# Exome-wide association study in 54,698 south Asians identifies novel type 2 diabetes associations with *HNF4A* and *GP2*

Sam Hodgson [1]*, Vi Bui [1]*, Siqi Hu[2]*, Margherita Bigossi[1], Daniel Stow[1], Cyrielle Maroteau[3], Alice Williamson[4,5], Alexandra M Blee[2], Adem Y. Dawed[3], Julia Carrasco-Zanini Sanchez[4], Viswanathan Baskar[6], Saravanan Jebarani[6], Benjamin M Jacobs[1], Georgios Kalantzis[7], Stuart Rison[1], Klaudia Walter[7], Erwan Pennarun[1], Genes & Health Research Team, MDRF Research Team[6], Hilary C Martin[7], Inês Barroso[8], Matteo Fumagalli[9], Venkatesan Radha[6], Rajendra Pradeepa[6], Claudia Langenberg[3,4], Viswanathan Mohan[6,10], Ranjit Mohan Anjana[6,10], David A van Heel[11], Yalda Jamshidi[3], Sarah Finer[1,12], Amit R. Majithia[2±], Moneeza K Siddiqui [1,6±]

*These authors contributed equally to this work

[±]These authors jointly supervised this work

1. Wolfson Institute of Population Health, Queen Mary University of London, United Kingdom

2. Division of Endocrinology & Metabolism, University of San Diego School of Medicine

3. Novo Nordisk Research Centre Oxford, Oxford, United Kingdom

4. Precision Healthcare University Research Institute, Queen Mary University of London, United Kingdom

5. Berlin Institute of Health at Charite, Universitätsmedizin, Berlin, Germany

6. Madras Diabetes Research Foundation (ICMR- Collaborating Centre of Excellence), Chennai, India

7. Wellcome Sanger Institute, Hinxton, United Kingdom

8. Exeter Centre of Excellence for Diabetes Research (EXCEED), University of Exeter Medical School, Exeter, United Kingdom

9. School of Biological and Behavioural Sciences, Queen Mary University of London, United Kingdom

10. Dr. Mohan's Diabetes Specialities Centre (IDF Centre of Excellence in Diabetes Care), Chennai, India

11. Blizard Institute, Queen Mary University of London, United Kingdom

12. Barts Health NHS Trust, London, UK

**Address for correspondence:**
Moneeza K Siddiqui
Wolfson Institute of Population Health,

Queen Mary, University of London
email: moneeza.siddiqui@qmul.ac.uk

# Abstract

Type 2 diabetes (T2D) disproportionately affects individuals of South Asian ancestry (SAS), yet they remain underrepresented in genetic studies. We performed an exome-wide association study in 54,698 SAS T2D case:controls and follow-up metabolic trait evaluation. We identified ancestry-specific genes and protein-coding variants, including a SAS-specific variant in the known monogenic diabetes gene *HNF4A* (rs150776703, Pro437Ser), which was associated with protection from T2D (OR = 0.48, $p = 2.8 \times 10^{-1}$), diabetic eye disease, and gestational diabetes. Experimental interrogation of *HNF4A* Pro437Ser revealed context-dependent enhancement of *HNF4A* transcriptional activity, suggesting gain-of-function. We additionally describe a T2D risk increasing variant in *GP2* (rs78193826, Val429Met, OR = 1.21, $p = 5.14 \times 10^{-6}$), which was also associated with beta-cell dysfunction. These findings underscore how studying ancestrally distinct populations disproportionately affected by T2D can reveal novel disease genes, therapeutic hypotheses, and biological insight.

# Introduction

Type 2 diabetes (T2D) is a common complex metabolic condition with significant heterogeneity both within and across ancestries[1–4]. Compared to individuals of European ancestry (EUR), people of South Asian ancestry (SAS) are 2-4 fold higher risk of T2D, and develop disease at younger ages and lower Body Mass Index (BMI), and have more severe clinical progression to complications [5–9]. Applying recently-developed genetic methods to characterise heterogeneity in T2D aetiology has shown that insulin deficiency and unfavourable fat distribution are key drivers of early-onset T2D in SAS and are associated with worse outcomes and higher rates of complications[10–12].

Despite contributing disproportionately to the global burden of T2D, [13] and comprising over 25% of the global population, they represent less than 0.4% of populations in genetic studies of T2D[1] . The most recent multi-ancestry T2D case-control exome-wide association study in 45,231 individuals contained only 5,833 SAS ancestry samples after aggregation of four cohorts[14]. This lack of representation limits efforts to understand disease mechanisms and develop effective prevention and management strategies, not only in South Asian countries but also in countries with large diasporic south Asian populations. Increasing representation in genomic studies also creates opportunities to identify novel genes and variants relevant to T2D pathophysiology. Indeed, this was illustrated by a recent exome-wide association study of 646 traits in 44,028 British Pakistanis and Bangladeshis in the Genes & Health Study[15]. In this population, 100 novel gene-phenotype associations and 546 genes with never previously reported homozygous loss-of-function variants were discovered.

Building on these observations, we conducted an exome-wide study of T2D in South Asians that represents a substantial increase in ancestry-specific sample size and case numbers relative to prior efforts. We validate and characterise identified associations through replication in independent South Asian cohorts, multi-ancestry meta-analyses, *in vitro* functional assays, and integration of proteomic data, enabling more detailed interrogation of the biological mechanisms underlying T2D risk in this population.

# Results

We included 54,698 SAS individuals (13,674 electronic health record-defined cases and 41,024 controls) from Genes & Health[16] in an exome-wide association study **(Figure 1)**. Descriptive details of included participants are given in **Table 1,** including event rates for diabetes-related complications, stratified by sex and genetically-inferred ancestry. We performed replication and phenotyping of variants of interest in 19,898 T2D cases and 4,711 controls of SAS ancestry from the Madras Diabetes Research Foundation[17].

## Gene-level testing for type 2 diabetes associations identifies a novel protective association in *HNF4A* driven by an ancestry-specific single variant

We performed analyses across 18,956 genes by aggregating rare variants with minor allele frequency (MAF) < 1% into functional categories (i.e. masks, as previously described[15] and in **Table S1**) and testing the association of these aggregates with type 2 diabetes. After correcting for multiple-testing ($p<9.35 \times 10^{-8}$ , **methods**) we identified three significant gene/mask combinations (*MAP3K15*: pLoF-HC|pDM; *RNF19A*:pLoF-HC; and *HNF4A*: pLoF-HC|pDM; variants constituting each mask given in **Table S2**) (**Fig 2A/B, S1)**.

Consistent with our findings, rare *MAP3K15* variants have previously been reported to associate with decreased T2D risk[18,19,20].(**Fig S2A/B)**. However, *RNF19A* has not been previously associated with T2D [21]. *RNF19A,* a highly constrained gene, has previously been linked to pancreatic body size and pancreatic cancer[21], but not diabetes-related phenotypes. In Genes & Health, the *RNF19A*:pLoF-HC mask was also nominally associated with HbA1c, BMI, HDL-C, and macrovascular diabetic complications including heart failure, peripheral arterial disease, myocardial infarction and transient ischaemic attack (**Fig S2C/D, Table S3)**. Further information on *RNF19A* and *MAP3K15* results are provided in the **supplementary text** and **Fig S2.**

Protein coding variant in HNF4A have been previously reported to increase T2D risk[18,19], but we identified the opposite direction of effect in our analysis. (*HNF4A*:pLoF-HC|pDM mask T2D OR = 0.56, 95% CI = 0.48 – 0.67, $p = 1.20 \times 10^{-11}$ ;**Fig 2A/B, S1,S3A)**. Given that LoF mutations in *HNF4A* are a known cause of monogenic diabetes (maturity onset diabetes of the young MODY)[18,22] and gene burden tests with masks containing low frequency, deleterious variants have been shown to increase T2D risk[21], we dissected our signal by conducting leave one variant out (LOVO) analysis. (**Fig 3A)**.

The LOVO analysis revealed that the gene-level association with type 2 diabetes was driven predominantly by a single missense variant, rs150776703 (*HNF4A* Pro437Ser) (**Fig. 3A**). Removal of this variant markedly attenuated the protective

signal (OR = 1.39, 95% CI = 0.78 – 2.45, p = 0.25), indicating that the aggregate association did not reflect a consistent effect across multiple rare variants.

Complementary single-variant association testing showed that the initial burden signal was driven by a single variant encoding p.Pro437Ser which associated with a reduced risk of type 2 diabetes (OR= 0.48, 95% CI = 0.40–0.59, P = $2.8 \times 10^{-1}$). The variant was observed at low frequency in South Asians (MAF = 0.008) and West Asians (MAF = 0.0003) and was extremely rare or absent in other ancestry groups (MAF < $1 \times 10^{-}$) (**Fig. 3B and supplementary text**). Among 13 homozygous carriers identified in Genes & Health, none had a recorded diagnosis of type 2 diabetes, compared with a prevalence of 24.9% in the full cohort.

*HNF4A* Pro437Ser was also associated with multiple glycaemic traits. In Genes & Health, carriers exhibited lower lifetime median HbA1c (beta = −2.23, 95% CI = −3.10 to −1.37, P = $4.0 \times 10^{-}$) and lower non-fasting glucose levels (beta = −0.44, 95% CI = −0.63 to −0.26, P = $2.4 \times 10^{-}$), as well as reduced risk of diabetic eye disease (beta = −0.04, P = $3.0 \times 10^{-}$) (**Fig. 3C; Table S4**). Nominal associations were observed with lower 120-minute oral glucose tolerance test values (beta = −2.35, 95% CI = −4.12 to −0.58, P = 0.009) and later age at type 2 diabetes diagnosis (ANOVA P = 0.01).

In addition to glycaemic phenotypes, *HNF4A* Pro437Ser was associated with higher lifetime total cholesterol (beta = 0.21, 95% CI = 0.12–0.32, P = $1.3 \times 10^{-}$) and nominally with non-HDL cholesterol (beta = 0.09, 95% CI = 0.02–0.17, P = 0.009) **(Fig. 3C; Table S4).** A non-significant trend towards increased coronary artery disease risk was observed (OR = 1.16, 95% CI = 0.96–1.40, P = 0.12). In longitudinal analyses using linked healthcare records, Pro437Ser was associated with lower risk of type 2 diabetes but was nominally associated with increased hazard of myocardial infarction (hazard ratio = 1.62, 95% CI = 1.09–2.44, P = 0.01) (**Fig. S3F,G; Table S4**). Further information on metabolic trait associations is given in **Supplementary Results.**

*HNF4A* Pro437Ser lies within the repressor domain of *HNF4A* (Fig. 3A)[23], whereas previously reported MODY-causing variants predominantly localise to upstream non-repressor regions. To assess whether variant location within *HNF4A* influenced direction of effect, we constructed domain-specific rare variant masks, aggregating coding variants within either non-repressor (n = 75) or repressor (n = 19) domains, and tested their association with T2D (**Fig. 3D; Table S5**). Associations across these analyses were modest and did not consistently reach statistical significance. Missense variants within the repressor domain showed a trend towards reduced type 2 diabetes risk, driven largely by inclusion of Pro437Ser. In contrast, masks capturing MODY-causing variants outside the repressor domain—variants that occur at lower frequency in South Asians than in Europeans—were associated with increased diabetes risk, with several masks reaching nominal significance.

Repressor domain variants were rare or ultra-rare in South Asians and largely absent or even rarer in Europeans, limiting independent replication in European-ancestry cohorts such as UK Biobank **(Table S5).**

*HNF4A* is a transcription factor expressed predominantly in the liver, with additional expression in the colon, kidney, and pancreas. To assess the transactivation potential of Pro437Ser variant in hepatocytes, *HNF4A* wild-type (WT), Pro437Ser, or the known MODY associated variant Arg89Trp[24] were transfected into Huh7 cells. Western blotting revealed robust overexpression of *HNF4A* for each variant, compared to minimal detection of endogenous HNF4A in the control cells (**Fig 3E**). WT *HNF4A* induced expression of a known hepatic HNF4A transcriptional target, G6PC, by approximately 28-fold (**Fig 3F**). *HNF4A* P437S exhibited a similar degree of activity to WT while the MODY associated R89W variant, which inhibits nuclear localization[24], displayed the lowest transactivation (**Fig 3F**).

To further compare the transcriptional activity, we performed dual-luciferase reporter assays using either a mouse G6PC promoter or a synthetic PPRE×3 minimal promoter. Both reporters showed dose-dependent activation by *HNF4A* variants, but the magnitude and dynamic range of the signal intensities differed substantially across promoter contexts. The mG6PC promoter, which contains non-consensus *HNF4A* motifs and requires additional co-factors for full transcriptional activation[25], exhibited a nonlinear activation profile with no clear separation between WT and Pro437Ser across the tested doses (**Fig 3G**). In contrast, the PPRE×3 construct, which contains three consensus DR1-type motifs, displayed a near-linear activation across the dilution series, and modestly higher transcriptional activity of the Pro437Ser variant relative to WT (**Fig 3H**). Overall, these gene transactivation assays in human cell lines indicate that *HNF4A* Pro437Ser is likely a context-dependent gain-of-function variant that enhances *HNF4A* transcriptional activity when DNA-binding and cofactor requirements are minimal, but the effect is masked on native metabolic promoters such as *G6PC*, where transcriptional output is constrained by low-affinity motifs and the absence of required cofactors.

## Single variant exome-wide association study (ExWAS) identifies coding variant in *GP2 and CDKAL1*

We performed single variant exome-wide association testing of type 2 diabetes status on 13,674 cases and 41,024 controls in Genes & Health using REGENIE, including only 25,236 missense coding variants with MAF > 0.01 (aligning with masks included in gene-based testing, **Table S1)**.

We identified one variant passing exome-wide significance (chr8:117172544:C:T, (rs13266634), a missense variant in *SLC30A8* which has been previously identified in imputed genotype data and known to be strongly associated with T2D (previously reported OR = 0.91, p = 8.6x $10^{-250}$ ) [21]. We additionally identified suggestive signals (P < 5e-5) mapping to four genes: *GP2, CDKAL1, CCNL2, and WFS1,* which we assessed using meta-analyses (**Fig 2C, Table S6); f**urther summary information for

these variants is shown in **Table 2**. After meta-analysis with existing associations obtained from the T2DKP, we observed stronger associations for *GP2 Val429Met (*OR = 1.10, 95% CI 1.08 – 1.11, p = 2.04 x $10^{-37}$*), CDKAL1 Pro409Arg (*OR = 1.02, 95% CI 1.02 – 1.03, p = 1.88x10x$^{-12}$*)* and *WFS1 Val333Ile (*OR = 1.063, 95% CI = 1.059 – 1.067, p = 6.5X$10^{-232}$, further detail given in **Supplementary text***),* and therefore focused downstream analyses on these coding variants with external evidence for replication.

**Ancestry-specific variant in pancreatic acinar cell-expressed *GP2* is associated with T2D via a beta-cell deficiency pathway**

We identified a missense variant, rs78193826 (C>T; Val429Met), in the *GP2* locus that was associated with increased risk of T2D (OR = 1.21, p = 5.14x10-6). To date, this variant has been reported only in individuals of East Asian ancestry in the Asian Genetic Epidemiology Network (AGEN) consortium (OR = 1.09, p = 1.43 × $10^{-}$ , n = 254,373)[26], despite inclusion in larger multi-ancestry studies. This likely reflects substantial differences in allele frequency across populations: *GP2* Val429Met is most common in East Asians (MAF 0.07), followed by South Asians (MAF 0.03; Genes & Health Bangladeshis 0.047 vs Pakistanis 0.025; **Fig. 4B**), and is rare or ultra-rare in other ancestries.

Using population genetic simulation methods similar to those used to understand genetic drift in Greenlanders [27], we found that under a simple model of neutral evolution, the observed allele frequency differentiation between Europeans and South Asians (p = 0.0012), and between Europeans and East Asians (p = 0.0011), was unlikely to have arisen by chance alone.

We then meta-analysed the effect of the variant on glycaemic traits, using data from G&H, MDRF and EAS cohorts in T2DKP[26,28], and insulin response from the MAGIC consortium[29]. Val429Met was significantly associated with higher fasting glucose (beta = 0.03, 95% CI 0.02 – 0.04, p = 7.4e-11) and nominally with lower 2hr C-peptide levels (beta = -0.08, 95% CI = -0.13 - -0.03, p = 0.002, n=14,704)(**Fig 4A, Fig S4, Table S4).** In contrast, the variant was not associated with any measures of insulin resistance, e.g. Triglycerides: HDL-C ratio (p = 0.23), and was only nominally associated with lower HDL-C (beta = -0.009, 95% CI = -0.015 - -0.002, p = 0.008) and BMI (beta = -0.009, 95% CI = -0.016 - -0.002, p = 0.02). Consistent with these findings, *GP2* variants have previously been reported to be associated with use of diabetes-controlling medication in Biobank Japan[30].

Expression quantitative trait data from GTEx[31], the Translational Human Pancreatic Islet Genotype Tissue-Expression Resource (TIGER)[32], and Pankbase[33] indicated *GP2* expression is largely restricted to the pancreas, with minimal expression in islet cells, and predominant expression in acinar cells (**Figure S5A/B).**

Our proteomic analyses show an increase in plasma abundance of the GP2 *cis*-protein on the Olink platform (beta = 0.29, SE = 0.09, p = 0.001) but not SomaScan (beta = 0.05, SE = 0.10, p = 0.63, **Fig 4C**). To support interpretation of plasma protein origin, we used a recently published human-proteome distribution atlas from 29 tissue and cell types and providing a Global Label Score (GLS) [ranging from 1-4], where a higher score denotes greater confidence for inferring the tissue of origin for plasma proteins[34] . *GP2* was assigned a high GLS of 3, consistent with pancreatic secretion (**Fig S5C).**

To further contextualise these findings within the heterogeneous genetic architecture of T2D, we examined associations between previously defined genetic endotypes called partitioned polygenic scores (pPS)[11] and plasma *GP2* levels in individuals of European ancestry from UK Biobank and in G&H (Olink), irrespective of diabetes status. Development of pPS and curation of Olink data have been described previously[1,10,35.]

Meta-analysis demonstrated that higher genetic burden related to beta-cell dysfunction was associated with higher plasma *GP2* levels [beta-cell 1 (insulin sensing): beta = 0.02, SE = 0.004, p = 1.74 × $10^{-}$□; beta-cell 2 (insulin production): beta = 0.02, SE = 0.004, p = 2.35 × $10^{-}$□] (**Fig. 4D, Fig. S5D**). In contrast, higher genetic burden related to BMI - and cholesterol-mediated T2D was associated with lower plasma *GP2*. A further association was observed for an insulin-resistance-related pPS characterized by lower alkaline phosphatase and cholesterol levels (beta = 0.37, SE = 0.08, p = 7.2 × $10^{-}$□); however, this score comprises only nine variants and the association was driven predominantly by a single variant in ABO (rs529565), previously reported to associate with alkaline phosphatase (**Fig. S5D**). Meta-analysis demonstrated that higher genetic burden related to beta-cell dysfunction was associated with higher plasma *GP2* levels [beta-cell 1 (insulin sensing): beta = 0.02, SE = 0.004, p = 1.74 × $10^{-}$□; beta-cell 2 (insulin production): beta = 0.02, SE = 0.004, p = 2.35 × $10^{-}$□] (Fig. 4D, Fig. S5D). In contrast, higher genetic burden related to BMI - and cholesterol-mediated T2D was associated with lower plasma *GP2*. A further association was observed for an insulin-resistance-related pPS characterised by lower alkaline phosphatase and cholesterol levels (beta = 0.37, SE = 0.08, p = 7.2 × $10^{-}$□); however, this score comprises only nine variants and the association was driven predominantly by a single variant in ABO (rs529565), previously reported to associate with alkaline phosphatase (Fig. S5D). (**Fig S5D**).

**A missense variant Pro409Arg may be associated with T2D independently of previously reported intronic variants in *CDKAL1***

To date, T2D-associated variants in *CDKAL1* have been intronic [36–38]. Here, we identify a coding missense variant, Pro409Arg (rs77152992), associated with T2D (meta-analysed OR = 1.02, 95% CI 1.02–1.03, p = $1.88\times10^{-12}$) and glycaemic traits, including fasting glucose (beta = 0.012, p = $7.7\times10^{-}$□) and HbA1c (beta = 0.002, p =

0.001), as well as earlier T2D onset (beta = -0.47, p = 0.001), progression to insulin therapy, and microvascular complications (HR = 1.17 and 1.10, p = $1.0\times10^{-}$ and 0.001, respectively) **(Fig S4, Table S4)**.

In MDRF, Pro409Arg was associated with lower BMI (beta = -0.22, p = $8.1\times10^{-}$) and nominally with higher 1h OGTT glucose, lower HOMA-IR, HOMA-B, and fasting/stimulated C-peptide, supporting a role in beta-cell function and glucose sensing. Burden testing in UK Biobank shows that rare damaging variants (excluding Pro409Arg) in Europeans are associated with earlier-onset diabetes (B = -0.57, p = $6.2\times10^{-}$).

Fine-mapping with SuSIE[39] and stratified analyses among carriers of the lead intronic variant suggest that Pro409Arg may contribute independently of previously reported intronic variants **(Fig 5A–E)**. Further detail is provided in the **Supplementary Results**.

# Discussion

## Summary

Irrespective of ancestry, this represents the largest exome sequencing study of type 2 diabetes to date, incorporating both single-variant and gene-level analyses[14]. We identify a protective variant in *HNF4A*, a gene classically implicated in monogenic diabetes, and highlight associations with *RNF19A*, as well as coding variants in *GP2* and *CDKAL1*. Integration of longitudinal health records with genomic and proteomic data enabled us to examine downstream health impacts and gain mechanistic insight into variant function.

Focusing on an under-represented South Asian population allowed the discovery of genetic effects not observable in European cohorts, exemplified by the SAS-specific, likely gain-of-function variant in *HNF4A* (Pro437Ser) and the EAS- and SAS-specific variant in *GP2* (Val429Met).

The discovery of Pro437Ser in *HNF4A*, underscores the value of studying under-represented ancestries with longitudinal health data. We identify Pro437Ser as a South Asian–enriched missense variant associated with reduced risk of type 2 diabetes and improved glycaemic traits, alongside divergent associations with lipid-related and cardiovascular outcomes. These contrasting effects provide important context for interpretation and highlight the need for caution when considering *HNF4A*-related pathways as therapeutic targets for type 2 diabetes.

Pro437Ser is located in the repressor domain of *HNF4A*[23], and is likely exerting its putative gain-of-function effect by interrupting the repression of the transcription factor. Beyond single-variant associations, our domain-specific burden analyses illustrate how incorporating functional annotation within transcription factors can

reveal patterns that are not apparent in conventional gene-level tests. By stratifying coding variation according to regulatory domains within *HNF4A*, we observed opposing directional effects that are consistent with known differences in transcriptional regulation across protein regions. These findings highlight the value of biologically informed aggregation strategies for interpreting rare variation in regulatory genes, particularly in ancestries where genetic architecture differs from well-studied European populations.

Functional analyses indicate that Pro437Ser preserves wild-type levels of transcriptional activity on a native metabolic promoter, while exhibiting modestly higher activity in simplified reporter contexts with reduced regulatory complexity. Together, these observations suggest that the phenotypic effects associated with Pro437Ser may reflect context-dependent differences in transcriptional regulation related to its localization within the repressor domain, rather than a uniform gain of function across all targets.

The Val429Met variant in *GP2* similarly illustrates the value of population-enriched discovery. This missense variant is more frequent in South and East Asian populations and is associated with increased risk of type 2 diabetes and glycaemic traits reflecting impaired insulin secretion, without associations with established measures of insulin resistance. The marked differences in allele frequency of Val429Met across ancestries further emphasise the value of population-specific studies for identifying biologically informative variation. While neutral evolutionary models were insufficient to explain the observed degree of differentiation between European, South Asian and East Asian populations, the relative contributions of selection and alternative demographic processes remain unresolved.

*GP2* encodes glycoprotein 2, an integral membrane protein highly expressed in the pancreas and a marker of pancreatic progenitor cells capable of differentiating into beta cells in vitro and in vivo[40]. Expression data indicate minimal *GP2* expression in pancreatic islet cells and predominant expression in acinar cells. While this pattern argues against a direct role within mature beta cells, progenitor studies support the possibility that genetic variation may influence diabetes risk indirectly through pancreatic development or exocrine–endocrine interaction[40].

Beyond genetic associations, Val429Met was associated with increased circulating *GP2* levels measured using one proteomic platform. Plasma *GP2* levels were associated with specific genetic endotypes of type 2 diabetes in both South Asians and white Europeans: higher beta-cell dysfunction burden was associated with higher GP2 levels, whereas obesity- and insulin-resistance–related burden was associated with lower *GP2* levels. These opposing patterns suggest that circulating *GP2* may reflect heterogeneity in diabetes pathophysiology, particularly with respect to beta-cell sensing and secretory dysfunction.

However, our data do not permit inference of causality, due to limited availability of proteomic data which precludes approaches such as Mendelian randomisation.

Therefore, it remains unclear whether elevated *GP2* levels reflect altered protein function, compensatory secretion, or downstream consequences of pancreatic dysfunction. Further functional and longitudinal studies will therefore be required to clarify the role of GP2 in diabetes pathogenesis and to assess its potential utility as a biomarker across diverse populations.

## Limitations

The sample size of 54,698 remains modest for the discovery of rare and low-frequency variants with small effect sizes. We therefore increased power where possible through meta-analysis with external datasets. Phenotype ascertainment was based on routinely collected electronic healthcare records, which may be subject to miscoding, misclassification, or delayed diagnosis. To mitigate this, we leveraged the UK National Diabetes Audit[41], which uses structured data entry and guideline-based diagnostic and monitoring criteria, and incorporated longitudinal health records to reduce case–control misclassification relative to cross-sectional designs.

A further limitation is the limited availability of large-scale genomic datasets including individuals of South Asian ancestry, which constrained opportunities for independent replication. Where direct replication was not feasible, we sought supporting evidence through meta-analysis with ancestrally distinct but biologically concordant variants; however, additional studies in diverse populations will be required to confirm and extend these findings.

In this exome-wide association study of type 2 diabetes in South Asian populations, we demonstrate that integrating exome sequencing and multi-omics with longitudinal clinical in under-represented ancestries can uncover population-enriched genetic associations and provide insight into biologically relevant pathways. Our findings highlight the value of expanding genetic studies beyond well-characterised European populations to better capture the diversity of diabetes risk and pathophysiology, with potential relevance for understanding disease mechanisms across global populations.

# Methods

## Study design

We performed a case-control genome- and exome-wide association study using participant data from the Genes & Health study based in the UK, and investigated phenome-wide associations of variants of interest. We extended phenome association analyses to the Madras Diabetes Research Federation (MDRF)[17] biobank based in India, and replicated and meta-analysed findings using UK Biobank data via the AstraZeneca PheWAS portal[42], and the type 2 diabetes knowledge portal (T2DKP)[21], a live, open access resource collating and meta-analysing summary statistics across ancestries for type 2 diabetes and related traits. An overview of study design and features in line with the STREGA reporting guideline[43] is presented in the accompanying STREGA checklist.

## Setting

Genes & Health is a long-term community-based study, funded to recruit 100,000 British Bangladeshi and British Pakistani individuals in the UK aged 16 and over, linking genetic testing to routine electronic health record (EHR) data. To date, Genes & Health has recruited 70k volunteers, among whom 55,273, have whole exome sequencing data, and 51,170 have imputed genotype data, with linkage to primary and secondary care clinical codes, laboratory test results, prescribing data, and multi-platform proteomic and metabolic data.

We replicated and meta-analysed findings in a separate setting, the MDRF biobank in India[17]. This is a national-level, multicentre, clinic-based disease registry, combining phenotypic data with genetic testing in a nationally-representative sample of individuals with T2D (n = 19,898) and healthy controls (n = 4,711).

## Participants

*Inclusion criteria*: We used no specific inclusion criteria beyond those of the Genes & Health study.

*Exclusion criteria*: We excluded from analyses individuals diagnosed with type 2 diabetes before 18 years of age, and those with a clinical code of type 1 diabetes, maturity onset diabetes of the young (MODY), neonatal diabetes, secondary diabetes (eg drug-induced diabetes), or clinical conditions strongly predisposing to secondary diabetes (eg cystic fibrosis, pancreatectomy). Clinical code lists used for exclusion criteria are provided in **Table S9**.

***Case definition (Genes & Health):*** We defined type 2 diabetes as an entry of type 2 diabetes in the National Diabetes Audit (NDA)[41], an annual England-wide registry of type 2 diabetes patients. Because NDA data are not yet available from 2023 onwards, we also defined as type 2 diabetes individuals with a first clinical code of type 2 diabetes in the EHR from 2023 onwards. For cases, we defined “age at diagnosis” as the age at the earliest clinical code of type 2 diabetes in the EHR.

*Control definition (Genes & Health)*: We defined controls as individuals *with no record of type 2 diabetes in the* NDA, and with no records of *type 2 diabetes* in the EHR at any time.

*Case & Control Definitions (MDRF):* The MDRF clinical cohort comprises of data extracted from electronic medical records (EMR) of Dr,Mohan's Diabetes Specialities Centre, a large tertiary diabetes care network with over 50 centers across 32 cities in 8 states of India. For individuals who underwent GWAS, demographic variables including sex, age, diabetes type (eg. type 2 diabetes, MODY, etc.), age at diagnosis and duration of diabetes as well as anthropometric details, biochemical data were extracted. Fasting and stimulated (post-breakfast) serum C-peptide concentrations were quantified using an electrochemiluminescence immunoassay on the Elecsys 2010 analyser (Hitachi). Venous blood samples were obtained after an overnight fast of >8 hours, followed by collection of a postprandial sample 90 minutes after consumption of a standardized mixed breakfast providing approximately 250 kcal. HOMA-B is derived using C-peptide serves as an index of pancreatic beta cell secretory function and is calculated from fasting plasma C-peptide and plasma glucose concentrations using the Homeostasis Model Assessment (HOMA) calculator (University of Oxford, UK), as previously described. T2D was defined by the absence of ketosis, good beta-cell functional reserve as evidenced by stimulated C-peptide (>0.6 pmol/ml), absence of pancreatic calculi (on X-ray abdomen) and good response to oral hypoglycaemic agents for more than two years.

## Variables

*Exposures:* We performed genetic association testing using both whole exome sequencing (exome-wide association study, ExWAS) and imputed genotype data (genome-wide association study, GWAS).

*Outcomes:* We defined one primary outcome: type 2 diabetes state. This was a binary variable used in regression-based association testing.

We defined a larger number of secondary outcomes (20 quantitative and 10 binary traits, **Table S10).** We examined the association of these traits with genetic exposures of interest passing a pre-determined significance threshold in either single variant or gene-based testing. These included binary traits (ie disease codes in the EHR), and quantitative traits (ie measured variables in the EHR, such as HbA1c or weight). Further detail is provided in the "statistical methods" section below.

*Confounders*: We adjusted analyses for confounders typically included in multivariable genetic association models: age at recruitment, age at recruitment squared, sex, and the first 20 genetic principal components.

## Data sources and measurements

*Sample Genotyping:* Genotyping information from Illumina Infinium Global Screening Array v3 (with additional multi-disease variants) was performed as previously described in 51k individuals, and imputed to the TOPMed-r3 panel. We excluded

variants with low INFO scores (below 0.7), minor allele frequency below 1%, and limited analyses to single nucleotide polymorphisms only.

*Sample Whole Exome Sequencing:* Exome sequencing was performed on 55k individuals by the Broad Institute using Twist Reagents and Illumina 150bp PE Novaseq-6000 sequencing, mapped to hg38 to produce crams. Crams with less than 85% of based at >20x coverage were excluded. UMAP was performed on the first 7 genetically determined principal components to define Pakistani and Bangladeshi ancestry individuals, which correlated highly with 1000 genomes population references; 198 outlier samples not corresponding to Pakistani or Bangladeshi ancestry were not included in analyses. Variant QC was performed using a random forest model trained on a set of genotype-corresponding variants identified as true positives.

## Bias

We identify several sources of bias, including ascertainment bias, selection bias, immortal time bias, phenotype misclassification, and indirect parental genetic effects. Because the Genes & Health cohort is enriched for autozygosity – a measure of the relatedness of an individual's immediate ancestors – we may not have fully accounted for population stratification.

## Study size

We included 54,698 individuals (13,674 cases and 41,024 controls, 575 excluded) in the exome-wide association study in Genes & Health, and 45,643 in the genome-wide association study (12,011 cases and 33,623 controls, 5,536 excluded); further details are given in **Figure 1.** Because we fine-mapped variants identified in the whole-exome study using imputed genotype data, we restricted genotype analyses including GWAS to only those individuals who were included in the ExWAS; there is not complete overlap between individuals included in exome sequencing and imputed genotype samples, and therefore there is a high level of exclusion (~10%) from the imputed genotype sample.

## Quantitative variables

We included quantitative trait variables defined by the Genes & Health QUANT-PY pipeline, which has been described in more detail elsewhere. In brief, this identifies all primary and secondary care quantitative treat readings in the EHR; harmonises units; removes duplicates (including serial readings within narrow windows); excludes outlier values above clinically curated reasonable upper and lower bounds; and presents individual and study-level mean, median, min and max values.

## Statistical methods

**Genome-wide association study (GWAS):**

We examined the association of single nucleotide polymorphisms (SNP) with minor allele frequency (MAF) > 1% and imputation quality metric (INFO) score > 0.7 in

45,643 participants using REGENIE v4.0, principally to enable fine-mapping of single variants identified in WES analysis. Association models were adjusted for age at recruitment, age at recruitment squared, sex, and the first 20 genetic principal components. Models were run as REGENIE "binary trait" models with approximate Firth correction below p = 0.05. We defined genome-wide significant loci as loci with a lead SNP p value < 5e-8. A summary Manhattan plot for this data is given in **Figure S7**. We did not identify any novel lead SNPs at a genome-wide significance threshold of p<5e-8, but include the summary statistics in our data deposit.

Because we intended to perform simultaneous fine-mapping in both whole exome sequence and imputed genotype data, we restricted our GWAS analyses to individuals who had both imputed genotype and whole exome sequence data. Because these samples do not overlap, we saw high numbers of participants excluded (n = 5,536, ~10%) from GWAS, which may have limited power to a small extent.

**Exome-wide association study (ExWAS)– single variant:**

We examined the association of 25,326 genetic variants with MAF > 1% in 54,698 participants using REGENIE v4.0. Because of high inbreeding coefficients among variants with MAF ~ 0.50, we cautiously excluded all variants in the window 0.46<MAF<0.54, and restricted analysis to single nucleotide polymorphisms. The methods for this analysis were otherwise identical to the GWAS methods above. We defined exome-wide significance as a Bonferroni-corrected p value of $p < 1.97 \times 10^{-6}$ (ie, 0.05 / 25,326 tests).

**Exome-wide association study (ExWAS) – gene-based testing:**

We used REGENIE v4.0 to examine the association of 534,315 gene-based "masks" previously defined[15] as sets of variants below three MAF thresholds (0.01, 0.001, 0.0001) or singleton, with specific functional annotations (LOFTEE_High_confidence_loss_of_function (LOFTEE-HC-LoF), deleterious missense variants (CADD > 20 & PolyPhen > 0.466 and SIFT<0.05), other missense variants (not meeting the above criteria), and synonymous variants); full definitions are given in **Table S1**. We defined gene testing exome-wide significance as a Bonferroni-corrected p value of $p < 9.35 \times 10^{-8}$. (ie, 0.05 / 534,315 tests).

For *HNF4A*, we additionally constructed bespoke masks defined using the variant categorisation above, but stratified by whether each variant lay within an *HNF4A* repressor or non-repressor domain, to explore whether the effects of identified variants differed by proposed exonic repressor-based location.

**Sex-stratified analysis**

For each analysis, we additionally ran sensitivity sex-stratified models (defined as the presence or absence of a Y chromosome) to explore the possibility of sex-specific effects (**Table S4)**.

**Fine mapping:**

To identify likely causal variants, we performed fine mapping using SuSIE v0.12.35 [39]. We defined linkage disequilibrium (LD) windows of 2Mb (1Mb either side) of the lead SNP, and present posterior inclusion probabilities (PIPs) for variants within these windows. We performed fine mapping in both imputed genotype and whole exome sequencing data, using the same individuals for both analyses. We additionally presented locuszoom plots using the R package locuszoomr.

**Association testing with metabolic traits**

For selected variants of interest which passed significance thresholds and appeared novel, we aimed to better characterise associations with metabolic traits through an association-testing pipeline in Genes & Health. This aimed to explore the per-minor-allele association of each variant with relevant binary and quantitative traits. Traits selected were based on availability and relevance: glycaemic, anthropometric, metabolically related, and lipid traits at the time of type 2 diabetes diagnosis, and clinically relevant insulin resistance phenotypes and diabetic complications (**Table S10)**. For variants within a gene for which data was available, we explored association of the variant with protein expression levels from two platforms (Olink and SomaLogic), and meta-analysed results between these platforms. Further detail on proteomic quality control is described in[35].

We explored the association of variants of interest with binary and quantitative traits in the Genes & Health study defined using robust, open-source pipelines using multivariable regression models adjusted for age at study recruitment, sex, ancestry, and the first 20 principal components. These include binary trait phenotypes defined as ICD10 codes, including codes mapped from SNOMED (used in Genes & Health primary care data) to ICD10 (which are used for Genes & Health secondary care data), as described previously. For quantitative traits, we explored the association of variants of interest with lifetime median values.

We explored the association of variants of interest with traits related to insulin resistance at the time of diagnosis in 13,468 type 2 diabetes cases. We defined traits at the time of diagnosis as a reading within 1 year (either before or after) the time of earliest date of type 2 diabetes diagnosis in the EHR, as described previously. Where multiple measures were available within this window, we selected the value closest in time to diagnosis data . Associations were reported from regression models adjusted for age at study recruitment, sex, ancestry, and the first 20 principal components.

We additionally examined the association of one variant of interest in the *HNF4A* gene with survival from birth to coronary artery disease, having observed a non-significant but biologically plausible trend towards an association in the multivariable regression models defined above **(**case counts given in **Table 1)**. We performed this analysis using a Cox proportional hazards model adjusted for age at study recruitment, sex, ancestry, and the first 20 principal components. Individuals were

censored at the earliest occurrence of: death, de-enrolment from the study, date of last data extraction, or time of coronary artery disease coding.

### Meta-analysis

We meta-analysed single variant summary statistics for type 2 diabetes case-control status and phenotypes of interest identified at nominal significance ($p<0.05$) in Genes & Health trait association testing with publicly available summary statistics aggregated via the T2DKP.

For gene-based results, we meta-analysed with gene-burden summary statistics from UK Biobank available through the AstraZeneca PheWAS portal. This resource provides gene-burden summary statistics across multiple masks, stratified by ancestry; we selected the most closely aligned mask to the Genes & Health discovery mask for each gene (eg, protein-truncating variants in UKB for high-confidence predicted loss of function variants in GnH). Because not all REGENIE gene burden masks produce a direction of effect with associated beta and standard error (eg GENE-P), for phenotypes with nominal significance ($p < 0.05$) in internal association testing but no associated beta in the most significant mask*test combination, we selected the most significant gene-burden mask*test combination within Genes & Health with a direction of effect, measured as the smallest p value. Meta-analysis was performed using fixed-effects meta-analysis with the R package metafor v4.6-0.

Where possible, we meta-analysed with traits from MDRF, including meta-analysis between MDRF and the T2DKP for traits not included in the “discovery” sample Genes & Health, and particularly for traits related to HOMA estimates of insulin resistance and production.

### Population genetics analyses

To test how likely it is to observe differenced in allele frequency among populations in target variants, we simulated the neutral evolution of 10,000 independent SNPs under a previously proposed joint demographic model for Europeans, South Asians and East Asians with 1,000 sampled individuals each using the *ms* software. We then retained simulations that yielded European allele frequencies lower than 0.2% and calculated the empirical rank p-value as the proportions of simulations generating frequencies in South and East Asians greater than the ones observed.

## Functional characterization of HNF4A variant

### qRT-PCR analysis of transcriptional activities across HNF4A variants

pTwist-CMV-puro expression vectors were generated by Twist Bioscience (San Francisco, CA, USA) for the following *HNF4A* variants: wild-type (WT, Uniprot isoform P41235-1, NM_000457.6), Pro437Ser (NM_000457.6:c.1309C>T), and

Arg89Trp (NM_000457.6:c.265C>T), a confirmed loss of function variant[24]. Each construct included a C-terminal Myc tag. The *HNF4A* constructs were transfected into WT HUH7 hepatocytes using TransIT-LT1 transfection reagent (#MIR2304, Mirus Bio, Madison, WI, USA) following the manufacturer's protocol. Variant overexpression was confirmed 24 hours after transfection using Western blotting with primary antibodies against c-Myc (#3-2500, Invitrogen, Thermo Fisher Scientific, Waltham, MA, USA) or HNF4A (#3113SCell Signaling Technology, Danvers, MA, USA) in three independent replicates. Expression of the HNF4A transcriptional target *G6PC* was assayed in HUH7 cells transiently transfected with an empty control vector (pLenti SpBsmBI sgRNA Puro (#62207, Addgene, Watertown, MA, USA)) or HNF4A WT, Pro437Ser, or Arg89Trp Myc-tagged constructs after 24 hours by qPCR as we have published previously. The following primers were used for *G6PC:* FWD 5'- GCTGTGATTGGAGACTGGCTCA-3', REV REV 5'- CAGGTTGGGGGTCAGTACC-3'; and the housekeeping control gene *GAPDH*: FWD 5'- CATCTTCTTTTGCGTCGCCA-3', REV 5'- TTAAAAGCAGCCCTGGTGACC-3'.

**Dual-luciferase analysis of transactivation activities across HNF4A variants**

HEK293 cells were co-transfected with either a mouse G6PC promoter–driven firefly luciferase reporter (10 ng/well)[24]or a PPRE×3-TK-Luc reporter (10 ng/well), together with a Renilla luciferase plasmid (10 ng/well) (pRL-SV40 for mG6PC or pGL4.75 for PPRE×3) as a transfection efficiency control. A 2-fold serial dilution of *HNF4A* variant expression plasmids (0–32 ng per well in a 96-well format) was used to assess dose-dependent transactivation. To maintain a constant total DNA amount across all conditions, a neutral pLenti-SpBsmBI-Puro backbone plasmid was added as filler DNA, and the final DNA mass and TransIT-LT1 reagent volume were kept constant for all conditions. Cells were harvested 48 h post-transfection, and firefly and Renilla luciferase activities were measured using the Dual-Luciferase Reporter Assay System (#E1910, Promega, Madison, WI, US) on a Tecan Spark multimode plate reader (Tecan Group Ltd., Männedorf, Switzerland). Firefly luciferase readings were normalized to Renilla to control for transfection efficiency. All experimental conditions were performed in three biological replicates.

# Acknowledgements

SH is funded by Wellcome Health Advances in Underrepresented Populations (HARP) Doctoral Fellowship 227532/Z/23/Z. VB is funded by the Wellcome Trust PhD programme - health data in practice: human-centred science (218584/Z/19/Z). ARM, AB and S Hu are supported by I01BX006293 from VA Office of Research and Development. DS and SF are funded by the Tackling Multimorbidity at Scale Strategic Priorities Fund programme (MR/W014416/1) delivered by the Medical Research Council and the National Institute for Health Research in partnership with

the Economic and Social Research Council and in collaboration with the Engineering and Physical Sciences Research Council. IB and SF are funded by a Wellcome Discovery Award (227897/Z/23/Z) and IB acknowledges support from the National Institute for Health and Care Research Exeter Biomedical Research Centre. The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care. GK & HCM's involvement in this research was funded in part by Wellcome Trust (grant no. 220540/Z/20/A, "Wellcome Sanger Institute Quinquennial Review 2021–2026"). MKS is supported by Barts Charity grants MGU0504 and G-002995

We acknowledgement the support of Indian Council of Medical Research (ICMR) for setting up the MDRF – Biobank.

Analyses in UKBiobank by AYD were performed under study accession number 53639.

Genes & Health is/has recently been core-funded by Wellcome (WT102627, WT210561), the Medical Research Council (UK) (M009017, MR/X009777/1, MR/X009920/1), Higher Education Funding Council for England Catalyst, Barts Charity (845/1796), Health Data Research UK (for London substantive site), and research delivery support from the NHS National Institute for Health Research Clinical Research Network (North Thames). We acknowledge the support of the National Institute for Health and Care Research Barts Biomedical Research Centre (NIHR203330); a delivery partnership of Barts Health NHS Trust, Queen Mary University of London, St George's University Hospitals NHS Foundation Trust and St George's University of London

Genes & Health is/has recently been funded by Alnylam Pharmaceuticals, Genomics PLC; and a Life Sciences Industry Consortium of AstraZeneca PLC, Bristol-Myers Squibb Company, GlaxoSmithKline Research and Development Limited, Maze Therapeutics Inc, Merck Sharp & Dohme LLC, Novo Nordisk A/S, Pfizer Inc, Takeda Development Centre Americas Inc.

We thank Social Action for Health, Centre of The Cell, members of our Community Advisory Group, and staff who have recruited and collected data from volunteers. We thank the NIHR National Biosample Centre (UK Biocentre), the Social Genetic & Developmental Psychiatry Centre (King's College London), Wellcome Sanger Institute, and Broad Institute for sample processing, genotyping, sequencing and variant annotation. This work uses data provided by patients and collected by the NHS as part of their care and support. This research utilised Queen Mary University of London's Apocrita HPC facility, supported by QMUL Research-IT, http://doi.org/10.5281/zenodo.438045

We thank: Barts Health NHS Trust, NHS Clinical Commissioning Groups (City and Hackney, Waltham Forest, Tower Hamlets, Newham, Redbridge, Havering, Barking and Dagenham), East London NHS Foundation Trust, Bradford Teaching Hospitals

NHS Foundation Trust, Public Health England (especially David Wyllie), Discovery Data Service/Endeavour Health Charitable Trust (especially David Stables), Voror Health Technologies Ltd (especially Sophie Don), NHS England (for what was NHS Digital) - for GDPR-compliant data sharing backed by individual written informed consent.

Most of all we thank all of the volunteers participating in Genes & Health.

A favourable ethical opinion for the main Genes & Health research study was granted by NRES Committee London - South East (reference 14/LO/1240) on 16 Sept 2014. Queen Mary University of London is the Sponsor, and Data Controller.

**Data availability:**

Genes & Health: Individual-level participant data are available to researchers and industry partners worldwide via application to and review by the Genes & Health Executive (https://www.genesandhealth.org/); applications are reviewed monthly. Approved researchers have access to individual-level data in the Genes & Health Trusted Research Environment (TRE) and can request the data files used in this study from the corresponding author(s). All data exports from the Genes & Health TRE are reviewed to prevent release of identifiable individual-level data. Summary data may be exported for cross-cohort meta-analysis or replication and for publication, subject to review

**Competing interests:**

CM, AD and YJ are current employees of Novo Nordisk Foundation.

**Author contribution**

SH, VB, SH, ARM, MKS: conceptualisation, formal analyses, visualisation, writing original draft and manuscript review. MB: replication data analysis, visualisation, manuscript review. RMA, VM, GP, RV, SF, DvH, CL, AB: supervision, data generation, project administration, resources, funding acquisition, manuscript review. CM, YJ, AYD: methodology, analysis, visualisation, and reviewing manuscript. DS, JS, VB: data curation, methodology, manuscript revision. All authors reviewed, edited and approved the final version of the manuscript.

# References

1. Suzuki, K. *et al.* Genetic drivers of heterogeneity in type 2 diabetes pathophysiology. *Nature* (2024) doi:10.1038/s41586-024-07019-6.

2. Mahajan, A. *et al.* Multi-ancestry genetic study of type 2 diabetes highlights the power of diverse populations for discovery and translation. *Nat. Genet.* **54**, 560–572 (2022).

3. Loh, M. *et al.* Identification of genetic effects underlying type 2 diabetes in South Asian and European populations. *Commun Biol* **5**, 329 (2022).

4. Srinivasan, S. *et al.* Common and Distinct Genetic Architecture of Age at Diagnosis of Diabetes in South Indian and European Populations. *Diabetes Care* **46**, 1515–1523 (2023).

5. Gujral, U. P., Pradeepa, R., Weber, M. B., Narayan, K. M. V. & Mohan, V. Type 2 diabetes in South Asians: similarities and differences with white Caucasian and other populations. *Ann. N. Y. Acad. Sci.* **1281**, 51–63 (2013).

6. Unnikrishnan, R., Gupta, P. K. & Mohan, V. Diabetes in south Asians: Phenotype, clinical presentation, and natural history. *Curr. Diab. Rep.* **18**, 30 (2018).

7. Narayan, K. M. V. & Kanaya, A. M. Why are South Asians prone to type 2 diabetes? A hypothesis based on underexplored pathways. *Diabetologia* **63**, 1103–1109 (2020).

8. Anjana, R. M. *et al.* Prevalence of diabetes and prediabetes in 15 states of India: results from the ICMR–INDIAB population-based cross-sectional study. *Lancet Diabetes Endocrinol.* **5**, 585–596 (2017).

9. Deepa, M. *et al.* Evaluation of Madras Diabetes Research Foundation-Indian Diabetes Risk Score in detecting undiagnosed diabetes in the Indian population: Results from the Indian Council of Medical Research-INdia DIABetes population-based study (INDIAB-15). *Indian J. Med. Res.* **157**, 239–249 (2023).

10. Hodgson, S. *et al.* Genetic basis of early onset and progression of type 2 diabetes in South Asians. *Nat. Med.* **31**, 323–331 (2025).

11. Smith, K. *et al.* Multi-ancestry polygenic mechanisms of type 2 diabetes. *Nat. Med.* (2024) doi:10.1038/s41591-024-02865-3.

12. Siddiqui, M. K. *et al.* Young-onset diabetes in Asian Indians is associated with lower measured and genetically determined beta cell function. *Diabetologia* **65**, 973–983 (2022).

13. Williams, R. & Diabetes Atlas Committee, I. D. F. … and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: Results from the International Diabetes Federation Diabetes Atlas. *Diabetes Res.* (2019).

14. Flannick, J. *et al.* Exome sequencing of 20,791 cases of type 2 diabetes and 24,440 controls. *Nature* **570**, 71–76 (2019).

15. Kim, H. I. *et al.* Exome sequencing and analysis of 44,028 British South Asians enriched for high autozygosity. *medRxiv* 2025.06.05.25329068 (2025) doi:10.1101/2025.06.05.25329068.

16. Finer, S. *et al.* Cohort Profile: East London Genes & Health (ELGH), a community-based population genomics and health study in British Bangladeshi and British Pakistani people. *Int. J. Epidemiol.* **49**, 20–21i (2020).

17. Anjana, R. M. *et al.* ICMR-MDRF diabetes biosamples: Cohort profile. *Indian J. Med. Res.* **160**, 514–526 (2024).

18. Warncke, K. *et al.* Frequency and characteristics of MODY 1 (HNF4A mutation) and MODY 5 (HNF1B mutation): Analysis from the DPV database. *J. Clin. Endocrinol. Metab.* **104**, 845–855 (2019).

19. Kooner, J. S. *et al.* Genome-wide association study in individuals of South Asian ancestry identifies six new type 2 diabetes susceptibility loci. *Nat. Genet.* **43**, 984–989 (2011).

20. Nag, A. *et al.* Human genetics uncovers MAP3K15 as an obesity-independent therapeutic target for diabetes. *Sci. Adv.* **8**, eadd5430 (2022).

21. Costanzo, M. C. *et al.* The Type 2 Diabetes Knowledge Portal: An open access genetic resource dedicated to type 2 diabetes and related traits. *Cell Metab.* **35**, 695-710.e6 (2023).

22. Kind, L. *et al.* Molecular mechanism of HNF-1A-mediated HNF4A gene regulation and promoter-driven HNF4A-MODY diabetes. *JCI Insight* **9**, (2024).

23. Lambert, É. *et al.* Human hepatocyte nuclear factor 4-α encodes isoforms with distinct transcriptional functions. *Mol. Cell. Proteomics* **19**, 808–827 (2020).

24. Kaci, A. *et al.* Functional characterization of HNF4A gene variants identify promoter and cell line specific transactivation effects. *Hum. Mol. Genet.* **33**, 894–904 (2024).

25. Yoon, J. C. *et al.* Control of hepatic gluconeogenesis through the transcriptional coactivator PGC-1. *Nature* **413**, 131–138 (2001).

26. Kim, Y. J. *et al.* The contribution of common and rare genetic variants to variation in metabolic traits in 288,137 East Asians. *Nat. Commun.* **13**, 6642 (2022).

27. Moltke, I. *et al.* A common Greenlandic TBC1D4 variant confers muscle insulin resistance and type 2 diabetes. *Nature* **512**, 190–193 (2014).

28. Spracklen, C. N. *et al.* Identification of type 2 diabetes loci in 433,540 East Asian individuals. *Nature* **582**, 240–245 (2020).

29. Chen, J. *et al.* The trans-ancestral genomic architecture of glycemic traits. *Nat. Genet.* **53**, 840–860 (2021).

30. Sakaue, S. *et al.* A cross-population atlas of genetic associations for 220 human phenotypes. *Nat. Genet.* **53**, 1415–1424 (2021).

31. GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* **45**, 580–585 (2013).

32. Alonso, L. *et al.* TIGER: The gene expression regulatory variation landscape of human pancreatic islets. *Cell Rep.* **37**, 109807 (2021).

33. Vu, H. A. T. H. *et al.* 2121-LB: PanKbase integrated single-cell map—A comprehensive atlas of human pancreatic islets unlocking insights into Type 1 and Type 2 diabetes. *Diabetes* **74**, 2121-LB (2025).

34. Malmström, E. *et al.* Human proteome distribution atlas for tissue-specific plasma proteome dynamics. *Cell* **188**, 2810-2822.e16 (2025).

35. Carrasco Zanini, J. *et al.* Proteomic and clinical impact of human knockouts in British South Asians. *medRxiv* 2025.10.15.25337977 (2025) doi:10.1101/2025.10.15.25337977.

36. Chidambaram, M., Radha, V. & Mohan, V. Replication of recently described type 2 diabetes gene variants in a South Indian population. *Metabolism* **59**, 1760–1766 (2010).

37. Steinthorsdottir, V. *et al.* A variant in CDKAL1 influences insulin response and risk of type 2 diabetes. *Nat. Genet.* **39**, 770–775 (2007).

38. Palmer, C. J. *et al.* Cdkal1, a type 2 diabetes susceptibility gene, regulates mitochondrial function in adipose tissue. *Mol. Metab.* **6**, 1212–1225 (2017).

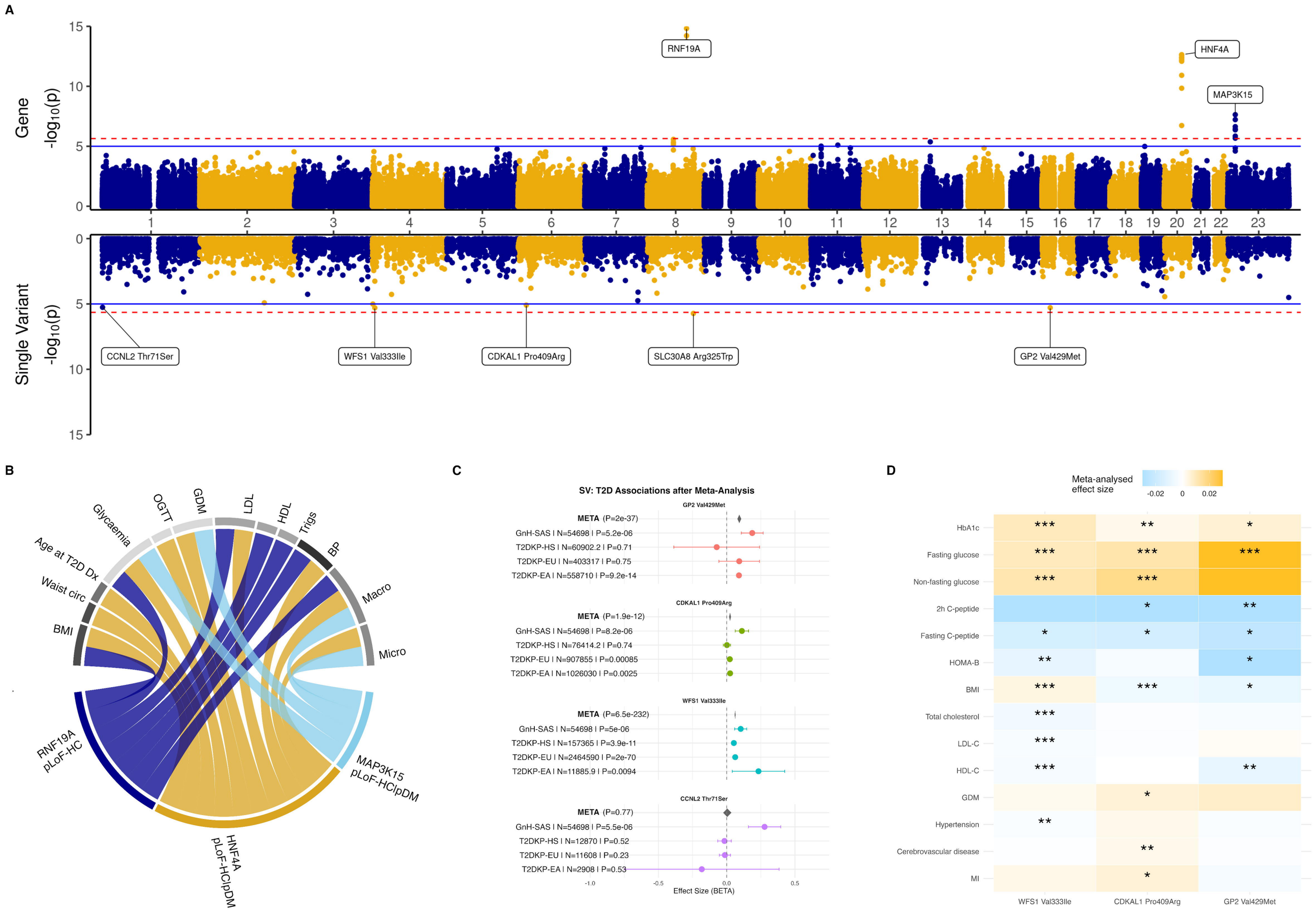
39. Zou, Y., Carbonetto, P., Wang, G. & Stephens, M. Fine-mapping from summary data with the “Sum of Single Effects” model. *PLoS Genet.* **18**, e1010299 (2022).

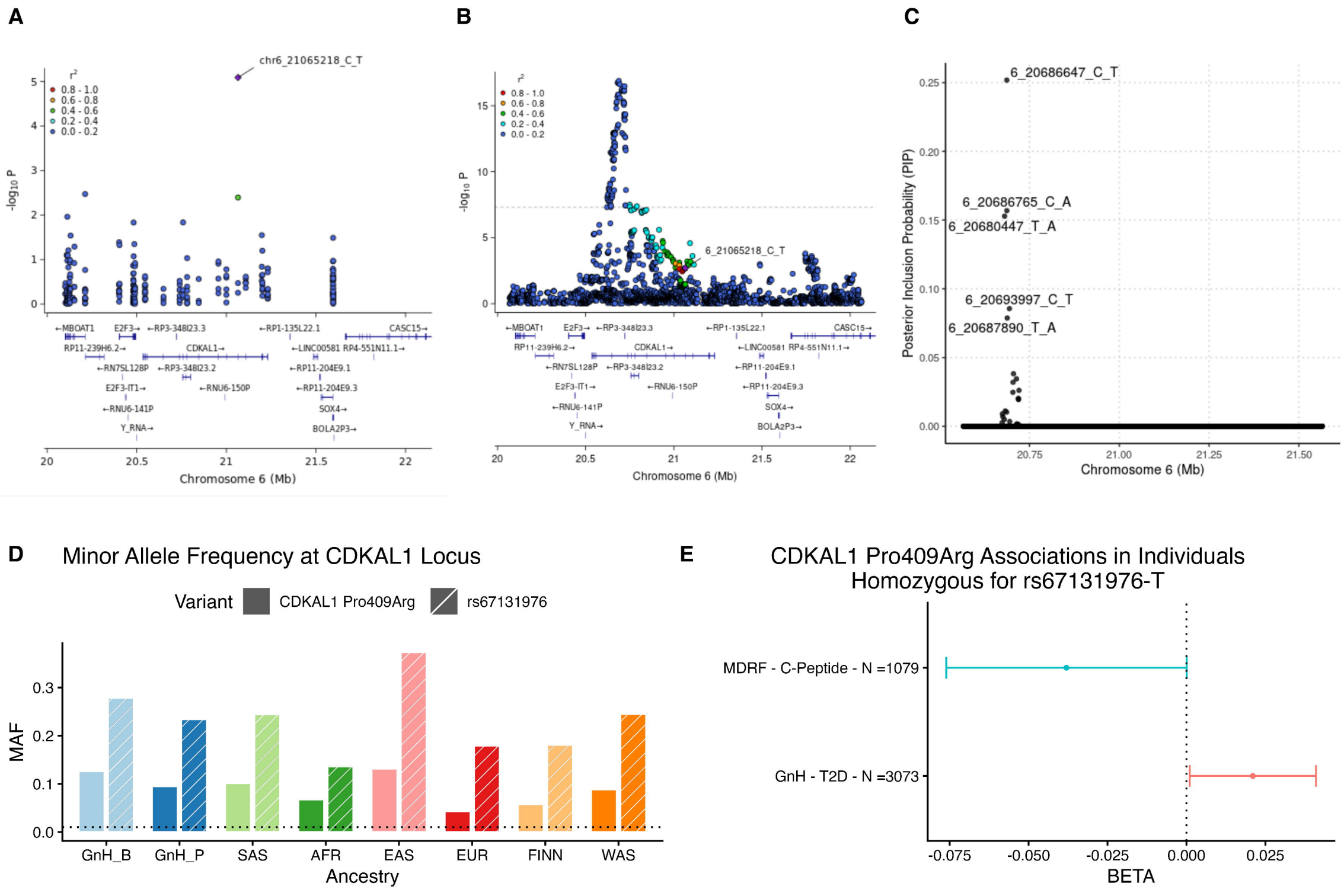
40. Aghazadeh, Y. *et al.* GP2-enriched pancreatic progenitors give rise to functional beta cells in vivo and eliminate the risk of teratoma formation. *Stem Cell Reports* **17**, 964–978 (2022).

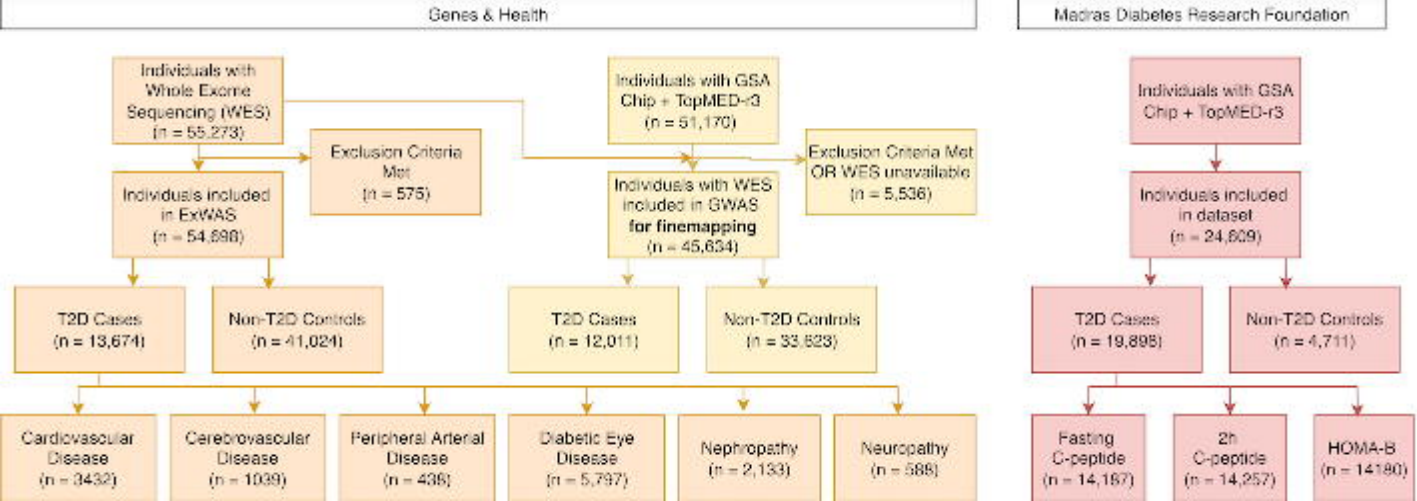
41. Holman, N. *et al.* Completion of annual diabetes care processes and mortality: A cohort study using the National Diabetes Audit for England and Wales. *Diabetes Obes. Metab.* **23**, 2728–2740 (2021).

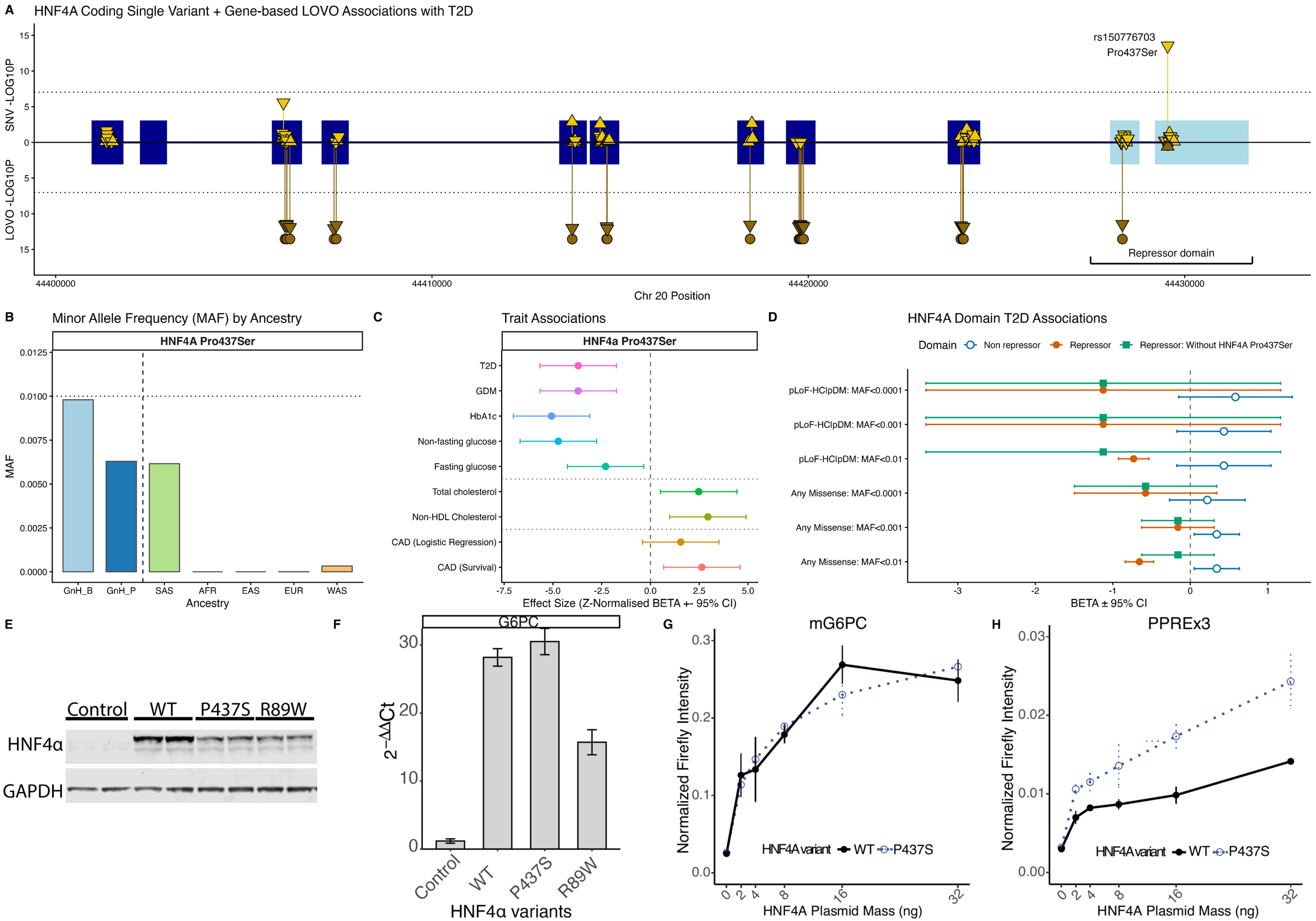
42. Wang, Q. *et al.* Rare variant contribution to human disease in 281,104 UK Biobank exomes. *Nature* **597**, 527–532 (2021).

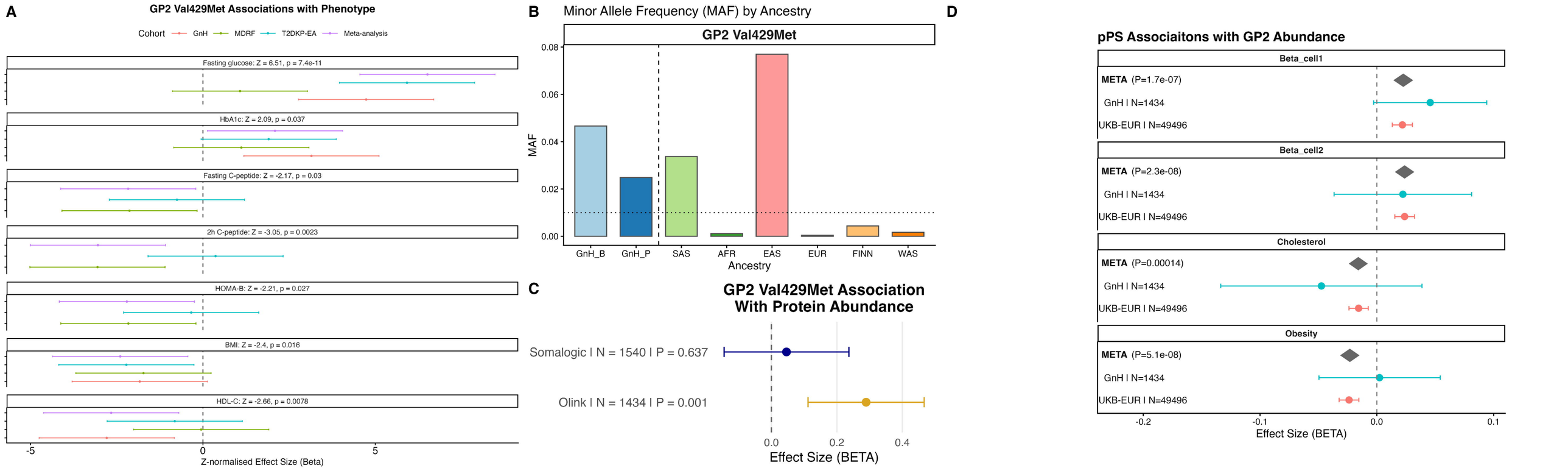
43. Little, J. *et al.* STrengthening the REporting of Genetic Association Studies (STREGA)--an extension of the STROBE statement. *Genet. Epidemiol.* **33**, 581–598 (2009).


A
Gene
-log10(p)
RNF19A
HNF4A
MAP3K15
Single Variant
CCNL2 Thr71Ser
WFS1 Val333Ile
CDKAL1 Pro409Arg
SLC30A8 Arg325Trp
GP2 Val429Met
B
Glycaemia
OGTT
GDM
LDL
HDL
Trigs
BP
Macro
Micro
Age at T2D Dx
Waist circ
BMI
RNF19A pLoF-HC
HNF4A pLoF-HC|pDM
MAP3K15 pLoF-HC|pDM
C
SV: T2D Associations after Meta-Analysis
GP2 Val429Met
META (P=2e-37)
GnH-SAS | N=54698 | P=5.2e-06
T2DKP-HS | N=60902.2 | P=0.71
T2DKP-EU | N=403317 | P=0.75
T2DKP-EA | N=558710 | P=9.2e-14
CDKAL1 Pro409Arg
META (P=1.9e-12)
GnH-SAS | N=54698 | P=8.2e-06
T2DKP-HS | N=76414.2 | P=0.74
T2DKP-EU | N=907855 | P=0.00085
T2DKP-EA | N=1026030 | P=0.0025
WFS1 Val333Ile
META (P=6.5e-232)
GnH-SAS | N=54698 | P=5e-06
T2DKP-HS | N=157365 | P=3.9e-11
T2DKP-EU | N=2464590 | P=2e-70
T2DKP-EA | N=11885.9 | P=0.0094
CCNL2 Thr71Ser
META (P=0.77)
GnH-SAS | N=54698 | P=5.5e-06
T2DKP-HS | N=12870 | P=0.52
T2DKP-EU | N=11608 | P=0.23
T2DKP-EA | N=2908 | P=0.53
Effect Size (BETA)
D
Meta-analysed effect size
-0.02 0 0.02
HbA1c
Fasting glucose
Non-fasting glucose
2h C-peptide
Fasting C-peptide
HOMA-B
BMI
Total cholesterol
LDL-C
HDL-C
GDM
Hypertension
Cerebrovascular disease
MI
WFS1 Val333Ile
CDKAL1 Pro409Arg
GP2 Val429Met

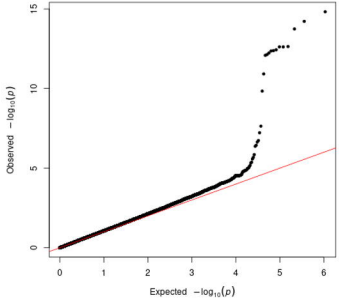

A
chr6_21065218_C_T
r²
0.8 - 1.0
0.6 - 0.8
0.4 - 0.6
0.2 - 0.4
0.0 - 0.2
-log10 P
Chromosome 6 (Mb)
B
6_21065218_C_T
C
6_20686647_C_T
6_20686765_C_A
6_20680447_T_A
6_20693997_C_T
6_20687890_T_A
Posterior Inclusion Probability (PIP)
Chromosome 6 (Mb)
D
Minor Allele Frequency at CDKAL1 Locus
Variant
CDKAL1 Pro409Arg
rs67131976
MAF
GnH_B
GnH_P
SAS
AFR
EAS
EUR
FINN
WAS
Ancestry
E
CDKAL1 Pro409Arg Associations in Individuals Homozygous for rs67131976-T
MDRF - C-Peptide - N =1079
GnH - T2D - N =3073
BETA


Genes & Health
Madras Diabetes Research Foundation
Individuals with Whole Exome Sequencing (WES) (n = 55,273)
Exclusion Criteria Met (n = 575)
Individuals included in ExWAS (n = 54,698)
T2D Cases (n = 13,674)
Non-T2D Controls (n = 41,024)
Cardiovascular Disease (n = 3432)
Cerebrovascular Disease (n = 1039)
Peripheral Arterial Disease (n = 438)
Individuals with GSA Chip + TopMED-r3 (n = 51,170)
Exclusion Criteria Met OR WES unavailable (n = 5,536)
Individuals with WES included in GWAS for finemapping (n = 45,634)
T2D Cases (n = 12,011)
Non-T2D Controls (n = 33,623)
Diabetic Eye Disease (n = 5,797)
Nephropathy (n = 2,133)
Neuropathy (n = 588)
Individuals with GSA Chip + TopMED-r3
Individuals included in dataset (n = 24,609)
T2D Cases (n = 19,898)
Non-T2D Controls (n = 4,711)
Fasting C-peptide (n = 14,167)
2h C-peptide (n = 14,257)
HOMA-B (n = 14180)

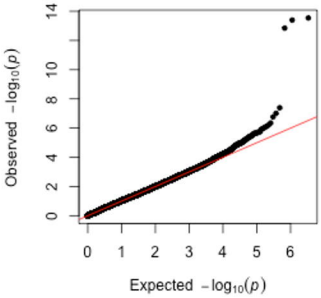

A
HNF4A Coding Single Variant + Gene-based LOVO Associations with T2D
SNV -LOG10P
LOVO -LOG10P
rs150776703
Pro437Ser
Repressor domain
44400000
44410000
44420000
44430000
Chr 20 Position
B
Minor Allele Frequency (MAF) by Ancestry
HNF4A Pro437Ser
MAF
GnH_B
GnH_P
SAS
AFR
EAS
EUR
WAS
Ancestry
C
Trait Associations
HNF4a Pro437Ser
T2D
GDM
HbA1c
Non-fasting glucose
Fasting glucose
Total cholesterol
Non-HDL Cholesterol
CAD (Logistic Regression)
CAD (Survival)
Effect Size (Z-Normalised BETA +- 95% CI)
D
HNF4A Domain T2D Associations
Domain
Non repressor
Repressor
Repressor: Without HNF4A Pro437Ser
pLoF-HCIpDM: MAF<0.0001
pLoF-HCIpDM: MAF<0.001
pLoF-HCIpDM: MAF<0.01
Any Missense: MAF<0.0001
Any Missense: MAF<0.001
Any Missense: MAF<0.01
BETA ± 95% CI
E
Control
WT
P437S
R89W
HNF4α
GAPDH
F
G6PC
2−ΔΔCt
Control
WT
P437S
R89W
HNF4α variants
G
mG6PC
Normalized Firefly Intensity
HNF4A variant
WT
P437S
HNF4A Plasmid Mass (ng)
H
PPREx3
Normalized Firefly Intensity
HNF4A variant
WT
P437S
HNF4A Plasmid Mass (ng)


A
GP2 Val429Met Associations with Phenotype
Cohort
GnH
MDRF
T2DKP-EA
Meta-analysis
Fasting glucose: Z = 6.51, p = 7.4e-11
HbA1c: Z = 2.09, p = 0.037
Fasting C-peptide: Z = -2.17, p = 0.03
2h C-peptide: Z = -3.05, p = 0.0023
HOMA-B: Z = -2.21, p = 0.027
BMI: Z = -2.4, p = 0.016
HDL-C: Z = -2.66, p = 0.0078
-5
0
5
Z-normalised Effect Size (Beta)
B
Minor Allele Frequency (MAF) by Ancestry
GP2 Val429Met
MAF
0.08
0.06
0.04
0.02
0.00
GnH_B
GnH_P
SAS
AFR
EAS
EUR
FINN
WAS
Ancestry
C
GP2 Val429Met Association
With Protein Abundance
Somalogic I N = 1540 I P = 0.637
Olink I N = 1434 I P = 0.001
0.0
0.2
0.4
Effect Size (BETA)
D
pPS Associaitons with GP2 Abundance
Beta_cell1
META (P=1.7e-07)
GnH I N=1434
UKB-EUR I N=49496
Beta_cell2
META (P=2.3e-08)
GnH I N=1434
UKB-EUR I N=49496
Cholesterol
META (P=0.00014)
GnH I N=1434
UKB-EUR I N=49496
Obesity
META (P=5.1e-08)
GnH I N=1434
UKB-EUR I N=49496
-0.2
-0.1
0.0
0.1
Effect Size (BETA)

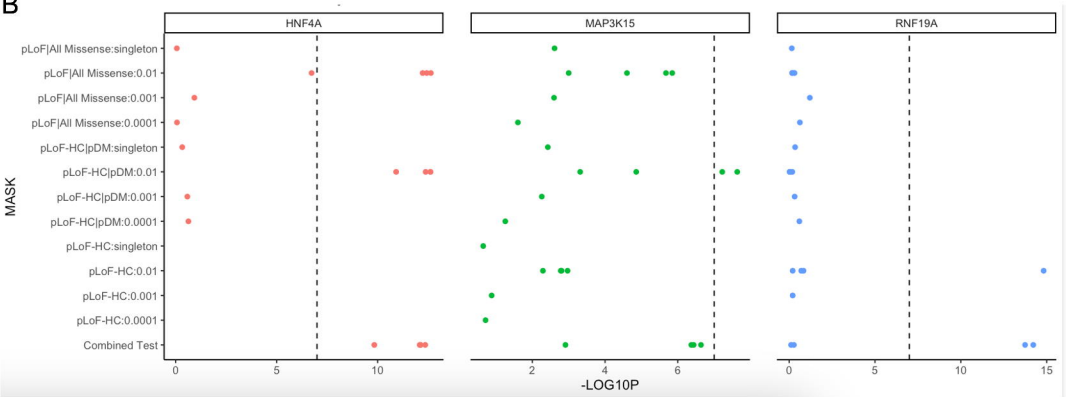
A

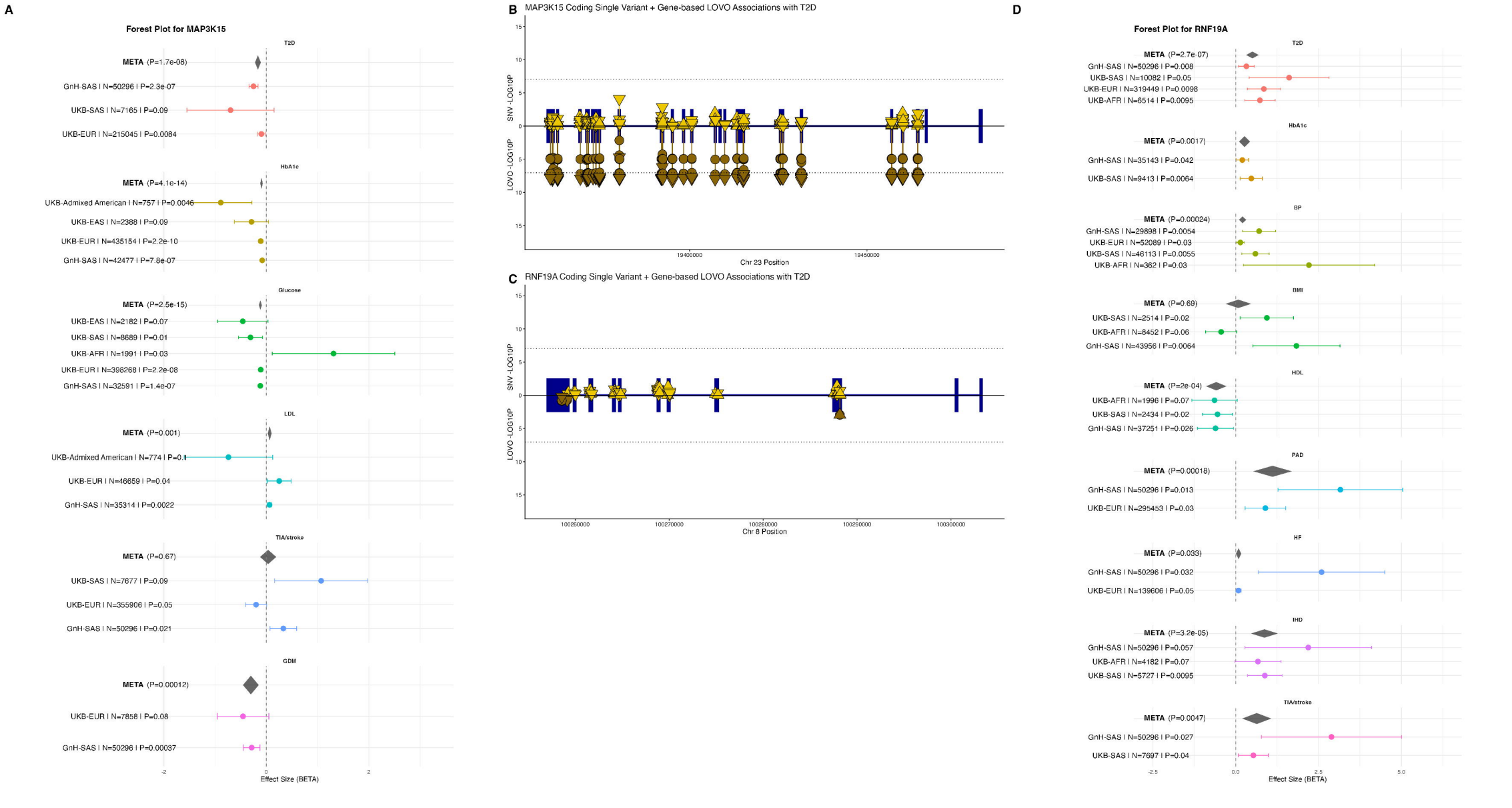

Gene-Based QQ: LambdaGC = 1.11
Observed $-\log_{10}(p)$
Expected $-\log_{10}(p)$
0
5
10
15
0
1
2
3
4
5
6

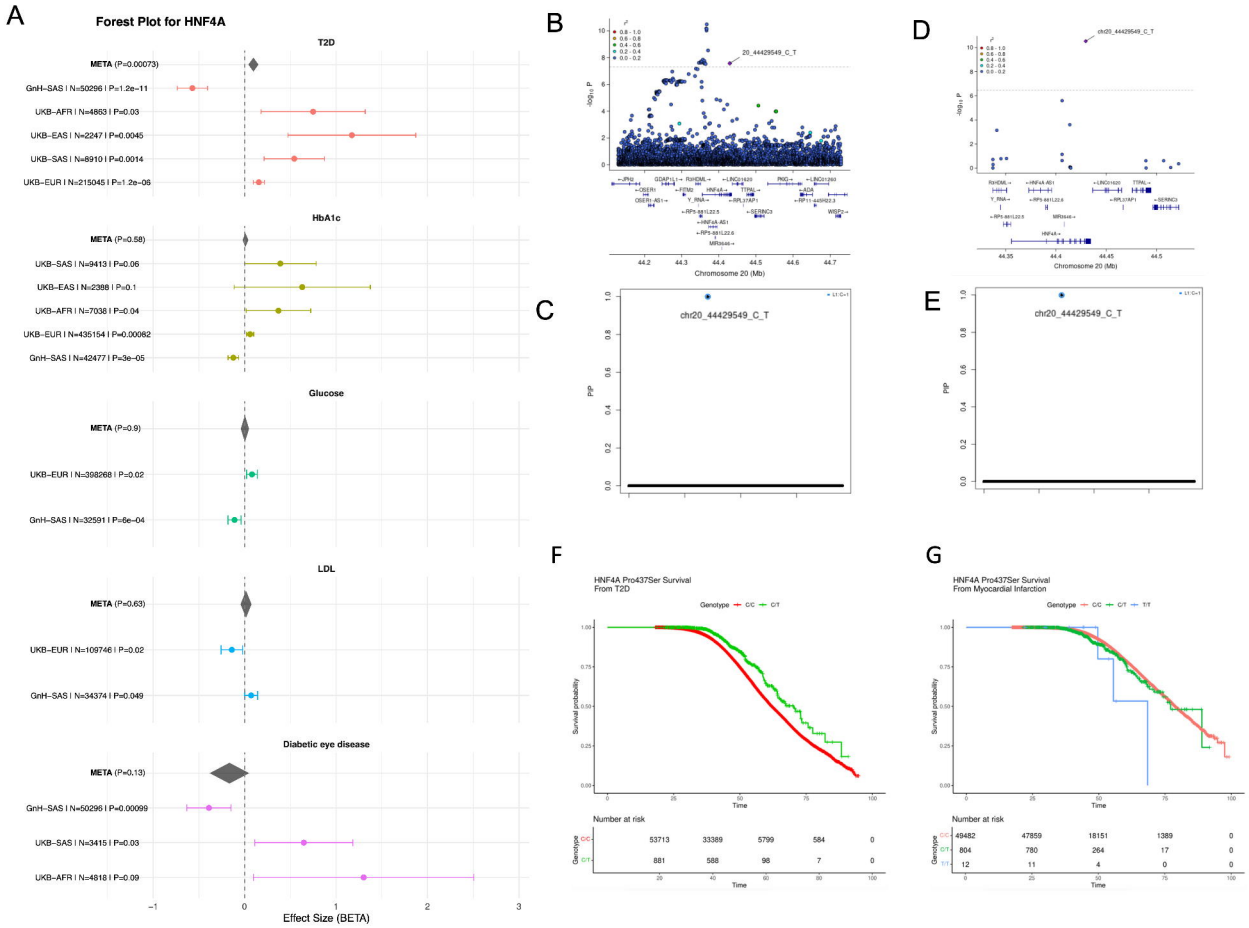
C

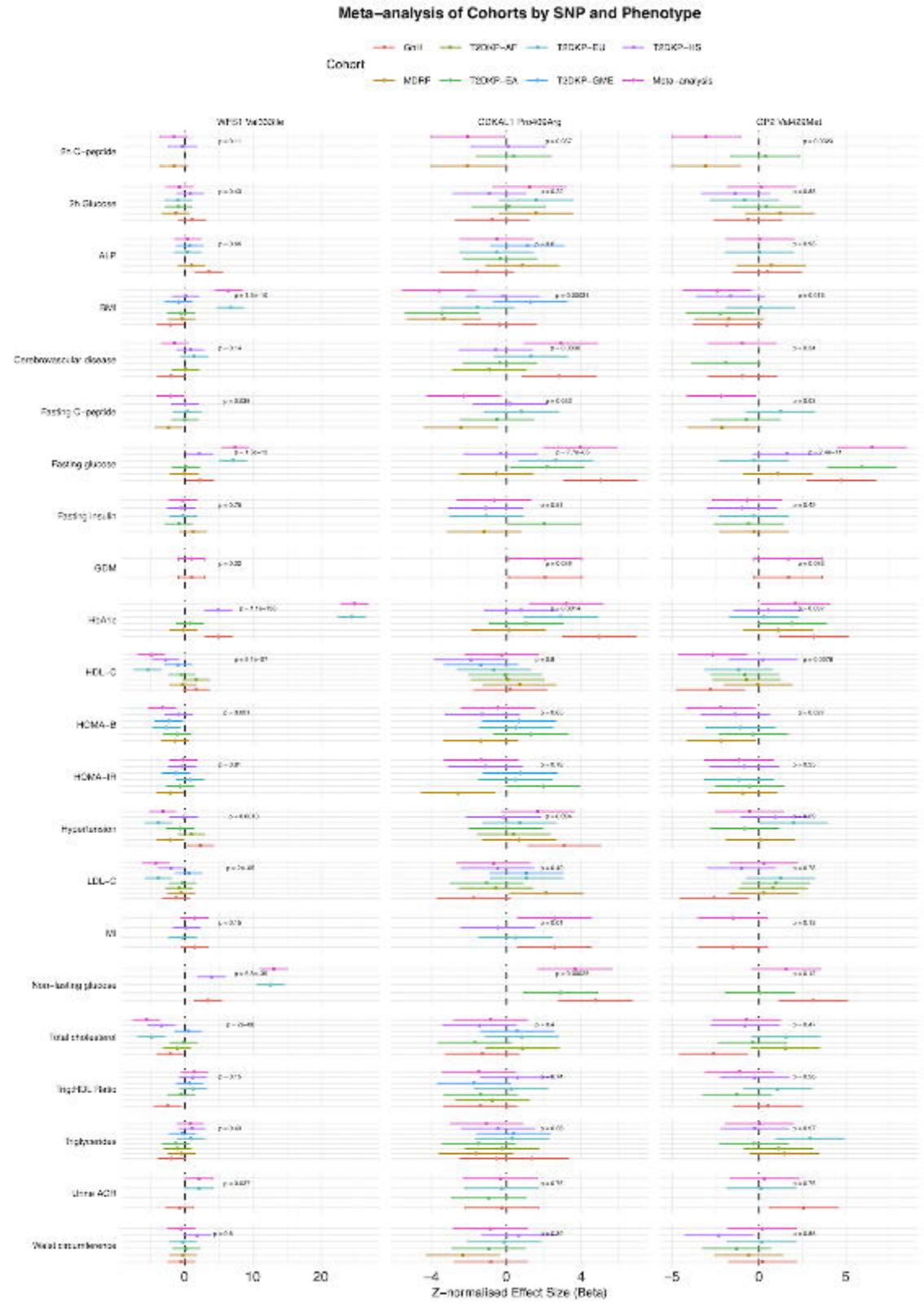

Single Variant QQ: LambdaGC = 1.10
Observed $-\log_{10}(p)$
Expected $-\log_{10}(p)$
0
2
4
6
8
10
14
0
1
2
3
4
5
6

B

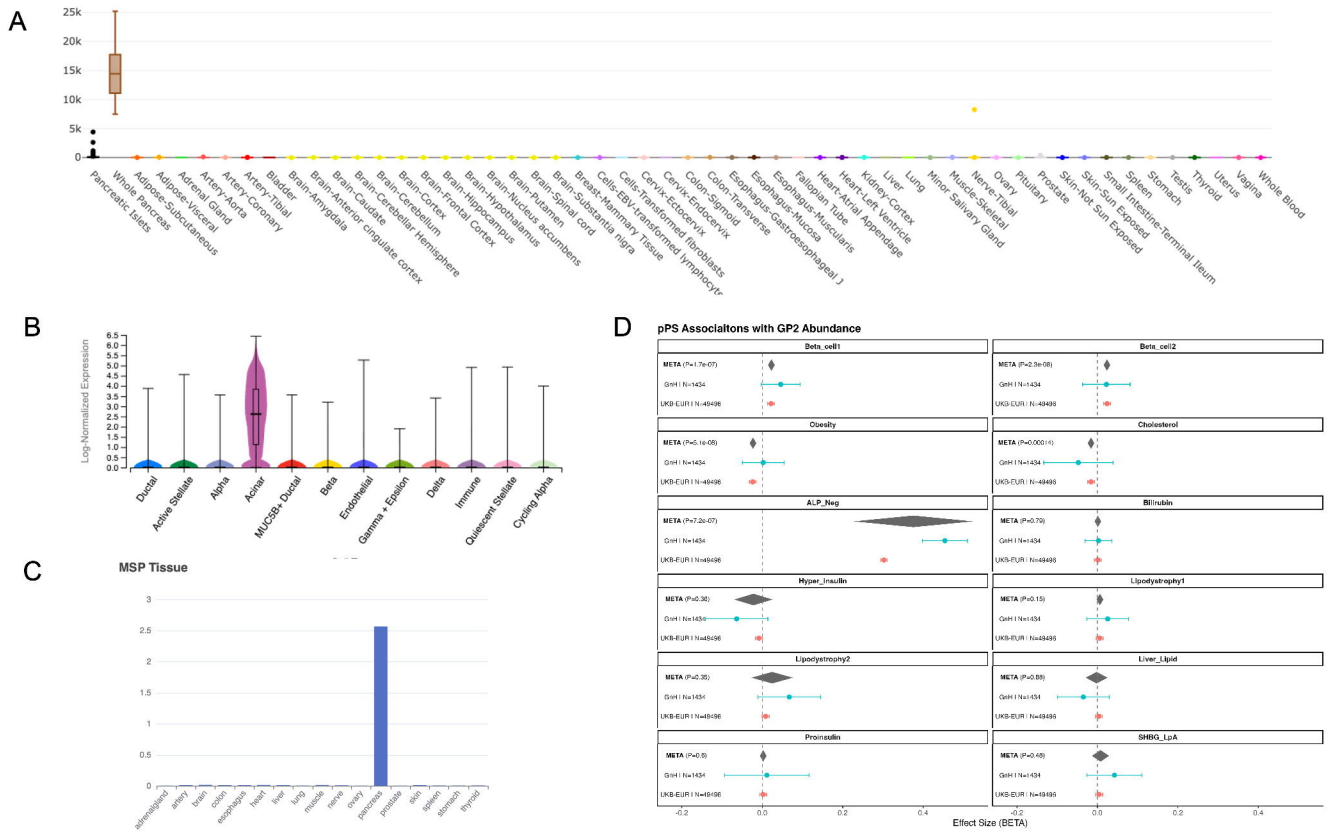

HNF4A
MAP3K15
RNF19A
MASK
pLoF|All Missense:singleton
pLoF|All Missense:0.01
pLoF|All Missense:0.001
pLoF|All Missense:0.0001
pLoF-HC|pDM:singleton
pLoF-HC|pDM:0.01
pLoF-HC|pDM:0.001
pLoF-HC|pDM:0.0001
pLoF-HC:singleton
pLoF-HC:0.01
pLoF-HC:0.001
pLoF-HC:0.0001
Combined Test
0
5
10
2
4
6
0
5
10
15
-LOG10P


A
Forest Plot for MAP3K15
T2D
META (P=1.7e-08)
GnH-SAS I N=50296 I P=2.3e-07
UKB-SAS I N=7165 I P=0.09
UKB-EUR I N=215045 I P=0.0084
HbA1c
META (P=4.1e-14)
UKB-Admixed American I N=757 I P=0.0046
UKB-EAS I N=2388 I P=0.09
UKB-EUR I N=435154 I P=2.2e-10
GnH-SAS I N=42477 I P=7.8e-07
Glucose
META (P=2.5e-15)
UKB-EAS I N=2182 I P=0.07
UKB-SAS I N=8689 I P=0.01
UKB-AFR I N=1991 I P=0.03
UKB-EUR I N=398268 I P=2.2e-08
GnH-SAS I N=32591 I P=1.4e-07
LDL
META (P=0.001)
UKB-Admixed American I N=774 I P=0.1
UKB-EUR I N=46659 I P=0.04
GnH-SAS I N=35314 I P=0.0022
TIA/stroke
META (P=0.67)
UKB-SAS I N=7677 I P=0.09
UKB-EUR I N=355906 I P=0.05
GnH-SAS I N=50296 I P=0.021
GDM
META (P=0.00012)
UKB-EUR I N=7858 I P=0.08
GnH-SAS I N=50296 I P=0.00037
-2
0
2
Effect Size (BETA)
B
MAP3K15 Coding Single Variant + Gene-based LOVO Associations with T2D
SNV -LOG10P
LOVO -LOG10P
15
10
5
0
5
10
15
19400000
19450000
Chr 23 Position
C
RNF19A Coding Single Variant + Gene-based LOVO Associations with T2D
SNV -LOG10P
LOVO -LOG10P
15
10
5
0
5
10
15
100260000
100270000
100280000
100290000
100300000
Chr 8 Position
D
Forest Plot for RNF19A
T2D
META (P=2.7e-07)
GnH-SAS I N=50296 I P=0.008
UKB-SAS I N=10082 I P=0.05
UKB-EUR I N=319449 I P=0.0098
UKB-AFR I N=6514 I P=0.0095
HbA1c
META (P=0.0017)
GnH-SAS I N=35143 I P=0.042
UKB-SAS I N=9413 I P=0.0064
BP
META (P=0.00024)
GnH-SAS I N=29898 I P=0.0054
UKB-EUR I N=52089 I P=0.03
UKB-SAS I N=46113 I P=0.0055
UKB-AFR I N=362 I P=0.03
BMI
META (P=0.69)
UKB-SAS I N=2514 I P=0.02
UKB-AFR I N=8452 I P=0.06
GnH-SAS I N=43956 I P=0.0064
HDL
META (P=2e-04)
UKB-AFR I N=1996 I P=0.07
UKB-SAS I N=2434 I P=0.02
GnH-SAS I N=37251 I P=0.026
PAD
META (P=0.00018)
GnH-SAS I N=50296 I P=0.013
UKB-EUR I N=295453 I P=0.03
HF
META (P=0.033)
GnH-SAS I N=50296 I P=0.032
UKB-EUR I N=139606 I P=0.05
IHD
META (P=3.2e-05)
GnH-SAS I N=50296 I P=0.057
UKB-AFR I N=4182 I P=0.07
UKB-SAS I N=5727 I P=0.0095
TIA/stroke
META (P=0.0047)
GnH-SAS I N=50296 I P=0.027
UKB-SAS I N=7697 I P=0.04
-2.5
0.0
2.5
5.0
Effect Size (BETA)


A
Forest Plot for HNF4A
T2D
META (P=0.00073)
GnH−SAS | N=50296 | P=1.2e−11
UKB−AFR | N=4863 | P=0.03
UKB−EAS | N=2247 | P=0.0045
UKB−SAS | N=8910 | P=0.0014
UKB−EUR | N=215045 | P=1.2e−06
HbA1c
META (P=0.58)
UKB−SAS | N=9413 | P=0.06
UKB−EAS | N=2388 | P=0.1
UKB−AFR | N=7038 | P=0.04
UKB−EUR | N=435154 | P=0.00082
GnH−SAS | N=42477 | P=3e−05
Glucose
META (P=0.9)
UKB−EUR | N=398268 | P=0.02
GnH−SAS | N=32591 | P=6e−04
LDL
META (P=0.63)
UKB−EUR | N=109746 | P=0.02
GnH−SAS | N=34374 | P=0.049
Diabetic eye disease
META (P=0.13)
GnH−SAS | N=50296 | P=0.00099
UKB−SAS | N=3415 | P=0.03
UKB−AFR | N=4818 | P=0.09
−1
0
1
2
3
Effect Size (BETA)
B
20_44429549_C_T
Chromosome 20 (Mb)
C
chr20_44429549_C_T
PIP
D
chr20_44429549_C_T
Chromosome 20 (Mb)
E
chr20_44429549_C_T
PIP
F
HNF4A Pro437Ser Survival
From T2D
Genotype C/C C/T
Survival probability
Time
Number at risk
C/C 53713 33389 5799 584 0
C/T 881 588 98 7 0
G
HNF4A Pro437Ser Survival
From Myocardial Infarction
Genotype C/C C/T T/T
Survival probability
Time
Number at risk
C/C 49482 47859 18151 1389 0
C/T 804 780 264 17 0
T/T 12 11 4 0 0

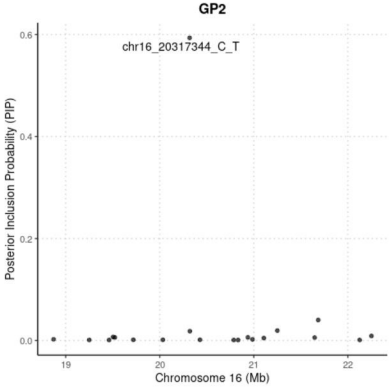


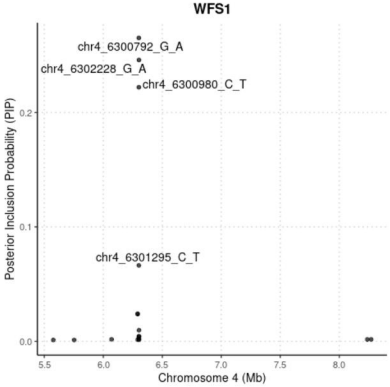
Meta–analysis of Cohorts by SNP and Phenotype
Cohort
GoH
T2DKP–AF
T2DKP–EU
T2DKP–HS
MDRF
T2DKP–EA
T2DKP–GME
Meta–analysis
WFS1 Val333Ile
CDKAL1 Pro409Arg
CP2 Val429Met
Ph C–peptide
Ph Glucose
ALP
BMI
Cerebrovascular disease
Fasting C–peptide
Fasting glucose
Fasting insulin
GDM
HbA1c
HDL–C
HOMA–B
HOMA–IR
Hypertension
LDL–C
MI
Non–fasting glucose
Total cholesterol
Trig/HDL Ratio
Triglycerides
Urinal ACR
Waist circumference
0
10
20
−4
0
4
−5
0
5
Z–normalised Effect Size (Beta)


A
25k
20k
15k
10k
5k
0
Pancreatic-Islets
Whole Pancreas
Adipose-Subcutaneous
Adipose-Visceral
Adrenal Gland
Artery-Aorta
Artery-Coronary
Artery-Tibial
Bladder
Brain-Amygdala
Brain-Anterior cingulate cortex
Brain-Caudate
Brain-Cerebellar Hemisphere
Brain-Cerebellum
Brain-Cortex
Brain-Frontal Cortex
Brain-Hippocampus
Brain-Hypothalamus
Brain-Nucleus accumbens
Brain-Putamen
Brain-Spinal cord
Brain-Substantia nigra
Breast-Mammary Tissue
Cells-EBV-transformed lymphocyte
Cells-Transformed fibroblasts
Cervix-Ectocervix
Cervix-Endocervix
Colon-Sigmoid
Colon-Transverse
Esophagus-Gastroesophageal J
Esophagus-Mucosa
Esophagus-Muscularis
Fallopian Tube
Heart-Atrial Appendage
Heart-Left Ventricle
Kidney-Cortex
Liver
Lung
Minor Salivary Gland
Muscle-Skeletal
Nerve-Tibial
Ovary
Pituitary
Prostate
Skin-Not Sun Exposed
Skin-Sun Exposed
Small Intestine-Terminal Ileum
Spleen
Stomach
Testis
Thyroid
Uterus
Vagina
Whole Blood
B
Log-Normalized Expression
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
Ductal
Active Stellate
Alpha
Acinar
MUC5B+ Ductal
Beta
Endothelial
Gamma + Epsilon
Delta
Immune
Quiescent Stellate
Cycling Alpha
C
MSP Tissue
3
2.5
2
1.5
1
0.5
0
adrenalgland
artery
brain
colon
esophagus
heart
liver
lung
muscle
nerve
ovary
pancreas
prostate
skin
spleen
stomach
thyroid
D
pPS Associaltons with GP2 Abundance
Beta_cell1
META (P=1.7e-07)
GrrH I N=1434
UKB-EUR I N=49496
Beta_cell2
META (P=2.3e-08)
GrrH I N=1434
UKB-EUR I N=49496
Obesity
META (P=5.1e-08)
GrrH I N=1434
UKB-EUR I N=49496
Cholesterol
META (P=0.00014)
GrrH I N=1434
UKB-EUR I N=49496
ALP_Neg
META (P=7.2e-07)
GrrH I N=1434
UKB-EUR I N=49496
Bilirubin
META (P=0.79)
GrrH I N=1434
UKB-EUR I N=49496
Hyper_insulin
META (P=0.38)
GrrH I N=1434
UKB-EUR I N=49496
Lipodystrophy1
META (P=0.15)
GrrH I N=1434
UKB-EUR I N=49496
Lipodystrophy2
META (P=0.35)
GrrH I N=1434
UKB-EUR I N=49496
Liver_Lipid
META (P=0.89)
GrrH I N=1434
UKB-EUR I N=49496
Proinsulin
META (P=0.5)
GrrH I N=1434
UKB-EUR I N=49496
SHBG_LpA
META (P=0.45)
GrrH I N=1434
UKB-EUR I N=49496
-0.2
0.0
0.2
0.4
Effect Size (BETA)

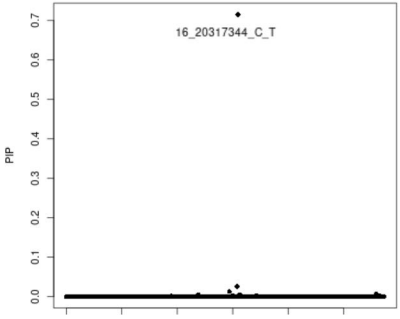
A

Whole exome sequencing

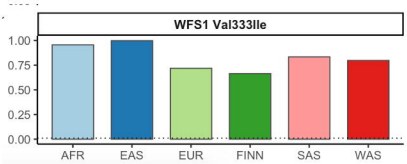


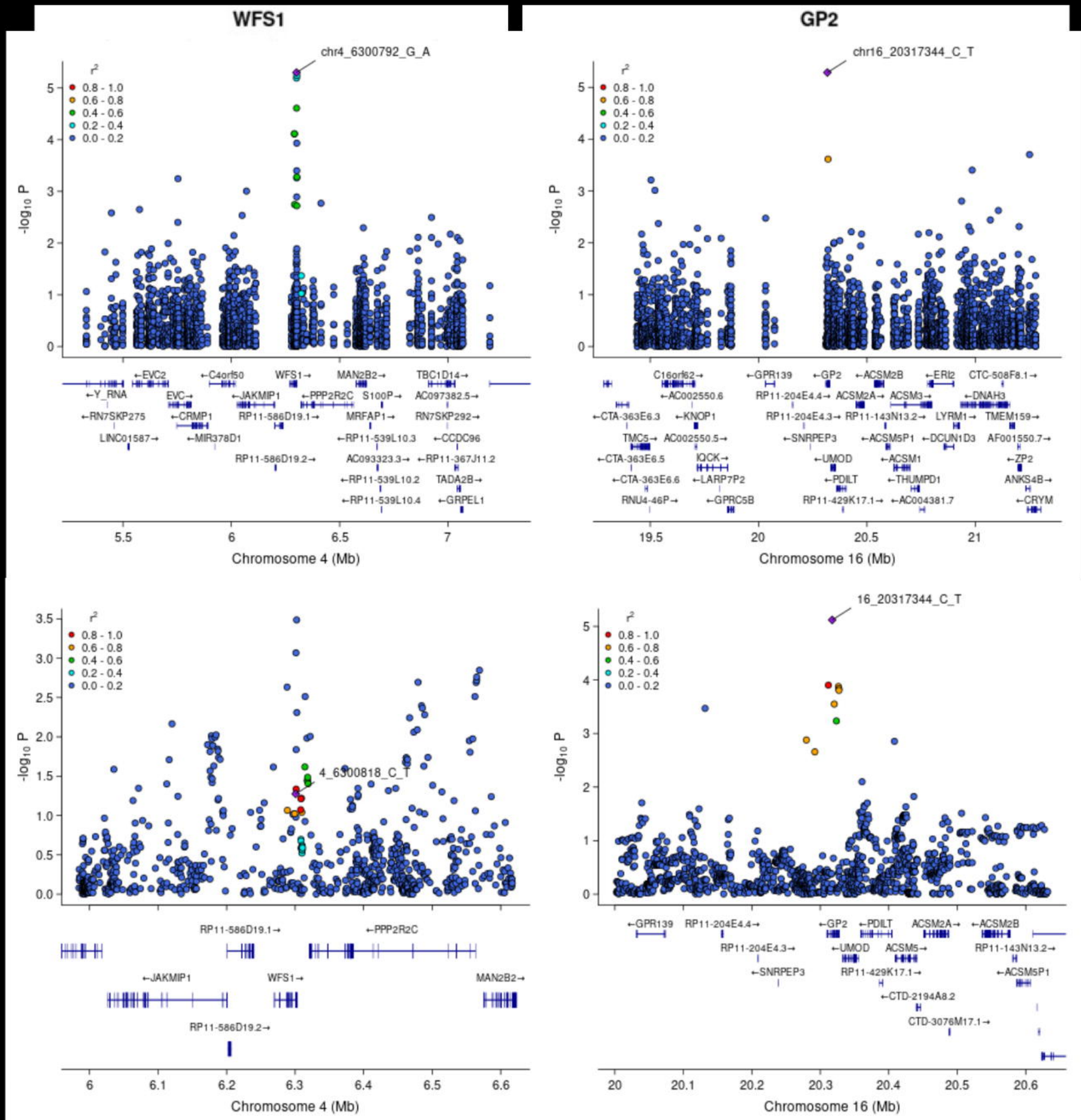



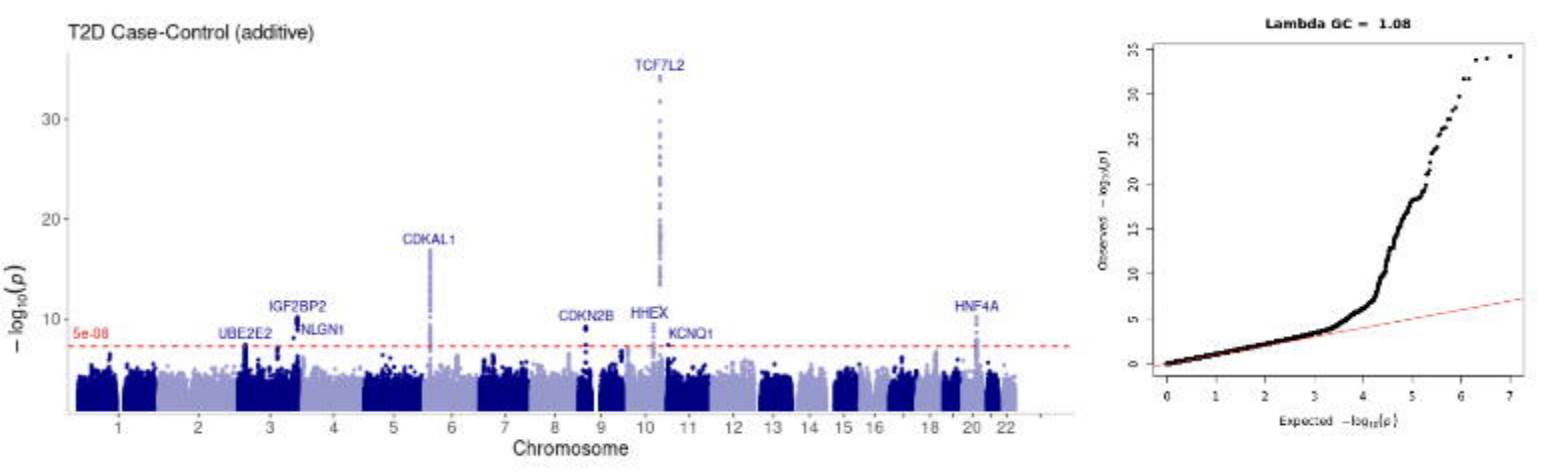

Imputed Genotype




B




WFS1
GP2
chr4_6300792_G_A
chr16_20317344_C_T
4_6300818_C_T
16_20317344_C_T
r²
0.8 - 1.0
0.6 - 0.8
0.4 - 0.6
0.2 - 0.4
0.0 - 0.2
-log10 P
Chromosome 4 (Mb)
Chromosome 16 (Mb)
EVC2
C4orf50
WFS1
MAN2B2
TBC1D14
Y_RNA
EVC
JAKMIP1
PPP2R2C
S100P
AC097382.5
RN7SKP275
CRMP1
RP11-586D19.1
MRFAP1
RN7SKP292
LINC01587
MIR378D1
RP11-539L10.3
CCDC96
RP11-586D19.2
AC093323.3
RP11-367J11.2
RP11-539L10.2
TADA2B
RP11-539L10.4
GRPEL1
C16orf62
GPR139
GP2
ACSM2B
ERI2
CTC-508F8.1
AC002550.6
RP11-204E4.4
ACSM2A
ACSM3
DNAH3
CTA-363E6.3
KNOP1
RP11-204E4.3
RP11-143N13.2
LYRM1
TMEM159
TMC5
AC002550.5
SNRPEP3
ACSM5P1
DCUN1D3
AF001550.7
CTA-363E6.5
IQCK
UMOD
ACSM1
ZP2
CTA-363E6.6
LARP7P2
PDILT
THUMPD1
ANKS4B
RNU4-46P
GPRC5B
RP11-429K17.1
AC004381.7
CRYM
ACSM5
CTD-2194A8.2
CTD-3076M17.1


T2D Case-Control (additive)
TCF7L2
CDKAL1
IGF2BP2
NLGN1
UBE2E2
CDKN2B
HHEX
KCNQ1
HNF4A
5e-08
$-\log_{10}(p)$
Chromosome
Lambda GC = 1.08
Observed $-\log_{10}(p)$
Expected $-\log_{10}(p)$