



Published in final edited form as:

*Nat Genet.* 2024 February ; 56(2): 212–221. doi:10.1038/s41588-023-01625-2.

## Comprehensive genetic study of the insulin resistance marker TG:HDL-C in the UK Biobank

Antonino Oliveri<sup>1,4</sup>, Ryan J Rebernick<sup>1,4</sup>, Annapurna Kuppa<sup>1</sup>, Asmita Pant<sup>1</sup>, Yanhua Chen<sup>1</sup>, Xiaomeng Du<sup>1</sup>, Kelly C. Cushing<sup>1</sup>, Hannah N Bell<sup>2</sup>, Chinmay Raut<sup>1</sup>, Ponnandy Prabhu<sup>1</sup>, Vincent L. Chen<sup>1</sup>, Brian D. Halligan<sup>1</sup>, Elizabeth K. Speliotes<sup>1,3,✉</sup>

<sup>1</sup>Division of Gastroenterology and Hepatology, University of Michigan Health System, Ann Arbor, MI, USA.

<sup>2</sup>Department of Molecular and Integrative Physiology, University of Michigan, Ann Arbor, MI, USA.

<sup>3</sup>Department of Computational Medicine and Bioinformatics, University of Michigan Medical School, Ann Arbor, MI, USA.

<sup>4</sup>These authors contributed equally: Antonino Oliveri, Ryan Rebernick.

### Abstract

Insulin resistance (IR) is a well-established risk factor for metabolic disease. The ratio of triglycerides to high-density lipoprotein cholesterol (TG:HDL-C) is a surrogate marker of IR. We conducted a genome-wide association study of the TG:HDL-C ratio in 402,398 Europeans within the UK Biobank. We identified 369 independent SNPs, of which 114 had a false discovery rate-adjusted  $P$  value  $< 0.05$  in other genome-wide studies of IR making them high-confidence IR-associated loci. Seventy-two of these 114 loci have not been previously associated with IR. These 114 loci cluster into five groups upon phenome-wide analysis and are enriched for candidate genes important in insulin signaling, adipocyte physiology and protein metabolism. We created a polygenic-risk score from the high-confidence IR-associated loci using 51,550 European individuals in the Michigan Genomics Initiative. We identified associations with

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**✉ Correspondence and requests for materials** should be addressed to Elizabeth K. Speliotes. [espeliot@med.umich.edu](mailto:espeliot@med.umich.edu).

#### Author contributions

E.K.S. contributed to the concept development, study design, data analysis and interpretation, writing of the manuscript and critical revision of the manuscript. A.O. and R.R. contributed to the study design, data analysis and interpretation, statistical analysis, drafting of the manuscript and critical review of the manuscript. X.D., Y.C. and C.R. contributed to data analysis and interpretation, statistical analysis and critical review of the manuscript. A.K., K.C.C., V.L.C., P.P., A.P., H.B. and B.H. contributed to data analysis and interpretation, drafting of the manuscript and critical review of the manuscript. All authors have critically reviewed the paper and approved the final version of this paper.

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41588-023-01625-2>.

#### Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41588-023-01625-2>.

#### Code availability

The code to assign the gene labels is publicly available at <https://doi.org/10.5281/zenodo.10182519> (ref. 89) and [https://github.com/oliveriantonino/annotation\\_TGHDL](https://github.com/oliveriantonino/annotation_TGHDL).

#### Reporting summary

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

diabetes, hyperglyceridemia, hypertension, nonalcoholic fatty liver disease and ischemic heart disease. Collectively, this study provides insight into the genes, pathways, tissues and subtypes critical in IR.

Insulin resistance (IR) is closely linked to numerous cardiometabolic risk factors and is thought to be the origin of many metabolic diseases<sup>1–3</sup>. IR is characterized by a diminished cellular response to insulin, leading to dyslipidemia<sup>4</sup> and higher circulating levels of insulin and glucose<sup>5</sup>. The gold standard method to measure IR requires the usage of glucose clamp—an invasive, expensive, and time-consuming technique that is impractical for routine clinical use<sup>6</sup>. Simpler methods, such as the insulin sensitivity index or the homeostatic model assessment for insulin resistance (HOMA-IR), involve the direct measurement of insulin and/or glucose levels and have been shown to correlate strongly with the gold-standard glucose-clamp technique<sup>7,8</sup>. These methods have been previously applied to identify 130 loci independently associated with IR across several studies<sup>9–14</sup>. These loci have been linked to genes with functions important in insulin receptor signaling (*GRB14* and *IRS1*), glycogen metabolism (*PPP1R3B*) and adipogenesis (*LYPLAL1* and *FAM13A*) among other pathways. However, these genetic studies of IR have been limited in scale when compared to analysis of other traits originating from groups such as the Genetic Investigation of Anthropometric Traits (GIANT) Consortium or the UK Biobank (UKBB). Although the UKBB did not quantify HOMA-IR or other direct measures of insulin, it did collect information on nonfasting lipids including triglycerides (TGs) and high-density lipoprotein cholesterol (HDL-C).

The TG:HDL-C ratio has been previously validated as a surrogate measure of IR<sup>15–19</sup>. To further investigate the genetic basis of IR, we performed a genome-wide association study (GWAS) of the TG:HDL-C using data from 402,398 Europeans within the UKBB<sup>20</sup>. Of the TG:HDL-C loci reaching genome-wide significance, we specifically focus on the subset of high-confidence IR-associated SNPs that reach significance in external studies of IR. These high-confidence loci are explored in the context of known insulin biology, interrogated for previously uncharacterized roles in IR, and examined for their contribution to disease in external datasets. Collectively, this study identifies numerous previously uncharacterized loci for IR in the context of metabolic pathways, traits and diseases.

## Results

### GWAS identifies 369 independent loci for TG:HDL-C

We extracted serum TG and HDL-C levels around the time of enrollment for 402,398 Europeans from the UKBB (Fig. 1). These values were used to calculate the TG:HDL-C ratio (1.07 (0.66, 1.73), median (Q1–Q3); Supplementary Table 1). We then performed a GWAS fitting a linear mixed model using Scalable and Accurate Implementation of GEneralized mixed model<sup>21</sup> (SAIGE). The dependent variable was the rank-based inverse normalized TG:HDL-C, and the independent variables were age, age squared, sex and the first ten principal components (PCs). The estimated intercept of the linkage disequilibrium (LD) score regression for this model was 1.5188. We corrected for this inflation in the test statistic. Notably, incorporating 20 PCs into the model did not substantially reduce

the value of the LD score regression intercept (1.4622). In total, 32,573 variants reached genome-wide significance ( $P < 5 \times 10^{-8}$ ) for TG:HDL-C after the exclusion of insertions and deletions (INDELs), multiallelic/ambiguous SNPs and SNPs not available in the Michigan Genomics Initiative (MGI). To extract a list of independent SNPs, we then applied conditional and joint multiple-SNP analysis (COJO)<sup>22</sup>. COJO identified 369 genome-wide significant SNPs with a minor allele frequency (MAF)  $\geq 0.01$  (Supplementary Tables 2 and 3). For each of these 369 SNPs, we determined the most likely causal gene using an algorithm incorporating proximity, Data-driven Expression Prioritization Integration for Complex Traits (DEPICT) prioritization, tissue expression and expression quantitative trait loci (eQTLs). Twenty-two of these 369 SNPs were nonsynonymous. For the remaining 347 SNPs, we checked whether they were in high LD with a nonsynonymous variant associated with TG:HDL-C at genome-wide significance. Using an  $r^2 > 0.8$  and a distance criterion of 500 kb as threshold, we identified 35 synonymous and intergenic SNPs in high LD with a nonsynonymous variant. To determine whether these 35 nonsynonymous loci in high LD with a synonymous or intergenic variant were identified by chance, we randomly sampled 347 SNPs with effect allele frequency (EAF)  $\geq 0.01$  and EAF  $\leq 0.99$  and quantified the number of nonsynonymous SNPs with an  $r^2 > 0.8$  and within 500 kb. Across 100 iterations of this process, we recovered 19.63 (95% confidence interval (CI): 12.47–28.53) nonsynonymous SNPs on average. This suggests that (1) the likelihood of recovering 35 nonsynonymous variants in high LD at random is very low ( $P < 0.01$ ) and (2) that many of these nonsynonymous SNPs are biologically relevant variants. Thus, we identified 57 total nonsynonymous loci (Supplementary Table 4). Three of these 57 SNPs lie within genes previously identified as likely causal genes through GWAS of other IR markers<sup>9–14</sup>. This includes *GCKR*, *FAM13A* and *JMJD1C*. *FAM13A* is closely tied to the regulation of adipogenesis and adipocyte function<sup>23,24</sup>. *GCKR* is known to regulate glucose metabolism<sup>25</sup>. Additionally, many of the identified nonsynonymous variants represent links between glucogenic and lipogenic metabolism. *APOA4*, *APOB* and *PNPLA2* regulate TG levels and are directly controlled by insulin<sup>26–29</sup>. Thus, many of the 57 nonsynonymous loci correspond to genes with prominent roles in nutrient metabolism.

### Of the 369 independent SNPs, 318 have not been previously reported for IR

To establish which of the 369 identified TG:HDL-C loci have previously unreported associations with IR, we carried out a conditional analysis incorporating the 130 variants from IR-associated traits reported by the Meta-Analyses of Glucose and Insulin-related Traits Consortium (MAGIC) studies (Supplementary Table 5)<sup>9–14</sup>. We ran SAIGE on the inverse normal transformed TG:HDL-C including the MAGIC variants as independent variables in the model in addition to the same covariates used in the original analysis. The estimated intercept of the LD score regression for this model was 1.4975, and the distribution of the test statistic was adjusted accordingly. In total, 322 of the 369 SNPs remained genome-wide significant ( $P < 5 \times 10^{-8}$ ) after conditional analysis. To confirm none of these 322 SNPs were previously reported for IR, we systematically reviewed previous studies of fasting insulin (FI), HOMA-IR, insulin sensitivity, insulin secretion or other surrogate measures of IR for the presence of these 322 SNPs. No associations were found between 319 of the 322 SNPs, while three SNPs (rs13107325, rs72959041 and rs8101064) were reported in the literature (Supplementary Table 6). Furthermore, rs3810291

was assumed to be a known signal because we could not verify whether this variant was in LD with rs200172871, a variant previously associated with fasting insulin adjusted for body mass index (BMI)<sup>9</sup>. Thus, of the 369 original SNPs reaching genome-wide significance for TG:HDL-C, 318 have not been previously reported for IR.

### **Of 369 independent loci, 114 are high-confidence IR-associated loci**

To verify the ability of the TG:HDL-C to capture known IR biology, we tested whether the TG:HDL-C encompassed the 130 IR-associated loci previously reported in the MAGIC studies<sup>9-14</sup> (Supplementary Table 5). Of the 130 loci, 127 were present in the UKBB. 92 of the 127 (72%) previously reported variants showed a false discovery rate (FDR)-adjusted  $P$  value  $< 0.05$  in the summary statistics of our study (Supplementary Tables 7 and 8). Fifty-seven of 127 variants (45%) reached genome-wide significance (Supplementary Table 8). Next, we tested whether any of the 369 independent TG:HDL-C variants met an FDR-adjusted  $P$  value  $< 0.05$  in the summary statistics of the MAGIC or Genetics of Insulin Sensitivity (GENESIS) consortia studies<sup>9-14,30</sup>. If a SNP was not available in the summary statistics, a proxy was used when available (Supplementary Table 9). In total, 114 of the 369 independent loci met an FDR-adjusted  $P$  value  $< 0.05$  in at least one of the analyzed traits related to IR. These 114 SNPs are thus high-confidence IR-associated loci having met genome-wide significance ( $P < 5 \times 10^{-8}$ ) for TG:HDL-C and an FDR-adjusted  $P$  value  $< 0.05$  in at least one independent study of an IR-related trait (Fig. 1a, Fig. 2a and Supplementary Table 2). In total, 111 of these 114 high-confidence loci overlapped specifically with fasting insulin, fasting insulin adjusted for BMI or HOMA-IR (Fig. 2b and Supplementary Tables 7, 10 and 11). Of these 114 loci, 72 have not been previously reported for IR.

Predicted causal genes have been previously associated with the constituent traits of metabolic syndrome (MetS) including obesity, waist-hip ratio (WHR), dyslipidemia and type 2 diabetes (T2D; Supplementary Table 6). Additionally, we examined how many of the 114 SNPs had a Bonferroni-adjusted  $P$  value  $< 0.05$  in other publicly available studies of metabolic traits<sup>9,31-38</sup>. Over 40% of these SNPs were significantly associated with low-density lipoprotein cholesterol (LDL-C, 47/114), T2D (62/114), WHR (69/114), ALT (61/114) and systolic blood pressure (SBP, 49/114; Supplementary Table 12). Although these 114 high-confidence IR-associated loci reached significance in the summary statistics of the MAGIC/GENESIS consortia, the remaining 255 SNPs not meeting significance remain promising targets for future study as well, given their association with numerous metabolic traits (Supplementary Table 6). Thus, although all 369 SNPs represent promising targets for IR biology, the 114 high-confidence IR-associated SNPs were prioritized for further analysis.

### **Phenome-wide association study identifies distinct effects within metabolic traits**

Although these 114 high-confidence IR-associated loci are associated with an increased TG:HDL-C ratio, whether they display uniform effects on other IR-associated phenotypes is unknown. We carried out a Phenome-wide association study (PheWAS) of the 114 SNPs on 14 traits related to metabolism, body composition and MetS-associated phenotypes in publicly available cohorts<sup>9,31-39</sup> (Supplementary Table 12). We then applied unsupervised

clustering and found five distinct subgroups of SNPs with variable effects on subsets of these traits (Fig. 3a, Supplementary Fig. 1 and Supplementary Tables 2 and 13a,b). The biological function of these clusters was interrogated using gene-set enrichment analysis in several databases (Supplementary Table 14) and named based on the significant biological processes identified.

Next, for each subgroup of SNPs, we created a polygenic-risk score (PRS) summing the dosages of the European individuals in the UKBB weighted by the effect sizes of the SNPs in MGI. Subsequently, we examined the ability of those PRSs to predict the constituent traits in the UKBB (Fig. 3b). We found that the subgroups had different patterns of effects on insulin-related traits such as BMI, WHR, serum lipids, NAFLD (measured as proton density fat fraction), and estimated glomerular filtration rate (eGFR). All subgroups were associated with increased TGs and lowered HDL, which was our primary phenotype. The insulin/growth group associated with increased LDL-C, SBP, WHR, T2D and alanine aminotransferase (ALT) but decreased BMI. The carbohydrate homeostasis subgroup had a similar pattern but associated with decreased T2D risk. The adipogenesis subgroup was also similar to the insulin/growth subgroup but had nonsignificant effects on LDL-C and WHR. The lipid homeostasis and brain processes subgroups associated with increased BMI, T2D and ALT. However, the lipid homeostasis subgroup and brain processes associated with increased and decreased WHR, respectively. The differential effects of each variant across traits recapitulate the known complexity of disease within insulin-resistant individuals.

### High-confidence IR loci enrich for insulin-related biology

The 114 high-confidence IR-associated SNPs were annotated using DEPICT to identify the enrichment in tissues, cell types and gene sets (Fig. 4 and Supplementary Table 15). Consistent with previous findings for SNPs associated with  $FI^9$ , the 114 loci show robust enrichment in adipose tissues (Fig. 4a,b). Enriched physiological systems correspond to those traditionally affected by MetS including the cardiovascular system (aortic/heart valves), digestive system (liver/pancreas), musculoskeletal system (joints/synovial membrane) and the female urogenital system (myometrium/uterus/fallopian tubes; Fig. 4c). Interestingly, unlike previous genetic studies of IR, the liver was identified in tissue enrichment. As IR is the result of an altered response of hepatocytes to insulin and IR is critical in the pathogenesis of nonalcoholic fatty liver disease (NAFLD), this association points toward a shared genetic underpinning.

DEPICT was also used to perform pathway analysis that identified three distinct subnetworks related to growth, metabolism and abnormal lipid homeostasis (Fig. 5 and Supplementary Table 16). Insulin is a key regulator of energy metabolism and growth. The growth-related subnetwork is enriched for several pro-growth protein–protein interaction networks including *EP300*, *CREBBP*, *SMAD1*, *SMAD3*, *ESR1* and *IGF1R*. Another subnetwork present contains nodes related to the metabolism of proteins and phosphorus-containing compounds. Protein synthesis is traditionally promoted by insulin in an insulin-sensitive state<sup>40</sup>. However, this anabolic process can be impaired in individuals with obesity<sup>41</sup>. Finally, the last subnetwork centers on abnormal lipid homeostasis and encompasses established links between glucogenic metabolism, lipogenic metabolism and

body mass. Collectively, these enrichment studies highlight the biological mechanisms underlying IR. Using the 369 TG:HDL-C SNPs provided increased power further highlighting enrichment in these pathways and tissues in addition to others (Supplementary Tables 15 and 16).

### Thirty-one high-confidence loci have sex-specific effects

To evaluate whether any of the 114 high-confidence IR-associated loci displayed sex-specific effects, we ran GWASs for TG:HDL-C in males and females separately. The estimated intercept of the LD score regression was 1.1985 and 1.3047 for males and females, respectively.

The relative distributions of the test statistics were adjusted for each sex. The GWAS results were then meta-analyzed (Supplementary Table 17). Seventy-six of the 369 independent TG:HDL-C loci showed a statistical heterogeneous effect ( $P_{\text{Het}} < 0.05$ ), while 31 of the 114 high-confidence IR-associated loci met this criterion. Interestingly, 24 of 31 (77%) high-confidence IR-associated loci with sex-specific effects displayed a stronger effect on TG:HDL-C in females when compared to males. The top loci showed a stronger sex-specific effect in females mapped to genes including *KLF14*, *ZCCHC8*, *LINC01625* and *RSPO3*. Conversely, loci mapping to *LPL/SLC18A1*, *LOC646736*, *ARL15* and *FNIP1* showed a stronger effect in males. The 24 SNPs with stronger sex-specific effects in females were enriched for loci significantly associated with WHR adjusted for BMI (FDR-adjusted  $P = 7.59 \times 10^{-20}$ ). One of these loci (rs10260148) maps to the transcription factor *KLF14* and also shows the strongest sex-specific effect in females. Previous studies of other SNPs mapping to *KLF14* have reported sex-specific associations with metabolic traits including T2D, WHR, TGs, HDL-C and LDL<sup>32,33,42–44</sup>. The stronger association in females is hypothesized to be driven by modulation of *KLF14* expression, rather than through hormonal means<sup>42,45</sup>. The mechanisms underlying the sex-specific effects in the remaining 30 loci are less well characterized. Further study of these sex-specific, high-confidence IR-associated loci may help explain the observed differences in metabolic phenotypes between men and women.

### Eleven TG:HDL-C SNPs are identified in non-European ancestries

To determine the role of the 114 high-confidence IR-associated SNPs in non-European ancestries, we carried out a GWAS of the TG:HDL-C for individuals of South Asian (SAS), African (AFR) and Chinese (CHI) ancestry in the UKBB (Supplementary Table 18). The estimated genomic inflation factor was 1.0494 (SAS), 1.0393 (AFR) and 1.0007 (CHI), and  $P$  values were adjusted accordingly. We identified the independent SNPs using COJO for the GWAS results of the SAS and AFR cohorts. As the CHI cohort is below the recommended sample size for applying COJO, a distance criterion of 500 kb was used instead. Between the three ancestries, we identified 11 total loci (SAS:6, AFR:4 and CHI:1; Supplementary Table 19). All the 11 loci identified in non-European ancestries were located within 500 kb from one of the 369 independent loci for TG:HDL-C, and 6 of the 11 loci were located within 500 kb of one of the high-confidence IR-associated loci. This suggests that these loci contribute to IR-associated phenotypes across varied ancestries. The majority (7 of 11) of the SNPs identified from non-European ancestries were in LD ( $r^2 > 0.1$ ) with the nearest



of the European ancestry-derived independent loci for TG:HDL-C (Supplementary Table 20). The four loci (rs15285, rs326, rs3135506 and rs12721054) that failed to meet the  $r^2$  threshold for LD did not associate with insulin traits in the MAGIC ( $P > 0.05$ ), so they do not meet the criteria for being high-confidence IR-associated loci. Therefore, there were no statistically significant ancestry-specific high-confidence loci identified.

### High-confidence IR-associated PRS associates with cardiometabolic traits

To evaluate the joint effect of the high-confidence 114 IR-associated loci overall on the risk of disease, we created a PRS and tested its association with Phecodes on 51,550 individuals of European ancestry of the MGI (Fig. 6 and Supplementary Table 21). The PRS was significantly associated with phenotypes used to identify MetS including hyperglyceridemia ( $P_{\text{adj}} = 8.91 \times 10^{-41}$ ), hyperlipidemia ( $P_{\text{adj}} = 1.26 \times 10^{-30}$ ) and hypertension ( $P_{\text{adj}} = 1.27 \times 10^{-17}$ ). Additionally, the PRS was associated with well-established sequelae of MetS including coronary atherosclerosis ( $P_{\text{adj}} = 1.27 \times 10^{-7}$ ) and chronic liver disease/cirrhosis ( $P_{\text{adj}} = 1.43 \times 10^{-7}$ ). Obesity failed to reach significance ( $P = 1$ ). Other notable associations included disorders of lipid metabolism ( $P_{\text{adj}} = 7.01 \times 10^{-31}$ ) and T2D ( $P_{\text{adj}} = 1.07 \times 10^{-19}$ ). The association of the PRS with IR and MetS-related sequelae in an independent cohort further solidifies the contribution of these loci to metabolic disease and its subsequent morbidity.

## Discussion

In our study we leveraged the readily measurable IR marker TG:HDL-C to identify previously unreported loci, genes and pathways central to IR pathology. Specifically, we performed a GWAS of TG:HDL-C in 402,398 Europeans within the UKBB. We identified 369 independent SNPs that encompass 51 of 130 loci previously reported for IR. Furthermore, 57 of the previously reported IR loci reached genome-wide significance in our study and this number increases to 92 if we consider variants with an FDR-adjusted  $P$  value  $< 0.05$ . Of the 369 loci independently associated with TG:HDL-C, 318 have not been previously reported for IR, 22 are nonsynonymous and 35 are in high LD with a nonsynonymous variant. In total, 114 of these SNPs (72 previously unreported, 6 nonsynonymous and 9 in high LD with a nonsynonymous variant) met an FDR-adjusted  $P$  value  $< 0.05$  in at least one other study of a marker of IR, thus making them high-confidence IR-associated loci. The 114 high-confidence SNPs explain 3.2% of the variance in TG:HDL-C levels and associate with IR-related traits in an independent cohort. Furthermore, these high-confidence loci enrich for new and established tissues, cell types, pathways and genes that are relevant to the patho-physiology of IR. These loci will help guide future studies and contribute to the mechanistic understanding of IR.

Consistent with the 2021 study by MAGIC on IR<sup>9</sup>, our findings confirm relationships between IR-associated loci and adipocyte biology, the endocrine system and pathways related to growth and cancer. Additionally, our study uncovers relationships with genes expressed in the liver and female reproductive system. The liver produces fat from glucose in response to insulin and insulin-resistant individuals can produce excess fat leading to hepatic steatosis. Liver-related genes represented by the 114 high-confidence IR-associated

loci include key metabolic enzymes such as alcohol dehydrogenase (*ADH4*), apolipoprotein H (*APOH*), ketohexokinase (*KHK*) and glycogen synthase kinase (*GLYCTK*). Additionally, *TM6SF2*, a known regulator of lipoprotein excretion, may lead to retention of fat in the liver and increased IR when mutated<sup>46</sup>. Mechanistically, dysregulation of these metabolic enzymes may contribute to IR. The association between IR and tissues from the female reproductive system is consistent with findings from MAGIC's pathway enrichment highlighting female infertility and reproductive structure development. Genes identified by DEPICT as being related to the female reproductive system include many regulators of metabolism including alcohol dehydrogenase (*ADH5*), carbamoylphosphate synthetase, aspartate transcarbamoylase, and dihydroorotase (*CAD*), pyridoxal-dependent decarboxylase domain containing 1 (*PDXDC1*) and phosphoinositide-3-kinase regulatory subunit 1 (*PIK3RI*). As IR is mechanistically linked with polycystic ovarian syndrome (PCOS), it is possible that these enzymes may contribute to the pathogenesis of PCOS as well.

IR is also closely tied to MetS. Previous genetic studies outside of the UKBB identified 28 loci associated with MetS<sup>47–50</sup>. Notably, our 114 high-confidence IR-associated loci capture 7 of these 28 loci (25%; Supplementary Table 22). Furthermore, 17 of these 28 loci (61%) reach genome-wide significance in our analysis (Supplementary Table 23). The 7 MetS loci captured within our 114 high-confidence IR-associated loci map to *GCKR*, *LOC157273*, *MLXIPL* and *LPL*. Numerous variants mapping to *GCKR* have been associated with metabolic phenotypes as this gene controls glucose utilization in many tissues. *MLXIPL* encodes the carbohydrate response element-binding protein (ChREBP). ChREBP is a transcription factor expressed primarily in the liver that mediates the conversion of glucose into lipids<sup>51–53</sup>. ChREBP also regulates many metabolic enzymes including *GCKR*<sup>51,54</sup>. Additionally, a study of 291,107 individuals from the UKBB sought to examine loci associated with MetS as defined by harmonized National Cholesterol Education Program criteria<sup>55</sup>. Of the 93 loci this study identified, 34 (37%) of these loci overlap with our 114 high-confidence IR-associated SNPs (Supplementary Table 24). Furthermore, 69 of the 93 loci (74%) reach genome-wide significance in our analysis and 90 associate with an FDR-adjusted  $P < 0.05$  (Supplementary Table 25). Our 114 high-confidence loci also associate with the phenotypes used to diagnose MetS in an external cohort with the notable exception of obesity (Fig. 6). This is interesting as (1) healthy weight individuals develop MetS and have poor outcomes, (2) not all individuals with obesity develop MetS, and (3) it is possible to be both obese and metabolically 'healthy'<sup>56–58</sup>. Collectively, the high extent of correlation between the TG:HDL-C-associated loci and MetS loci provides further evidence for the role of IR in MetS.

Our study also identifies five subgroups of variants with distinct effects across metabolic diseases. The 'insulin/growth' subgroup includes variants mapping to genes including *INSR*, *VEGFA*, *FGFR2* and *BMP7*, which are key regulators of growth and development. *INSR* encodes the insulin receptor, which is the primary mediator of the cellular response to insulin<sup>59</sup>. VEGFs are highly conserved proangiogenic factors<sup>60</sup>. *FGFR2/BMP7* have critical roles in the adipogenesis and the maintenance of adipose tissue<sup>61,62</sup>. This known biology in addition to the positive association with WHR suggests a role for these variants in fat distribution. Other variants within the insulin/growth subgroup map



to genes including *RSPO3*, *FAM13A* and *PPARG* that when altered may result in subcutaneous lipodystrophy<sup>63,64</sup>. Individuals with subcutaneous lipodystrophy have high central deposition of fat, liver fat and cholesterol<sup>65</sup>. Dysregulating the function of these genes impairs the ability to store energy as subcutaneous fat leading to IR at a lower BMI. This same pattern but with a decreased risk of T2D and glucose can be seen in the ‘carbohydrate homeostasis’ subgroup, with variants mapping to *MLXIPL*, *FGF21* and *GCKR*. *MLXIPL* and *FGF21* may have a role in reducing the deposition of TGs and increasing energy metabolism<sup>66,67</sup>. *GCKR* encodes glucokinase regulatory protein that inhibits glucokinase, the enzyme that phosphorylates glucose to catalyze the first step in hepatic glycogen synthesis/glycolysis and  $\beta$ -cell insulin secretion. It has been shown that the P446L *GCKR* variant increases glycolytic flux to decrease serum glucose, increase glycolysis and increase hepatic de novo lipogenesis<sup>68</sup>. *GCKR* variants have previously been shown to confer decreased risk of T2D but increased risk of other metabolic sequelae including NAFLD, gout and familial combined hyperlipidemia<sup>69</sup>. Both the ‘lipid homeostasis’ and ‘brain processes’ subgroups decrease LDL-C while increasing BMI, T2D and ALT. These groups may represent a diversion of energy from cholesterol to TG formation.

Key limitations of our study include (1) that UKBB participants consist of primarily middle-aged and older Europeans, and (2) that we use TG:HDL-C ratio as a surrogate measure of IR. While the ease of TG:HDL-C measurement facilitates the collection of large sample sizes, it is limited by the strength of the relationship between TG:HDL-C and the physiological state of IR. Studies examining the correlation between TG:HDL-C and IR in different populations have yielded mixed results<sup>70–75</sup>. Mechanistic studies of IR have shown close ties between dyslipidemia and IR<sup>4</sup>. To account for the likely imperfect IR–TG:HDL-C relationship and increase our confidence in capturing true IR pathology, we prioritized the 114 SNPs that met significance in other studies of IR markers. These SNPs represent high-confidence loci that narrow in on the physiology of IR within the broader context of nutrient excess and MetS. As the median age of the primary cohort is 69 and the bulk of analyses were performed in individuals of European ancestry, the relevance of these loci to younger, non-European individuals may be limited. Our subgroup analyses in the non-European individuals of the UKBB were significantly limited in scale and thus underpowered in comparison to the European cohort. Thus, future studies should evaluate the contribution of these 114 high-confidence IR-associated loci to disease in younger and more ancestrally diverse cohorts than are available in the UKBB.

In conclusion, our study leverages data from 402,398 individuals to characterize the genetics of IR. We identify numerous loci, pathways and genes that were previously unreported for IR. We define groups of genes with variable effects on metabolic traits that better define disease subtypes. Future studies of these implicated loci and genes will help us better understand the causes of IR and its relationship to other metabolic diseases.

## Methods

### Ethics statement

The UKBB protocols were approved by the National Research Ethics Service Committee. The MGI protocols were approved by the University of Michigan Medical School Institutional Review Board. Analyses in the UKBB were conducted under approved project 18120 (E.K.S.). Participants signed written informed consent, specifically applicable to health-related research. All ethical regulations were followed.

### Data and genotyping

The UKBB contains genotype, clinical and demographic data of over 400,000 individuals aged 40–69 years at the time of study recruitment. Protocols for participant genotyping, data collection and quality control have been described previously<sup>20,76</sup>. In brief, participants were genotyped on one of two purpose-designed arrays (UK BiLEVE Axiom Array ( $n = 50,520$ ) and UKBB Axiom Array ( $n = 438,692$ )) with 95% maker overlap. The Haplotype Reference Consortium (HRC) was used as a reference panel to phase and impute the data. EasyQC (version 9.2) was used for quality using an imputation quality cutoff of 0.85.

The MGI is a hospital-based cohort containing genetic data and clinical phenotypes<sup>77</sup>. Participants were genotyped using the University of Michigan Advanced Genomics Core on one of two customized versions of the Illumina Infinium CoreExome-24 bead array platform. Imputation has been previously described<sup>78</sup>. Briefly, genotypes were imputed to both the HRC reference panel and the Trans-Omics for Precision Medicine reference panel.

### GWAS and COJO

The TG:HDL-C was calculated based on the serum TGs and HDL-C at the time of enrollment, respectively, in the European individuals from the UKBB. European ancestry was genetically defined. No statistical method was used to predetermine the sample size. First, a subgroup of individuals was chosen as European using the field 22006 of the UKBB ( $n_1 = 409,605$ ). This subgroup consisted of a list of participants who self-identified as ‘White British’ with similar genetic ancestry based on PCs. The individuals of the UKBB excluded at the previous step were then projected, based on their genotype data, on a common ancestry space together with a reference sample of individuals of different ancestries using TRACE<sup>79</sup>. The application of a k-nearest neighbors algorithm by TRACE classified a further subset of individuals as European ( $n_2 = 52,702$ ). Our primary cohort consisted of the sum of the two groups ( $n_1 + n_2 = 462,307$ ). Europeans were included in the analysis if their records did not show any missing information about TGs, HDL-C, age, sex and PCs 1–10, and if the genetic data were available ( $n = 402,398$ ). We fit a linear mixed model using SAIGE<sup>21</sup> (version 0.29) with rank-based inverse normal transformed TG:HDL-C as the dependent variable and age, age squared, sex and PCs 1–10 and SNPs as independent variables. The distribution of the TG:HDL-C was normal given the rank-based inverse normal transformation. The effects of the SNPs were tested under an additive genetic model. We excluded variants with an imputation cutoff  $<0.85$  or minor allele count  $<3.5$ . The LD score regression intercept was quantified using ldsc<sup>80</sup> (version 1.0.1) and used to adjust the test statistic of our GWAS for inflation. After excluding multiallelic or ambiguous

SNPs, INDELS, variants with a minor allele frequency  $<0.01$  and variants not available in the MGI, we ran a COJO analysis<sup>22</sup> using GCTA (version 1.91.2) to extract independent SNPs having  $r^2 < 0.1$ . All data are presented for the TG:HDL-C increasing allele. Another linear mixed model for the rank-based inverse normal transformed TG:HDL-C was fit using the same variables and also including the 11–20 PCs. The LD score regression intercept for the latter model was estimated as described above.

### Independent synonymous and intergenic SNPs in high LD with nonsynonymous variants

Starting from the results of the GWAS, we filtered out variants 1) having  $P > 5 \times 10^{-8}$ , 2) having a distance  $>500$  kb from lead-independent synonymous and intergenic SNPs, 3) having an imputation cutoff  $<0.85$ , and 4) which were multiallelic. We annotated the remaining variants using ANNOVAR<sup>81</sup> (build hg19, dbSNP150) and calculated the  $r^2$  between nonsynonymous variants and the lead-independent synonymous or intergenic SNPs. Two variants were considered in high LD if  $r^2 > 0.8$ . In case an independent synonymous or intergenic SNP was in high LD with more nonsynonymous variants, only the nonsynonymous variant with the highest  $r^2$  was reported. To check whether the total number of nonsynonymous loci in high LD was due to chance, we randomly sampled a number of SNPs throughout the genome equal to the number of intergenic and synonymous SNPs in our study 100 times. Rare variants (MAF  $< 0.01$ ) were excluded. For each iteration, we treated the randomly selected SNPs as if they were lead-independent variants, repeated the process previously described with the exception of the  $P$  value filter and quantified the number of nonsynonymous SNPs in high LD. Mean and 2.5 and 97.5 quantiles were estimated from the empirical distribution of the number of nonsynonymous SNPs in high LD.

### Variant and gene annotation

The nearest gene to each variant was assigned using ANNOVAR. We also reported prioritized genes by DEPICT (FDR  $< 0.05$ ), whether the gene is expressed in adipose subcutaneous, adipose visceral, adrenal gland, liver, muscle-skeletal, pancreas and uterus, whether the SNP had an eQTL with the indicated gene in subcutaneous fat, visceral fat, internal mammary artery, liver, aortic wall, skeletal muscle and blood and whether the SNP was nonsynonymous (or in high LD with a nonsynonymous variant). A gene was considered expressed in a tissue if the median expression of the gene in the tissue was greater than twice the median of the expression of the gene across all the tissues. Median expression of the genes was obtained by the Genotype-Tissue Expression (GTEx) project<sup>82</sup> (v8). A SNP was considered in eQTL with a gene in a given tissue if the adjusted  $P$  value provided by STARNET<sup>83</sup> (dbGaP accession [phs001203.v1.p1](#)) was less than 0.05.

For each variant, we assigned the most likely causal gene using the following algorithm. First, nonsynonymous variants were assigned to their constituent gene. If the SNP was synonymous or intergenic and not in LD with a nonsynonymous variant, we constructed a candidate list consisting of the nearest gene and the genes prioritized by DEPICT. Subsequently, if DEPICT prioritized the nearest gene, this gene was selected. Alternatively, each gene within the gene list was assigned 1 point (maximum 16 possible points) for (1) expression in adipose subcutaneous, adipose visceral, adrenal gland, liver, muscle-skeletal,

pancreas and uterine tissue, or (2) an eQTL in adipose subcutaneous, adipose visceral, adrenal gland, liver, muscle-skeletal, aortic wall, blood or internal mammary artery tissue. The gene with the most points was selected, and in case of ties, both genes were reported.

### **Overlap between the variants associated with TG:HDL-C ratio and other genetic studies**

To identify high-confidence IR-associated loci, we determined which independent TG:HDL-C SNPs were associated with insulin-related traits from the MAGIC and GENESIS Consortia<sup>9–14,30</sup>. We chose studies where at least one trait under investigation was a quantity related to insulin, IR, insulin sensitivity or insulin secretion, and the summary statistics for individuals of European ancestry were available. The analyzed traits were fasting insulin<sup>12–14</sup>, fasting insulin adjusted for BMI<sup>9,12,13</sup>, HOMA-IR<sup>13,14</sup>, HOMA-IR adjusted for BMI<sup>13</sup>, insulin sensitivity index<sup>11</sup>, modified Stumvoll insulin sensitivity index<sup>10</sup>, insulin sensitivity measured by hyperinsulinemic-euglycemic clamp<sup>30</sup>, insulin sensitivity measured by hyperinsulinemic-euglycemic clamp adjusted for BMI<sup>30</sup>, corrected insulin response<sup>11</sup> and overall insulin response to glucose estimated as area under the curve for insulin over a total area under the curve for glucose<sup>11</sup>. If a SNP associated with TG:HDL-C was not available in the summary statistics of a study, we identified SNPs with an  $r^2 > 0.8$  in UKBB and used the one with the highest  $r^2$  that was also present in the study as a proxy. To verify which allele of a proxy paired with the effect allele of a missing SNP, we used LDlink<sup>84</sup>. Once the proxies were identified, the  $P$  values of proxies and available SNPs in the summary statistics were FDR-adjusted. A SNP was considered to overlap with a trait of the other study if its FDR-adjusted  $P$  value was less than 0.05 in the other study.

### **Overlap between the reported variants related to insulin in the MAGIC studies and the SNPs associated with TG:HDL-C**

To check whether the previously reported insulin-related variants<sup>9–14</sup> from the MAGIC Consortium associated with TG:HDL-C, we examined the summary statistics of those variants in our GWAS. Because the reported variants of the GENESIS Consortium study did not reach genome-wide significance, the GENESIS Consortium study was excluded. All other reported variants were available except for rs73343765, rs200172871 and rs200678953, which were derived from ref. 9, a multi-ancestry meta-analysis. No proxy SNPs were found through LDlink; therefore, we continued the analysis with the remaining 127. We extracted the summary statistics of the available variants from our GWAS results and then adjusted the  $P$  values using FDR. A variant was considered to overlap with the TG:HDL-C if its FDR-adjusted  $P$  value in our GWAS was less than 0.05.

### **Review of the unreported SNPs for IR traits in previous studies**

We carried out a conditional analysis of the inverse normal transformed TG:HDL-C for the TG:HDL-C independent SNPs running SAIGE (version 0.29), where we included age, age squared, sex, PCs 1–10 and the dosages of the 127 previous reported variants associated with fasting insulin, HOMA-IR or any other index used to measure IR, insulin sensitivity or insulin secretion from the MAGIC<sup>9–14</sup> as independent variables of the model. The reported variants of the GENESIS Consortium study were not included as none of them reached genome-wide significance. The independent SNPs reaching genome-wide significance after the conditional analysis were considered potential unreported loci for IR. To confirm the

SNPs were unreported, we systematically reviewed previous studies of fasting insulin, HOMA-IR, insulin sensitivity, insulin secretion or other surrogate measures of IR for the presence of those SNPs and the nonsynonymous SNPs with an  $r^2 > 0.8$  to the lead SNPs.

### Enrichment analysis of the high-confidence SNPs associated with the TG:HDL-C

The high-confidence IR-associated loci were analyzed using DEPICT (version 1, release 173) to highlight the enrichment of tissues, cell types and gene sets and carry out a pathway analysis. Tissue and gene-set enrichments with an FDR  $< 0.20$  and pathways with a  $P$  value  $< 1 \times 10^{-4}$  were considered statistically significant. The results of the pathway analysis were plotted using Cytoscape<sup>85</sup> (version 3.7.1).

### Associations between the high-confidence SNPs and cardiometabolic traits

We assessed whether the high-confidence IR-associated variants associated with cardiometabolic traits related to IR. We chose the Global Lipids Genetics Consortium (GLGC)<sup>31</sup> for TG, HDL-C and LDL-C; the GIANT<sup>32</sup> Consortium for BMI and WHR adjusted for BMI; the GOLDPlus<sup>38</sup> Consortium for NAFLD measured as Pdff; the DIAGRAM<sup>33</sup> Consortium for T2D; the MAGIC<sup>9</sup> for fasting glucose adjusted for BMI (glucose) and FI adjusted for BMI (insulin) and the summary statistics from refs. 34–37,39, respectively, for alanine aminotransferase (ALT), myocardial infarction (MI), SBP, eGFR and PCOS. We calculated a  $z$  score for all the available variants in a study. If a variant was not available, we set the  $z$  score equal to 0. To cluster variants, we applied complete-linkage hierarchical clustering using Pearson correlation as distance metric. We ran a GWAS for the rank-based inverse normal transformed TG:HDL-C in MGI to extract the effect sizes for the 114 high-confidence loci using SAIGE. Age, age squared, sex and 1–10 PCs were included in the model. For each cluster of variants, a PRS was created summing the dosages of the unrelated European individuals of the UKBB weighted by the effect sizes from MGI. In case of relatedness, only one participant per family was randomly chosen to create the PRS. Relatedness up to the second degree was estimated using KING<sup>86</sup> (version 2.2.6). Individuals were subdivided based on the quartiles of PRS, and only the participants in the top quartile were compared to the bottom quartile for associations with traits adjusted for sex, age, age squared and PCs 1–10. Outcomes were reported in s.d. and log odds ratio for continuous and binary traits, respectively. A gene-set enrichment analysis was performed on the genes of each cluster using FUMA<sup>87</sup> and used to assign a name to each cluster. Only the relevant databases for the name assignments were reported.

### TG:HDL-C PRS in the MGI

The high-confidence IR-associated SNPs were combined in a PRS, which was calculated by summing the dosages of 51,550 unrelated individuals of European ancestry from MGI weighted by the effect sizes of the loci from UKBB. A rank-based inverse normal transformation was applied to the PRS. We studied the association between the PRS and the Phecodes in MGI fitting Firth's logistic regression model using the PheWAS R package (version 0.99.5). The Phecodes were created from International Classification of Diseases (ICD) codes. Age, age squared, sex and the first ten PCs were included as predictors. An association was considered significant using a Bonferroni level significance adjusted for the number of traits tested ( $\alpha^* = 0.05/1,659 = 3 \times 10^{-5}$ ).

### Variance explained by PRS in MGI

To estimate the percentage of variance explained by the high-confidence TG:HDL-C loci, we fit a linear regression with inverse normal transformed TG:HDL-C as the outcome and the inverse normal transformed 114 SNP PRS from above and PCs 1–10 as the predictors. The adjusted  $r^2$  of the model was used as an estimate of the explained variance.

### Sex-specific analysis

To verify whether the SNPs might have a heterogeneous effect between male and female individuals, we carried out a sex-stratified GWAS separately for males ( $n = 185,749$ ) and females ( $n = 216,649$ ) in the UKBB using SAIGE. The outcome of the models was the inverse normal transformed TG:HDL-C, and the predictors were age, age squared and the 1–10 PCs. In both the GWASes, the intercept from the LD score regression was used to adjust the  $P$  values for population stratification. The results were then meta-analyzed using METAL<sup>88</sup> (28 August 2018 release). A SNP was considered to have a heterogeneous effect if the heterogeneous  $P$  value of Cochran's  $Q$  test was less than 0.05. The genes of the SNPs that showed sex-heterogeneous effect were annotated separately by sex using FUMA<sup>87</sup>.

### Non-European ancestry analysis in the UKBB

We carried out a GWAS of the TG:HDL-C for the individuals of SAS ( $n = 8,158$ ), AFR ( $n = 6,632$ ) and CHI ( $n = 1,300$ ) ancestries in the UKBB. We used the same outcome and predictors described for the GWAS of the Europeans, and the same quality control was applied for both the individuals and the variants. For each GWAS, we estimated the genomic inflation factor and used it to adjust the  $P$  values. The independent SNPs for the SAS and AFR ancestries were extracted using COJO. The independent SNPs for the CHI ancestry were extracted using a 500 kb distance criterion, given that the small sample size of the cohort ( $n < 4,000$ ) did not meet the recommended size to apply COJO. The  $r^2$  was calculated among the independent SNPs having a distance  $< 500$  kb across the different ancestries using LDlink. When more independent SNPs had a distance  $< 500$  kb within the same ancestry, the closest SNP to the hits of the other ancestries was chosen. LD between two SNPs was defined as  $r^2 > 0.1$ .

### Statistics and reproducibility

All significant variants in a UKBB cohort reached genome-wide significance ( $P < 5 \times 10^{-8}$ ). To be reported as significant in a non-UKBB study, a variant must reach an FDR-adjusted  $P$  value  $< 0.05$ . We adjusted the  $P$  values using the Benjamini–Hochberg FDR procedure.  $Z$  scores were used to visualize the significance of variants in external studies.  $Z$  scores were calculated from unadjusted  $P$  values across all reported variants. PRSs were calculated for the UKBB as described in the Methods section. The effect size for each PRS represents the association in s.d. and log odds ratio for continuous and binary traits, respectively. PRS significance in the UKBB was determined using linear and logistic regression models. To identify significant tissues, cell types and pathways, an FDR-adjusted cutoff of 0.20 was used. A cutoff of  $P < 1 \times 10^{-4}$  was used for DEPICT-enriched gene sets. A PRS was calculated for the MGI, as described in the Methods. Associations between the PRS and MGI Phecodes were assessed using Firth's logistic regression model. A Bonferroni-adjusted



cutoff ( $\alpha = 3 \times 10^{-5}$ ) on a  $-\log_{10}$  scale was used to assess significance. Please see the Methods section for more comprehensive descriptions of the statistical methods used.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgements

The authors would like to acknowledge the MGI participants, Precision Health at the University of Michigan, the University of Michigan Medical School Central Biorepository and the University of Michigan Advanced Genomics Core for providing data and specimen storage, management, processing and distribution services, as well as the Center for Statistical Genetics in the Department of Biostatistics at the School of Public Health for genotype data curation, imputation and management in support of the research reported in this publication. A.O., A.K., A.P., X.D., Y.C., K.C.C., C.R., P.P., V.L.C., B.D.H. and E.K.S. supported in part by R01 DK106621 (to E.K.S.), R01 DK107904 (to E.K.S.), R01 DK128871 (to E.K.S.), R01 DK131787 (to E.K.S.) and/or the University of Michigan Department of Internal Medicine and/or The University of Michigan MBioFAR Award. R.J.R. is supported by F30 CA275039-02. H.B. is supported by F30 CA257292. Analyses in the UKBB were conducted under approved project 18120 (to E.K.S.). The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

## Competing interests

V.L.C. received grant funding from KOWA and AstraZeneca. The Regents of the University of Michigan and E.K.S. have a pending patent on the use of systems and methods for analysis of samples associated with IR and related conditions. The rest of the authors do not have any conflicts of interest.

## Data availability

GWAS results from this study are available at the GWAS Catalog (study accessions GCST90295949–GCST90295954, all intervening numbers). UKBB genomic and phenotypic data supporting this publication are available upon application (<https://ukbiobank.ac.uk>). MGI individual-level data are not currently available to the public due to patient privacy requirements. Otherwise, all data used to generate figures can be found in supplementary tables, source data or in the above publicly available datasets. Source data are provided with this paper.

## References

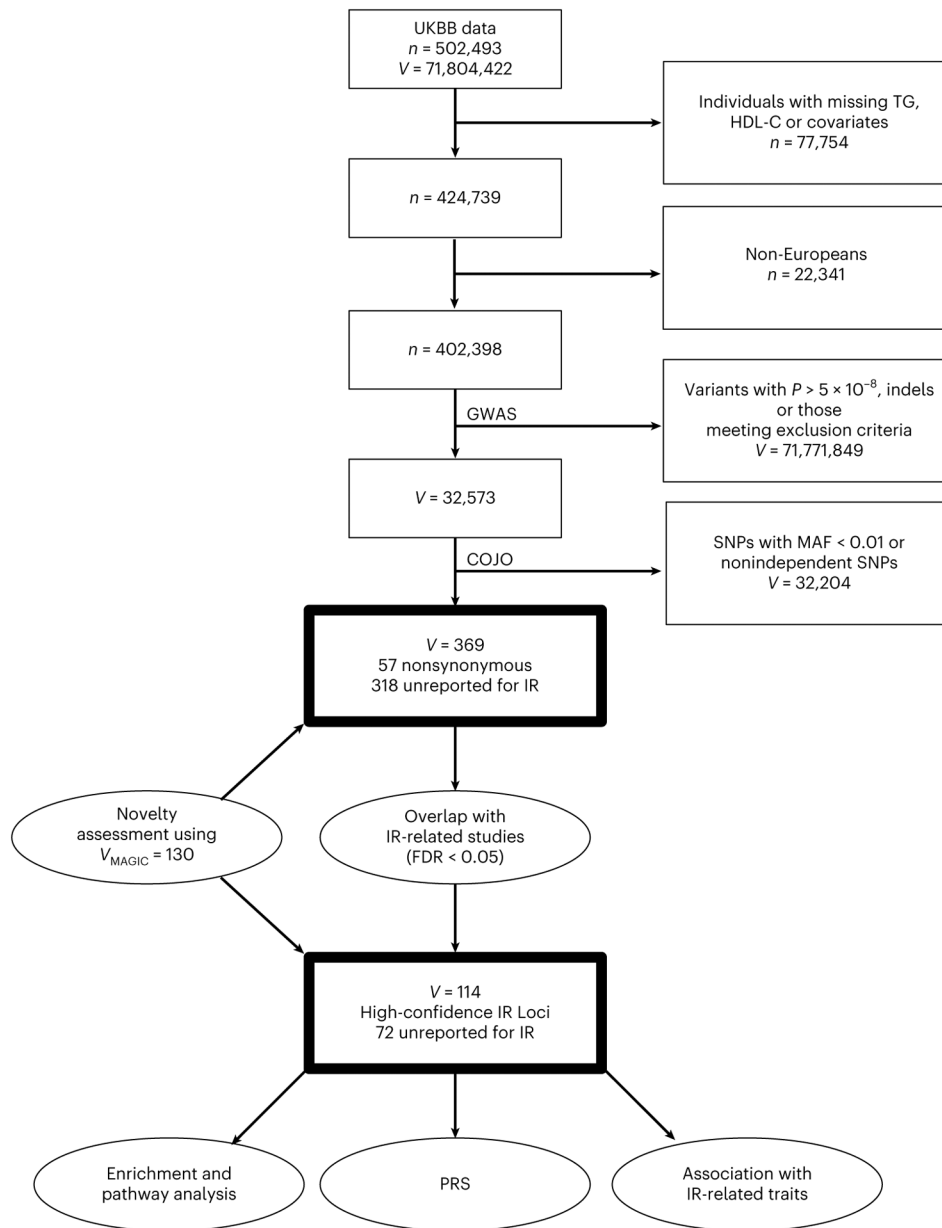
1. Brown AE & Walker M Genetics of insulin resistance and the metabolic syndrome. *Curr. Cardiol. Rep* 18, 75 (2016). [PubMed: 27312935]
2. Melvin A, O’Rahilly S & Savage DB Genetic syndromes of severe insulin resistance. *Curr. Opin. Genet. Dev* 50, 60–67 (2018). [PubMed: 29477938]
3. Mundi MS et al. Evolution of NAFLD and its management. *Nutr. Clin. Pract* 35, 72–84 (2020). [PubMed: 31840865]
4. Ormazabal V et al. Association between insulin resistance and the development of cardiovascular disease. *Cardiovasc. Diabetol* 17, 122 (2018). [PubMed: 30170598]
5. Lee JM, Okumura MJ, Davis MM, Herman WH & Gurney JG Prevalence and determinants of insulin resistance among U.S. adolescents: a population-based study. *Diabetes Care* 29, 2427–2432 (2006). [PubMed: 17065679]
6. Ren X et al. Association between triglyceride to HDL-C ratio (TG/HDL-C) and insulin resistance in Chinese patients with newly diagnosed type 2 diabetes mellitus. *PLoS ONE* 11, e0154345 (2016). [PubMed: 27115999]

7. Bonora E et al. Homeostasis model assessment closely mirrors the glucose clamp technique in the assessment of insulin sensitivity: studies in subjects with various degrees of glucose tolerance and insulin sensitivity. *Diabetes Care* 23, 57–63 (2000). [PubMed: 10857969]
8. Stühlinger MC et al. Relationship between insulin resistance and an endogenous nitric oxide synthase inhibitor. *JAMA* 287, 1420–1426 (2002). [PubMed: 11903029]
9. Chen J Meta-Analysis of Glucose and Insulin-related Traits Consortium (MAGIC) et al. The trans-ancestral genomic architecture of glycemic traits. *Nat. Genet* 53, 840–860 (2021). [PubMed: 34059833]
10. Walford GA et al. Genome-wide association study of the modified Stumvoll insulin sensitivity index identifies BCL2 and FAM19A2 as novel insulin sensitivity loci. *Diabetes* 65, 3200–3211 (2016). [PubMed: 27416945]
11. Prokopenko I et al. A central role for GRB10 in regulation of islet function in man. *PLoS Genet.* 10, e1004235 (2014). [PubMed: 24699409]
12. Scott RA et al. Large-scale association analyses identify new loci influencing glycemic traits and provide insight into the underlying biological pathways. *Nat. Genet* 44, 991–1005 (2012). [PubMed: 22885924]
13. Manning AK et al. A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. *Nat. Genet* 44, 659–669 (2012). [PubMed: 22581228]
14. 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). [PubMed: 20081858]
15. Iwani NA et al. Triglyceride to HDL-C ratio is associated with insulin resistance in overweight and obese children. *Sci. Rep* 7, 40055 (2017). [PubMed: 28059134]
16. McLaughlin T et al. Use of metabolic markers to identify overweight individuals who are insulin resistant. *Ann. Intern. Med* 139, 802–809 (2003). [PubMed: 14623617]
17. Pantoja-Torres B et al. High triglycerides to HDL-cholesterol ratio is associated with insulin resistance in normal-weight healthy adults. *Diabetes Metab. Syndr* 13, 382–388 (2019). [PubMed: 30641729]
18. Chiang JK, Lai NS, Chang JK & Koo M Predicting insulin resistance using the triglyceride-to-high-density lipoprotein cholesterol ratio in Taiwanese adults. *Cardiovasc. Diabetol* 10, 93 (2011). [PubMed: 22004541]
19. 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). [PubMed: 34678755]
20. Sudlow C et al. UK Biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.* 12, e1001779 (2015). [PubMed: 25826379]
21. Zhou W et al. Efficiently controlling for case–control imbalance and sample relatedness in large-scale genetic association studies. *Nat. Genet* 50, 1335–1341 (2018). [PubMed: 30104761]
22. Yang J et al. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat. Genet* 44, 369–375 (2012). [PubMed: 22426310]
23. Tang J et al. Obesity-associated family with sequence similarity 13, member A (FAM13A) is dispensable for adipose development and insulin sensitivity. *Int. J. Obes. (Lond.)* 43, 1269–1280 (2019). [PubMed: 30301961]
24. Fathzadeh M et al. FAM13A affects body fat distribution and adipocyte function. *Nat. Commun* 11, 1465 (2020). [PubMed: 32193374]
25. Fernandes Silva L, Vangipurapu J, Kuulasmaa T & Laakso M An intronic variant in the GCKR gene is associated with multiple lipids. *Sci. Rep* 9, 10240 (2019). [PubMed: 31308433]
26. Li X, Wang F, Xu M, Howles P & Tso P ApoA-IV improves insulin sensitivity and glucose uptake in mouse adipocytes via PI3K-Akt signaling. *Sci. Rep* 7, 41289 (2017). [PubMed: 28117404]
27. Nowak M et al. Insulin-mediated down-regulation of apolipoprotein A5 gene expression through the phosphatidylinositol 3-kinase pathway: role of upstream stimulatory factor. *Mol. Cell. Biol* 25, 1537–1548 (2005). [PubMed: 15684402]
28. Haas ME, Attie AD & Biddinger SB The regulation of ApoB metabolism by insulin. *Trends Endocrinol. Metab* 24, 391–397 (2013). [PubMed: 23721961]

29. Kim JY, Tillison K, Lee JH, Rearick DA & Smas CM The adipose tissue triglyceride lipase ATGL/PNPLA2 is downregulated by insulin and TNF- $\alpha$  in 3T3-L1 adipocytes and is a target for transactivation by PPAR $\gamma$ . *Am. J. Physiol. Endocrinol. Metab* 291, E115–E127 (2006). [PubMed: 16705060]
30. Knowles JW et al. Identification and validation of N-acetyltransferase 2 as an insulin sensitivity gene. *J. Clin. Investig* 125, 1739–1751 (2015). [PubMed: 25798622]
31. Graham SE et al. The power of genetic diversity in genome-wide association studies of lipids. *Nature* 600, 675–679 (2021). [PubMed: 34887591]
32. Pulit SL 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). [PubMed: 30239722]
33. 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). [PubMed: 30297969]
34. Chen VL et al. Genome-wide association study of serum liver enzymes implicates diverse metabolic and liver pathology. *Nat. Commun* 12, 816 (2021). [PubMed: 33547301]
35. Hartiala JA et al. Genome-wide analysis identifies novel susceptibility loci for myocardial infarction. *Eur. Heart J* 42, 919–933 (2021). [PubMed: 33532862]
36. Evangelou E et al. Genetic analysis of over 1 million people identifies 535 new loci associated with blood pressure traits. *Nat. Genet* 50, 1412–1425 (2018). [PubMed: 30224653]
37. Stanzick KJ et al. Discovery and prioritization of variants and genes for kidney function in >1.2 million individuals. *Nat. Commun* 12, 4350 (2021). [PubMed: 34272381]
38. Chen Y et al. Genome-wide association meta-analysis identifies 17 loci associated with nonalcoholic fatty liver disease. *Nat. Genet* 55, 1640–1650 (2023). [PubMed: 37709864]
39. Day F et al. Large-scale genome-wide meta-analysis of polycystic ovary syndrome suggests shared genetic architecture for different diagnosis criteria. *PLoS Genet.* 14, e1007813 (2018). [PubMed: 30566500]
40. Proud CG Regulation of protein synthesis by insulin. *Biochem. Soc. Trans* 34, 213–216 (2006). [PubMed: 16545079]
41. Guillet C, Masgrau A, Walrand S & Boirie Y Impaired protein metabolism: interlinks between obesity, insulin resistance and inflammation. *Obes. Rev* 13, 51–57 (2012). [PubMed: 23107259]
42. Yang Q & Civelek M Transcription factor KLF14 and metabolic syndrome. *Front. Cardiovasc. Med* 7, 91 (2020). [PubMed: 32548128]
43. Teslovich TM et al. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* 466, 707–713 (2010). [PubMed: 20686565]
44. Shungin D et al. New genetic loci link adipose and insulin biology to body fat distribution. *Nature* 518, 187–196 (2015). [PubMed: 25673412]
45. Small KS 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). [PubMed: 29632379]
46. Mahdessian H et al. TM6SF2 is a regulator of liver fat metabolism influencing triglyceride secretion and hepatic lipid droplet content. *Proc. Natl Acad. Sci. USA* 111, 8913–8918 (2014). [PubMed: 24927523]
47. Zabaneh D & Balding DJ A genome-wide association study of the metabolic syndrome in Indian Asian men. *PLoS ONE* 5, e11961 (2010). [PubMed: 20694148]
48. Zhu Y et al. Susceptibility loci for metabolic syndrome and metabolic components identified in Han Chinese: a multi-stage genome-wide association study. *J. Cell. Mol. Med* 21, 1106–1116 (2017). [PubMed: 28371326]
49. Kristiansson K et al. Genome-wide screen for metabolic syndrome susceptibility loci reveals strong lipid gene contribution but no evidence for common genetic basis for clustering of metabolic syndrome traits. *Circ. Cardiovasc. Genet* 5, 242–249 (2012). [PubMed: 22399527]
50. Kraja AT et al. A bivariate genome-wide approach to metabolic syndrome: STAMPEED consortium. *Diabetes* 60, 1329–1339 (2011). [PubMed: 21386085]

51. Agius L, Chachra SS & Ford BE The protective role of the carbohydrate response element binding protein in the liver: the metabolite perspective. *Front. Endocrinol* 11, 594041 (2020).
52. Abdul-Wahed A, Guilmeau S & Postic C Sweet sixteenth for ChREBP: established roles and future goals. *Cell Metab.* 26, 324–341 (2017). [PubMed: 28768172]
53. Ortega-Prieto P & Postic C Carbohydrate sensing through the transcription factor ChREBP. *Front. Genet* 10, 472 (2019). [PubMed: 31275349]
54. Arden C et al. Elevated glucose represses liver glucokinase and induces its regulatory protein to safeguard hepatic phosphate homeostasis. *Diabetes* 60, 3110–3120 (2011). [PubMed: 22013014]
55. Lind L Genome-wide association study of the metabolic syndrome in UK Biobank. *Metab. Syndr. Relat. Disord* 17, 505–511 (2019). [PubMed: 31589552]
56. O'Donovan G et al. Fat distribution in men of different waist girth, fitness level and exercise habit. *Int. J. Obes. (Lond.)* 33, 1356–1362 (2009). [PubMed: 19752873]
57. Paley CA & Johnson MI Abdominal obesity and metabolic syndrome: exercise as medicine? *BMC Sports Sci. Med. Rehabil* 10, 7 (2018). [PubMed: 29755739]
58. Shi TH, Wang B & Natarajan S The influence of metabolic syndrome in predicting mortality risk among US adults: importance of metabolic syndrome even in adults with normal weight. *Prev. Chronic Dis* 17, E36 (2020). [PubMed: 32441641]
59. Wang K et al. Differential roles of insulin like growth factor 1 receptor and insulin receptor during embryonic heart development. *BMC Dev. Biol* 19, 5 (2019). [PubMed: 30909860]
60. Holmes DI & Zachary I The vascular endothelial growth factor (VEGF) family: angiogenic factors in health and disease. *Genome Biol.* 6, 209 (2005). [PubMed: 15693956]
61. Kim S, Ahn C, Bong N, Choe S & Lee DK Biphasic effects of FGF2 on adipogenesis. *PLoS ONE* 10, e0120073 (2015). [PubMed: 25790378]
62. Blázquez-Medela AM, Jumabay M & Boström KI Beyond the bone: bone morphogenetic protein signaling in adipose tissue. *Obes. Rev* 20, 648–658 (2019). [PubMed: 30609449]
63. Yaghootkar H et al. Genetic evidence for a normal-weight 'metabolically obese' phenotype linking insulin resistance, hypertension, coronary artery disease, and type 2 diabetes. *Diabetes* 63, 4369–4377 (2014). [PubMed: 25048195]
64. Bond ST, Calkin AC & Drew BG Sex differences in white adipose tissue expansion: emerging molecular mechanisms. *Clin. Sci. (Lond.)* 135, 2691–2708 (2021). [PubMed: 34908104]
65. Brown RJ et al. The diagnosis and management of lipodystrophy syndromes: a multi-society practice guideline. *J. Clin. Endocrinol. Metab* 101, 4500–4511 (2016). [PubMed: 27710244]
66. Huang Z, Xu A & Cheung BMY The potential role of fibroblast growth factor 21 in lipid metabolism and hypertension. *Curr. Hypertens. Rep* 19, 28 (2017). [PubMed: 28337713]
67. Iizuka K, Takao K & Yabe D ChREBP-mediated regulation of lipid metabolism: involvement of the gut microbiota, liver, and adipose tissue. *Front. Endocrinol* 11, 587189 (2020).
68. Santoro N et al. Hepatic de novo lipogenesis in obese youth is modulated by a common variant in the GCKR gene. *J. Clin. Endocrinol. Metab* 100, E1125–E1132 (2015). [PubMed: 26043229]
69. Brouwers MCGJ, Jacobs C, Bast A, Stehouwer CDA & Schaper NC Modulation of glucokinase regulatory protein: a double-edged sword? *Trends Mol. Med* 21, 583–594 (2015). [PubMed: 26432016]
70. Chauhan A, Singhal A & Goyal P TG/HDL ratio: a marker for insulin resistance and atherosclerosis in prediabetics or not? *J. Fam. Med. Prim. Care* 10, 3700–3705 (2021).
71. Cordero A & Alegria-Ezquerria E TG/HDL ratio as surrogate marker for insulin resistance. *E. J. Cardiol. Pract* 8, (2009).
72. Giannini C et al. The triglyceride-to-HDL cholesterol ratio: association with insulin resistance in obese youths of different ethnic backgrounds. *Diabetes Care* 34, 1869–1874 (2011). [PubMed: 21730284]
73. Behiry EG, El Nady NM, AbdEl Haie OM, Mattar MK & Magdy A Evaluation of TG-HDL ratio instead of HOMA ratio as insulin resistance marker in overweight and children with obesity. *Endocr. Metab. Immune Disord. Drug Targets* 19, 676–682 (2019). [PubMed: 30663576]

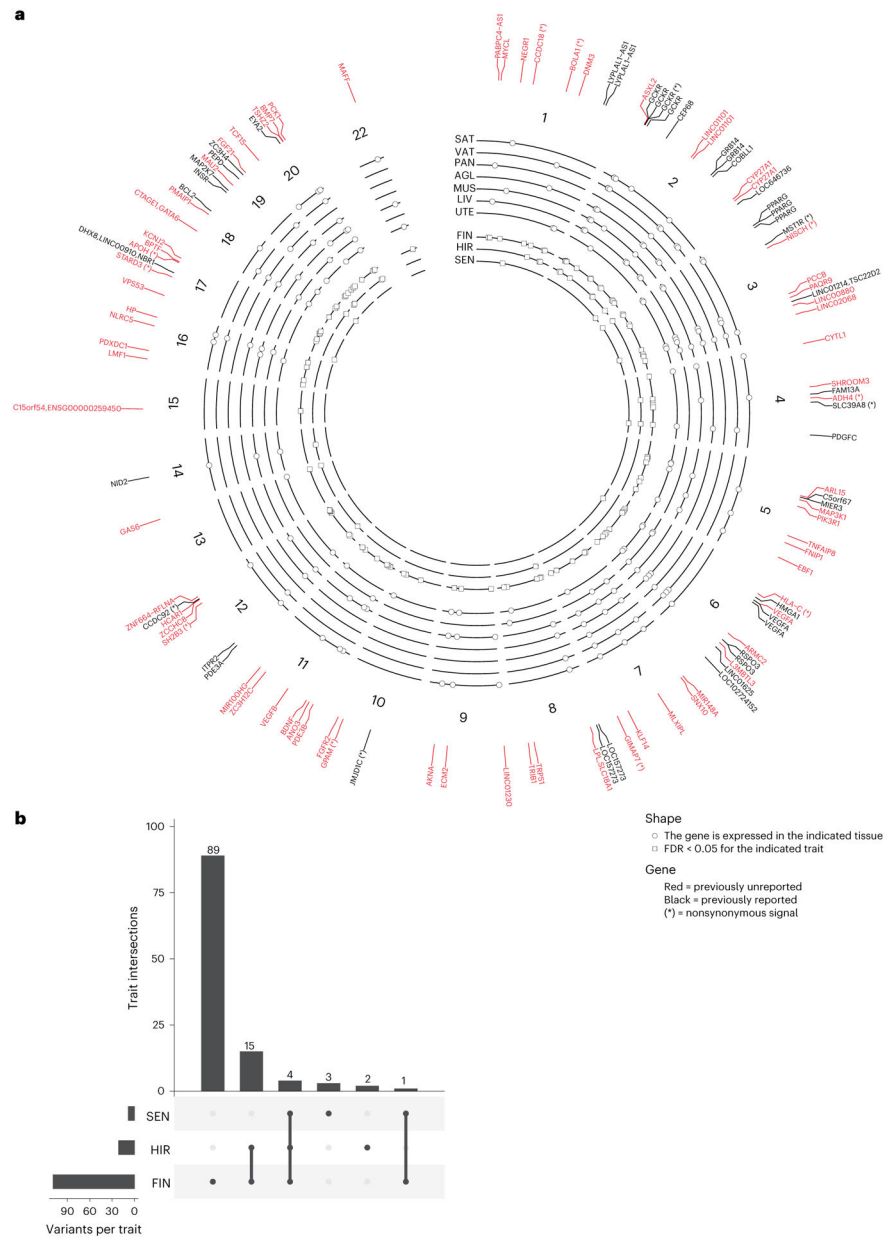
74. Knight MG et al. The TG/HDL-C ratio does not predict insulin resistance in overweight women of African descent: a study of South African, African American and West African women. *Ethn. Dis* 21, 490–494 (2011). [PubMed: 22428356]
75. Young KA et al. The triglyceride to high-density lipoprotein cholesterol (TG/HDL-C) ratio as a predictor of insulin resistance,  $\beta$ -cell function, and diabetes in Hispanics and African Americans. *J. Diabetes Complications* 33, 118–122 (2019). [PubMed: 30522790]
76. Maguire LH et al. Genome-wide association analyses identify 39 new susceptibility loci for diverticular disease. *Nat. Genet* 50, 1359–1365 (2018). [PubMed: 30177863]
77. Dey R, Schmidt EM, Abecasis GR & Lee S A fast and accurate algorithm to test for binary phenotypes and its application to PheWAS. *Am. J. Hum. Genet* 101, 37–49 (2017). [PubMed: 28602423]
78. Zawistowski M et al. The Michigan Genomics Initiative: a biobank linking genotypes and electronic clinical records in Michigan Medicine patients. *Cell Genom.* 3, 100257 (2023). [PubMed: 36819667]
79. Taliun D et al. LASER server: ancestry tracing with genotypes or sequence reads. *Bioinformatics* 33, 2056–2058 (2017). [PubMed: 28200055]
80. Bulik-Sullivan BK et al. LD score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet* 47, 291–295 (2015). [PubMed: 25642630]
81. Wang K, Li M & Hakonarson H ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* 38, e164 (2010). [PubMed: 20601685]
82. GTEx Consortium The Genotype-Tissue Expression (GTEx) project. *Nat. Genet* 45, 580–585 (2013). [PubMed: 23715323]
83. Franzén O et al. Cardiometabolic risk loci share downstream *cis*- and *trans*-gene regulation across tissues and diseases. *Science* 353, 827–830 (2016). [PubMed: 27540175]
84. Machiela MJ & Chanock SJ LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. *Bioinformatics* 31, 3555–3557 (2015). [PubMed: 26139635]
85. Shannon P et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504 (2003). [PubMed: 14597658]
86. Manichaikul A et al. Robust relationship inference in genome-wide association studies. *Bioinformatics* 26, 2867–2873 (2010). [PubMed: 20926424]
87. Watanabe K, Taskesen E, van Bochoven A & Posthuma D Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun* 8, 1826 (2017). [PubMed: 29184056]
88. Willer CJ, Li Y & Abecasis GR METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* 26, 2190–2191 (2010). [PubMed: 20616382]
89. Oliveri A Code used to annotate the TG:HDL-C loci in the paper ‘comprehensive genetic study of the insulin resistance marker TG:HDL-C in the UK Biobank.’ Zenodo. 10.5281/zenodo.10182519 (2023).



**Fig. 1 |. Study design: TG:HDL-C in the UK Biobank.**

Inclusion and exclusion criteria for individuals, variants and selection of the 114 high-confidence IR-associated loci. Includes information on previously unreported genes and nonsynonymous variants. The term ‘covariates’ includes age, sex and 1–10 PCs. Variants were excluded from the study if they had an imputation score  $>0.85$ , were multiallelic or ambiguous or were available in the UKBB but not in the MGI.  $V$  = variants;  $n$  = sample size.





**Fig. 2 |. Overlap of the 114 high-confidence IR-associated loci with insulin-related traits.**  
**a**, Overlap of each SNP with insulin traits and annotations. Fasting insulin and fasting insulin adjusted for BMI (PMID: 34059833, 22581228, 22885924 and 20081858) were represented in the category ‘FIN’; HOMA-IR and HOMA-IR adjusted for BMI (PMID: 22581228 and 20081858) were represented in the category ‘HIR’; the modified Stumvoll insulin sensitivity index (PMID: 27416945) was represented in the category ‘SEN’. Points in black indicate the SNP is nonsynonymous or in high LD ( $r^2 > 0.8$ ) with a nonsynonymous variant. Gene labels in black indicate the SNP has been previously reported for IR in literature. Gene labels in red indicate the SNP has not been previously reported for IR. Binarized expression of each loci’s predicted causal gene is shown for the following seven tissues: SAT, VAT, PAN, AGL, MUL, LIV and UTE. An expressed value corresponds to

meeting a threshold of twice the median expression for all other GTEx tissues. **b.** Overlap of high-confidence 114 high-confidence IR-associated loci with insulin-related traits in an aggregate form. FIN, fasting insulin; SEN, insulin sensitivity; HIR, HOMA-IR; SAT, subcutaneous adipose tissue; VAT, visceral adipose tissue; PAN, pancreas; AGL, adrenal gland; MUL, musculoskeletal; LIV, liver; UTE, uterus.

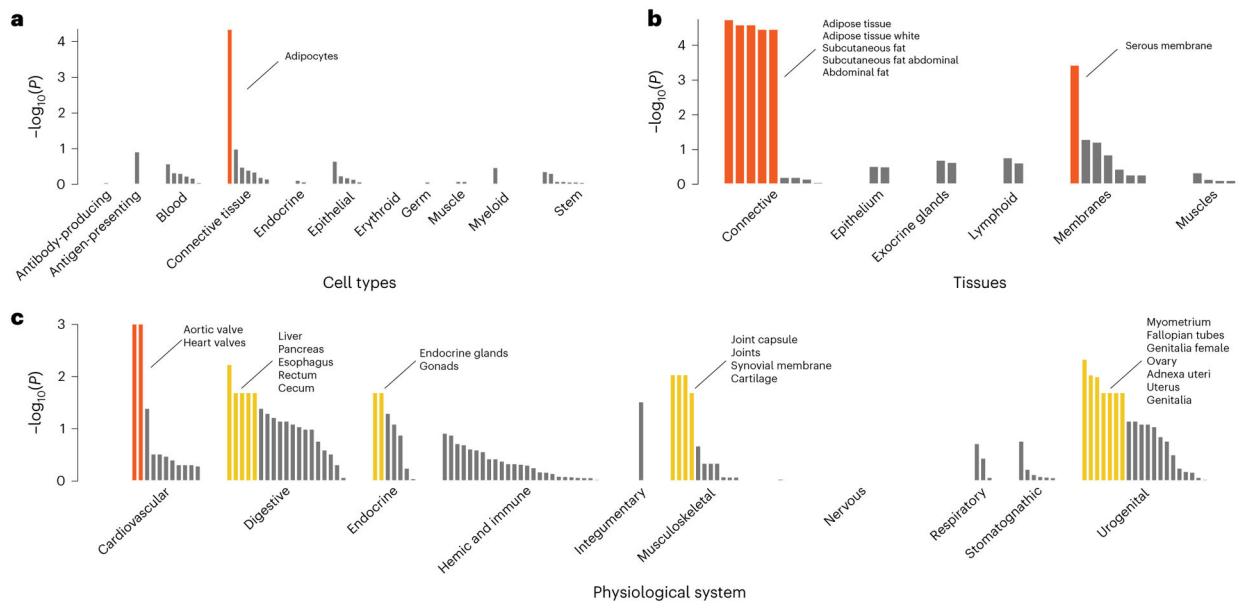
Author Manuscript

Author Manuscript

Author Manuscript

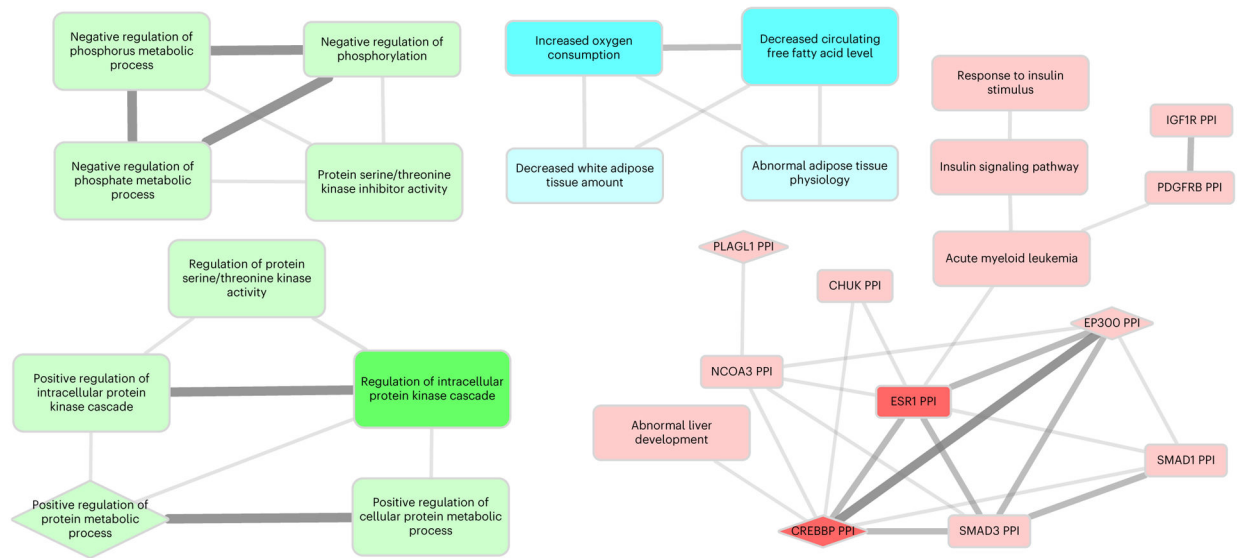
Author Manuscript





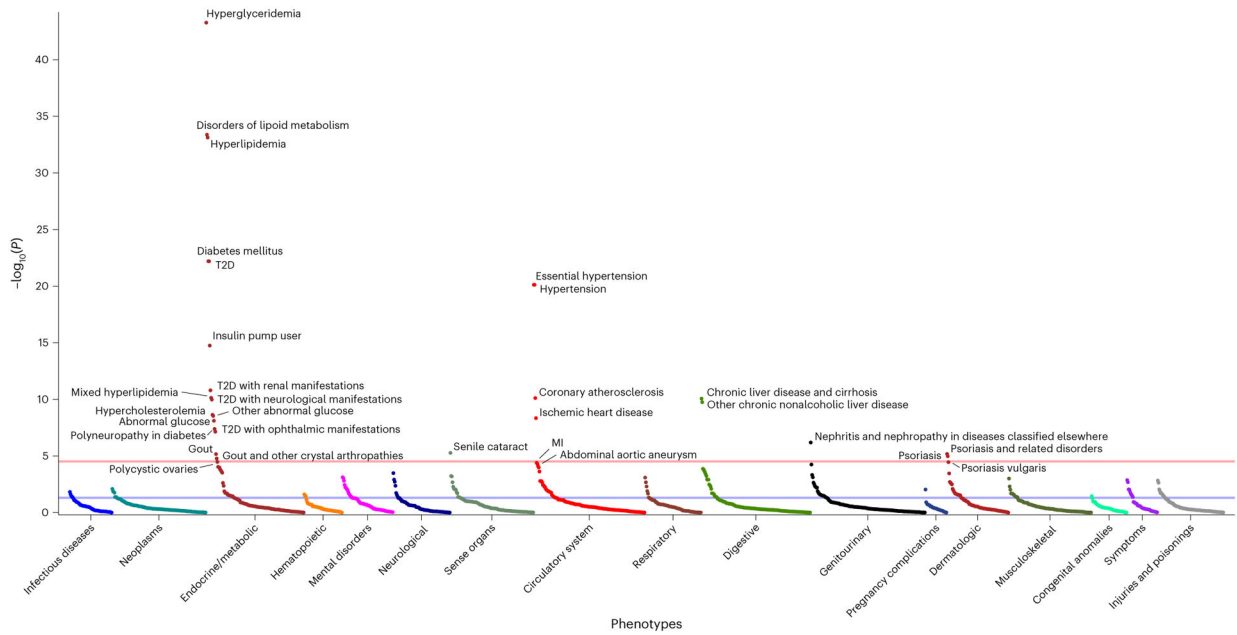
**Fig. 4 |. Tissue, cell type, and physiological system enrichment of the 114 high-confidence IR-associated loci.**

**a–c**, DEPICT enrichment for high-confidence IR-associated loci for cell types (**a**), tissues (**b**) and physiological systems (**c**). The *x* axis represents the MeSH terms that were analyzed by DEPICT, also shown in Supplementary Table 15. The *y* axis represents the nominal *P* value of the association presented on a  $-\log_{10}(P)$  scale. Associations used two-tailed *z* scores adjusted with the FDR procedure. Orange and yellow bars represent associations with an FDR less than 0.05 and between 0.05 and 0.20, respectively. MeSH, medical subject heading.



**Fig. 5 | Gene-set enrichment of the 114 high-confidence IR-associated loci.**

Gene sets reaching statistical significance ( $P < 1 \times 10^{-4}$ ) in the gene-set enrichment analysis of the 114 high-confidence IR-associated loci. Associations used two-tailed z scores adjusted with the FDR procedure. A darker color was used to highlight the node(s) with the most edges within each group. Edges represent Pearson's correlation between two nodes. Nodes with a rectangular shape have a  $P$  value of association between  $1 \times 10^{-4}$  and  $1 \times 10^{-5}$ ; nodes with a diamond shape have a  $P$  value of association between  $1 \times 10^{-5}$  and  $1 \times 10^{-6}$ . Please refer to Supplementary Table 16 for the exact  $P$  values of association.



**Fig. 6 | PRS analysis in the MGI using the 114 high-confidence IR-associated loci.** PheWAS Manhattan plot showing the association between the 114 SNP PRS and traits in the MGI using Firth’s logistic regression model. Points are colored based on the different phenotype categories. The blue and the red lines represent the suggestive ( $\alpha = 0.05$ ) and Bonferroni-adjusted ( $\alpha = 3 \times 10^{-5}$ ) significance threshold on a  $-\log_{10}$  scale, respectively. Traits with borderline Bonferroni significance threshold were also reported.