

Distinct and shared genetic architectures of Gestational diabetes mellitus and Type 2 Diabetes Mellitus

A. Elliott^{1,2,3}, R. K. Walters^{1,2,3}, M. Pirinen^{4,5,6}, M. Kurki^{1,2}, N. Junna⁴, J. Goldstein², M.P. Reeve⁴, H. Siirtola⁷, S. Lemmelä^{4,8}, P. Turley^{9,10}, FinnGen, A. Palotie^{1,2,3,4}, M. Daly^{1,2,3,4#}, E. Widén^{4#}

¹Analytic and Translational Genetics Unit, Massachusetts Gen. Hosp., Boston, MA,

²Stanley Ctr. for Psychiatric Res., Broad Inst. of Harvard and MIT, Cambridge, MA,

³Harvard Med. Sch., Boston, MA,

⁴Institute for Molecular Med. Finland, Helsinki Institute of Life Sciences., University of Helsinki, Helsinki, Finland,

⁵Department of Mathematics and Statistics, University of Helsinki, Helsinki, Finland

⁶Department of Public Health, University of Helsinki, Helsinki, Finland

⁷TAUCHI Research Center, Faculty of Information Technology and Communication Sciences (ITC), Tampere University, Tampere, Finland,

⁸Finnish Institute for Health and Welfare (THL), Helsinki, Finland

⁹Center for Economic and Social Research, University of Southern California, Los Angeles, CA, USA.

¹⁰Department of Economics, University of Southern California, Los Angeles, CA, USA.

Corresponding authors

Elisabeth Widén

Email: elisabeth.widen@helsinki.fi

Address: Institute for Molecular Medicine (FIMM), Biomedicum 2U, Tukholmankatu 8, 00290, Helsinki, Finland

Mark Daly

Email: mark.daly@helsinki.fi

Address: Institute for Molecular Medicine (FIMM), Biomedicum 2U, Tukholmankatu 8, 00290, Helsinki, Finland

Abstract

Gestational diabetes mellitus (GDM) affects more than 16 million pregnancies annually worldwide and is related to an increased lifetime risk of Type 2 diabetes (T2D). The diseases are hypothesized to share a genetic predisposition, but there are few GWAS studies of GDM and none of them is sufficiently powered to assess whether any variants or biological pathways are specific to GDM. We conducted the largest genome-wide association study of GDM to date in 12,332 cases and 131,109 parous female controls in the FinnGen Study and identified 13 GDM-associated loci including 8 novel loci. Genetic features distinct from T2D were identified both at the locus and genomic scale. Our results suggest that the genetics of GDM risk falls into two distinct categories – one part conventional T2D polygenic risk and one part predominantly influencing mechanisms disrupted in pregnancy. Loci with GDM-predominant effects map to genes related to islet cells, central glucose homeostasis, steroidogenesis, and placental expression. These results pave the way for an improved biological understanding of GDM pathophysiology and its role in the development and course of T2D.

Main text

Gestational diabetes mellitus (GDM) is a common heritable metabolic disorder characterized by new-onset glucose intolerance occurring during pregnancy and affecting 7-9% of women in the United States¹ and 14% of pregnancies globally². This, combined with adverse maternal and fetal outcomes in gestation, and subsequent high risk of maternal metabolic and cardiovascular disease with its associated mortality and morbidity, make GDM a major public health concern globally.

Clinically, GDM is related to an increased lifetime risk of Type 2 diabetes (T2D)³⁻⁵ – with over a third of women developing T2D within 15 years of their GDM diagnosis. GDM also aggregates in families and is associated with a family history of T2D⁶⁻⁸. Thus, it has been proposed that both diseases share a common etiology and a common genetic predisposition and that GDM would then merely represent a perturbation revealing an existing predisposition to T2D⁹. Yet, few studies have sought to uncover the genetic underpinnings of GDM and most of these studies have aimed at evaluating the impact of T2D loci¹⁰⁻¹².

To date, only three Genome-Wide Association Studies (GWASes) have focused on the genetics of GDM¹³⁻¹⁵. The largest of them with 5,485 cases and 347,856 female controls revealed 5 genome-wide significant loci, of which 4 have also been previously associated with T2D¹⁵. While the results seem to support the hypothesis of shared etiology, none of the studies were sufficiently powered to directly

assess the degree to which genetic risk is shared in GDM vs. T2D and whether any variants and biological pathways are specific to, or have differential effects in, GDM.

Thus, to elucidate the genetic underpinnings of GDM we conducted a genome-wide association study¹⁶ of GDM in 12,332 cases and 131,109 parous female controls. Participants were of Finnish ancestry from the FinnGen study. The association analysis was performed as described in Kurki et al¹⁷. Cases were identified using Finnish health and population registry sources including registry data from inpatient hospitalizations, outpatient specialty clinics, and birth registry. Cases were confirmed to have a diagnosis within a pregnancy window and those with diagnoses of diabetes prior to the index pregnancy were excluded (Online Methods, Supplementary Note).

Our GWAS nearly tripled the previously known loci for GDM, identifying 13 distinct associated chromosomal regions (Figure 1, Supplementary Figures 1-13). Fine mapping¹⁸ pinpointed 14 independent signals (the region near CDKN2A/B containing two independent signals) of which 9 regions had a 95% credible set containing 5 or fewer SNPs (Table 1, Supplementary Table 1, Online Methods). Nine regions represented novel GDM associations not reported in previous GDM GWAS. We further characterize the GDM GWAS results through annotation and colocalization of credible sets with >3,800 GWAS (Supplementary Tables 2-3), QTLs for gene expression, biomarkers and metabolites (Supplementary Tables 4-9), and chromatin interactions (Supplementary Table 10), along with tests of enrichment by functional consequence, gene expression, or canonical gene sets (Supplementary Tables 11-14; Supplementary Figures 14-17, Online Methods).

We next performed analyses to evaluate the shared genetic etiology with T2D. Assessment of genome-wide significant signals using our algorithm Significant Cross-trait Outliers and Trends in JOint York regression (SCOUTJOY; Online Methods) indicated that the 13 GDM-associated loci showed significant heterogeneity in their relationship to T2D ($p < 0.001$, Supplementary Table 21). Five of the 13 GDM-associated loci were not significantly associated ($p < 5 \times 10^{-8}$) with T2D in either a previously published large T2D meta-analysis¹⁹ or in FinnGen, while the remaining loci are established T2D hits (Table 1, Supplementary Figure 18). At the genomic level GDM and T2D were genetically correlated ($r_g = 0.71$, $se = 0.06$, $p = 6.8 \times 10^{-37}$), which is significantly greater than zero ($p = 6.8 \times 10^{-37}$) but less than 1 ($p = 1.2 \times 10^{-7}$; Online Methods). Significant genetic correlations were also seen with 12 diseases or traits and 8 blood laboratory values in cases where the disorder or value was phenotypically related to GDM (Figure 1; Supplementary Figures 19-20, Supplementary Tables 15-19). In both the genomic correlation and top hits comparison, GDM was significantly associated with fasting glucose, HbA1C, and 2 hour

glucose result on oral glucose tolerance testing but was not associated with fasting insulin level. None of these glycemic traits or related disorders, however, appeared to stratify the 13 GDM-associated loci into distinct groups similar to T2D (Supplementary Figure 21, Supplementary Table 20).

To explore the relationship between GDM and T2D effects in more detail, we applied a Bayesian classification algorithm to the top associations for GDM and top associations for T2D (13 loci for GDM and 15 for T2D; Online Methods). The included T2D loci were the strongest signals obtained in a GWAS of male FinnGen participants (cases 27,607; controls 118,687) selected to have comparable statistical evidence for association (Supplementary Note). We utilized the male T2D scan to avoid confounding by sample overlap, though similar T2D results are seen in men and parous and nonparous women (Supplementary Note, Supplementary Tables 21 and 23-24, Supplementary Figure 22-24), suggesting the comparison is not confounded by overlap or sex-specific effects on T2D.

The shared variants analysis suggested that the genetics of GDM risk falls into two categories, one shared with T2D risk and the other uniquely gestational (Figure 2, Supplementary Table 22). Specifically, comparison of effect sizes between GDM and T2D does not support the existence of a single, consistent relationship between GDM and T2D across loci, but instead proposes two distinct classes of significant variants in this scan (Figure 2) – Class G, with GDM-predominant effects, and Class T, with T2D-predominant effects with the two-class model of relationship between GDM and T2D better fits the observed distribution of ORs by $\log_{10}(\text{likelihood-ratio})$ of approximately 30. Class G contains 8 of the 13 GDM-associated loci that have GDM-predominant SNP effects with an effect size roughly 3 times greater in GDM than T2D (Figure 2, Table 1). In comparison, the GDM-associated SNPs contained in Class T had effects in the two disorders that were consistent with T2D-signals significantly associated with diabetes only in the T2D GWAS – namely a reduced effect size in GDM versus T2D – a pattern of effects that was observed for all SNPs in Class T. The existence of the GDM-predominant class of effects, Class G, distinct from those traditionally seen in T2D, raises the possibility of physiologic mechanisms of glycemic control specifically important during pregnancy (Supplementary Note, Supplementary Table 25, Supplementary Figure 25).

As presented in Table 1, the eight class G loci have a peak SNP that is either intronic to a protein-coding gene, a missense mutation or, a 5' UTR variant. Although the effects of a locus don't always operate through the nearest gene, several of the loci implicate genes involved in plausible cellular processes e.g., signal transduction and hormone processing. (Table 1). For example, the missense mutation protective against GDM resides in *MAP3K15* on the X-chromosome (G838S,

chrX_19380197_C_T_b38), which encodes a protein kinase (MAP3K15) that regulates apoptotic-mediated cell death and stress response. The gene has high expression in the adrenal glands and has previously been linked to steroidogenesis²⁰ and polycystic ovarian syndrome²¹. The GDM-associated missense mutation is rare outside of Finland but other rare coding mutations in *MAP3K15* have previously been associated with T2D in UK Biobank²², i.e. female carriers of such rare nonsynonymous variants had a 30% reduced risk of T2D and reduced blood glucose and HbA1C levels, and hemizygous male carriers of rare protein-truncating variants had a 40% reduced risk of T2D. Further characterization of the GDM-associated mutation by PheWas analyses in FinnGen indicated that the mutation is associated with increased risk for hypertension ($\beta = 0.11$, $p = 2.0e-8$), but in contrast, there is no association with T2D ($\beta = -0.09$, $p = 1.8e-3$).

Other hormone-related class G signals include a variant mapping to the 5'UTR of the estrogen receptor gene, *ESR1*, and a variant nearby *PCSK1*, a gene encoding for prohormone convertase 1/3, which critically regulates endocrine and neuronal prohormone processing. Previous data show that homozygous loss of function of *PCSK1* results in a generalized and pleiotropic prohormone conversion defect characterized by severe obesity, impaired adrenal and thyroid function, reactive hypoglycemia, elevated levels of proinsulin, and low levels of insulin²³, whereas common gene variants have been associated with BMI²⁴, fasting proinsulin, fasting glucose, and T2D. Interestingly the GDM-associated allele identified in our study is associated with lower BMI ($\beta = -0.02$, $p = 5.3e-11$), lower weight ($\beta = -0.02$, $p = 5.3e-11$), and lower height ($\beta = -0.01$, $p = 3.9e-6$).

Finally, to gain further insight into potential functional differences between GDM and T2D we examined the cell type specific expression patterns associated with the GWAS summary statistics²⁵ (Figure 3, Online Methods, Supplementary Tables 26-29, Supplementary Figures 26-28). We evaluate cell-type specific enrichment despite the lack of significant tissue-level enrichment because pregnancy induces major adaptive changes to specific cell populations within maternal tissues that might not be reflected in bulk tissue expression. Analyses integrating multiple large single cell RNA expression datasets indicated that pancreatic β cells are significantly associated both with GDM and T2D. However, only GDM had significant associations with the hypothalamus, i.e., hypothalamic GABAergic neurons (*GABA2*), hypothalamic glutaminergic neurons (*GLU7*), and neurons in the VMH arcuate nucleus (*NR5a1_Adcyap1*; Figure 3; Supplementary Table 28).

Taken together, in this paper we present data from the largest GDM GWAS to date, identifying 14 independent signals in 13 associated chromosomal regions. GDM is a common disorder of pregnancy

that has significantly increased in prevalence in all racial and ethnic groups in the last 15 years²⁶. Despite conferring substantial morbidity to both mother and child, relatively little is known about the genetics of GDM outside of a proposed shared genetic etiology with T2D. Our key finding is that GDM in fact has a partially distinct genetic etiology, i.e., while GDM and T2D in part share a polygenic predisposition, there is a second category of GDM genetic risk factors that are predominantly gestational contributors to disease. This contextualizes the substantial effect of the *MTNR1B*, which had been reported previously as an outlier¹¹, but our data now show that *MTNR1B* is representative of a whole group of GDM-predominant loci, characterized by a 3 times larger effect on GDM than on T2D.

Further studies will be required to characterize the precise GDM-specific molecular effects, but our current results suggest plausible mechanisms related to maternal adaptive physiological responses to pregnancy. Broadly, pregnancy increases circulating gestational hormones (e.g., human placental lactogen progesterone, and estrogen) causing alternations in normal homeostatic glycemetic pathways in the brain and pancreas as well as impaired insulin sensitivity in maternal peripheral tissues. The brain and pancreas both show clear enrichment of signal in our cell-type specificity analysis of GDM, with our results in brain showing specific associations with hypothalamic and arcuate (ARC) neurons in GDM that are not seen in T2D (Figure 3, Supplementary Tables 26-28). Hypothalamic and ARC neurons have been implicated in adaptive glycemetic response during pregnancy²⁷. In that context, our *ESR1* locus is particularly interesting given that the ventromedial hypothalamus (VMH) contains glucose-sensing neurons that express the estrogen receptor- α ($ER\alpha$, encoded by *ESR1*) and act to regulate glucose levels²⁸. Moreover, $ER\alpha$ neurons in the ARC have altered expression of several of our class G genes (e.g. *PCSK1*, *MTNR1B*, *SPC25*) in response to $ER\alpha$ knockout or perturbation of estrogen levels (which occurs in pregnancy)²⁹. The involvement of these cells in GDM is further supported by our cell-type specificity results highlighting *Nr5a1_Adcyap1* in ARC which project from the VMH³⁰ and appear to correspond to a similar gene expression pattern to *Glu7³¹* in our conditional analysis (Supplementary Figure 27, Supplementary Table 28, Supplementary Note). In contrast, while pregnancy is characterized by impaired insulin sensitivity and some of our class G genes (e.g. *PCSK1* and *MTNR1B*) affect insulin biosynthesis and secretion^{32,33}, no evidence of a genetic relationship between GDM and insulin levels is observed either in top hits or genome-wide at the current sample sizes. Nonetheless, given the complexity of GDM the limited sample size of our dataset allows us to glean insight into only a few potential mechanisms. Much larger studies are warranted for a comprehensive view of the overall in-depth molecular underpinnings of GDM susceptibility.

The current study design in the rather homogeneous Finnish population carries specific strengths and weaknesses associated with this analysis approach. On one hand GWAS discovery is enhanced by population homogeneity¹⁷ and the linkage of national birth, inpatient and outpatient medical registries enables robust phenotyping (see Online Methods). The generalizability of the results may suffer however as some detected loci may be for rare alleles specifically enriched in the Finnish population. In our analyses of GDM two loci mapped to rare alleles enriched in Finland, which may be difficult to replicate elsewhere, while 70% of the loci correspond to variants that are common (MAF>10%), in non-Finnish European ancestry individuals (Table 1). Nonetheless, additional studies prioritizing ancestrally diverse populations are needed for a better understanding of the genetic underpinnings of GDM in all populations at risk.

In summary, we discovered 9 novel loci associated with GDM and demonstrate that GDM genetic risk is distinct from T2D both at the locus and genomic scale. Our results suggest that the genetics of GDM risk falls into two categories – one part T2D risk and one part predominantly gestational contributors to disease. Tissue characterization of GDM genetics further implicates tissues previously identified in adaptive pregnancy responses, raising hypotheses regarding genetic effects in these tissues during pregnancy. Broadly this work underscores the benefits of focusing on resources on disorders of pregnancy as pregnancy is a natural perturbation that offers leverage to discover loci with novel physiologic mechanisms of glycemic or homeostatic control.

Methods

Cohort

The FinnGen Study is a public–private partnership project combining data from Finnish biobanks and electronic health records from national registries. The linked national health registers include data on hospital and outpatient visits, primary care, cause of death, and medication records. Approval from the FinnGen Study was received to use the data in the present work. After a 1-year embargo, the FinnGen summary stats are available for download. In this study, we used the results from the FinnGen release R8, which includes data from 342,499 individuals and more than 4500 disease endpoints.

Phenotyping

Full details of phenotyping are described in the Supplementary Note. Briefly, clinical endpoints with corresponding dates were constructed for gestational diabetes and related diagnoses for exclusions for all FinnGen participants as described in the Supplementary Note. Temporal phenotyping was then performed to phenotype each pregnancy for presence of glycemic disease and then assign individuals as cases or controls. Beginning with 330,000 pregnancies among genotyped FinnGen participants, we defined a “pregnancy window” of 40 before delivery until 5 weeks after delivery. A pregnancy met inclusion criteria for “gestational diabetes” if it had (I1) gestational diabetes ICD codes occurring in the pregnancy window, (I2) any diabetes codes occurring in the pregnancy window (e.g. for ICD8), or (I3) abnormal blood glucose test results in the Medical birth register, which contains data on the mother’s diseases during pregnancy. Pregnancies were then excluded for: (E1) any previous diabetes diagnosis code occurring outside a pregnancy window; (E2) any previous significant pancreatic disease, including chronic pancreatitis, pancreatic necrosis, pancreatic cancer, or cystic fibrosis; or (E3) any previous Type 1 or Type 2 diabetes code. Pregnancies passing these exclusion criteria and without any inclusion criteria for gestational diabetes were designated as “unaffected”. Then to phenotype individuals, cases were identified among the 151,000 genotyped females with a history of pregnancy as those at least 1 pregnancy meeting inclusion criteria for gestational diabetes and passing exclusion criteria. Controls were defined as females with only “unaffected” pregnancies (i.e. no diabetes or significant pancreatic diseases occurring prior to or during any pregnancy, and no abnormal blood glucose in the Medical birth register).

Genotyping and GWAS

A detailed description of the study design and analytical methods are available in the online documentation.

In brief, FinnGen individuals have been genotyped with Illumina and Affymetrix chip arrays. QC was performed to remove samples and variants of poor quality. Imputation was performed using population-specific SISu v3 imputation reference panel.

A subset of unrelated individuals of genetically confirmed Finnish ancestry was identified. GWAS was performed using REGENIE 2.2.4. Sex, age, 10 PCs, and genotyping batch were included as covariates in the analysis.

Finemapping

Finemapping of a 1.5MB locus around any GWAS lead SNP was performed using the SuSiE algorithm¹⁸ which reports causal variants and a 95% credible set for each independent signal. Details described previously¹⁷ and here, <https://finngen.gitbook.io/documentation/methods/finemapping>. As LD, we used in-sample dosages (i.e cases and controls used for each phenotype) computed with LDStore. Independent signals were those that either represent the primary strongest signal with lead $p < 5 \times 10^{-8}$ or as secondary signals that must have genome-wide significance and log bayes factor (BF) > 2.

Annotation

Variants were annotated with Ensembl Variant Effect Predictor (VEP) version 104, (<https://www.ensembl.org/info/docs/tools/vep/index.html>) data to give the projected variant consequence. Each variant was also annotated for enrichment in Finland compared to compared to non-Finnish-Swedish-Estonian Europeans, as described previously¹⁷. Annotation with known prior GWAS loci was performed as previously described¹⁷. In brief, for each independent association we annotated every phenotype in GWAS Catalog that was significantly associated with either (1) the lead PIP variant or (2) any variant in the credible set. Similar annotation was performed for metabolite associations from the MetSIM study³⁴ (Supplementary Note). Each locus was also annotated with SNP2GENE in FUMA version v1.3.7 (fuma.ctglab.nl/snp2gene/) for chromatin interactions (Supplementary Table 10), eQTL associations (Supplementary Table 4), and prior GWAS hits (Supplementary Table 3).

Colocalization

Colocalization was performed on all fine mapped regions as previously described for the FinnGen study¹⁷. In brief, the probabilistic model to intergrate GWAS and eQTL data was eCAVIAR³⁵ however the input posterior inclusion probabilities (PIP) estimated by the SuSiE algorithm¹⁸. The eCAVIAR method uses PIPs for variants in each region to compute a colocalization posterior probability (CLPP,

see Supplementary Note). The intersection of variants in credible sets was then checked across multiple phenotypes from FinnGen (Supplementary Table 2), GTEx (Supplementary Table 5), eQTL Catalogue (Supplementary Table 6), GeneRisk (Supplementary Table 7), and the UK Biobank (Supplementary Table 8).

Gene enrichment analysis

Gene-level association results from MAGMA were used to identify tissue and pathway enrichments using the SNP2GENE and GENE2FUNC modules of FUMA (version v1.4.1). The MAGMA results were tested for (a) association with gene expression levels in GTEx v8 (Supplementary Table 12, Supplementary Figure 14), (b) enrichment in sets of differentially expressed genes identified across tissues from GTEx v8 (Supplementary Table 13, Supplementary Figure 15), (c) enrichment in gene sets for pathways or other biological processes including those defined by KEGG (MsigDB c2), GO biological processes (MsigDB c5), or WikiPathways (Supplementary Table 14), and (d) enrichment in gene sets defined by reported associations in GWAS Catalog (Supplementary Table 14, Supplementary Figure 17).

Genetic correlation

We estimated the SNP heritability (h^2_{SNP}) of GDM and pairwise genetic correlations (SNP- r_g) between GDM and diabetes-related diseases and traits using LDSC Version 1.0.1. Testing difference of r_g from perfect correlation was performed using a one tailed Z-score test:

$$Z = \frac{1 - r_g}{se(r_g)}$$

See Supplementary Note for details on additional genetic correlation analyses.

Significant Cross-trait Outliers and Trends in JOint York regression (SCOUTJOY)

To compare heterogeneity of GDM-associated loci's genetic effects in any two disorders we developed SCOUTJOY (Supplementary Note), extending the heterogeneity detection algorithm introduced by MR-PRESSO³⁶ to handle sample overlap and estimation error in GWAS of each phenotype. To estimate the primary relationship in effect sizes between the two disorders while accounting for estimation error we derive estimators for York regression³⁷ with a fixed intercept. Global heterogeneity testing was then performed using simulated null replicates to test the hypothesis that the observed effect sizes across top hits are consistent with a single uniform relationship. Outlier variants were similarly identified as having larger than expected residuals from the York regression. We extend this outlier detection

method from MR-PRESSO to allow iterative assessment of each variant based on the outlier status of all other variants. Code for SCOUTJOY and York regression with a fixed intercept is available on github (link to be provided on acceptance).

Shared variants analysis

We applied linemodels package (<https://github.com/mjpirinen/linemodels>) to summary statistics from T2D GWAS and GDM GWAS. The analysis included 28 lead variants from the GWASes (13 from GDM and 15 from T2D). We classified the variants into two classes based on their bivariate effect sizes. The classes were represented by line models whose slopes were estimated using an EM-algorithm, resulting in values 1.53 (labelled as class T) and 0.25 (labelled as class G). For both models, the scale parameters determining the magnitude of effect sizes were set to 0.2 and the correlation parameters determining the allowed deviation from the lines were set to 0.99. The membership probabilities in the two classes were computed separately for each variant by assuming that the classes were equally probable a priori. Since the two GWAS did not have overlapping samples, the correlation of their effect estimators was set to 0.

Cell type specificity analyses

To get better resolution on specific cell types, we performed cell-type specificity analyses with high quality single cell murine datasets using FUMA (<https://fuma.ctglab.nl/tutorial#celltype>; Supplementary Note). First, we identified tissue-level associations with Tabula Muris data³⁸ identifying significant associations (FDR < .05) with expression in brain and pancreas after Benjamini-Hochberg multiple testing correction (Supplementary Figure 27). We then performed cell-type specificity analyses as previously described²⁵, augmenting Tabula Muris with additional high quality scRNA-seq of hypothesized involved brain regions (Supplementary Note). Analysis was performed on genetic summary statistics for both our Gestational Diabetes GWAS and for a recent Type 2 Diabetes European meta-analysis dataset¹⁹. We also compare the pancreatic results to analysis of high quality scRNA-seq of pancreas in humans to assess the impact of known differences in human vs. mouse pancreatic cellular function and physiology (Supplementary Table 29, Supplementary Figure 28).

Ethics statement

Patients and control subjects in FinnGen provided informed consent for biobank research, based on the Finnish Biobank Act. Alternatively, separate research cohorts, collected prior the Finnish Biobank Act

came into effect (in September 2013) and start of FinnGen (August 2017), were collected based on study-specific consents and later transferred to the Finnish biobanks after approval by Fimea, the National Supervisory Authority for Welfare and Health. Recruitment protocols followed the biobank protocols approved by Fimea. The Coordinating Ethics Committee of the Hospital District of Helsinki and Uusimaa (HUS) approved the FinnGen study protocol Nr HUS/990/2017.

The FinnGen study is approved by Finnish Institute for Health and Welfare (THL), approval number THL/2031/6.02.00/2017, amendments THL/1101/5.05.00/2017, THL/341/6.02.00/2018, THL/2222/6.02.00/2018, THL/283/6.02.00/2019, THL/1721/5.05.00/2019, Digital and population data service agency VRK43431/2017-3, VRK/6909/2018-3, VRK/4415/2019-3 the Social Insurance Institution (KELA) KELA 58/522/2017, KELA 131/522/2018, KELA 70/522/2019, KELA 98/522/2019, and Statistics Finland TK-53-1041-17.

Tables

Table 1: 14 genome-wide significant fine-mapped signals for GDM.

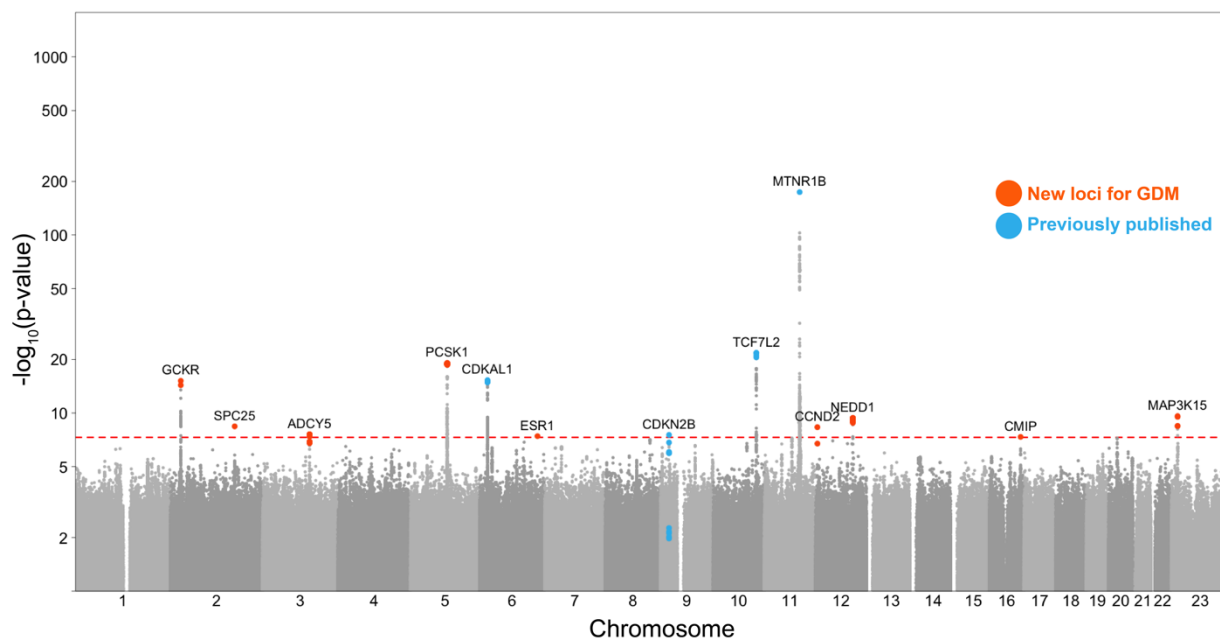
Region	Lead variant	Ref/Alt	AF	Beta (SE)	p	Nearest gene	Mutation ^a	Cred. size ^b	pClassG ^c	Known ^d
<i>ClassG Loci - GDM-predominant effect</i>										
2:27508073-27519736	rs780093	T/C	0.649	0.12(0.0149)	6.75E-16	GCKR	intron	3	0.995	T2D
2:16890084	rs1402837	C/T	0.17	0.11(0.0186)	3.87E-09	SPC25	intron	1	0.981	
5:96357306-96392261	rs1820176	T/C	0.314	-0.14(0.0154)	7.86E-20	PCSK1	intron	26	>0.999	
6:151805650	rs537224022	C/G	0.00984	-0.447(0.0812)	3.82E-08	ESR1	5' UTR	1	0.999	
11:92975544	rs10830963	C/G	0.358	0.403(0.0143)	8.65E-175	MTNR1B	intron	1	>0.999	GDM,T2D
12:97449565-97470365	rs74628648	C/T	0.0786	-0.169(0.027)	4.03E-10	NEDD1	intron	15	0.978	T2D
16:81488676	rs2926003	C/T	0.337	-0.0824(0.0151)	4.52E-08	CMIP	intron	48	0.987	
X:19266251-19485409	rs56381411	C/T	0.0153	-0.404(0.0638)	2.44E-10	MAP3K15	missense	4	>0.999	
<i>ClassT Loci - T2D-predominant effect</i>										
6:20673649-20703721	rs34499031	T/TAA	0.332	0.12(0.0148)	5.10E-16	CDKAL1	intron	8	<0.001	GDM,T2D
10:112994312-113014674	rs34872471	T/C	0.203	0.168(0.0173)	1.69E-22	TCF7L2	intron	4	<0.001	GDM,T2D
12:4275678-4367206	rs76895963	T/G	0.0305	-0.26(0.0445)	4.69E-09	CCND2	intron	2	<0.001	T2D
<i>Unclassified Loci</i>										
3:123346931-123405666	rs6798189	G/A	0.184	-0.103(0.0186)	2.60E-08	ADCY5	intron	16	0.138	T2D
9:22129580-22136490	rs1333051	A/T	0.115	-0.126(0.0228)	2.92E-08	CDKN2B	regulatory	5	0.465	GDM,T2D
9:22133646-22134652	rs7019437	C/G	0.438	0.0394(0.0142)	5.49E-03	CDKN2B	intergenic	5	n/a	

For each independent association identified by finemapping with SuSiE, the lead variant (highest posterior inclusion probability [PIP]) and the region spanned by the credible set is reported. Loci are grouped according to their classification in the shared variants analysis (Online Methods, Supplementary Table 22). Genomic positions are on GRCh38. Reference (ref) and alternative (alt) alleles, alternative allele frequency (AF), GWAS results, and nearest coding gene are given for the lead variant. β (log odds ratio), its standard error (SE), and corresponding two-sided P-value from logistic regression using REGENIE. ^aMost severe annotated consequence among variants in the credible set. ^bNumber of variants in the credible set for the region. ^cPosterior probability that the lead variant is in the GDM-predominant class identified in the shared variants analysis. The secondary finemapped

association on chromosome 9 (rs7019437) was omitted from that analysis. ^dWhether the locus has been reported as significantly associated with GDM or T2D in previous GWAS^{15,19,39}.

Figures

A.



B.

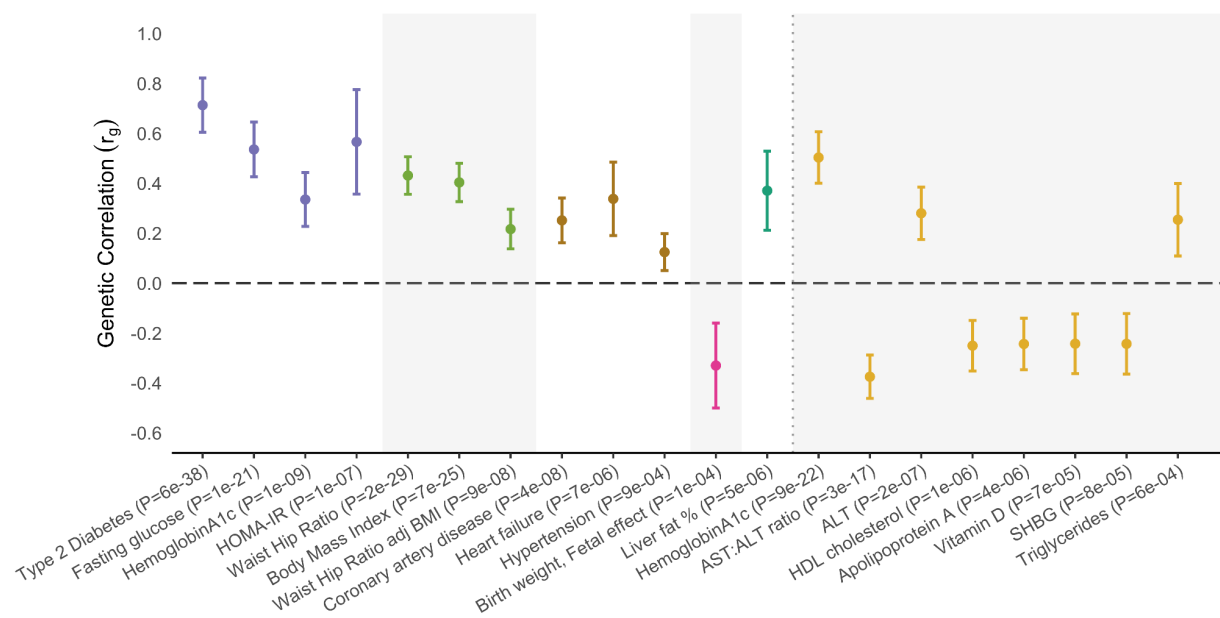


Figure 1: Genome-wide association results for GDM.

(A) Manhattan plot of GWAS of GDM in 12,332 cases and 131,109 parous female controls of Finnish ancestry. The x-axis reflects chromosomal positions and the y-axis reflects $-\log_{10}(P)$ values for the two-tailed association test for each variant, presented on a log scale. Red dotted line indicates the

significance threshold ($P = 5 \times 10^{-8}$). Colored SNPs represent the credible set members for the 13 genome-wide significant loci, with blue indicating loci previously associated with GDM and orange indicating novel associations. Labels indicate the gene nearest to the fine-mapped lead SNP. (B) Genetic correlations (SNP-rg) between GDM and other diseases, traits and biomarkers estimated using LD score regression. Depicted traits are significant after Bonferroni correction for 53 traits ($p < 9.4e-4$); results for all tested traits are reported in Supplementary Tables 15-16. Error bars show +/- 1 standard error. Colors indicate phenotype category.

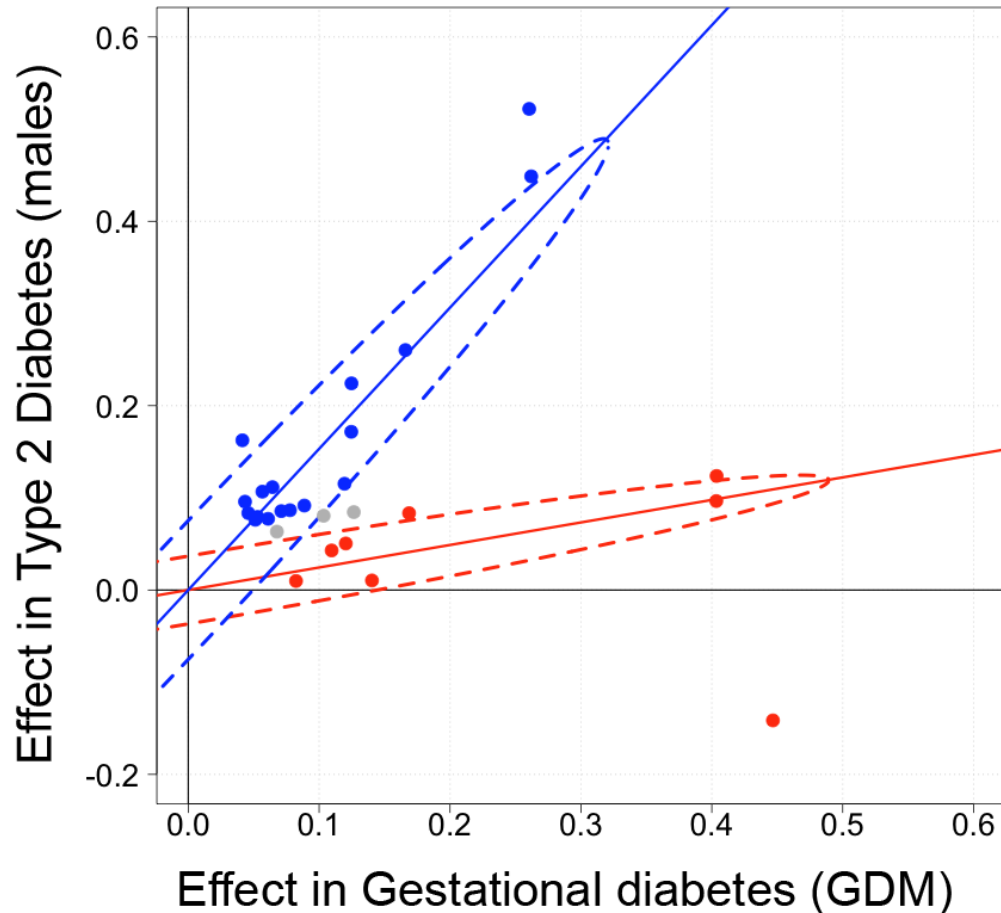


Figure 2: Classification of the genetic effects of SNPs in GDM and T2D.

Comparison of log odds ratios in GWAS of GDM (x-axis) and T2D in males (y-axis) for top-associated SNPs from GDM (13 SNPs) and (15 SNPs). Two distinct classes of SNP effects were identified by a Bayesian classifier in shared variants analysis: Class T (blue) containing SNPs with T2D-predominant genetic effects and Class G (red) with GDM-predominant effects (Supplementary Table 22). Grey SNPs were not confidently assigned to either class (posterior probability > 95%). Dotted ellipses indicate the 95% probability regions of the fitted bivariate effect size distributions with each class.

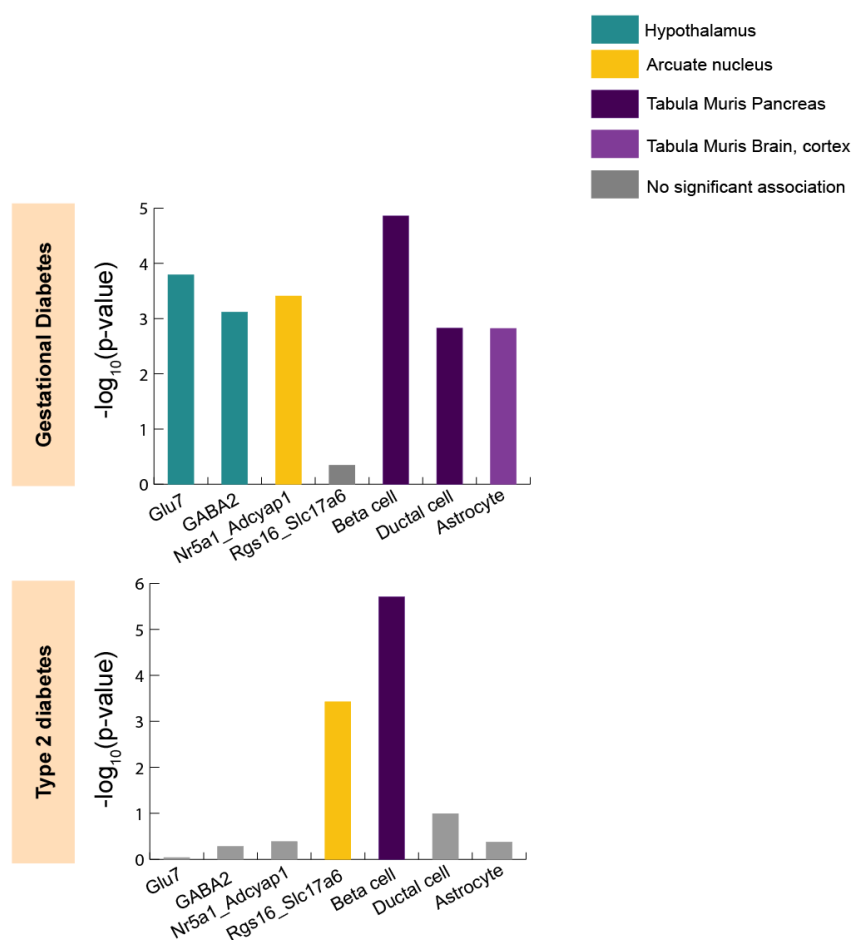


Figure 3: Cell type specificity analysis of GDM and T2D highlights different cell associations.

Cell type specificity analysis was performed for GDM and for prior meta-analysis of T2D from Mahajan et al.¹⁹ using high quality murine single-cell RNA-seq datasets with FUMA (Supplementary Tables 26-29). Unadjusted P-values are reported for the association between relative gene expression in the given cell type and MAGMA gene-level associations in the GWAS. Results are shown for cell types that both (a) are significantly associated with at least one GWAS after correction for multiple testing of all cell types in all datasets, and (b) have putatively independent association conditional on other cell types in the same RNA-seq dataset. Colors indicate RNA-seq dataset source and significance.

References

1. ACOG Practice Bulletin No. 190: Gestational Diabetes Mellitus. *Obstetrics and Gynecology* **131**, e49-e64 (2018).
2. Wang, H., Li, N., Chivese, T., Werfalli, M., Sun, H., Yuen, L., Hoegfeldt, C.A., Elise Powe, C., Immanuel, J., Karuranga, S., Divakar, H., Levitt, N., Li, C., Simmons, D., Yang, X. & Group, I.D.F.D.A.C.H.i.P.S.I. IDF Diabetes Atlas: Estimation of Global and Regional Gestational Diabetes Mellitus Prevalence for 2021 by International Association of Diabetes in Pregnancy Study Group's Criteria. *Diabetes Research and Clinical Practice* **183**, 109050 (2022).
3. Bellamy, L., Casas, J.P., Hingorani, A.D. & Williams, D. Type 2 diabetes mellitus after gestational diabetes: a systematic review and meta-analysis. *Lancet* **373**, 1773-9 (2009).
4. Dennison, R.A., Chen, E.S., Green, M.E., Legard, C., Kotecha, D., Farmer, G., Sharp, S.J., Ward, R.J., Usher-Smith, J.A. & Griffin, S.J. The absolute and relative risk of type 2 diabetes after gestational diabetes: A systematic review and meta-analysis of 129 studies. *Diabetes Res Clin Pract* **171**, 108625 (2021).
5. Auvinen, A.M., Luro, K., Jokelainen, J., Jarvela, I., Knip, M., Auvinen, J. & Tapanainen, J.S. Type 1 and type 2 diabetes after gestational diabetes: a 23 year cohort study. *Diabetologia* **63**, 2123-2128 (2020).
6. McIntyre, H.D., Catalano, P., Zhang, C., Desoye, G., Mathiesen, E.R. & Damm, P. Gestational diabetes mellitus. *Nat Rev Dis Primers* **5**, 47 (2019).
7. Williams, M.A., Qiu, C., Dempsey, J.C. & Luthy, D.A. Familial aggregation of type 2 diabetes and chronic hypertension in women with gestational diabetes mellitus. *J Reprod Med* **48**, 955-62 (2003).
8. Condon, J., Shaw, J.E., Luciano, M., Kyvik, K.O., Martin, N.G. & Duffy, D.L. A study of diabetes mellitus within a large sample of Australian twins. *Twin Res Hum Genet* **11**, 28-40 (2008).
9. Robitaille, J. & Grant, A.M. The genetics of gestational diabetes mellitus: evidence for relationship with type 2 diabetes mellitus. *Genet Med* **10**, 240-50 (2008).
10. Ding, M., Chavarro, J., Olsen, S., Lin, Y., Ley, S.H., Bao, W., Rawal, S., Grunnet, L.G., Thuesen, A.C.B., Mills, J.L., Yeung, E., Hinkle, S.N., Zhang, W., Vaag, A., Liu, A., Hu, F.B. & Zhang, C. Genetic variants of gestational diabetes mellitus: a study of 112 SNPs among 8722 women in two independent populations. *Diabetologia* **61**, 1758-1768 (2018).
11. Powe, C.E. & Kwak, S.H. Genetic Studies of Gestational Diabetes and Glucose Metabolism in Pregnancy. *Curr Diab Rep* **20**, 69 (2020).
12. Kawai, V.K., Levinson, R.T., Adefurin, A., Kurnik, D., Collier, S.P., Conway, D. & Stein, C.M. A genetic risk score that includes common type 2 diabetes risk variants is associated with gestational diabetes. *Clin Endocrinol (Oxf)* **87**, 149-155 (2017).
13. Kwak, S.H., Kim, S.H., Cho, Y.M., Go, M.J., Cho, Y.S., Choi, S.H., Moon, M.K., Jung, H.S., Shin, H.D., Kang, H.M., Cho, N.H., Lee, I.K., Kim, S.Y., Han, B.G., Jang, H.C. & Park, K.S. A genome-wide association study of gestational diabetes mellitus in Korean women. *Diabetes* **61**, 531-41 (2012).
14. Wu, N.N., Zhao, D., Ma, W., Lang, J.N., Liu, S.M., Fu, Y., Wang, X., Wang, Z.W. & Li, Q. A genome-wide association study of gestational diabetes mellitus in Chinese women. *J Matern Fetal Neonatal Med* **34**, 1557-1564 (2021).
15. Pervjakova, N., Moen, G.H., Borges, M.C., Ferreira, T., Cook, J.P., Allard, C., Beaumont, R.N., Canouil, M., Hatem, G., Heiskala, A., Joensuu, A., Karhunen, V., Kwak, S.H., Lin, F.T.J., Liu, J., Rifas-Shiman, S., Tam, C.H., Tam, W.H., Thorleifsson, G., Andrew, T., Auvinen, J., Bhowmik, B., Bonnefond, A., Delahaye, F., Demirkan, A.,

- Froguel, P., Haller-Kikkatalo, K., Hardardottir, H., Hummel, S., Hussain, A., Kajantie, E., Keikkala, E., Khamis, A., Lahti, J., Lekva, T., Mustaniemi, S., Sommer, C., Tagoma, A., Tzala, E., Uibo, R., Vaarasmaki, M., Villa, P.M., Birkeland, K.I., Bouchard, L., Duijn, C.M., Finer, S., Groop, L., Hamalainen, E., Hayes, G.M., Hitman, G.A., Jang, H.C., Jarvelin, M.R., Jenum, A.K., Laivuori, H., Ma, R.C., Melander, O., Oken, E., Park, K.S., Perron, P., Prasad, R.B., Qvigstad, E., Sebert, S., Stefansson, K., Steinthorsdottir, V., Tuomi, T., Hivert, M.F., Franks, P.W., McCarthy, M.I., Lindgren, C.M., Freathy, R.M., Lawlor, D.A., Morris, A.P. & Magi, R. Multi-ancestry genome-wide association study of gestational diabetes mellitus highlights genetic links with type 2 diabetes. *Hum Mol Genet* **31**, 3377-3391 (2022).
16. Mbatchou, J., Barnard, L., Backman, J., Marcketta, A., Kosmicki, J.A., Ziyatdinov, A., Benner, C., O'Dushlaine, C., Barber, M., Boutkov, B., Habegger, L., Ferreira, M., Baras, A., Reid, J., Abecasis, G., Maxwell, E. & Marchini, J. Computationally efficient whole-genome regression for quantitative and binary traits. *Nature Genetics* **53**, 1097-1103 (2021).
17. Kurki, M.I., Karjalainen, J., Palta, P., Sipila, T.P., Kristiansson, K., Donner, K.M., Reeve, M.P., Laivuori, H., Aavikko, M., Kaunisto, M.A., Loukola, A., Lahtela, E., Mattsson, H., Laiho, P., Della Briotta Parolo, P., Lehisto, A.A., Kanai, M., Mars, N., Ramo, J., Kiiskinen, T., Heyne, H.O., Veerapen, K., Rueger, S., Lemmela, S., Zhou, W., Ruotsalainen, S., Parn, K., Hiekkalinna, T., Koskelainen, S., Paajanen, T., Llorens, V., Gracia-Tabuenca, J., Siirtola, H., Reis, K., Elnahas, A.G., Sun, B., Foley, C.N., Aalto-Setälä, K., Alasoo, K., Arvas, M., Auro, K., Biswas, S., Bizaki-Vallaskangas, A., Carpen, O., Chen, C.Y., Dada, O.A., Ding, Z., Ehm, M.G., Eklund, K., Farkkila, M., Finucane, H., Ganna, A., Ghazal, A., Graham, R.R., Green, E.M., Hakanen, A., Hautalahti, M., Hedman, A.K., Hiltunen, M., Hinttala, R., Hovatta, I., Hu, X., Huertas-Vazquez, A., Huilaja, L., Hunkapiller, J., Jacob, H., Jensen, J.N., Joensuu, H., John, S., Julkunen, V., Jung, M., Junttila, J., Kaarniranta, K., Kahonen, M., Kajanne, R., Kallio, L., Kalviainen, R., Kaprio, J., FinnGen, Kerimov, N., Kettunen, J., Kilpelainen, E., Kilpi, T., Klinger, K., Kosma, V.M., Kuopio, T., Kurra, V., Laisk, T., Laukkanen, J., Lawless, N., Liu, A., Longerich, S., Magi, R., Makela, J., Makitie, A., Malarstig, A., Mannermaa, A., Maranville, J., Matakidou, A., Meretoja, T., Mozaffari, S.V., Niemi, M.E.K., Niemi, M., Niiranen, T., CJ, O.D., Obeidat, M.E., Okafo, G., Ollila, H.M., Palomaki, A., Palotie, T., Partanen, J., Paul, D.S., Pelkonen, M., Pendergrass, R.K., Petrovski, S., Pitkaranta, A., Platt, A., Pulford, D., Punkka, E., Pussinen, P., Raghavan, N., Rahimov, F., Rajpal, D., Renaud, N.A., Riley-Gillis, B., Rodosthenous, R., Saarentaus, E., Salminen, A., Salminen, E., Salomaa, V., Schleutker, J., Serpi, R., Shen, H.Y., Siegel, R., Silander, K., Siltanen, S., Soini, S., Soininen, H., Sul, J.H., Tachmazidou, I., Tasanen, K., Tienari, P., Toppila-Salmi, S., Tukiainen, T., Tuomi, T., Turunen, J.A., Ulirsch, J.C., Vaura, F., Virolainen, P., Waring, J., Waterworth, D., Yang, R., Nelis, M., Reigo, A., Metspalu, A., Milani, L., Esko, T., Fox, C., Havulinna, A.S., Perola, M., Ripatti, S., Jalanko, A., Laitinen, T., Makela, T.P., Plenge, R., McCarthy, M., Runz, H., Daly, M.J. & Palotie, A. FinnGen provides genetic insights from a well-phenotyped isolated population. *Nature* **613**, 508-518 (2023).
18. Wang, G., Sarkar, A., Carbonetto, P. & Stephens, M. A simple new approach to variable selection in regression, with application to genetic fine mapping. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **82**, 1273-1300 (2020).
19. Mahajan, A., Taliun, D., Thurner, M., Robertson, N.R., Torres, J.M., Rayner, N.W., Payne, A.J., Steinthorsdottir, V., Scott, R.A., Grarup, N., Cook, J.P., Schmidt, E.M., Wuttke, M., Sarnowski, C., Magi, R., Nano, J., Gieger, C., Trompet, S., Lecoeur, C., Preuss, M.H., Prins, B.P., Guo, X., Bielak, L.F., Below, J.E., Bowden, D.W., Chambers,

- J.C., Kim, Y.J., Ng, M.C.Y., Petty, L.E., Sim, X., Zhang, W., Bennett, A.J., Bork-Jensen, J., Brummett, C.M., Canouil, M., Ec Kardt, K.U., Fischer, K., Kardia, S.L.R., Kronenberg, F., Lall, K., Liu, C.T., Locke, A.E., Luan, J., Ntalla, I., Nylander, V., Schonherr, S., Schurmann, C., Yengo, L., Bottinger, E.P., Brandslund, I., Christensen, C., Dedoussis, G., Florez, J.C., Ford, I., Franco, O.H., Frayling, T.M., Giedraitis, V., Hackinger, S., Hattersley, A.T., Herder, C., Ikram, M.A., Ingelsson, M., Jorgensen, M.E., Jorgensen, T., Kriebel, J., Kuusisto, J., Ligthart, S., Lindgren, C.M., Linneberg, A., Lyssenko, V., Mamakou, V., Meitinger, T., Mohlke, K.L., Morris, A.D., Nadkarni, G., Pankow, J.S., Peters, A., Sattar, N., Stancakova, A., Strauch, K., Taylor, K.D., Thorand, B., Thorleifsson, G., Thorsteinsdottir, U., Tuomilehto, J., Witte, D.R., Dupuis, J., Peyser, P.A., Zeggini, E., Loos, R.J.F., Froguel, P., Ingelsson, E., Lind, L., Groop, L., Laakso, M., Collins, F.S., Jukema, J.W., Palmer, C.N.A., Grallert, H., Metspalu, A., Dehghan, A., Kottgen, A., Abecasis, G.R., Meigs, J.B., Rotter, J.I., Marchini, J., Pedersen, O., Hansen, T., Langenberg, C., Wareham, N.J., Stefansson, K., Gloyn, A.L., Morris, A.P., Boehnke, M. & McCarthy, M.I. 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).
20. Del Valle, I., Buonocore, F., Duncan, A.J., Lin, L., Barenco, M., Parnaik, R., Shah, S., Hubank, M., Gerrelli, D. & Achermann, J.C. A genomic atlas of human adrenal and gonad development. *Wellcome Open Res* **2**, 25 (2017).
 21. Bouazzi, L., Sproll, P., Eid, W. & Bignon-Laubert, A. The transcriptional regulator CBX2 and ovarian function: A whole genome and whole transcriptome approach. *Sci Rep* **9**, 17033 (2019).
 22. Nag, A., Dhindsa, R.S., Harper, A.R., Vitsios, D., Ahnmark, A., Bilican, B., Madeyski-Bengtson, K., Zarrouki, B., Wang, Q., Smith, K., Smith, D., Challis, B., Paul, D.S., Bohlooly-Y, M., Snowden, M., Baker, D., Fritsche-Danielson, R., Pangalos, M.N. & Petrovski, S. Human genetic evidence supports MAP3K15 inhibition as a therapeutic strategy for diabetes. *medRxiv*, 2021.11.14.21266328 (2021).
 23. Stijnen, P., Ramos-Molina, B., O'Rahilly, S. & Creemers, J.W. PCSK1 Mutations and Human Endocrinopathies: From Obesity to Gastrointestinal Disorders. *Endocr Rev* **37**, 347-71 (2016).
 24. Benzinou, M., Creemers, J.W., Choquet, H., Lobbens, S., Dina, C., Durand, E., Guerardel, A., Boutin, P., Jouret, B., Heude, B., Balkau, B., Tichet, J., Marre, M., Potoczna, N., Horber, F., Le Stunff, C., Czernichow, S., Sandbaek, A., Lauritzen, T., Borch-Johnsen, K., Andersen, G., Kiess, W., Korner, A., Kovacs, P., Jacobson, P., Carlsson, L.M., Walley, A.J., Jorgensen, T., Hansen, T., Pedersen, O., Meyre, D. & Froguel, P. Common nonsynonymous variants in PCSK1 confer risk of obesity. *Nat Genet* **40**, 943-5 (2008).
 25. Watanabe, K., Umicovic Mirkov, M., de Leeuw, C.A., van den Heuvel, M.P. & Posthuma, D. Genetic mapping of cell type specificity for complex traits. *Nat Commun* **10**, 3222 (2019).
 26. Shah, N.S., Wang, M.C., Freaney, P.M., Perak, A.M., Carnethon, M.R., Kandula, N.R., Gunderson, E.P., Bullard, K.M., Grobman, W.A., O'Brien, M.J. & Khan, S.S. Trends in Gestational Diabetes at First Live Birth by Race and Ethnicity in the US, 2011-2019. *JAMA* **326**, 660-669 (2021).
 27. Ladyman, S.R. & Grattan, D.R. Region-Specific Suppression of Hypothalamic Responses to Insulin To Adapt to Elevated Maternal Insulin Secretion During Pregnancy. *Endocrinology* **158**, 4257-4269 (2017).
 28. He, Y., Xu, P., Wang, C., Xia, Y., Yu, M., Yang, Y., Yu, K., Cai, X., Qu, N., Saito, K., Wang, J., Hyseni, I., Robertson, M., Piyarathna, B., Gao, M., Khan, S.A., Liu, F., Chen,

- R., Coarfa, C., Zhao, Z., Tong, Q., Sun, Z. & Xu, Y. Estrogen receptor-alpha expressing neurons in the ventrolateral VMH regulate glucose balance. *Nat Commun* **11**, 2165 (2020).
29. Yang, J.A., Stires, H., Belden, W.J. & Roepke, T.A. The Arcuate Estrogen-Regulated Transcriptome: Estrogen Response Element-Dependent and -Independent Signaling of ERalpha in Female Mice. *Endocrinology* **158**, 612-626 (2017).
30. Campbell, J.N., Macosko, E.Z., Fenselau, H., Pers, T.H., Lyubetskaya, A., Tenen, D., Goldman, M., Verstegen, A.M., Resch, J.M., McCarroll, S.A., Rosen, E.D., Lowell, B.B. & Tsai, L.T. A molecular census of arcuate hypothalamus and median eminence cell types. *Nat Neurosci* **20**, 484-496 (2017).
31. Chen, R., Wu, X., Jiang, L. & Zhang, Y. Single-Cell RNA-Seq Reveals Hypothalamic Cell Diversity. *Cell Rep* **18**, 3227-3241 (2017).
32. Taylor, A.J., Chen, Y.-C. & Verchere, C.B. Beta-cell Prohormone Convertase 1/3 Plays a Larger Role in Glucose Homeostasis Than Beta-cell Prohormone Convertase 2 in Mice. *The FASEB Journal* **34**, 1-1 (2020).
33. Lyssenko, V., Nagorny, C.L., Erdos, M.R., Wierup, N., Jonsson, A., Spegel, P., Bugliani, M., Saxena, R., Fex, M., Pulizzi, N., Isomaa, B., Tuomi, T., Nilsson, P., Kuusisto, J., Tuomilehto, J., Boehnke, M., Altshuler, D., Sundler, F., Eriksson, J.G., Jackson, A.U., Laakso, M., Marchetti, P., Watanabe, R.M., Mulder, H. & Groop, L. Common variant in MTNR1B associated with increased risk of type 2 diabetes and impaired early insulin secretion. *Nat Genet* **41**, 82-8 (2009).
34. Laakso, M., Kuusisto, J., Stancakova, A., Kuulasmaa, T., Pajukanta, P., Lusi, A.J., Collins, F.S., Mohlke, K.L. & Boehnke, M. The Metabolic Syndrome in Men study: a resource for studies of metabolic and cardiovascular diseases. *J Lipid Res* **58**, 481-493 (2017).
35. Hormozdiani, F., van de Bunt, M., Segre, A.V., Li, X., Joo, J.W.J., Bilow, M., Sul, J.H., Sankararaman, S., Pasaniuc, B. & Eskin, E. Colocalization of GWAS and eQTL Signals Detects Target Genes. *Am J Hum Genet* **99**, 1245-1260 (2016).
36. Verbanck, M., Chen, C.-Y., Neale, B. & Do, R. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. *Nature Genetics* **50**, 693-698 (2018).
37. York, D. Least squares fitting of a straight line with correlated errors. *Earth and Planetary Science Letters* **5**, 320-324 (1968).
38. Tabula Muris Consortium. Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. *Nature* **562**, 367-372 (2018).
39. Vujkovic, M., Keaton, J.M., Lynch, J.A., Miller, D.R., Zhou, J., Tcheandjieu, C., Huffman, J.E., Assimes, T.L., Lorenz, K., Zhu, X., Hilliard, A.T., Judy, R.L., Huang, J., Lee, K.M., Klarin, D., Pyarajan, S., Danesh, J., Melander, O., Rasheed, A., Mallick, N.H., Hameed, S., Qureshi, I.H., Afzal, M.N., Malik, U., Jalal, A., Abbas, S., Sheng, X., Gao, L., Kaestner, K.H., Susztak, K., Sun, Y.V., DuVall, S.L., Cho, K., Lee, J.S., Gaziano, J.M., Phillips, L.S., Meigs, J.B., Reaven, P.D., Wilson, P.W., Edwards, T.L., Rader, D.J., Damrauer, S.M., O'Donnell, C.J., Tsao, P.S., Consortium, H., Regeneron Genetics, C., Program, V.A.M.V., Chang, K.-M., Voight, B.F. & Saleheen, D. Discovery of 318 new risk loci for type 2 diabetes and related vascular outcomes among 1.4 million participants in a multi-ancestry meta-analysis. *Nature Genetics* **52**, 680-691 (2020).

Competing Interest Statement

R.W. has received honoraria from The Jackson Laboratory and sponsored travel from the Russell Sage Foundation A.P. is the chief scientific officer for the FinnGen project that has funding from 13 pharmaceutical companies. M.J.D is a founder of Maze Therapeutics. The remaining authors have no competing interests to declare.

Funding Statement

The FinnGen project is funded by two grants from Business Finland (HUS 4685/31/2016 and UH 4386/31/2016) and by eleven industry partners (AbbVie Inc, AstraZeneca UK Ltd, Biogen MA Inc, Celgene Corporation, Celgene International II Sàrl, Genentech Inc, Merck Sharp & Dohme Corp, Pfizer Inc., GlaxoSmithKline, Sanofi, Maze Therapeutics Inc., Janssen Biotech Inc). This work received the following support: A.E. was a research Scholar supported by Sarnoff Cardiovascular Research Foundation, R.W. was supported by R01MH101244, M.P. was supported by the Academy of Finland (336825, 338507), P.T. received support by NIH/NIA (R00-AG062787), E.W. received support by the Academy of Finland Center of Excellence program (352796).

Availability Statement

FinnGen summary data (genome-wide association summary statistics and phenotype data) are made freely available every 6 month. The FinnGen data may be accessed through Finnish Biobanks' FinnBB portal (www.finbb.fi) and THL Biobank data through THL Biobank (<https://thl.fi/en/web/thl-biobank>).