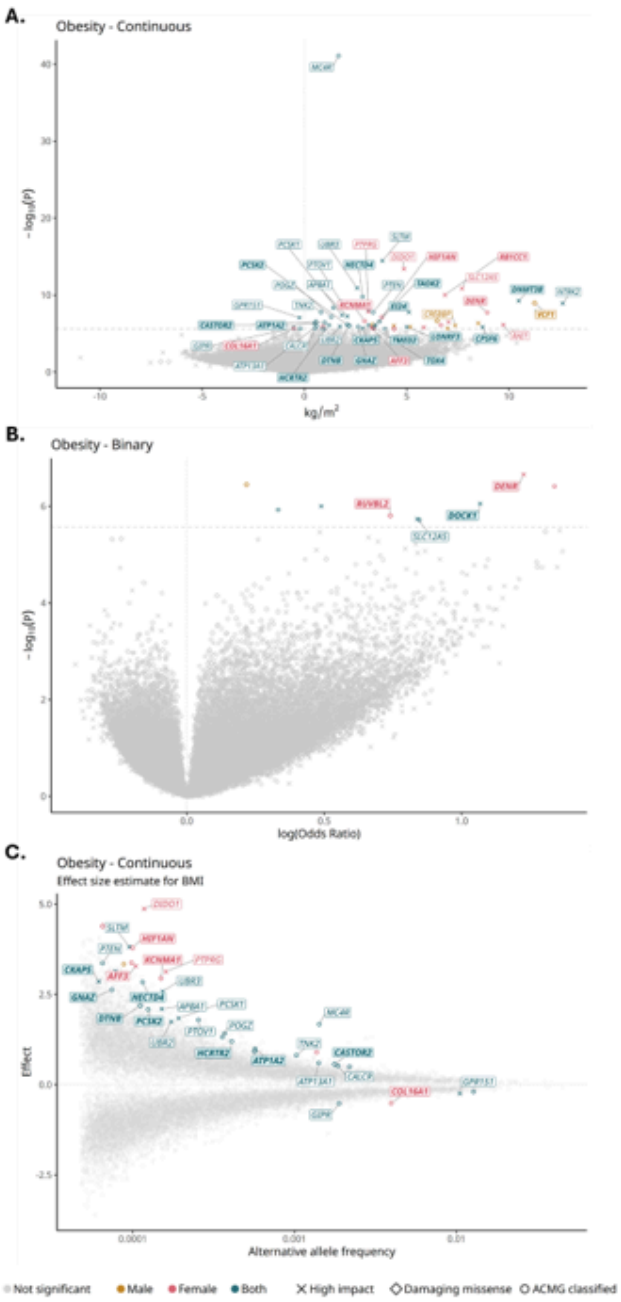



**Figure 1**

Study design and summary of the results.

Variants from UK Biobank and the Turkish Variome were classified using AAVC and FuncVEP, grouped into variant masks, and tested in gene-based burden analyses of binary and continuous cardiometabolic phenotypes using REGENIE and remeta. Analyses were stratified by genetic sex, meta-analysed across different cohorts based on genetic ancestry, and summarized by replicated, novel, sex-specific, and protective gene–phenotype associations.




## Figure 2

Gene-burden results for obesity and gene-based prevalence in UK Biobank and the Turkish Variome.

Volcano plots show association results for obesity-related phenotypes: (A) BMI as a continuous phenotype and (B) obesity as a binary phenotype. (C) The trumpet plot shows the relationship between effect size and combined UK Biobank–Turkish Variome allele frequency for genome-wide significant obesity-associated genes. The trumpet plot includes associations with at least 50 carriers in the combined-sex analysis or at least 25 carriers in sex-specific analyses. Known obesity-related genes and novel genes with supporting evidence are labelled. Novel genes are shown in bold and highlighted with boxed labels. For each gene–phenotype association, the plotted point corresponds to the most significant combination of allele-frequency threshold, variant mask, and sex-specific analysis.

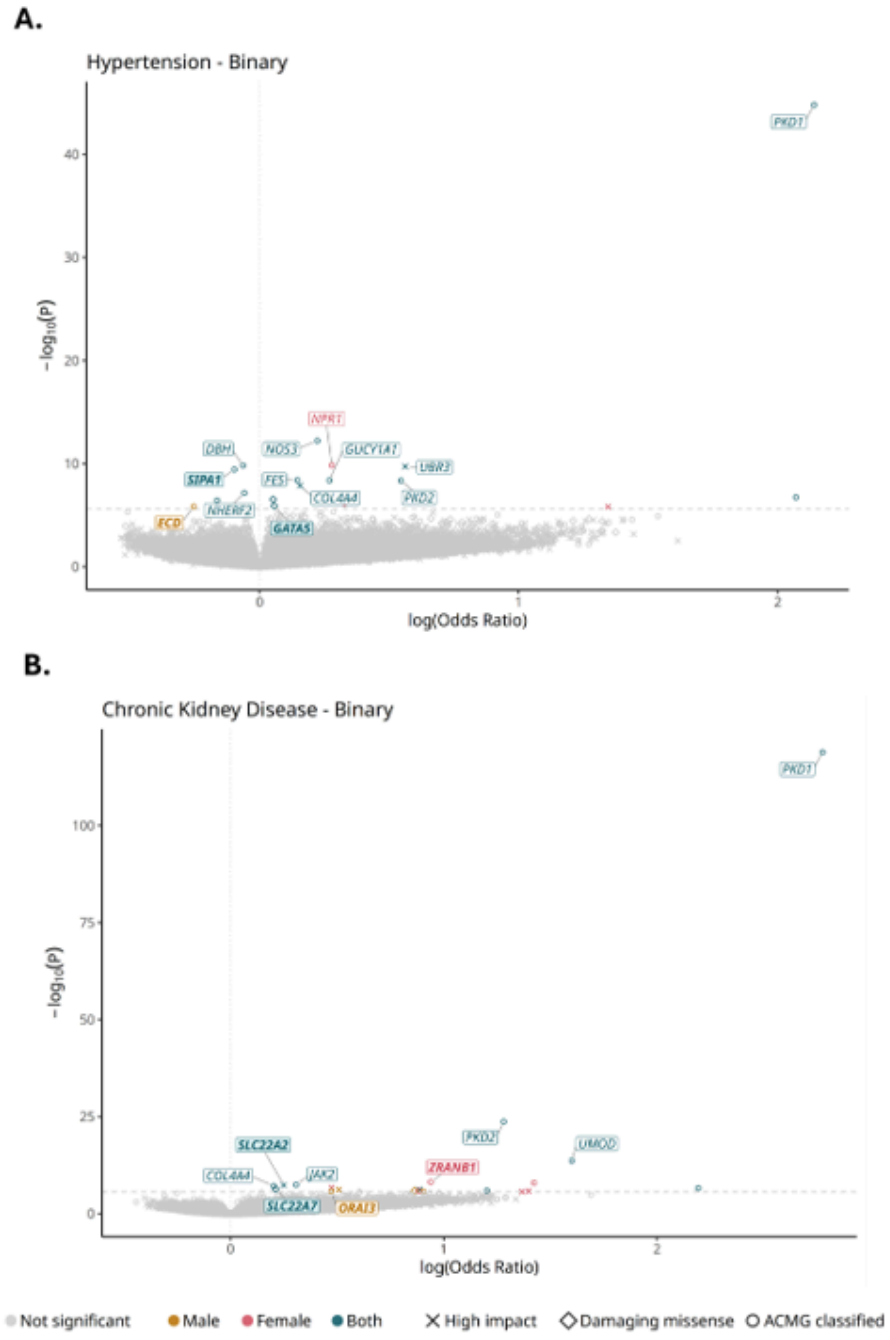
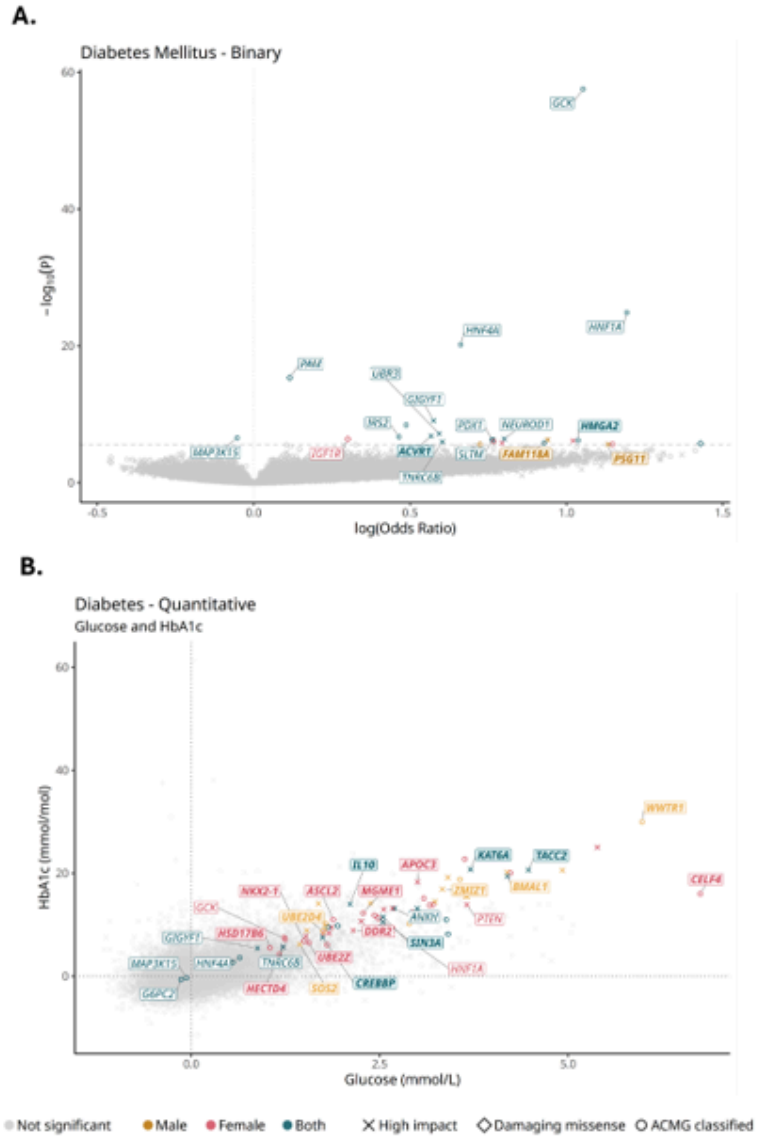



**Figure 3**

Gene-based burden test results for hypertension and chronic kidney disease.

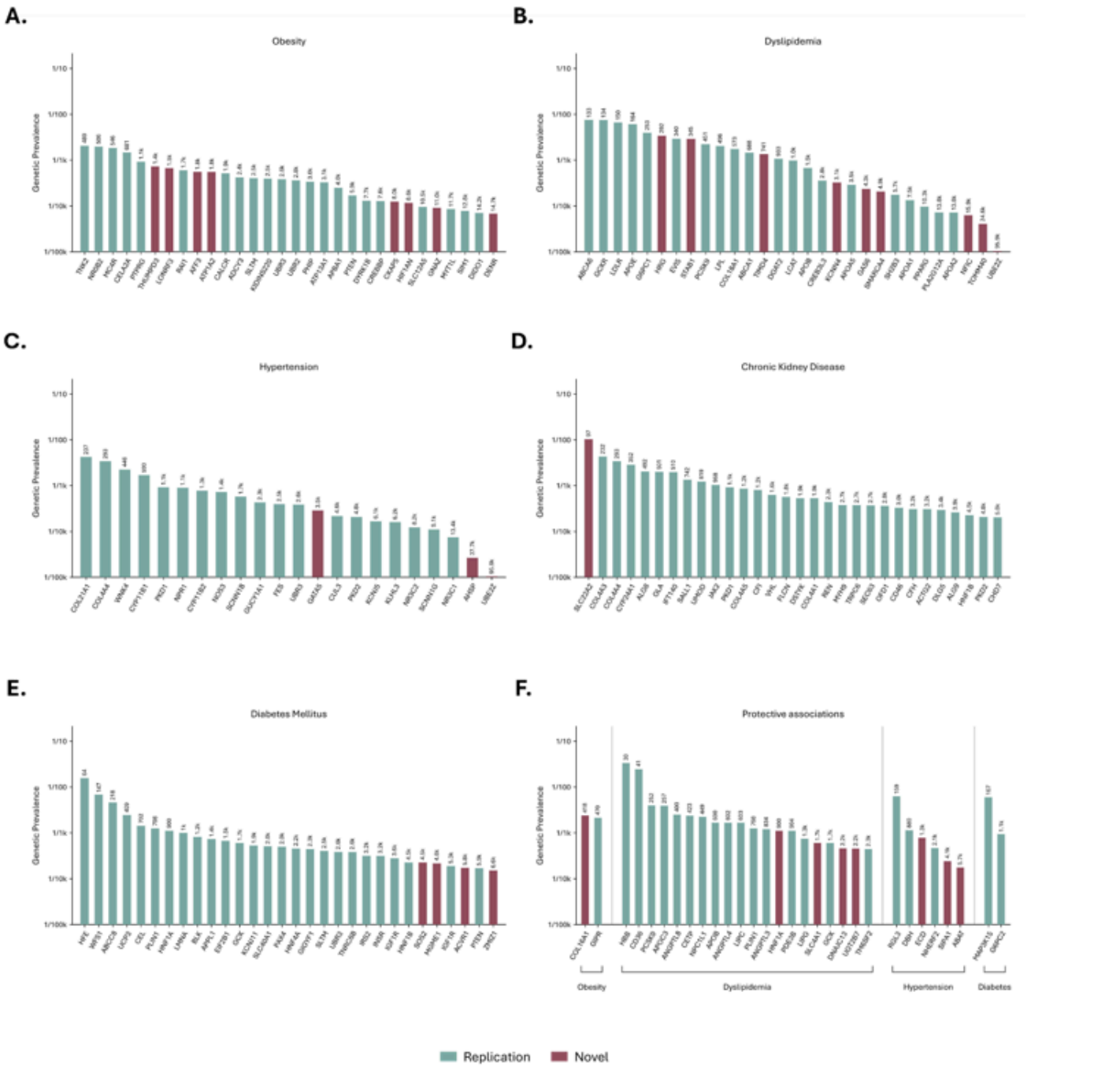
Volcano plots show association results for (A) hypertension and chronic (B) kidney disease. Known hypertension- and chronic kidney disease-related genes and novel genes with supporting evidence are labelled. Novel genes are shown in bold and highlighted with boxed labels. For each gene–phenotype association, the plotted point corresponds to the most significant combination of allele-frequency threshold, variant mask, and sex-specific analysis.




## Figure 4

Gene-based burden test results for diabetes mellitus.

(A) Volcano plot showing association results for diabetes mellitus. (B) Scatter plot showing effect sizes of genome-wide significant associations for HbA1c and glucose levels. Known diabetes-related genes and novel genes with supporting evidence are labelled. Novel genes are shown in bold and highlighted with boxed labels. For each gene–phenotype association, the plotted point corresponds to the most significant combination of allele-frequency threshold, variant mask, and sex-specific analysis.




## Figure 5

Genetic epidemiology of genes associated with cardiometabolic phenotypes.

Bar plots show the estimated genetic prevalence of AAVC-classified pathogenic or likely pathogenic variants in genes associated with cardiometabolic phenotypes. Results are shown for genes implicated in (A) obesity, (B) dyslipidemia, (C) hypertension, (D) chronic kidney disease, (E) diabetes mellitus, and (F) protective cardiometabolic associations. Carrier frequencies were calculated across the eight genetic ancestry groups represented in gnomAD and summarized as the mean ancestry-specific frequency.

## Supplementary Files

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

- RSSupplementaryInformationGenebasedmapofcardiometabolicdiseases.docx
- RSSupplementaryTablesGenebasedmapofcardiometabolicdiseases.xlsx
- ExtendedDataFig.docx