nature genetics

ARTICLES

https://doi.org/10.1038/s41588-020-0637-y

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

Marijana Vujkovic[1,2,112], Jacob M. Keaton[3,4,5,6,112], Julie A. Lynch[7,8], Donald R. Miller[9,10], Jin Zhou[11,12], Catherine Tcheandjieu[13,14,15], Jennifer E. Huffman[16], Themistocles L. Assimes[13,14], Kimberly Lorenz[1,17,18], Xiang Zhu[13,19], Austin T. Hilliard[13,14], Renae L. Judy[1,20], Jie Huang[16,21], Kyung M. Lee[7], Derek Klarin[16,22,23,24], Saiju Pyarajan[16,25,26], John Danesh[27], Olle Melander[28], Asif Rasheed[29], Nadeem H. Mallick[30], Shahid Hameed[30], Irshad H. Qureshi[31,32], Muhammad Naeem Afzal[31,32], Uzma Malik[31,32], Anjum Jalal[33], Shahid Abbas[33], Xin Sheng[2], Long Gao[17], Klaus H. Kaestner[17], Katalin Susztak[2], Yan V. Sun[34,35], Scott L. DuVall[7,36], Kelly Cho[16,25], Jennifer S. Lee[13,14], J. Michael Gaziano[16,25], Lawrence S. Phillips[34,37], James B. Meigs[23,26,38], Peter D. Reaven[11,39], Peter W. Wilson[34,40], Todd L. Edwards[4,41], Daniel J. Rader[2,17], Scott M. Damrauer[1,20], Christopher J. O'Donnell[16,25,26], Philip S. Tsao[13,14], The HPAP Consortium*, Regeneron Genetics Center*, VA Million Veteran Program*, Kyong-Mi Chang[1,2,42,112], Benjamin F. Voight[1,17,18,112] ✉ and Danish Saleheen[29,43,44,112] ✉

We investigated type 2 diabetes (T2D) genetic susceptibility via multi-ancestry meta-analysis of 228,499 cases and 1,178,783 controls in the Million Veteran Program (MVP), DIAMANTE, Biobank Japan and other studies. We report 568 associations, including 286 autosomal, 7 X-chromosomal and 25 identified in ancestry-specific analyses that were previously unreported. Transcriptome-wide association analysis detected 3,568 T2D associations with genetically predicted gene expression in 687 novel genes; of these, 54 are known to interact with FDA-approved drugs. A polygenic risk score (PRS) was strongly associated with increased risk of T2D-related retinopathy and modestly associated with chronic kidney disease (CKD), peripheral artery disease (PAD) and neuropathy. We investigated the genetic etiology of T2D-related vascular outcomes in the MVP and observed statistical SNP-T2D interactions at 13 variants, including coronary heart disease (CHD), CKD, PAD and neuropathy. These findings may help to identify potential therapeutic targets for T2D and genomic pathways that link T2D to vascular outcomes.

T2D, a leading cause of morbidity globally, is projected to affect up to 629 million people by 2045 (ref. [1]). People with T2D are at increased risk of developing a wide range of macro- and microvascular outcomes[2] and there are large disparities in prevalence, severity and comorbidities across global populations. Over 400 common variants have been identified that confer disease susceptibility[3,4], yet because most studies have been performed in cohorts of European or Asian ancestry, the impact of these variants across all ancestry groups needs to be quantified. Identification of genetic factors and genes that underlie T2D-related complications could inform clinical management strategies, including patient stratification or optimization of study design of randomized controlled trials. The lack of large, multi-ancestry richly phenotyped cohorts linked to genetic data has made it difficult to address these questions.

We conducted a multi-ancestry association study of T2D risk comprising 228,499 individuals with T2D and 1,178,783 control individuals of European, African American, Hispanic, South Asian and East Asian ancestry. We investigated the association of a T2D PRS with major T2D-related macrovascular outcomes (CHD, ischemic stroke and PAD) and three microvascular diseases (CKD, retinopathy and neuropathy) in the MVP[5]. Subsequently, we conducted a genome-wide SNP–T2D interaction analysis in the MVP to identify genetic variants where the effect of the SNP on the vascular outcome depends on the context of T2D presence. We also performed association analyses of genetically predicted expression levels and expression quantitative trait–T2D colocalization analyses to identify the effects of gene–tissue pairs that influence T2D risk through inter-individual variation in expression.

This study complements prior genetic studies of T2D through the use of large-scale clinical data in conjunction with polygenic scores and evaluation of context specificity for genetic effects on T2D vascular sequelae, and by describing the regulatory circuits that influence T2D risk.

Results

Study populations. We performed a genome-wide, multi-ancestry T2D-association analysis (228,499 cases and 1,178,783 controls) that encompassed five ancestral groups (Europeans, African Americans,

A full list of affiliations appears at the end of the paper.

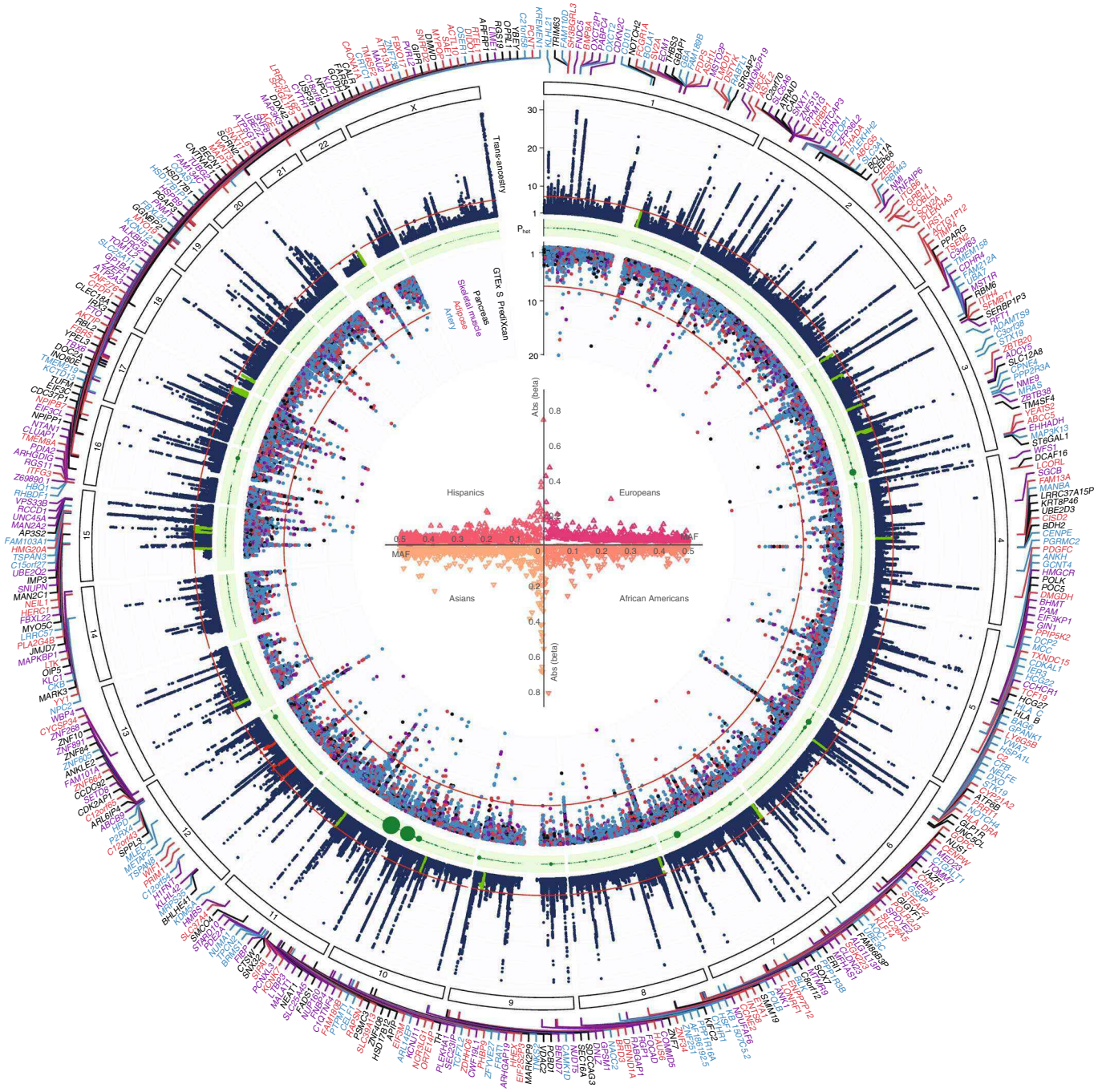


Fig. 1 | Trans-ancestry GWAS meta-analysis identifies 318 loci associated with T2D. The graph represents a circos plot performed in 228,499 T2D cases and 1,178,783 controls. The outer track corresponds to $-\log_{10}(P)$ for association with T2D in the trans-ancestry meta-analysis using a fixed-effects model with inverse-variance weighting of log odds ratios (*y* axis truncated at 30), by chromosomal position. The red line indicates genome-wide significance ($P = 5.0 \times 10^{-8}$). Purple gene labels indicate genes that were identified in skeletal muscle eQTLs by S-PrediXcan analysis, red gene names indicate genes identified in adipose eQTLs, black gene names indicate genes identified in pancreas eQTLs, and blue gene names indicate genes that were identified in eQTLs from arteries. The green band corresponds to measures of heterogeneity related to the index SNPs associated with T2D that were generated using Cochran's Q statistic. Dot sizes are proportional to I^2 or ancestry-related heterogeneity. The inner track corresponds to $-\log_{10}(P)$ for association with skeletal muscle, adipose, pancreas and artery tissue eQTLs from S-PrediXcan analysis (*y* axis truncated at 20), by chromosomal position. The red line indicates genome-wide significance ($P = 5.0 \times 10^{-8}$). Inset, effects of all 318 index SNPs on T2D by minor allele frequency, stratified and colored by ancestral group.

Hispanics, South Asians and East Asians) by meta-analyzing genome-wide association study (GWAS) summary statistics derived from the MVP[5] and other studies with non-overlapping participants: the DIAMANTE Consortium[3], Penn Medicine Biobank[6], the Pakistan Genomic Resource[7], Biobank Japan[4], the Malmö Diet and Cancer Study[8], Medstar[9] and PennCath[9] (Methods and Supplementary Tables 1 and 2). MVP participants ($n = 273{,}409$) comprised predominantly male subjects (91.6%) and were classified as Europeans (72.1%), African Americans (19.5%), Hispanics (7.5%) and Asians (0.9%) (Supplementary Table 3).

Table 1 | T2D locus discovery in African Americans

Description	Lead SNP	RSID	EA	NEA	EAF	Beta	SE	*P*	*n*	*n* Cojo	Established SNP
Novel AA	chr12:38710523	rs7315028	G	A	0.882	0.124	0.022	1.5×10^{-8}	56,150	1	-
	chr12:57968738	rs11172254	G	A	0.817	0.097	0.017	1.8×10^{-8}	56,150	1	-
	chr12:88338461	rs10745460	T	A	0.660	0.079	0.014	3.7×10^{-8}	56,150	0	-
Novel TE	chr7:50887174	rs7781440	C	T	0.284	−0.086	0.015	5.3×10^{-9}	56,150	0	-
	chr12:80985872	rs1528287	G	T	0.059	−0.494	0.080	8.2×10^{-10}	56,150	1	-
Established	chr3:123065778	rs11708067	G	A	0.151	−0.118	0.018	2.3×10^{-11}	56,150	0	chr3:123082398
	chr3:185534482	rs9859406	G	A	0.257	−0.115	0.015	5.7×10^{-14}	56,150	0	chr3:185829891
	chr5:55807370	rs464605	C	T	0.429	-0.077	0.013	1.1×10^{-9}	56,150	0	chr5:55860781
	chr6:39016636	rs10305420	C	T	0.920	0.142	0.025	8.5×10^{-9}	56,150	0	chr6:39282371
	chr7:15064896	-	G	T	0.565	0.101	0.013	2.7×10^{-15}	56,150	0	chr7:15060429
	chr7:28180556	rs864745	C	T	0.257	−0.083	0.014	1.1×10^{-8}	56,150	0	chr7:28198677
	chr7:44185088	rs2908274	G	A	0.359	−0.089	0.014	5.4×10^{-11}	56,150	1	chr7:44266184
	chr8:41510260	rs12550613	G	C	0.310	−0.114	0.014	5.5×10^{-16}	56,150	0	chr8:41537318
	chr8:118166327	rs60461843	T	A	0.939	0.172	0.028	1.3×10^{-9}	56,150	1	chr8:118024315
	chr9:139241595	rs28562046	G	C	0.709	0.080	0.014	2.8×10^{-8}	56,150	0	chr9:139737088
	chr10:114758349	rs7903146	C	T	0.706	−0.226	0.014	5.6×10^{-60}	56,150	0	chr10:114871594
	chr11:2691500	rs231361	G	A	0.656	−0.080	0.013	2.2×10^{-9}	56,150	2	chr11:2717680
	chr11:2858546	rs2237897	C	T	0.908	0.143	0.024	2.2×10^{-9}	56,150	1	chr11:2717680
	chr12:66215214	rs2583938	T	A	0.197	−0.123	0.018	3.3×10^{-12}	56,150	0	chr12:66358347
	chr15:77776498	rs952471	G	C	0.534	0.077	0.013	4.2×10^{-9}	56,150	0	chr15:77339496
	chr16:53811788	rs62033400	G	A	0.102	0.151	0.021	6.1×10^{-13}	56,150	1	chr16:53758720

Association between genetic variants and T2D in African Americans in the MVP was assessed through logistic regression assuming an additive model of variants with MAF > 1%. A meta-analysis was performed using a fixed-effects model with inverse-variance weighting of log odds ratios. Variants were considered to be genome-wide significant if they passed the conventional *P* value threshold of 5×10^{-8}. AA, African American; TE, trans-ancestry; RSID, RefSNP identification number; EA, effect allele; NEA, non-effect allele; EAF, effect allele frequency; beta, effect estimate; SE, standard error; *n*, sample size; *n* Cojo, additional number of conditionally independent variants identified at the respective locus (and listed in Supplementary Table 12).

Single-variant autosomal analyses. We identified 558 independent sentinel SNPs (286 previously unreported, >500 kb and r^2 linkage disequilibrium (LD) < 0.05 from previous reports; see Methods)[3,4,10,11] associated with T2D (Fig. 1, Table 1, Supplementary Tables 4–8 and Extended Data Fig. 1). Twenty-one additional SNPs were associated at genome-wide significance in an ancestry-specific analysis of Europeans only (Supplementary Table 6). We found that novel loci had smaller magnitudes of effect (average beta regression coefficient of 0.032 ± 0.012 per allele) than previously established SNPs (average beta of 0.054 ± 0.045 per allele; Supplementary Table 5), which presumably results from enhanced power to discover weaker effects due to the large sample size and ancestral diversity. Genome-wide chip heritability analysis explained 19% of the T2D risk on a liability scale[3].

In analysis focused on African American participants (Table 1), we observed a total of 21 loci associated with T2D susceptibility at genome-wide significance, 16 of which were in strong LD with established T2D variants. Three variants were novel and their effects on T2D appeared to be specific to African Americans. Single-variant analysis in the Hispanic subset identified two associated SNPs, both of which tagged previously reported T2D loci (Supplementary Table 7). No novel associations were observed among the individuals of Asian ancestry (Supplementary Table 8).

Polygenicity and population stratification. To evaluate whether the observed genomic inflation was due to the polygenic nature of T2D or due to underlying population stratification, LD score regression (LDSC)[12] was used in Europeans and Asians to compare lambda genomic control (GC)[13] and LDSC intercept (Methods). In Asians, a total of 1,077,427 SNPs were analyzed, which resulted in a lambda GC of 1.342 and intercept of 1.094 (SE = 0.012). In Europeans, 1,198,787 SNPs were analyzed, which resulted in a lambda GC of 1.863 and intercept of 1.139 (SE = 0.016). Admixture-adjusted LDSC[14] was used in African Americans and Hispanics. A total of 945,603 SNPs were analyzed in African Americans, with lambda GC of 1.180 and intercept of 1.048 (SE = 0.007). For Hispanics, 1,077,427 SNPs were analyzed, with lambda GC of 1.093 and intercept of 1.091 (SE = 0.113). Except perhaps for Hispanics (where the estimated error on the intercept is large), these results suggest that a substantial part of the observed inflation these populations is due to T2D polygenicity.

X-chromosome analyses. In a trans-ancestry analysis of the X chromosome, we identified a total of ten association signals for T2D, of which seven were novel (Table 2, Supplementary Table 9 and Extended Data Fig. 2). A European-restricted analysis identified four loci on the X chromosome, all of which were identified in the trans-ancestry meta-analysis. One novel X-chromosome locus was associated with T2D specifically in African Americans. Of note, one novel trans-ancestry association was identified near the androgen receptor (*AR*) gene and was in strong LD with a previously reported variant (rs4509480), which was previously shown to associate with male-pattern baldness (EUR $r^2 = 0.98$, rs200644307).

Effect heterogeneity between Europeans and African Americans. Whereas at most loci we found no evidence for heterogeneity of effect estimates between Europeans and African Americans, we did observe that 44 (7.9%) variants had significantly different effect size estimates between the two groups (Supplementary Table 10). Remarkably, four loci near *SLC30A8*, *PTPRQ*, *GRB10* and *SALL2* showed higher effect sizes for T2D at stronger levels of significance

Table 2 | T2D chromosome X analysis (overall results)

Population	Lead SNP	EA	NEA	EAF	Novel	Literature SNP	Nearest gene	*n* Cases	*n* Controls	Beta	SE	*P*
Trans-ancestry	chrX:19497290	A	G	0.968	1	-	*MAP3K15*	102,683	170,726	0.131	0.023	1.4×10^{-8}
	chrX:20009166	T	C	0.323	1	-	*BCLAF3;MAP7D2*	102,683	170,726	0.058	0.010	7.9×10^{-9}
	chrX:31851610	T	C	0.343	1	-	*DMD*	102,683	170,726	0.047	0.009	3.5×10^{-8}
	chrX:56902211	A	T	0.612	0	X:57170781	*SPIN2A;FAAH2*	102,683	170,726	−0.069	0.010	1.9×10^{-12}
	chrX:66168667	A	G	0.277	1	-	*AR;EDA2R*	102,683	170,726	0.082	0.011	1.9×10^{-13}
	chrX:109888390	A	C	0.364	1	-	*RGAG1;CHRDL1*	102,683	170,726	−0.048	0.008	7.7×10^{-9}
	chrX:117955250	T	C	0.231	0	X:117915163	*IL13RA1*	102,683	170,726	0.077	0.010	4.1×10^{-15}
	chrX:124390172	T	C	0.853	1	-	*TENM1*	102,683	170,726	−0.075	0.013	9.0×10^{-9}
	chrX:135859359	C	G	0.407	1	-	*ARHGEF6*	102,683	170,726	−0.049	0.008	7.3×10^{-9}
	chrX:153882606	C	G	0.026	0	X:152908887	*CCNQ;DUSP9*	102,683	170,726	−0.486	0.026	3.0×10^{-78}
European	chrX:56759371	T	G	0.218	0	X:57170781	*SPIN2A;FAAH2*	69,869	127,197	0.069	0.013	1.7×10^{-8}
	chrX:66316809	G	A	0.290	1	-	*EDA2R*	69,869	127,197	0.077	0.013	3.4×10^{-9}
	chrX:117877437	A	G	0.223	0	X:117915163	*IL13RA1*	69,869	127,197	0.118	0.013	5.5×10^{-20}
	chrX:152898928	C	A	0.247	0	X:152908887	*CCNQ;DUSP9*	69,869	127,197	−0.163	0.012	7.9×10^{-46}
African	chrX:67255974	C	T	0.189	1	-	*AR;OPHN1*	23,305	30,140	0.104	0.019	3.4×10^{-8}
American	chrX:132597984	C	T	0.282	1	-	*GPC3;GPC4*	23,305	30,140	0.135	0.024	1.4×10^{-8}
	chrX:153882606	C	G	0.026	0	X:152908887	*G6PD*	23,305	30,140	−0.500	0.027	1.6×10^{-76}

A sex-stratified (male or female) ancestry-separated (European, African American, Hispanic or Asian) analysis was performed with dosage (number of X-chromosome copies) as the independent variable and T2D as the outcome. Covariates included age and first ten PCs of ancestry. The ancestry-specific sex-stratified results are presented in Supplementary Table 9. Output from ancestry-separated male and female analyses were then meta-analyzed using a fixed-effects model with inverse-variance weighting of log odds ratios and are shown here. For the trans-ancestry meta-analysis, the ancestry-specific sex-meta-analyzed data were additionally meta-analyzed using a fixed-effects model with inverse-variance weighting of log odds ratios. Variants were considered to be genome-wide significant if they passed the conventional *P* value threshold of 5×10^{-8}. EA, effect allele; NEA, non-effect allele; EAF, effect allele frequency; Beta, effect estimate; SE, standard error; *n* cases, total number of T2D cases; *n* controls, total number of unaffected controls.

in African Americans compared with Europeans. Of these loci, associations with loss-of-function (LOF) variants in *SLC30A8* were previously reported in Europeans, African Americans and South Asians.

Secondary signal analysis. We detected a total of 233 conditionally independent SNPs that flanked 49 novel and 108 previously reported lead SNPs in Europeans (Supplementary Tables 11 and 12). We observed no novel conditionally independent variants in participants of South Asian, East Asian and Hispanic ancestry.

Fine mapping of lead SNPs with coding variants. To identify coding variants that may drive the association between the lead SNPs and T2D risk, we investigated predicted loss-of-function (pLOF) and missense variants near the identified T2D lead variants from the European-specific T2D summary statistics (Supplementary Table 13). We identified two pLOF (*LPL* and *ANKDD1B*) and 45 missense variants in 47 genes that were in LD with at least one of the T2D lead SNPs ($r^2 > 0.5$, MVP reference panel in Europeans) and were associated at $P < 1.0 \times 10^{-4}$. Of the 56 pLOF and missense variants, 14 missense variants were found to be the sentinel T2D SNPs, 19 variants were in LD with novel lead SNPs, and 37 variants were previously reported.

Genome-wide coding variant association analysis. We additionally performed a genome-wide screen of all PLOFs and missense variants (not bound by proximity to sentinel T2D lead variants) to enumerate potential T2D genes not captured by common variant tags (Supplementary Table 14). We identified one additional pLOF variant in *CCHCR1*, whereas 37 novel missense variants were associated with T2D at $P < 5 \times 10^{-8}$.

Rare coding variant phenome-wide association study. We next performed phenome-wide association analysis (PheWAS) for each of the three pLOF variants associated with T2D in MVP participants of European ancestry, UK Biobank data, and Biobank Japan (Table 3). These loci included *ANKDD1B* p.Trp480* (rs34358), *CCHCR1* p.Trp78* (rs3130453) and *LPL* p.*474Ser (rs328), and they were significantly associated with metabolic and inflammatory conditions. Klarin et al.[15] previously reported pheWAS associations for *LPL* p.*474Ser with dyslipidemia, coronary atherosclerosis and other chronic ischemic heart disease in MVP, and lipid and cardiometabolic associations for this variant were also observed in Biobank Japan and UK Biobank. In MVP, *ANKDD1B* p.Trp480* was associated with dyslipidemia, hypercholesterolemia and diabetic neurological manifestations. In Biobank Japan, this variant was a range of blood and immune cell traits, whereas in UK Biobank, the SNP was associated with metabolic and anthropometric traits. In MVP and UK Biobank, *CCHCR1* p.Trp78* was associated with a battery of autoimmune traits and in Biobank Japan this variant was associated with total cholesterol, LDL-C, BMI, NK cells and sodium electrolytes.

Transcriptome-wide association analyses. Next, we used common variants from the European T2D GWAS meta-analysis to evaluate the association of genetically predicted gene expression levels with T2D risk across 52 tissues, including kidney and islet cells, using S-PrediXcan (Supplementary Table 15 and Extended Data Fig. 3). We identified 4,468 statistically significant gene–tissue combination pairs that were genetically predictive of T2D risk, of which 4,211 transcript expression quantitative trait loci (eQTLs) were in LD ($r^2 > 0.5$) with T2D signals. We identified 873 genes in this analysis that would not have been identified by nearest-gene annotation alone. The strongest gene–tissue combination signals were for *NRAP* in the cerebellum and *TCF7L2* in the aortic artery.

We then used the coloc software package to identify the subset of significant genes in which there was a high posterior probability that the set of model SNPs in the S-PrediXcan analysis for each gene

Table 3 | PheWAS of two pLOF variants in MVP participants of European ancestry

Gene	RSID	Amino acid change	PheWAS phenotype	*P*	*n* Cases	*n* Controls	OR	95% CI lower	95% CI upper
ANKDD1B	rs34358	p.Trp480*	Diabetes mellitus	1.04×10^{-6}	62,930	104,442	0.96	0.95	0.98
			T2D	1.36×10^{-6}	62,531	104,442	0.96	0.95	0.98
			T2D with neurological manifestations	1.63×10^{-5}	14,159	104,442	0.94	0.92	0.97
			Disorders of lipid metabolism	5.03×10^{-8}	141,535	41,406	1.05	1.03	1.07
			Hyperlipidemia	4.66×10^{-8}	141,408	41,406	1.05	1.03	1.07
			Hypercholesterolemia	2.33×10^{-6}	32,008	41,406	1.06	1.03	1.08
CCHCR1	rs3130453	p.Trp78*	Diabetes mellitus	4.26×10^{-5}	62,930	104,442	0.97	0.96	0.98
			Type 1 diabetes	3.99×10^{-7}	6,566	104,442	0.91	0.88	0.95
			T2D	3.96×10^{-5}	62,531	104,442	0.97	0.96	0.98
			Epistaxis or throat hemorrhage	1.96×10^{-5}	2,751	110,902	1.12	1.07	1.19
			Celiac disease	2.72×10^{-19}	418	124,470	0.52	0.45	0.60
			Microscopic hematuria	1.83×10^{-5}	4,078	147,054	1.1	1.05	1.15
			Psoriatic arthropathy	7.82×10^{-10}	1,077	140,876	0.76	0.70	0.83

The pLOF variants were tested using logistic regression adjusting for age, sex and ten PCs in an additive effects model using the PheWAS R package in R (v3.2.0). Phenotypes were required to have a case count of over 25 in order to be included in the PheWAS, and a multiple testing threshold for statistical significance was set to the Bonferroni-corrected *P* value threshold of 2.8×10^{-5}. pLOF, predicted loss-of-function; RSID, RefSNP identification number; *n* Cases (number of cases with PheWAS phenotype); *n* Controls (number of unaffected controls for the respective PheWAS phenotype); OR, odds ratio; CI, confidence interval.

was associated with gene expression and with T2D. This analysis refined the results of the transcriptome-wide association scan and excluded some results that might be the consequence of LD between causal SNPs for gene expression and T2D. We detected 3,166 gene–tissue pairs where there was statistically significant association with T2D risk and high posterior probability (posterior probability for the fourth hypothesis, namely colocalized functional and GWAS association, $PP4 > 0.8$) of colocalization, covering a total of 695 distinct genes. When the 804 genes were compared to the GWAS catalog for mapped and reported genes for all prior studies of diabetes or diabetes complications, 687 had not been previously reported. Hypergeometric enrichment analysis showed that most enriched gene expression signals were in cervical spinal cord, basal ganglia and glomerular kidney (Supplementary Table 16).

Assessment of gene–drug relationships. Of the 695 genes identified in the S-PrediXcan analyses, 54 genes had documented interactions with a total of 283 FDA-approved drugs and chemical compounds that did not have an indication for T2D treatment or reported adverse drug events in patients with diabetes by using the Side Effect Resource (SIDER) database of drugs and side effects[16]. The Drug Gene Interaction Database (DGIdb v3.0) was used to identify a total of 322 gene–drug combinations for which it was predicted that the combination would modulate blood glucose, depending on its direction of effect on T2D risk with increasing gene expression and drug action (activator or inhibitor; Supplementary Table 17). Gene–drug combinations included several established T2D loci, such as *KCNJ11*, which was targeted by 15 compounds (for example, sulfonylureas, glinides and p-glycoprotein inhibitors); *SCN3A*, which was targeted by 57 compounds (for example, anti-arrythmetics and anti-epileptics); *PIK3CB*, which was targeted by 46 compounds (for example, cancer drugs); *ACE*, which was targeted by 36 compounds (for example, angiotensin-converting enzyme (ACE) inhibitors); *HMGCR*, which was targeted by 18 compounds (for example, HMG-CoA reductase inhibitors); *PIK3C2A*, which was targeted by 15 compounds (for example, anti-cancer drugs); *F2*, which was targeted by 11 compounds (for example, anticoagulants) and *BLK*, which was targeted by 9 compounds (for example, protein kinase inhibitors).

Tissue-specific and epigenetic enrichment of T2D heritability. To understand the contribution of disease-associated tissues to T2D heritability, we performed tissue-specific analysis using LDSC[17] (Supplementary Table 18). The strongest associations were observed in the genomic annotation surveyed in pancreas and pancreatic islets (for example, pancreatic islet H3K27ac and pancreatic chromatin accessibility, and so on; $P < 0.001$). We additionally tested for enrichment of epigenetic features using GREGOR[18], which compares overlap of T2D-associated locus variants relative to control variants matched for number of LD proxies, allele frequency and gene proximity[18] (Supplementary Tables 19–21). Similar to the results from LDSC, eight of the top ten associated hits mapped to the pancreas, including H3K27ac, pancreatic islet H3K27ac and pancreatic islet activated enhancer activity, among others.

Pathway and functional enrichment analysis. To explore whether our results recapitulate the pathophysiology of T2D, we performed gene set enrichment analysis with all of the variants using DEPICT ($P < 1 \times 10^{-5}$, Supplementary Table 22). Medical subject heading (MeSH)-based analysis showed that several different adipose tissues and sites were enriched (for example, abdominal subcutaneous fat and white adipose tissue). In addition, DEPICT analysis showed that the most significant gene set involved the AKT2 subnetwork, lung cancer, the GAB1 signalosome, protein kinase binding, signal transduction and epidermal growth factor receptor signaling (Supplementary Tables 23 and 24).

Genetic correlation between T2D and other phenotypes. Genome-wide genetic correlations of T2D were calculated with a total of 774 complex traits and diseases by comparing allelic effects using LDSC with the European-specific T2D summary statistics (Methods). A total of 270 significant associations were observed ($P < 5 \times 10^{-8}$; Supplementary Table 25). The strongest positive correlations were observed with waist circumference, overall health, BMI and fat mass of arms, legs, body and trunk, hypertension, coronary artery disease, dyslipidemia, alcohol intake, wheezing and cigarette smoking. There was also a strong negative correlation with years of education.

Table 4 | Genome-wide interaction analysis of vascular and non-vascular complications (not stratified by T2D status)

Outcome type	Outcome	SNP	RSID	NE	EA	EAF	*P* for interaction	Nearest gene
Vascular	CHD	chr9:22076071	rs1831733	T	C	0.482	1.6×10^{-13}	*CDKN2B;CDKN2A*
		chr1:109821511	rs602633	G	T	0.216	4.4×10^{-10}	*SORT1*
		chr12:20231526	rs71039916	TCTTA	T	0.034	8.2×10^{-9}	*PDE3A*
	AIS	chr1:15429233	rs491203	G	A	0.057	7.6×10^{-9}	*TMEM51*
		chr8:94056373	rs2134937	T	C	0.049	3.3×10^{-8}	*TRIQK*
	PAD	chr8:97331026	rs3104154	C	T	0.044	3.0×10^{-8}	*PTDSS1*
Non-vascular	Retinopathy	chr1:146606059	rs76754787	ATT	AT	0.030	1.2×10^{-11}	*GJA8*
		chr10:30992882	rs10733997	A	G	0.037	9.7×10^{-9}	*SVIL2P*
		chr10:119646217	rs2255624	T	G	0.032	1.6×10^{-8}	*SLC18A2*
		chr10:114767771	rs4132670	G	A	0.319	2.1×10^{-8}	*TCF7L2*
	CKD	chr16:20356012	rs34857077	G	GA	0.237	6.4×10^{-19}	*UMOD*
		chr4:181816870	rs2177223	T	C	0.038	2.8×10^{-8}	*TENM3*
	Neuropathy	chr2:206668118	rs78977169	CATA	C	0.023	3.4×10^{-8}	*NRP2*

The analysis included a total case count of 67,403 for CKD, 56,285 for CHD, 35,882 for PAD, 11,796 for AIS, 13,881 for retinopathy and 40,475 for neuropathy. Results stratified by T2D presence (yes or no) are presented in Supplementary Table 26. A logistic regression analysis was performed among MVP participants of European ancestry, in which the respective outcome was tested with SNP, T2D, SNPxT2D, age, sex and ten PCs as covariates. *P* values for interaction between SNP and T2D are noted in the column labeled *P* for interaction. Variants were considered to show a statistically different effect between people with and without T2D if the *P* value for interaction was genome-wide significant ($P < 5 \times 10^{-8}$) and at least one T2D-stratum showed nominal significance ($P < 0.001$; Supplementary Table 26). RSID, RefSNP identification number; NEA, non-effect allele; EA, effect allele; EAF, effect allele frequency.

T2D-related vascular outcomes. Next, we investigated SNP–T2D interaction effects associated with T2D-related vascular outcomes among MVP participants of European descent ($P < 5 \times 10^{-8}$; Methods, Table 4, and Supplementary Table 26). The analysis included a total case count of 67,403 for CKD, 56,285 for CHD, 35,882 for PAD, 11,796 for acute ischemic stroke (AIS), 13,881 for retinopathy and 40,475 for neuropathy. We identified several genome-wide-significant interactions where the genetic associations with T2D-related vascular outcomes were modified by T2D (Table 4 and Supplementary Table 26). We identified two loci for CHD (rs1831733 in chromosome 9p21 and rs602633 near *SORT1*) and one for CKD (rs34857077 in *UMOD*), for which the difference in the effect estimates between T2D strata was genome-wide significant ($P < 5 \times 10^{-8}$) and at least one T2D stratum was genome-wide significant. We identified one locus for CHD (rs71039916 near *PDE3A*), one for CKD (rs2177223 near *TENM3*), one for PAD (rs3104154 in *PTDSS1*), one for neuropathy (rs78977169 near *NRP2*), four for retinopathy (rs76754787 near *GJA8*, rs10733997 in *SVIL2P*, rs2255624 near *SLC18A2* and rs4132670 in *TCF7L2*), as well as two for AIS (rs491203 near *TMEM51* and rs2134937 near *TRIQK*) that showed genome-wide significance for difference in effect estimates between the T2D strata and nominal significance ($P < 0.001$) for at least one T2D stratum.

Polygenic risk scores and T2D-related vascular outcomes. Genome-wide polygenic risk scores (gPRSs) for T2D were calculated in Europeans according to the T2D effect estimates from the previously reported DIAMANTE consortium[3] and then categorized into deciles (Tables 5 and 6). As expected, participants with the highest T2D gPRS scores (90–100% T2D gPRS percentile) showed the highest risk for T2D (OR = 5.21 and 95% CI 4.94–5.49; Extended Data Fig. 5) when compared to the reference group (0–10% T2D gPRS percentile) in a cross-sectional study design.

We evaluated whether the T2D gPRS was associated with the risk of micro- and macrovascular outcomes in an analysis restricted to participants with T2D. The *P* values were calculated using gPRS as a continuous exposure, and odds ratios were calculated by contrasting the top to the bottom gPRS decile (Fig. 2 and Tables 5 and 6). We observed strong association between a T2D gPRS and microvascular complications, in particular with retinopathy, but to a lesser extent with neuropathy and CKD. For macrovascular outcomes, T2D gPRS was associated with the risk of PAD, but not with the risk of CHD or AIS.

Discussion

We report the discovery of 318 novel autosomal and X-chromosomal variants associated with T2D susceptibility in a trans-ancestry GWAS. We also report 13 variants associated with differences in T2D-related micro- and macrovascular outcomes between individuals with and without diabetes. The substantial locus discovery was achieved by combining data from several large-scale biobanks and consortia, where the MVP data constituted over 40% of all cases of T2D. Furthermore, we present the largest cohort of African Americans including over 56,000 participants, substantially larger than previous African-specific studies published to date.

Analyses of coding variants identified 44 variants associated with T2D, including three pLOF variants in the *LPL*, *ANKDD1B* and *CCHCR1* genes. We identified 804 putative causal genes at both novel and previously reported loci, including 54 genes that were found to be possible targets for FDA-approved drugs and chemical compounds. Our SNP–T2D interaction analyses identified several loci at which the association between a genetic variant and a vascular outcome differed between people with T2D as compared to those without. We also found that a high polygenic risk for T2D strongly increased the risk for retinopathy in individuals with T2D, as well as for CKD, neuropathy and PAD.

T2D is highly prevalent in people of African ancestry; however, there are a total of three published T2D GWAS reports in this ancestral group with only four definitely detected loci[19–21]. In our study with over 56,000 participants of recent African ancestry, we report four novel loci for T2D that are solely observed in this ancestral group, including one that is located on the X chromosome. Of the previously reported loci, only rs3842770 (*INS–IGF2*) was replicated here. We did not observe replication either with rs7560163 (ref. [21]) or rs73284431, reported from a large study conducted in sub-Saharan Africa. The reported *HLA-B* variant rs2244020 did not replicate in our study, but we did observe a significant association with another SNP in the HLA region (rs10305420, OR

Table 5 | Polygenic risk scores and vascular outcomes

Outcome type	Outcome	T2D PRS decile	*n* Cases	*n* Controls	OR	95% CI lower	95% CI upper	*P*	*P* for linear trend
Vascular	CHD	0–10%	2,913	3,924	1.00	Ref.	Ref.	–	0.636
		10–20%	2,940	3,924	1.01	0.92	1.12	0.811	
		20–30%	2,958	3,924	0.98	0.89	1.08	0.742	
		30–40%	2,934	3,924	0.99	0.90	1.09	0.835	
		40–50%	2,988	3,924	1.01	0.92	1.11	0.801	
		50–60%	3,001	3,924	0.98	0.90	1.08	0.744	
		60–70%	2,977	3,924	1.01	0.92	1.10	0.887	
		70–80%	2,916	3,924	1.02	0.93	1.12	0.632	
		80–90%	3,032	3,924	0.96	0.88	1.05	0.391	
		90–100%	3,038	3,924	1.03	0.94	1.12	0.537	
	AIS	0–10%	555	6,027	1.00	Ref.	Ref.	–	0.070
		10–20%	563	6,027	0.90	0.76	1.07	0.238	
		20–30%	583	6,027	0.98	0.83	1.15	0.782	
		30–40%	619	6,027	0.98	0.84	1.15	0.821	
		40–50%	530	6,027	0.99	0.85	1.16	0.924	
		50–60%	576	6,027	0.99	0.85	1.16	0.941	
		60–70%	645	6,027	0.97	0.83	1.13	0.672	
		70–80%	590	6,027	1.04	0.90	1.20	0.611	
		80–90%	558	6,027	1.05	0.91	1.22	0.494	
		90–100%	627	6,027	1.02	0.89	1.17	0.784	
	PAD	0–10%	1,966	4,871	1.00	Ref.	Ref.	–	2.0×10^{-7}
		10–20%	1,964	4,871	1.00	0.93	1.08	0.927	
		20–30%	1,948	4,871	1.01	0.93	1.08	0.890	
		30–40%	1,984	4,871	1.04	0.96	1.12	0.361	
		40–50%	1,964	4,871	1.03	0.96	1.11	0.425	
		50–60%	1,950	4,871	1.02	0.95	1.10	0.559	
		60–70%	1,972	4,871	1.05	0.98	1.14	0.165	
		70–80%	1,960	4,871	1.05	0.97	1.13	0.203	
		80–90%	2,019	4,871	1.10	1.02	1.19	0.010	
		90–100%	2,102	4,871	1.20	1.11	1.29	1.9×10^{-6}	

In the MVP participants of European ancestry with T2D, gPRSs for T2D were generated by calculating a linear combination of weights derived from the Europeans in the DIAMANTE Consortium, using the prune and threshold method in PRSice-2 software (pruning $r^2 = 0.8$, $P = 0.05$). The gPRSs were divided into deciles and the risk of T2D-related vascular outcomes was assessed using a logistic regression model, using the lowest decile (0–10%) as the reference category, together with the potential confounding factors of age, sex and the first ten PCs of European ancestry. The decile-specific *P* values are shown in the column labeled *P*. In a separate logistic regression analysis, the continuous PRS was set as the dependent variable together with age, sex and the first ten PCs, and the *P* value for linear trend is shown in the column labeled *P* for linear trend. For CHD, a CHD PRS (from CARDIoGRAMplusC4D and UK Biobank) is included in the regression model as an additional covariate. For AIS, a stroke PRS (from the MEGASTROKE Consortium) is included in the regression model as an additional covariate. *n* Cases, number of cases with the respective vascular outcome; *n* Controls, number of unaffected controls for the respective vascular outcome; OR, odds ratio; CI, confidence interval.

1.15 and $P = 8.5 \times 10^{-9}$). We observed that the major G-allele of chrX:153882606 (rs782270174) was associated with increased risk of T2D in African Americans. This variant is in high LD ($r^2 = 0.93$) with *G6PD* G202A (rs1050828), for which the minor allele is associated with lower HbA1c due to shorter RBC lifespan[22]. In a post hoc analysis, we examined the relationship of chrX:153882606 to the most recent value of HbA1c, obtained prior to MVP study enrollment in African American males, and did observe a strong negative association (beta = −0.072, SE = 0.0015, $n = 55{,}165$ and $P < 1.0 \times 10^{-322}$). Therefore, we cannot rule out the possibility that the apparent protective association of T2D at rs782270174 in reality reflects under-diagnosis of T2D due to reduced HbA1c in African Americans. We did not replicate the association of the *AGTR2* variant (rs146662075, chrX:115408811) as reported by Bonas-Guarch et al.[10], which might be the result of poor imputation of the 1000 Genomes reference panel for this variant.

The presence of a coding variant near a tagging SNP does not constitute enough evidence to infer a causal association. However, recent exome-array genotyping of over 350,000 individuals identified 40 coding variants associated with T2D, of which 26 mapped near known risk-associated loci[23]. Similarly, an exome sequencing study in over 40,000 participants reported 15 variants associated with T2D, of which only 2 were not previously reported by GWAS[24]. Sequencing efforts are indispensable for identifying causal variants and genes related to disease, as well as for providing insight into the contributions of ultra-rare alleles while adding to the value of array-based association studies.

Our transcriptome-wide analyses identified 804 putatively causal genes, including 54 genes that appear to be regulated by approved drugs and 687 genes that have not been reported previously. Some of these genes are already well established for T2D etiology (for example *KCNJ11*). Except for skeletal muscle, the tissues that

Table 6 | Polygenic risk scores and non-vascular outcomes

Outcome type	Outcome	T2D PRS decile	*n* Cases	*n* Controls	OR	95% CI lower	95% CI upper	*P*	*P* for linear trend
Non-vascular	Retinopathy	0–10%	792	4,533	1.00	Ref.	Ref.	–	3.1×10^{-32}
		10–20%	832	4,533	1.08	0.97	1.20	0.158	
		20–30%	795	4,533	1.05	0.94	1.17	0.364	
		30–40%	852	4,533	1.14	1.02	1.26	0.019	
		40–50%	814	4,533	1.08	0.97	1.20	0.152	
		50–60%	891	4,533	1.20	1.08	1.33	6.8×10^{-4}	
		60–70%	901	4,533	1.25	1.13	1.39	3.1×10^{-5}	
		70–80%	936	4,533	1.30	1.17	1.45	6.8×10^{-7}	
		80–90%	1,031	4,533	1.47	1.33	1.63	2.2×10^{-13}	
		90–100%	1,069	4,533	1.59	1.44	1.77	4.2×10^{-19}	
	CKD	0–10%	3,446	3,391	1.00	Ref.	Ref.	–	7.3×10^{-6}
		10–20%	3,490	3,391	1.03	0.93	1.15	0.508	
		20–30%	3,439	3,391	1.04	0.94	1.14	0.488	
		30–40%	3,463	3,391	1.05	0.95	1.16	0.323	
		40–50%	3,370	3,391	1.04	0.95	1.14	0.409	
		50–60%	3,362	3,391	1.07	0.97	1.17	0.166	
		60–70%	3,389	3,391	1.07	0.98	1.17	0.129	
		70–80%	3,285	3,391	1.07	0.98	1.17	0.121	
		80–90%	3,373	3,391	1.07	0.98	1.16	0.151	
		90–100%	3,326	3,391	1.16	1.07	1.26	5.9×10^{-4}	
	Neuropathy	0–10%	2,176	3,814	1.00	Ref.	Ref.	–	7.9×10^{-8}
		10–20%	2,193	3,814	1.03	0.96	1.11	0.436	
		20–30%	2,217	3,814	1.07	0.99	1.15	0.075	
		30–40%	2,218	3,814	1.06	0.99	1.15	0.110	
		40–50%	2,217	3,814	1.05	0.98	1.13	0.192	
		50–60%	2,293	3,814	1.11	1.03	1.20	0.006	
		60–70%	2,261	3,814	1.10	1.02	1.18	0.014	
		70–80%	2,253	3,814	1.10	1.02	1.19	0.009	
		80–90%	2,265	3,814	1.11	1.03	1.19	0.007	
		90–100%	2,377	3,814	1.21	1.12	1.30	9.7×10^{-7}	

In the MVP participants of European ancestry with T2D, gPRSs for T2D were generated by calculating a linear combination of weights derived from the Europeans in the DIAMANTE Consortium, using the prune and threshold method in PRSice-2 software (pruning $r^2 = 0.8$, $P = 0.05$). The gPRSs were divided into deciles and the risk of T2D-related non-vascular outcomes was assessed using a logistic regression model using the lowest decile (0–10%) as the reference category, together with the potential confounding factors of age, sex and the first ten PCs of European ancestry. The decile-specific *P* values are shown in the column labeled *P*. In a separate logistic regression analysis, the continuous PRS was set as the dependent variable together with age, sex and the first ten PCs, and the *P* value for linear trend is shown in the column labeled *P* for linear trend. For CKD, a CKD PRS (from the CKDgen Consortium) is included in the regression model as an additional covariate. *n* Cases, number of cases with the respective non-vascular outcome; *n* Controls, number of unaffected controls for the respective non-vascular outcome; OR, odds ratio; CI, confidence interval.

showed strongest associations are not known to be of importance in T2D etiology. However, this could be simply explained by the fact that eQTLs appear to be ubiquitous across tissues and that eQTL discovery across tissues may not be the same, given the eQTL effect sizes and sample sizes of T2D-relevant tissues. We did not observe any significant association in the islet α- and β-cells, which could be the result of the small sample size (for example 30 α-cells and 19 β-cells). In addition, whole islet transcriptomes are notoriously variable due to the large differences in islet composition among humans, and a few transcripts make up half the transcriptome[25].

Of particular clinical importance, we identified several genes that are therapeutic targets for medications in patients treated for cardiometabolic conditions. We identified two genes, *SCN3A* and *SV2A*, whose expression is modified by anti-epileptic agents, and evidence exists that anti-epileptic agents may influence glucose regulation. A randomized controlled trial has reported that the anti-convulsant valproic acid lowers blood glucose concentrations[26]. The information from the gene–drug analyses may facilitate future drug repurposing screens.

It is possible that the use of the T2D gPRS provides an opportunity to identify patients who are at the highest risk of developing microvascular complications, such as retinopathy. Here, we observed that among vascular outcomes, the T2D gPRS was most significantly associated with retinopathy. In addition, we observed significant associations with other T2D-related outcomes such as CKD, PAD and neuropathy. Studies at specific loci using both common and rare coding variants will be required to understand pathways leading to T2D-related vascular outcomes.

In a SNP–T2D interaction analysis on T2D-related vascular outcomes, we identified 13 loci at which the effect on outcome was different between the strata of T2D, of which 3 occurred at previously established variants and 10 had not been reported previously. Our findings have clinical translational potential for risk stratification and the identification of individuals with diabetes who are

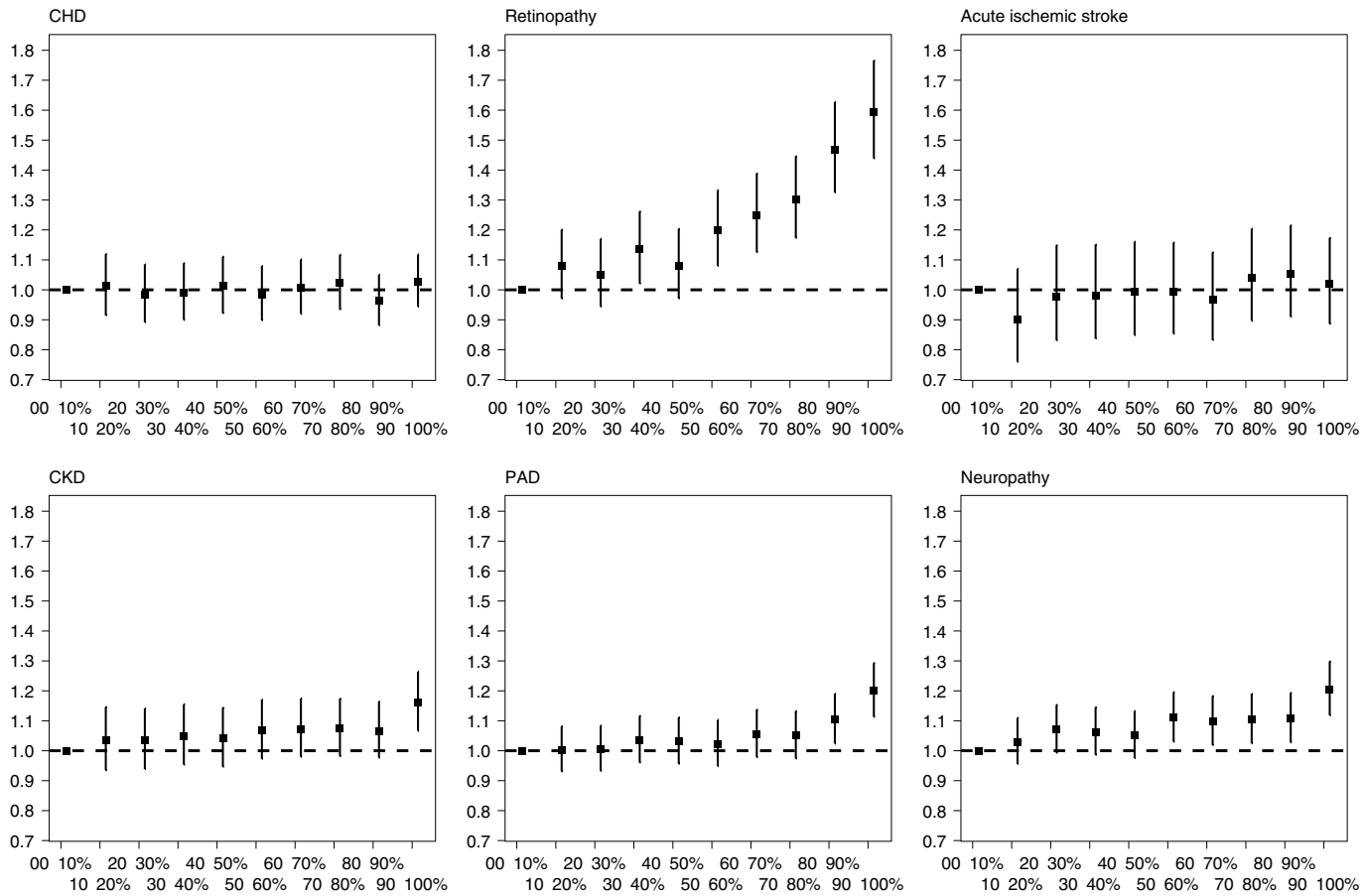


Fig. 2 | T2D genome-wide polygenic risk score is mainly predictive of microvascular outcomes. A T2D gPRS was calculated and categorized into deciles on the basis of the scores in controls. The PRS-outcome associations are shown for macrovascular outcomes (CKD: 67,403 cases and 129,827 controls; CHD: 56,285 cases and 140,945 controls; PAD: 35,882 cases and 161,348 controls) and for microvascular outcomes (AIS: 11,796 cases and 178,481 controls; retinopathy: 13,881 cases and 123,538 controls; neuropathy: 40,475 cases and 110,331 controls). Effect sizes and 95% confidence intervals are shown per decile per micro- or macrovascular outcome. For each of the complication outcomes, separate logistic regression models were fitted for people with T2D, and the models include the following independent variables: T2D PRS (from the DIAMANTE Consortium), age, sex, BMI and ten PCAs. For CHD, a CHD PRS (from CardiogramplusC4DplusUKBB) was included in the regression model as an additional covariate. For AIS, a stroke PRS (from the MEGASTROKE Consortium) was included in the regression model as an additional covariate. For CKD, a CKD PRS (from the CKDgen Consortium) was included in the regression model as an additional covariate.

predisposed to develop subsequent vascular outcomes, and present therapeutic opportunities to attenuate the risk of diabetes progression in individuals with T2D.

For T2D-related retinopathy, four variants were found to have different effect sizes between people with and without T2D. The strongest signal for interaction in relation to retinopathy was observed for *GJA8*. Deletion of this gene has been associated with eye abnormalities and retinopathy of prematurity in premature infants, inherited cataracts, visual impairment and cardiac defects and eye abnormalities[27–29]. *TCF7L2* is a known diabetes locus and its association with progression to retinopathy has been established previously[30]. *SLC18A2* is expressed in adult retina and retinal pigment epithelium tissues; the product of this gene is involved in the transport of monoamines into secretory vesicles for exocytosis[31]. *SVILP1* has been previously shown to be associated with thiamine (vitamin B1), which is frequently prescribed to people with blurry vision[32].

For CKD, we identified two loci, *UMOD* and *TENM3*, with gene–T2D interaction effects. *UMOD* encodes uromodulin, which is exclusively produced by the kidney tubule, where it plays an important role in kidney and urine function. A large-scale study in over 133,000 participants has shown that the serum creatinine-lowering allele in *UMOD* (rs12917707) is more prevalent in individuals with diabetes and with CKD as compared to participants without diabetes and without CKD[33]. Variation in *TENM3* has been associated with cholangitis and kidney disorders in the UK Biobank[34].

SNP-T2D interaction analysis of neuropathy identified one locus, *NRP2*. *NRP2* encodes neuropilin-2, which is an essential cell surface receptor involved in VEGF-dependent angiogenesis and sensory nerve regeneration.

For CHD, we identified several SNP–T2D interactions. Variation at chromosome 9p21 has previously been associated with CHD and T2D. *SORT1* is a lipid-associated locus; in our analyses, allelic variation at this locus that decreases CHD risk and decreases lipids conferred a stronger protection in people with T2D compared to those without T2D. Coupled with findings in mice that identified SORT1 as a novel target of insulin signaling, our findings raise the hypothesis that *SORT1* may contribute to altered hepatic apoB metabolism under insulin-resistant conditions.

The SNP rs71039916 is located near *PDE3A* and colocalizes with a SNP (rs3752728, $D' = 0.867$ and $r^2 = 0.08$) that is associated with diastolic blood pressure[35,36]. As a phosphodiesterase that reduces cAMP levels, the PDE3A protein limits protein kinase A/cAMP signaling and has been shown to affect proliferation of vascular

smooth muscle cells[37]. Research in cell lines has shown that cAMP levels might impact the regulation of insulin secretion in pancreatic β-cells, and more recent gene ablation studies in mice have established that cAMP/CREB signaling controls the insulinotropic and anti-apoptotic effects of GLP-1 signaling in adult mouse β-cells[38]. Subcutaneous adipose tissue of patients with T2D show increased PDE activity, and inverse correlations between total PDE3 activity and BMI have been reported in adipocytes[39].

In summary, we have identified 318 novel genetic variants associated with T2D risk and T2D-related vascular outcomes, including 3 population-specific autosomal loci in African Americans, 8 variants on the X chromosome, and an additional 13 variants associated with differences in T2D-related micro- and macrovascular outcomes across diabetic strata. Over 21% of our discovery sample comprised non-European participants; indeed, the African American component alone included over 56,000 subjects. We hope this baseline data set will provide a resource to better understand the genetic etiology of disease and maximize the benefits of polygenic risk prediction in these groups.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41588-020-0637-y.

Received: 16 November 2019; Accepted: 29 April 2020;
Published online: 15 June 2020

References

1. International Diabetes Federation. *IDF Diabetes Atlas* 8th edn (International Diabetes Federation, 2017).
2. American Diabetes Association Standards of medical care in diabetes—2018. *Diabetes Care* **41**, S1–S2 (2018).
3. Mahajan, A. et al. Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. *Nat. Genet.* **50**, 1505–1513 (2018).
4. Suzuki, K. et al. Identification of 28 new susceptibility loci for type 2 diabetes in the Japanese population. *Nat. Genet.* **51**, 379–386 (2019).
5. Gaziano, J. M. et al. Million Veteran Program: a mega-biobank to study genetic influences on health and disease. *J. Clin. Epidemiol.* **70**, 214–223 (2016).
6. Levin, M. G. et al. Genomic risk stratification predicts all-cause mortality after cardiac catheterization. *Circ. Genom. Precis. Med.* **11**, e002352 (2018).
7. Saleheen, D. et al. Human knockouts and phenotypic analysis in a cohort with a high rate of consanguinity. *Nature* **544**, 235–239 (2017).
8. Berglund, G., Elmstahl, S., Janzon, L. & Larsson, S. A. The Malmo Diet and Cancer Study. Design and feasibility. *J. Intern. Med.* **233**, 45–51 (1993).
9. Reilly, M. P. et al. Identification of *ADAMTS7* as a novel locus for coronary atherosclerosis and association of *ABO* with myocardial infarction in the presence of coronary atherosclerosis: two genome-wide association studies. *Lancet* **377**, 383–392 (2011).
10. Bonas-Guarch, S. et al. Re-analysis of public genetic data reveals a rare X-chromosomal variant associated with type 2 diabetes. *Nat. Commun.* **9**, 321 (2018).
11. Xue, A. et al. Genome-wide association analyses identify 143 risk variants and putative regulatory mechanisms for type 2 diabetes. *Nat. Commun.* **9**, 2941 (2018).
12. Bulik-Sullivan, B. K. et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
13. Devlin, B. & Roeder, K. Genomic control for association studies. *Biometrics* **55**, 997–1004 (1999).
14. Luo, Y. et al. Estimating heritability of complex traits in admixed populations with summary statistics. Preprint at *bioRxiv* https://doi.org/10.1101/503144 (2018).
15. Klarin, D. et al. Genetics of blood lipids among ~300,000 multi-ethnic participants of the Million Veteran Program. *Nat. Genet.* **50**, 1514–1523 (2018).
16. Kuhn, M., Letunic, I., Jensen, L. J. & Bork, P. The SIDER database of drugs and side effects. *Nucleic Acids Res.* **44**, D1075–D1079 (2016).
17. Finucane, H. K. et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015).
18. Schmidt, E. M. et al. GREGOR: evaluating global enrichment of trait-associated variants in epigenomic features using a systematic, data-driven approach. *Bioinformatics* **31**, 2601–2606 (2015).
19. Ng, M. C. et al. Meta-analysis of genome-wide association studies in African Americans provides insights into the genetic architecture of type 2 diabetes. *PLoS Genet.* **10**, e1004517 (2014).
20. Chen, J. et al. Genome-wide association study of type 2 diabetes in Africa. *Diabetologia* **62**, 1204–1211 (2019).
21. Palmer, N. D. et al. A genome-wide association search for type 2 diabetes genes in African Americans. *PLoS ONE* **7**, e29202 (2012).
22. Wheeler, E. et al. Impact of common genetic determinants of Hemoglobin A1c on type 2 diabetes risk and diagnosis in ancestrally diverse populations: a transethnic genome-wide meta-analysis. *PLoS Med.* **14**, e1002383 (2017).
23. Mahajan, A. et al. Refining the accuracy of validated target identification through coding variant fine-mapping in type 2 diabetes. *Nat. Genet.* **50**, 559–571 (2018).
24. Flannick, J. et al. Exome sequencing of 20,791 cases of type 2 diabetes and 24,440 controls. *Nature* **570**, 71–76 (2019).
25. Carrano, A. C., Mulas, F., Zeng, C. & Sander, M. Interrogating islets in health and disease with single-cell technologies. *Mol. Metab.* **6**, 991–1001 (2017).
26. Martin, C. K., Han, H., Anton, S. D., Greenway, F. L. & Smith, S. R. Effect of valproic acid on body weight, food intake, physical activity and hormones: results of a randomized controlled trial. *J. Psychopharmacol.* **23**, 814–825 (2009).
27. Buse, M. et al. Expanding the phenotype of reciprocal 1q21.1 deletions and duplications: a case series. *Ital. J. Pediatr.* **43**, 61 (2017).
28. Devi, R. R. & Vijayalakshmi, P. Novel mutations in *GJA8* associated with autosomal dominant congenital cataract and microcornea. *Mol. Vis.* **12**, 190–195 (2006).
29. Mackay, D. S., Bennett, T. M., Culican, S. M. & Shiels, A. Exome sequencing identifies novel and recurrent mutations in *GJA8* and *CRYGD* associated with inherited cataract. *Hum. Genomics* **8**, 19 (2014).
30. Luo, J. et al. *TCF7L2* variation and proliferative diabetic retinopathy. *Diabetes* **62**, 2613–2617 (2013).
31. Eiden, L. E., Schafer, M. K., Weihe, E. & Schutz, B. The vesicular amine transporter family (SLC18): amine/proton antiporters required for vesicular accumulation and regulated exocytotic secretion of monoamines and acetylcholine. *Pflugers Arch.* **447**, 636–640 (2004).
32. Sharma, P. & Sharma, R. Toxic optic neuropathy. *Indian J. Ophthalmol* **59**, 137–141 (2011).
33. Pattaro, C. et al. Genetic associations at 53 loci highlight cell types and biological pathways relevant for kidney function. *Nat. Commun.* **7**, 10023 (2016).
34. Canela-Xandri, O., Rawlik, K. & Tenesa, A. An atlas of genetic associations in UK Biobank. *Nat. Genet.* **50**, 1593–1599 (2018).
35. Ehret, G. B. et al. The genetics of blood pressure regulation and its target organs from association studies in 342,415 individuals. *Nat. Genet.* **48**, 1171–1184 (2016).
36. Sung, Y. J. et al. A large-scale multi-ancestry genome-wide study accounting for smoking behavior identifies multiple significant loci for blood pressure. *Am. J. Hum. Genet.* **102**, 375–400 (2018).
37. Maass, P. G. et al. *PDE3A* mutations cause autosomal dominant hypertension with brachydactyly. *Nat. Genet.* **47**, 647–653 (2015).
38. Shin, S. et al. CREB mediates the insulinotropic and anti-apoptotic effects of GLP-1 signaling in adult mouse β-cells. *Mol. Metab* **3**, 803–812 (2014).
39. Omar, B., Banke, E., Ekelund, M., Frederiksen, S. & Degerman, E. Alterations in cyclic nucleotide phosphodiesterase activities in omental and subcutaneous adipose tissues in human obesity. *Nutr. Diabetes* **1**, e13 (2011).

[1]Corporal Michael J. Crescenz VA Medical Center, Philadelphia, PA, USA. [2]Department of Medicine, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA. [3]Biomedical Laboratory Research and Development, Tennessee Valley Healthcare System, Nashville, TN, USA. [4]Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Nashville, TN, USA. [5]Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA. [6]Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN, USA. [7]VA Informatics and Computing

Infrastructure, VA Salt Lake City Health Care System, Salt Lake City, UT, USA. [8]College of Nursing and Health Sciences, University of Massachusetts, Lowell, MA, USA. [9]Edith Nourse Rogers Memorial VA Hospital, Bedford, MA, USA. [10]Center for Population Health, University of Massachusetts, Lowell, MA, USA. [11]Phoenix VA Health Care System, Phoenix, AZ, USA. [12]Mel and Enid Zuckerman College of Public Health, University of Arizona, Tucson, AZ, USA. [13]VA Palo Alto Health Care System, Palo Alto, CA, USA. [14]Department of Medicine, Stanford University School of Medicine, Stanford, CA, USA. [15]Department of Pediatric Cardiology, Stanford University School of Medicine, Stanford, CA, USA. [16]VA Boston Healthcare System, Boston, MA, USA. [17]Department of Genetics, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA. [18]Department of Systems Pharmacology and Translational Therapeutics, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA. [19]Department of Statistics, Stanford University, Stanford, CA, USA. [20]Department of Surgery, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA. [21]Department of Global Health, Peking University School of Public Health, Beijing, China. [22]Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA. [23]Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, USA. [24]Division of Vascular Surgery and Endovascular Therapy, University of Florida School of Medicine, Gainesville, FL, USA. [25]Department of Medicine, Brigham Women's Hospital, Boston, MA, USA. [26]Department of Medicine, Harvard Medical School, Boston, MA, USA. [27]Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK. [28]Department of Clinical Sciences Malmö, Lund University, Malmö, Sweden. [29]Center for Non-Communicable Diseases, Karachi, Sindh, Pakistan. [30]Punjab Institute of Cardiology, Lahore, Punjab, Pakistan. [31]Department of Medicine, King Edward Medical University, Lahore, Punjab, Pakistan. [32]Mayo Hospital, Lahore, Punjab, Pakistan. [33]Department of Cardiology, Faisalabad Institute of Cardiology, Faisalabad, Punjab, Pakistan. [34]Atlanta VA Medical Center, Decatur, GA, USA. [35]Department of Epidemiology, Emory University Rollins School of Public Health, Atlanta, GA, USA. [36]Department of Internal Medicine, University of Utah School of Medicine, Salt Lake City, UT, USA. [37]Division of Endocrinology, Emory University School of Medicine, Atlanta, GA, USA. [38]Division of General Internal Medicine, Massachusetts General Hospital, Boston, MA, USA. [39]College of Medicine, University of Arizona, Phoenix, AZ, USA. [40]Division of Cardiology, Emory University School of Medicine, Atlanta, GA, USA. [41]Nashville VA Medical Center, Nashville, TN, USA. [42]South Texas Veterans Health Care System, San Antonio, TX, USA. [43]Department of Medicine, Columbia University Irving Medical Center, New York, NY, USA. [44]Department of Cardiology, Columbia University Irving Medical Center, New York, NY, USA. [112]These authors contributed equally: Marijana Vujkovic, Jacob M. Keaton, Kyong-Mi Chang, Benjamin F. Voight, Danish Saleheen. *A full list of authors and their affiliations appears at the end of the paper. ✉e-mail: bvoight@pennmedicine.upenn.edu; ds3792@cumc.columbia.edu

The HPAP Consortium

Mark A. Atkinson[45,46], Al C. Powers[47,48,49], Ali Naji[20] and Klaus H. Kaestner[17]

[45]Department of Pathology, University of Florida Diabetes Institute, Gainesville, FL, USA. [46]Department of Pediatrics, University of Florida Diabetes Institute, Gainesville, FL, USA. [47]Department of Molecular Physiology and Biophysics, Vanderbilt University, Nashville, TN, USA. [48]Division of Diabetes, Endocrinology, and Metabolism, Vanderbilt University Medical Center, Nashville, TN, USA. [49]VA Tennessee Valley Healthcare System, Nashville, TN, USA.

Regeneron Genetics Center

Goncalo R. Abecasis[50], Aris Baras[50], Michael N. Cantor[50], Giovanni Coppola[50], Aris N. Economides[50], Luca A. Lotta[50], John D. Overton[50], Jeffrey G. Reid[50], Alan R. Shuldiner[50], Christina Beechert[50], Caitlin Forsythe[50], Erin D. Fuller[50], Zhenhua Gu[50], Michael Lattari[50], Alexander E. Lopez[50], Thomas D. Schleicher[50], Maria Sotiropoulos Padilla[50], Karina Toledo[50], Louis Widom[50], Sarah E. Wolf[50], Manasi Pradhan[50], Kia Manoochehri[50], Ricardo H. Ulloa[50], Xiaodong Bai[50], Suganthi Balasubramanian[50], Leland Barnard[50], Andrew L. Blumenfeld[50], Gisu Eom[50], Lukas Habegger[50], Alicia Hawes[50], Shareef Khalid[50], Evan K. Maxwell[50], William J. Salerno[50], Jeffrey C. Staples[50], Ashish Yadav[50], Marcus B. Jones[50] and Lyndon J. Mitnaul[50]

[50]Regeneron Pharmaceuticals, Inc., Tarrytown, NY, USA.

VA Million Veteran Program

Samuel M. Aguayo[11], Sunil K. Ahuja[42], Zuhair K. Ballas[51], Sujata Bhushan[52], Edward J. Boyko[53], David M. Cohen[54], John Concato[55], Joseph I. Constans[56], Louis J. Dellitalia[57], Joseph M. Fayad[58], Ronald S. Fernando[59], Hermes J. Florez[60], Melinda A. Gaddy[61], Saib S. Gappy[62], Gretchen Gibson[63], Michael Godschalk[64], Jennifer A. Greco[65], Samir Gupta[66], Salvador Gutierrez[67], Kimberly D. Hammer[68], Mark B. Hamner[69], John B. Harley[70], Adriana M. Hung[49], Mostaqul Huq[71], Robin A. Hurley[72], Pran R. Iruvanti[73], Douglas J. Ivins[74], Frank J. Jacono[75], Darshana N. Jhala[76], Laurence S. Kaminsky[77], Scott Kinlay[16], Jon B. Klein[78], Suthat Liangpunsakul[79],

Jack H. Lichy[80], Stephen M. Mastorides[81], Roy O. Mathew[82], Kristin M. Mattocks[83], Rachel McArdle[84], Paul N. Meyer[85], Laurence J. Meyer[7], Jonathan P. Moorman[86], Timothy R. Morgan[87], Maureen Murdoch[88], Xuan-Mai T. Nguyen[16], Olaoluwa O. Okusaga[89], Kris-Ann K. Oursler[90], Nora R. Ratcliffe[91], Michael I. Rauchman[92], R. Brooks Robey[93], George W. Ross[94], Richard J. Servatius[95], Satish C. Sharma[96], Scott E. Sherman[97], Elif Sonel[98], Peruvemba Sriram[99], Todd Stapley[100], Robert T. Striker[101], Neeraj Tandon[102], Gerardo Villareal[103], Agnes S. Wallbom[104], John M. Wells[9], Jeffrey C. Whittle[105], Mary A. Whooley[106], Junzhe Xu[107], Shing-Shing Yeh[108], Michaela Aslan[55], Jessica V. Brewer[16], Mary T. Brophy[16], Todd Connor[109], Dean P. Argyres[109], Nhan V. Do[16], Elizabeth R. Hauser[110], Donald E. Humphries[16], Luis E. Selva[16], Shahpoor Shayan[16], Brady Stephens[111], Stacey B. Whitbourne[16], Hongyu Zhao[55], Jennifer Moser[80], Jean C. Beckham[110], Jim L. Breeling[16], J. P. Casas Romero[16], Grant D. Huang[80], Rachel B. Ramoni[16], Saiju Pyarajan[16,25,26], Yan V. Sun[34,35], Kelly Cho[16,25], Peter W. Wilson[34,40], Christopher J. O'Donnell[16,25,26], Philip S. Tsao[13,14], Kyong-Mi Chang[1,2], J. Michael Gaziano[16,25] and Sumitra Muralidhar[80]

[51]Iowa City VA Health Care System, Iowa City, IA, USA. [52]VA North Texas Health Care System, Dallas, TX, USA. [53]VA Puget Sound Health Care System, Seattle, WA, USA. [54]Portland VA Medical Center, Portland, OR, USA. [55]VA Connecticut Healthcare System, West Haven, CT, USA. [56]Southeast Louisiana Veterans Health Care System, New Orleans, LA, USA. [57]Birmingham VA Medical Center, Birmingham, AL, USA. [58]VA Southern Nevada Healthcare System, North Las Vegas, NV, USA. [59]VA Loma Linda Healthcare System, Loma Linda, CA, USA. [60]Miami VA Health Care System, Miami, FL, USA. [61]VA Eastern Kansas Health Care System, Leavenworth, KS, USA. [62]John D. Dingell VA Medical Center, Detroit, MI, USA. [63]Fayetteville VA Medical Center, Fayetteville, AR, USA. [64]Richmond VA Medical Center, Richmond, VA, USA. [65]Sioux Falls VA Health Care System, Sioux Falls, SD, USA. [66]VA San Diego Healthcare System, San Diego, CA, USA. [67]Edward Hines Jr. VA Medical Center, Hines, IL, USA. [68]Fargo VA Health Care System, Fargo, ND, USA. [69]Ralph H. Johnson VA Medical Center, Charleston, SC, USA. [70]Cincinnati VA Medical Center, Cincinnati, OH, USA. [71]VA Sierra Nevada Health Care System, Reno, NV, USA. [72]W.G. (Bill) Hefner VA Medical Center, Salisbury, NC, USA. [73]Hampton VA Medical Center, Hampton, VA, USA. [74]Eastern Oklahoma VA Health Care System, Muskogee, OK, USA. [75]VA Northeast Ohio Healthcare System, Cleveland, OH, USA. [76]Philadelphia VA Medical Center, Philadelphia, PA, USA. [77]VA Health Care Upstate New York, Albany, NY, USA. [78]Louisville VA Medical Center, Louisville, KY, USA. [79]Richard Roudebush VA Medical Center, Indianapolis, IN, USA. [80]Washington DC VA Medical Center, Washington DC, USA. [81]James A. Haley Veterans Hospital, Tampa, FL, USA. [82]Columbia VA Health Care System, Columbia, SC, USA. [83]Central Western Massachusetts Healthcare System, Leeds, MA, USA. [84]Bay Pines VA Healthcare System, Bay Pines, FL, USA. [85]Southern Arizona VA Health Care System, Tucson, AZ, USA. [86]James H. Quillen VA Medical Center, Johnson City, TN, USA. [87]VA Long Beach Healthcare System, Long Beach, CA, USA. [88]Minneapolis VA Health Care System, Minneapolis, MN, USA. [89]Michael E. DeBakey VA Medical Center, Houston, TX, USA. [90]Salem VA Medical Center, Salem, VA, USA. [91]Manchester VA Medical Center, Manchester, NH, USA. [92]St. Louis VA Health Care System, St. Louis, MO, USA. [93]White River Junction VA Medical Center, White River Junction, VT, USA. [94]VA Pacific Islands Health Care System, Honolulu, HI, USA. [95]Syracuse VA Medical Center, Syracuse, NY, USA. [96]Providence VA Medical Center, Providence, RI, USA. [97]VA New York Harbor Healthcare System, New York, NY, USA. [98]VA Pittsburgh Health Care System, Pittsburgh, PA, USA. [99]North Florida/South Georgia Veterans Health System, Gainesville, FL, USA. [100]VA Maine Healthcare System, Augusta, ME, USA. [101]William S. Middleton Memorial Veterans Hospital, Madison, WI, USA. [102]Overton Brooks VA Medical Center, Shreveport, LA, USA. [103]New Mexico VA Health Care System, Albuquerque, NM, USA. [104]VA Greater Los Angeles Health Care System, Los Angeles, CA, USA. [105]Clement J. Zablocki VA Medical Center, Milwaukee, WI, USA. [106]San Francisco VA Health Care System, San Francisco, CA, USA. [107]VA Western New York Healthcare System, Buffalo, NY, USA. [108]Northport VA Medical Center, Northport, NY, USA. [109]Raymond G. Murphy VA Medical Center, Albuquerque, NM, USA. [110]Durham VA Medical Center, Durham, NC, USA. [111]Canandaigua VA Medical Center, Canandaigua, NY, USA.

Methods

Overview. We conducted a large-scale multi-ancestry T2D GWAS of common variants in over 1.4 million participants. We subsequently conducted analyses to facilitate the prioritization of these individual findings, including transcriptome-wide predicted gene expression, secondary signal analysis, T2D-related vascular outcomes analysis, coding variant mapping and a drug repurposing screen.

Discovery cohort. The MVP is a large cohort of fully consented veterans of the US military forces recruited from 63 participating Department of Veterans Affairs (VA) medical facilities[5]. Recruitment started in 2011, and all veterans were eligible for participation (Supplementary Table 3). We analyzed clinical data to July 2017 for participants who enrolled between January 2011 and October 2016. All study participants provided blood samples for DNA extraction and genotyping, and completed surveys about their health, lifestyle, and military experiences. Consent to participate and permission to re-contact was provided after counseling by research staff and mailing of informational materials. Study participation included providing consent to access the electronic health records of the participant for research purposes and data that captured a median follow-up time of 10 years at time of study enrollment. The electronic health care record of each veteran was integrated into the MVP biorepository, and included inpatient International Classification of Diseases (ICD-9-CM and ICD-10-CM) diagnosis codes, Current Procedural Terminology (CPT) procedure codes, clinical laboratory measurements and reports of diagnostic imaging modalities. Researchers were provided data that was de-identified except for dates. Blood samples were collected by phlebotomists and banked at the VA Central Biorepository in Boston, where DNA was extracted and shipped to two external centers for genotyping. The MVP received ethical and study protocol approval from the VA Central Institutional Review Board (cIRB) in accordance with the principles outlined in the Declaration of Helsinki.

Genotyping. DNA extracted from buffy coat was genotyped using a custom Affymetrix Axiom biobank array. The MVP 1.0 genotyping array contains a total of 723,305 SNPs, enriched for low frequency variants in African and Hispanic populations, and variants associated with diseases common to the VA population[5].

Genotype quality control. Standard quality control and genotype calling algorithms were applied using the Affymetrix Power Tools Suite (v1.18). Duplicate samples, samples with more heterozygosity than expected and samples with over 2.5% of missing genotype calls were excluded. We excluded related individuals (halfway between second- and third-degree relatives or closer) with KING software[40]. Before imputation, variants that were poorly called or that deviated from their expected allele frequency based on reference data from the 1000 Genomes Project were excluded[41]. After prephasing using EAGLE v2, genotypes were imputed via Minimac4 software[42] from the 1000 Genomes Project phase 3, version 5 reference panel. The top 30 principal components (PCs) were computed using FlashPCA in all MVP participants and an additional 2,504 individuals from 1000 Genomes. These PCs were used for the unification of self-reported race/ancestry and genetically inferred ancestry to compose ancestral groups[43].

Race and ancestry. Information on race and ancestry was obtained from self-reported data through centralized VA data collection methods using standardized survey forms, or through the use of information from the VA Corporate Data Warehouse or Observational Medical Outcomes Partnership data. Self-reported race/ancestry was missing in 3.67% of participants, and 39.4% of participants had some form of discordant information between the various data sources. Race and ancestry categories were merged to form the ancestral groups using a unifying classification algorithm based on self-identified race/ethnicity and genetically inferred ancestral information, termed HARE (Harmonized Ancestry and Race/Ethnicity)[43]. Using this approach, all but 6,257 (1.78%) participants were assigned to one of the four ancestral groups.

Phenotype classification. ICD-9-CM diagnosis codes from electronic health care records were available for MVP participants from as early as 1998. Participants were classified as a T2D case if they had 2 or more T2D-related diagnosis codes (ICD-9-CM 250.2x) from VA or fee basis inpatient stays or face-to-face primary care outpatient visits in the 731 days before the enrollment date up to 1 July 2017, excluding those with co-occurring diagnosis codes for T1D (250.1x), secondary or other diabetes, or a medical condition that may cause diabetes (249.xx). Participants were selected as controls if they had no ICD-9-CM diagnosis code for type 1, type 2 or secondary diabetes mellitus up to July 2017.

For T2D-related vascular outcomes, the following definitions were used. CHD was defined as at least one admission to a VA hospital with discharge diagnosis of admission for myocardial information, or at least one procedure code for revascularization (coronary artery bypass grafting or percutaneous coronary intervention), or at least 2 ICD-9-CM codes for CAD (410 to 414) registered on at least 2 separate encounters. PAD: was defined as the presence of 2 or more ICD-9-CM codes or CPT codes as outlined in Klarin et al.[15], or having 1 code and 2 or more visits to a vascular surgeon within a 14-month period. AIS was defined as the presence of at least 1 ICD-9-CM discharge diagnosis code for stroke excluding head injury or rehab (433.x1, 434 (excluding 434.x0), and 436)[44]; CKD was classified as an estimated glomerular filtration rate of <60 ml min^{-1} per 1.73 m^2 on two separate occasions 90 days apart, or ICD-9-CM diagnosis codes for chronic renal failure (585) and/or a history of kidney transplantation (ICD-9-CM code V42). Neuropathy was defined using the following ICD-9-CM diagnosis codes: diabetic neuropathy (356.9 or 250.6), amyotrophy (358.1), cranial nerve palsy (951.0, 951.1 or 951.3), mono-neuropathy (354.0–355.9), Charcot's arthropathy (713.5), polyneuropathy (357.2), neurogenic bladder (596.54), autonomic neuropathy (337.0 or 337.1) or orthostatic hypotension (458). Retinopathy was defined using the ICD-9-DM diagnosis codes for: T2D with ophthalmic manifestations (250.50 or 250.52), retinal detachments and defects (361.0 or 361.1), disorders of vitreous body (379.2), other retinal disorders (362.0, 362.1, 362.3, 362.81, 362.83 or 362.84), excluding ICD-9-CM codes associated with macular degeneration (362.5).

MVP analysis. We tested imputed SNPs that passed quality control (for example, HWE $> 1.0 \times 10^{-10}$, INFO > 0.3 and call rate > 0.975) for association with T2D through logistic regression assuming an additive model of variants with MAF $> 0.1\%$ in Europeans, and MAF $> 1\%$ in African Americans, Hispanics and Asians using PLINK2a[45]. Covariates included age, sex and ten PCs of genetic ancestry.

Meta-analysis. Summary statistics available from previously published T2D GWAS studies were obtained for meta-analysis (Supplementary Table 2). All cohorts were imputed using the 1000 Genomes Project phase 3, v5 reference panel, with the exception of the DIAMANTE Consortium, where genotype calls were imputed using the Haplotype Reference Consortium reference panel. Only SNPs with ancestry-specific MAF $> 1\%$ in these studies were used. Ancestry-specific and multi-ancestry meta-analysis were performed using in a fixed-effects model using METAL with inverse-variance weighting of log odds ratios[46]. Between-study allelic effect size heterogeneity was assessed with Cochran's Q statistic as implemented in METAL. Variants were considered to be genome-wide significant if they passed the conventional P value threshold of 5×10^{-8}. We excluded variants with a high amount of heterogeneity (I^2 statistic $> 75\%$) across the ancestral groups.

X-chromosome analyses. X-chromosome genotypes were processed separately. During prephasing and imputation an additional flag of -chrX was added. Post-imputation XWAS quality control included first removing variants in pseudo-autosomal regions, second, removing those not in HWE in females ($P > 1.0 \times 10^{-6}$) and third, removing those with differential allele frequencies or differential missingness ($P < 10^{-7}$) between male and female controls (Extended Data Fig. 2)[47]. For each ancestry-specific subset, we performed sex-stratified analysis where dosages (the number of X-chromosome copies) in individuals with T2D were equivalent to controls within each sex stratum. The ancestry-restricted sex-stratified X-chromosome analyses were first meta-analyzed into a multi-ancestry sex-stratified analysis. Then, the multi-ancestry results from males and multi-ancestry results from females were meta-analyzed, in which none of the analyzed variants was detected using the Cochran test for heterogeneity ($P < 5 \times 10^{-8}$). Results are presented in Table 2 and Supplementary Table 9.

Secondary signal analysis. GCTA software was used to conduct approximate conditional analyses to detect ancestry-specific distinct association signals at each of the lead SNPs. Race-stratified MVP cohorts (197,066 Europeans and 53,445 African Americans) were used to model LD patterns between variants as a reference panel. For each lead SNP, conditionally independent variants that reached locus-wide significance ($P < 1.0 \times 10^{-5}$) were considered as secondary signals of distinct association. If the minimum distance between any distinct signals from two separate loci was less than 500 kb, we performed additional conditional analysis including both regions and reassessed the independence of each signal. Finally, the predicted conditionally independent variants were tested in a logistic regression model in the MVP study only to empirically validate the signal, and results are shown in Supplementary Tables 11 and 12.

Coding variant mapping. All imputed variants in MVP were evaluated with the Ensemble Variant Effect Predictor, and predicted LOF and missense variants were extracted. LD was calculated with established variants, and the effect of the missense variant was calculated conditioning on the lead SNP to assess how much residual variance was explained by the SNP in T2D risk. A P value of 0.05 was considered to be statistically significant.

S-PrediXcan and colocalization analyses. Genetically predicted gene expression and its association with T2D risk was estimated using S-PrediXcan. Input included meta-analyzed summary statistics from the European T2D GWAS and reference eQTL summary statistics for 52 tissues including 48 tissues from GTEx, 2 cell types in kidney tissue (glomerulus and tubulus)[48], and 2 cell types in pancreatic islet tissue (α- and β-cells)[49]. Analyses incorporated genotype covariance matrices based on the 1000 Genomes European populations to account for LD structure. Colocalization analysis was performed to address the issue of LD contamination in S-PrediXcan analyses. The output is shown in Supplementary Table 15.

Polygenicity and population stratification. LDSC[12] was used to calculate population-specific LD scores in Europeans and Asians using SNPs selected from HapMap[50] after SNPs with INFO < 0.95 and SNPs in the major histocompatibility complex region were excluded. Of note, LDSC is likely to be biased in admixed populations, and therefore an admixture-adjusted LDSC was used in African Americans and Hispanics[14].

Tissue- and epigenetic-specific enrichment of T2D heritability. We analyzed cell-type-specific annotations to identify enrichments of T2D heritability. First, a baseline gene model was generated, which consisted of 53 functional categories including UCSC gene models, ENCODE functional annotations[51], Roadmap epigenomic annotations[52] and FANTOM5 enhancers[53]. Gene expression and chromatin data were also analyzed to identify disease-relevant tissues, cell types, and tissue-specific epigenetic annotations. We used LDSC[12,17,54] to test for enriched heritability in regions surrounding genes with the highest tissue-specific expression. Sources of data that were analyzed included 53 human tissue or cell-type RNA-seq data from GTEx; 152 human, mouse, or rat tissue or cell-type array data from the Franke lab[55]; 3 sets of mouse brain-cell-type array data from Cahoy et al.[56]; 292 mouse immune-cell-type array data from ImmGen[57] and 396 human epigenetic annotations from the Roadmap Epigenomics Consortium[52]. We tested for epigenomic enrichment of genetic variants using GREGOR[18]. We tested for enrichment of 2,747 genomic features selected by the T2D lead variants with $P < 5 \times 10^{-8}$, or their LD proxies ($r^2 > 0.7$) relative to control variants. Enrichment was considered significant if the enrichment P value was less than the Bonferroni-corrected threshold of 1.8×10^{-5} (nominal $P = 0.05$ of 2725 non-zero tested sites). Consortia annotations were obtained and processed as follows. Data from the consolidated epigenomes section of the Roadmap Epigenomics Project portal[52] were downloaded on 10 February 2016. All ENCODE consortium[51] data were downloaded on 6 January 2016 from the ENCODE project portal by limiting to *Homo sapiens* samples and selecting the named assay, except for the Uniform DNase files, which were downloaded on 28 March 2016. We used the FAIRE-seq ENCODE data, transcription profiling array data, ChIP–seq files and histone data. The complete list of 2,305 ENCODE and Roadmap Epigenomics features used is provided in Supplementary Table 20. In addition we performed a literature search on PubMed and in the GEO data archive, which focused on the five tissues that were most likely to be involved in T2D etiology: pancreas, liver, adipose, muscle and intestine. Most searches were performed from 15 August 2016 to 29 September 2016 and we identified a total of 442 features across 42 publications (Supplementary Table 21).

Phenome-wide association analysis. For the three LOF variants that were identified using coding variant analysis, we performed a PheWAS to fully leverage the diverse nature of MVP as well as the full catalog of relevant ICD-9-CM diagnosis and CPT procedure codes (Table 5). Of genotyped veterans, participants were included in the PheWAS if their respective electronic health record reflected two or more separate encounters in the VA Healthcare System in each of the two years prior to enrollment in the MVP. A total of 277,531 veterans spanning 21,209,658 available ICD-9-CM diagnosis codes were available. We restricted our analysis to the subgroup of 197,066 European participants. Diagnosis and procedure codes were collapsed to clinical disease groups and corresponding controls using predefined groupings[58]. Phenotypes were required to have a case count over 25 in order to be included in the PheWAS, and a multiple testing threshold for statistical significance was set to $P < 2.8 \times 10^{-5}$ (Bonferroni method). Each of the previously unpublished LOF variants were tested using logistic regression adjusting for age, sex and ten PCs in an additive effects model using the PheWAS R package in R v3.2.0. The results from these analyses are shown in Table 3 (Extended Data Fig. 4).

Analysis of T2D-related outcomes. Genetic data on European participants was separately analyzed using vascular outcomes as a binary outcome, and T2D as an interaction variable with SNPs. Interaction analysis with robust variance was applied to reduce the effect of heteroscedasticity[59] using SUGEN software (v8.8)[60]. We evaluated the interaction between SNPs and presence of T2D status using an interaction term for the two independent variables. Due to the binary nature of the outcome, the standard output from the interaction effect estimate was interpreted on a multiplicative scale. To obtain interaction on an additive scale, we calculated the relative excess risk due to interaction (RERI) metric. In case–control studies using the linear additive odds ratio model, as proposed by Richardson and Kaufman, in our study takes the form:

$$\text{Odds} = e^{\beta_0}(1 + \beta_1 \times \text{SNP} + \beta_2 \times \text{T2D} + \beta_3 \times \text{SNP} \times \text{T2D})$$

in which the coefficient β_3 measures the departure from additivity of exposure effect on an odds ratio scale; that is,

$$\text{RERI}_{\text{OR}} = \beta_3 = \text{OR}(\text{SNP} \times \text{T2D}) - \text{OR}(\text{T2D}) - \text{OR}(\text{SNP}) + 1$$

We performed analysis using a linear odds model to quantify the excess odds per unit of the given explanatory variables on the outcome. In this model, RERI is an estimate of the excess odds on a linear scale due to the interaction between two explanatory variables. In the SNP × T2D interaction analysis, we used a significance threshold of $P < 5 \times 10^{-8}$ to denote variants that had statistically different effect sizes. An additional filter was applied, and variants for which the effect size in at least one of the two T2D strata was nominally significant at $P < 0.001$ were included. Manhattan plots, as shown in Extended Data Fig. 4, Table 4 and Supplementary Table 26, represent the interaction coefficients on this scale.

Polygenic risk scores and risk of T2D and related outcomes. We constructed a gPRS for T2D in the MVP participants of European ancestry by calculating a linear combination of weights derived from the Europeans in the DIAMANTE Consortium[3] using the prune and threshold method in PRSice-2 software. After an initial sensitivity analysis, the r^2 threshold for pruning was set to 0.8, and the P value for significance threshold was set to 0.05. The gPRSs were divided into deciles and the risk of T2D was assessed using a logistic regression model using the lowest decile as a reference, together with the potential confounding factors of age, sex, BMI and the first ten PCs. An additional outcomes analysis was performed to evaluate to what extent a T2D gPRS is predictive of T2D-induced morbidities. The data set was restricted to participants with T2D, and stratum-restricted T2D gPRS deciles were generated. Logistic regression models were applied where the micro- and macrovascular conditions were modeled as outcomes, and independent variables included strata-restricted gPRS deciles, age, sex and the first ten PCs of European ancestry. The data were visualized using shape plots.

Heritability estimates and genetic correlations with other complex traits and diseases. LDSC was used to estimate the heritability coefficient, and subsequently population and sample prevalence estimates were applied to estimate heritability on the liability scale[61]. A genome-wide genetic correlation analysis was performed to investigate possible coregulation or a shared genetic basis between T2D and other complex traits and diseases. Pairwise genetic correlation coefficients were estimated between the meta-analyzed T2D GWAS summary output in Europeans and each of 774 precomputed and publicly available GWAS summary statistics for complex traits and diseases by using LDSC through LD Hub v1.9.3 (http://ldsc.broadinstitute.org). Statistical significance was set to a Bonferroni-corrected level of $P < 6.5 \times 10^{-5}$.

Enrichment and pathway analyses. Tissue enrichment for S-PrediXcan results was evaluated by calculating exact P values for under- or over-enrichment based on the cumulative distribution function of the hypergeometric distribution. The Bonferroni-corrected threshold for significance was $P < 0.001$ considering evaluation of 52 tissues. Enrichment analyses in DEPICT[62] were conducted using lead T2D SNPs. DEPICT is based on predefined phenotypic gene sets from multiple databases and Affymetrix HGU133a2.0 expression microarray data from over 37,000 subjects to build highly expressed gene sets for MeSH tissue and cell-type annotations. Output includes a P value for enrichment and a yes/no indicator of whether the false discovery rate q value is significant ($P < 0.05$).

Evaluation of drug classes for genes with associations with gene expression. To identify drug–gene pairs that may be leads for repurposing or may be attractive leads for novel inhibitory drugs, we identified drugs that targeted genes whose predicted expression was significantly associated with T2D risk in the S-PrediXcan analyses and which we predicted would lower blood glucose on the basis of their direction of effect on T2D risk with increasing gene expression and drug action (activator or inhibitor). Medications with a primary indication for diabetes and medications with adverse drug events for patients with diabetes were evaluated using the SIDER medications that targeted genes were queried using DGIdb. These drug targets represent a set of genes that are both likely to be involved in glucose regulation in one or more tissues and can be targeted by drugs. Genes and medications identified in this analysis are presented in Supplementary Table 17.

Ethics statement. The Central VA Institutional Review Board (IRB) and site-specific Research and Development Committees approved the MVP study. The Vanderbilt University Medical Center IRB approved the use of BioVU data for this study. All other cohorts participating in this meta-analysis have ethical approval from their local institutions. All relevant ethical regulations were followed.

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

Data availability

The full summary-level association data from the trans-ancestry, European, African American, Hispanic and Asian meta-analysis from this report are available through dbGAP under accession number phs001672.v3.p1 (Veterans Administration Million Veteran Program Summary Results from Omics Studies). Source data are provided with this paper. More specifically, dbGaP accession number pha004943.1 refers to the African American–specific summary statistics, pha004944.1 to the Asian-specific summary statistics, pha004945.1 refers to the European-specific summary statistics, pha004946.1 refers to the Hispanic-specific summary statistics, and pha004947.1 refers to the trans-ancestry summary statistics.

Code availability

Imputation was performed using MiniMac4 and EAGLE v2. Association analysis was performed using PLINK2A and XWAS v3.0. Post-GWAS processing software include: PRSice-2, LD Hub v1.9.3, FlashPCA v2.0, METAL v2011-03-25, GCTA-COJO v1.93, S-PrediXcan v0.6.1, SUGEN v8.9, DEPICT v140721, SIDER v4.1, DGIdb v3.0 and KING v2.1.6, as outlined in the Methods. Clear code for analysis is available at the associated website of each software package. Additional analyses were performed in R-3.2.

References

40. Manichaikul, A. et al. Robust relationship inference in genome-wide association studies. *Bioinformatics* **26**, 2867–2873 (2010).
41. 1000 Genomes Project Consortium et al. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
42. Das, S. et al. Next-generation genotype imputation service and methods. *Nat. Genet.* **48**, 1284–1287 (2016).
43. Fang, H. et al. Harmonizing genetic ancestry and self-identified race/ethnicity in genome-wide association studies. *Am. J. Hum. Genet.* **105**, 763–772 (2019).
44. Tirschwell, D. L. & Longstreth, W. T. Jr. Validating administrative data in stroke research. *Stroke* **33**, 2465–2470 (2002).
45. Chang, C. C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).
46. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
47. Gao, F. et al. XWAS: a software toolset for genetic data analysis and association studies of the X chromosome. *J. Hered.* **106**, 666–671 (2015).
48. Ko, Y. A. et al. Genetic-variation-driven gene-expression changes highlight genes with important functions for kidney disease. *Am. J. Hum. Genet.* **100**, 940–953 (2017).
49. Ackermann, A. M., Wang, Z., Schug, J., Naji, A. & Kaestner, K. H. Integration of ATAC-seq and RNA-seq identifies human alpha cell and beta cell signature genes. *Mol. Metab.* **5**, 233–244 (2016).
50. International HapMap Consortium et al. Integrating common and rare genetic variation in diverse human populations. *Nature* **467**, 52–58 (2010).
51. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
52. Roadmap Epigenomics Consortium et al. Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).
53. Andersson, R. et al. An atlas of active enhancers across human cell types and tissues. *Nature* **507**, 455–461 (2014).
54. Finucane, H. K. et al. Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *Nat. Genet.* **50**, 621–629 (2018).
55. Fehrmann, R. S. et al. Gene expression analysis identifies global gene dosage sensitivity in cancer. *Nat. Genet.* **47**, 115–125 (2015).
56. Cahoy, J. D. et al. A transcriptome database for astrocytes, neurons, and oligodendrocytes: a new resource for understanding brain development and function. *J. Neurosci.* **28**, 264–278 (2008).
57. Heng, T. S. & Painter, M. W., Immunological Genome Project Consortium. The Immunological Genome Project: networks of gene expression in immune cells. *Nat. Immunol.* **9**, 1091–1094 (2008).
58. Denny, J. C. et al. PheWAS: demonstrating the feasibility of a phenome-wide scan to discover gene-disease associations. *Bioinformatics* **26**, 1205–1210 (2010).
59. Voorman, A., Lumley, T., McKnight, B. & Rice, K. Behavior of QQ-plots and genomic control in studies of gene-environment interaction. *PLoS ONE* **6**, e19416 (2011).
60. Lin, D. Y. et al. Genetic association analysis under complex survey sampling: the Hispanic Community Health Study/Study of Latinos. *Am. J. Hum. Genet.* **95**, 675–688 (2014).
61. Bulik-Sullivan, B. et al. An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47**, 1236–1241 (2015).
62. Pers, T. H. et al. Biological interpretation of genome-wide association studies using predicted gene functions. *Nat. Commun.* **6**, 5890 (2015).

Acknowledgements

This research is based on data from the MVP, Office of Research and Development, Veterans Health Administration and was supported by award no. MVP000. This publication does not represent the views of the VA, the US Food and Drug Administration, or the US Government. This research was also supported by funding from: the VA award I01-BX003362 (P.S.T. and K.-M.C.) and the VA Informatics and Computing Infrastructure (VINCI) VA HSR RES 130457 (S.L.D.). B.F.V. acknowledges support for this work from the National Institutes of Health (NIH) National Institute of Diabetes and Digestive and Kidney Diseases (DK101478), the NIH National Human Genome Research Institute (HG010067) and a Linda Pechenik Montague Investigator award. K.-M.C., S.M.D., J.M.G., C.J.O., L.S.P., J.S.L., and P.S.T. are supported by the VA Cooperative Studies Program. S.M.D. is supported by the Veterans Administration [IK2-CX001780]. D.K. is supported by the National Heart, Lung, and Blood Institute of the NIH (T32 HL007734). K.H.K. is supported by NIH award UC4-DK-112217. K.S. is supported by NIH R01 DK087635. L.S.P. is supported in part by VA awards I01-CX001025, and I01CX001737, NIH awards R21DK099716, U01 DK091958, U01 DK098246, P30DK111024 and R03AI133172, and a Cystic Fibrosis Foundation award PHILLI12A0. We thank all study participants for their contribution. Data on T2D were contributed by investigators from the DIAMANTE Consortium, Biobank Japan, Malmö Diet and Cancer Study, PennCath, MedStar, Pakistan Genomic Resource, Penn Medicine Biobank, and Regeneron Genetics Center. Data on stroke were provided by MEGASTROKE investigators, and data on CKD were contributed by CKDgen investigators. Data on islet α- and β-cells were contributed by the HPAP Consortium (RRID:SCR_016202 and https://hpap.pmacs.upenn.edu/). Data on coronary artery disease were contributed by the CARDIoGRAMplusC4D investigators. We thank Josep Maria Mercader and Aaron Leong for careful review and comments.

Author contributions

M.V., J.M.K., K.-M.C., D.S., B.F.V., P.S.T. and C.J.O. were responsible for the concept and design. The acquisition, analysis or interpretation of data were performed by M.V., J.M.K., K.-M.C., D.S., B.F.V., P.S.T., R.L.J., C.T., T.L.A., J.E.H., J.Z., J.H., K.L., X.Z., J.A.L., A.T.H., K.M.L, D.K., S.P., J.D., O.M., A.R., N.H.M., S.H., I.H.Q., M.N.A., U.M., A.J., S.A., X.S., L.G., K.H.K., K.S., Y.V.S., S.L.D., K.C., J.S.L., J.M.G., L.S.P., D.R.M., J.B.M., P.D.R., P.W.W., T.L.E., D.J.R., S.M.D. and C.J.O. The authors M.V. and D.S. drafted the manuscript. The critical revision of the manuscript for important intellectual content was carried out by M.V., J.M.K., K.-M.C., D.S., B.F.V., J.A.L., P.S.T., C.T., J.Z., J.H., X.Z., D.K., X.S., L.G., K.H.K., K.S., L.S.P., J.B.M., P.D.R., T.L.E., S.M.D. and C.J.O. Finally, K.-M.C., D.S., and B.F.V. provided administrative, technical or material support.

Competing interests

None of the sponsors of the following authors had a role in the design and conduct of the study, in the collection, management, analysis and interpretation of the data, or in the preparation, review or approval of the manuscript. D.S. has received support from the British Heart Foundation, Pfizer, Regeneron, Genentech and Eli Lilly pharmaceuticals. L.S.P. has served on Scientific Advisory Boards for Janssen, and received research support from Abbvie, Merck, Amylin, Eli Lilly, Novo Nordisk, Sanofi, PhaseBio, Roche, Abbvie, Vascular Pharmaceuticals, Janssen, Glaxo SmithKline, Pfizer, Kowa and the Cystic Fibrosis Foundation. L.S.P. is a cofounder, officer, board member and stockholder of the diabetes management-related software company Diasyst. S.L.D. has received research grant support from the following for-profit companies through the University of Utah or the Western Institute for Biomedical Research (an affiliated non-profit of VA Salt Lake City Health Care System): AbbVie, Anolinx, Astellas Pharma, AstraZeneca Pharmaceuticals, Boehringer Ingelheim International, Celgene Corporation, Eli Lilly and Company, Genentech, Genomic Health, Gilead Sciences, GlaxoSmithKline, Innocrin Pharmaceuticals, Janssen Pharmaceuticals, Kantar Health, Myriad Genetic Laboratories, Novartis International and PAREXEL International Corporation. P.D.R. has received research grant support from the following for-profit companies: Bristol Myers Squib and Lysulin, and has consulted with Intercept Pharmaceuticals and Boston Heart Diagnostics. S.M.D. receives research support to the University of Pennsylvania from RenalytixAI and consults for Calico Labs.

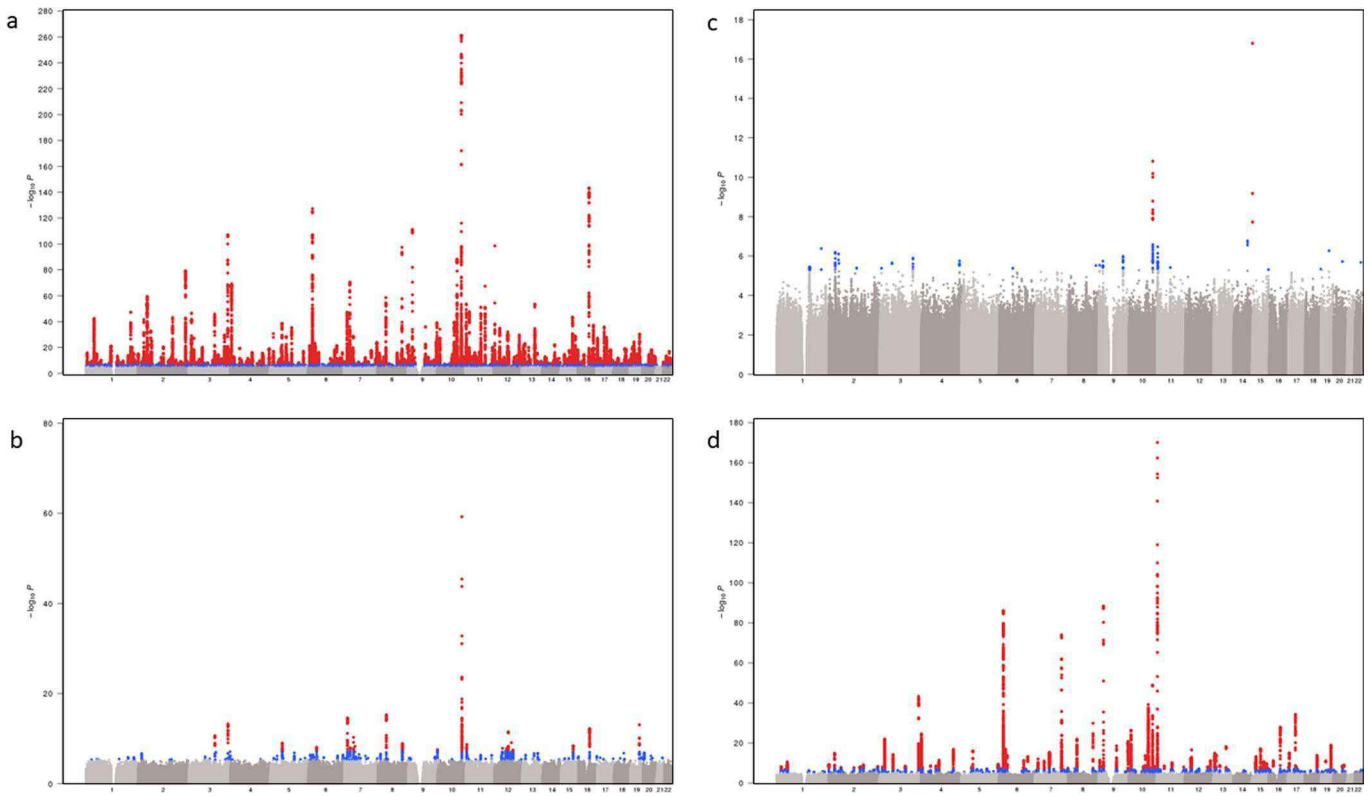
Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41588-020-0637-y.

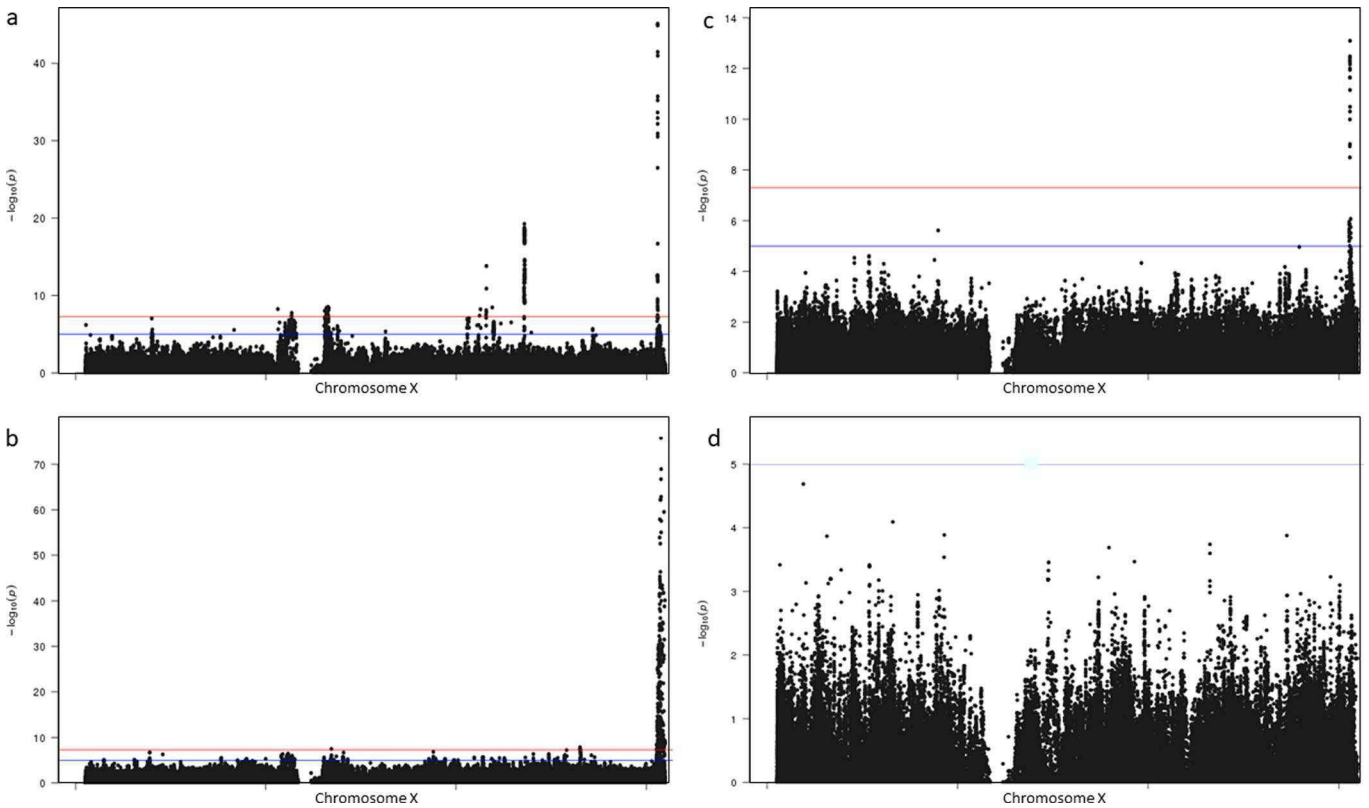
Supplementary information is available for this paper at https://doi.org/10.1038/s41588-020-0637-y.

Correspondence and requests for materials should be addressed to B.F.V. or D.S.

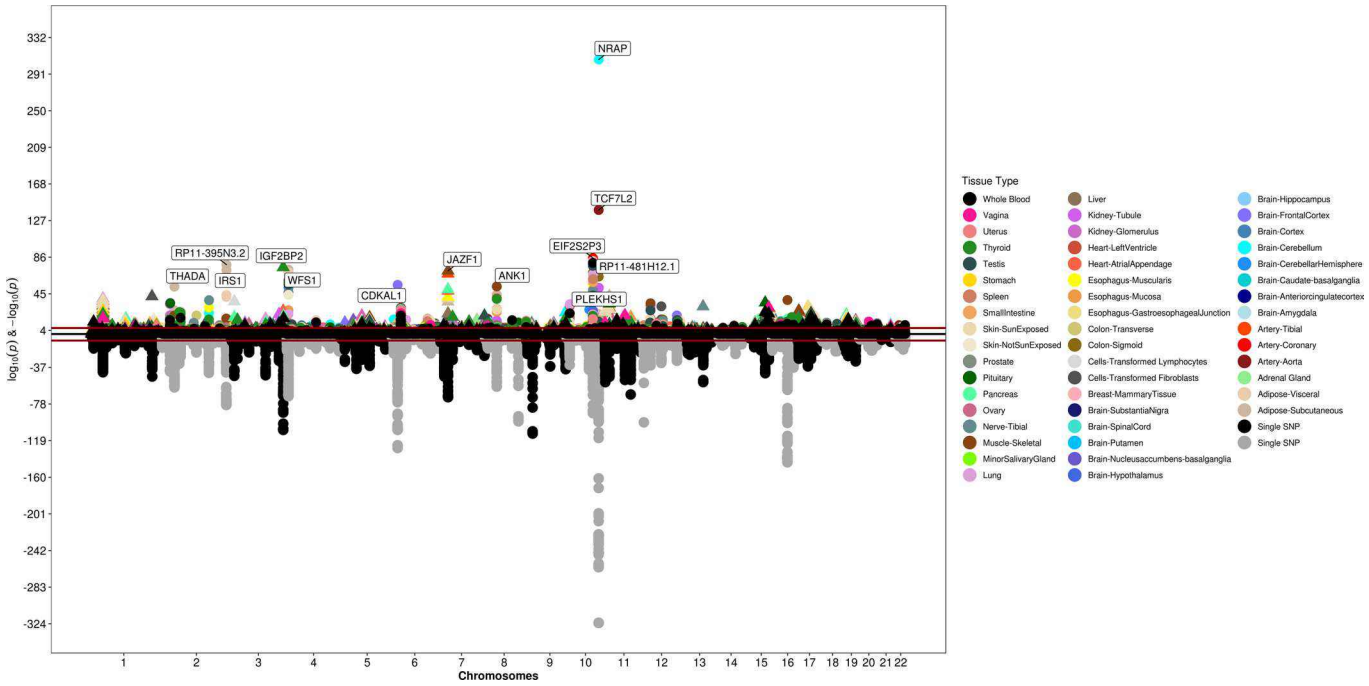
Reprints and permissions information is available at www.nature.com/reprints.



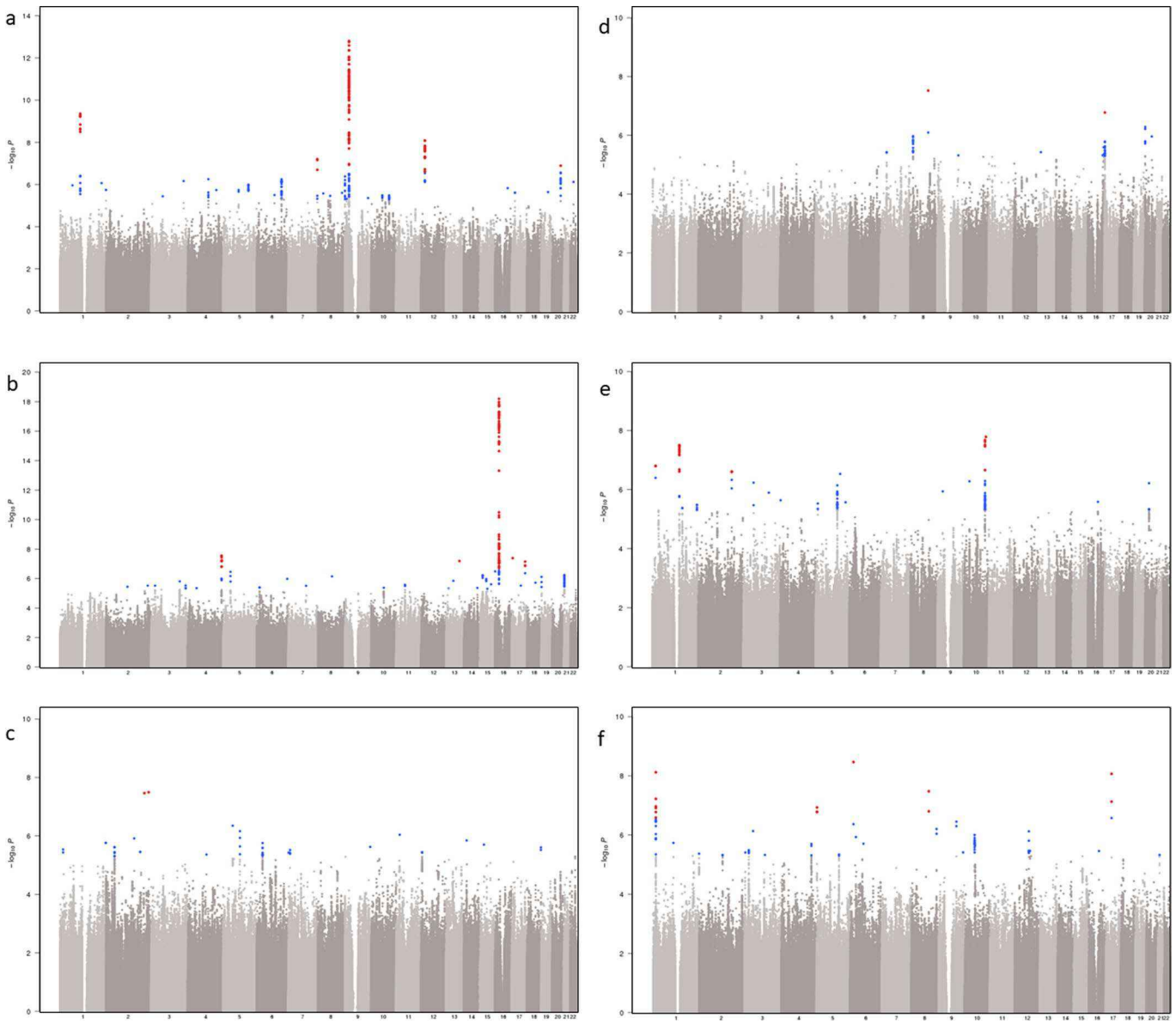
Extended Data Fig. 1 | Trans-ethnic and ancestry-specific GWAS Manhattan plots. a–d, Each graph represents a Manhattan plot. The *y*-axis corresponds to $-\log_{10}(P)$ for association with T2D in the respective ancestral group (**a**, Europeans (148,726 T2D cases, 965,732 controls, $\lambda = 1.21$); **b**, African American (24,646 T2D cases, 31,446 controls, $\lambda = 1.08$); **c**, Hispanics (8,616 T2D cases, 11,829 controls, $\lambda = 1.03$); **d**, Asians (46,511 T2D cases, 169,776 controls, $\lambda = 1.15$)). The *x*-axis represents chromosomal position on the autosomal genome. The *y*-axis truncated at 1×10^{-300}. Points that are color-coded blue correspond to a *P*-value between 5.0×10^{-8} and 1.0×10^{-6}. Points color-coded red indicate genome-wide significance ($P = 5.0 \times 10^{-8}$).



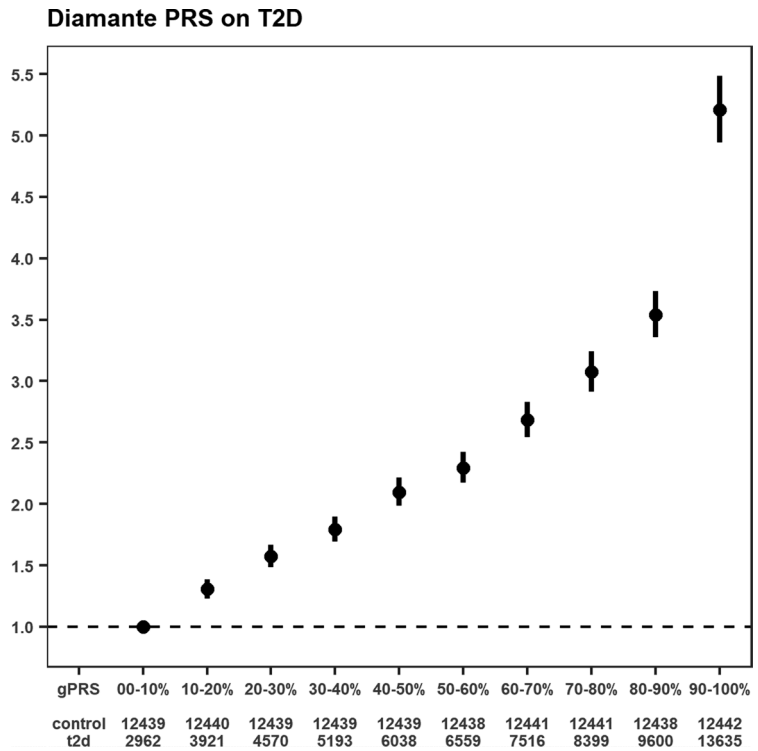
Extended Data Fig. 2 | Trans-ethnic and ancestry-specific chromosome X Manhattan plots. a–d, Each graph represents a Manhattan plot. The *y*-axis corresponds to $-\log_{10}(P)$ for association with T2D in the respective ancestral group (**a**, Europeans (69,869 T2D cases, 127,197 controls); **b**, African American (23,305 T2D cases, 30,140 controls); **c**, Hispanics (8,616 T2D cases, 11,829 controls); **d**, Asians (893 T2D cases, 1,560 controls)). The *x*-axis represents chromosomal position on chromosome X. The blue line corresponds with a significance threshold of $P = 5.0 \times 10^{-8}$. The red line corresponds with genome-wide significance ($P = 5.0 \times 10^{-8}$).



Extended Data Fig. 3 | Results from PrediXcan analysis using GTEX data. This graph represents an inverted Manhattan plot based on the output from the European T2D GWAS (148,726 T2D cases, 965,732 controls). The *y*-axis corresponds to $-\log_{10}(P)$ for association with genetically predicted gene expression in the respective tissue type (color coding shown on the right). Data were analyzed using S-PrediXcan software. The *x*-axis represents chromosomal position on the autosomal genome.



Extended Data Fig. 4 | Manhattan plots for T2D-related complications using interaction analysis in individuals of European ancestry. a–f, Each graph represents a Manhattan plot. The *y*-axis corresponds to $-\log_{10}$ (*P*) for association of SNP×T2D on T2D-related vascular outcome (**a**, coronary heart disease (56,285 cases, 140,945 controls, $\lambda = 1.06$); **b**, chronic kidney disease (67,403 cases, 129,827 controls, $\lambda = 1.02$); **c**, neuropathy (40,475 cases, 110,331 controls, $\lambda = 1.03$); **d**, peripheral artery disease (5,882 cases, 161,348 controls, $\lambda = 1.02$); **e**, retinopathy (13,881 cases, 123,538 controls, $\lambda = 1.02$); **f**, acute ischemic stroke (11,796 cases, 178,481 controls, $\lambda = 1.00$)). The *x*-axis represents chromosomal position on the autosomal genome. Points that are color-coded blue correspond to a *P*-value between 5.0×10^{-8} and 1.0×10^{-6}. Points color-coded red indicate genome-wide significance ($P = 5.0 \times 10^{-8}$).



Extended Data Fig. 5 | T2D PRS and the risk of T2D. A shape plot representing the risk of a T2D genome-wide PRS (gPRS) on the odds ratio of T2D in MVP participants of European ancestry (69,869 T2D cases, 127,197 controls). The weights for the PRS have been obtained from an external reference dataset, namely the DIAMANTE Consortium. The gPRS has been divided into 10 deciles based on gPRS values in MVP white participants without T2D. The reference group is the lowest decile (0-10%). Odds ratios are shown as red dots, with their respective 95th percent confidence intervals displayed as red vertical lines.

nature research

Corresponding author(s): Benjamin F. Voight, Danish Saleheen

Last updated by author(s): April 28, 2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Phenotypic data was collected from the electronic health record and genetic data using the Million Veteran Program (MVP) Axiom array.
Data analysis	Imputation was performed using MiniMac4 and EAGLE (v2), association analysis was performed using PLINK2A and XWAS (v3.0). Post-GWAS processing software include: PRSice-2, LD Hub (v1.9.3), FlashPCA (v2.0), METAL (v2011-03-25), GCTA-COJO (v1.93), S-PrediXcan (v0.6.1), SUGEN (v8.9), DEPICT (v140721), SIDER(v4.1), DGidb (v3.0), and KING (v2.1.6) as outlined in the Online Methods. Clear code for analysis is available at their associated websites. Additional analyses were performed in R-3.2.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The full summary level association data from the MVP Trans-Ancestry, European, and African American, Hispanic, and Asian T2D meta-analysis from this report are available through dbGAP (Accession number: pha004826.1).

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	All samples available of main ancestries (European, African, Hispanic, and Asian) were used for analysis (after quality control, see Supplementary Table 1 for full details). Sample size was determined based on using all genetic data available from MVP, DIAMANTE Consortium, Penn Medicine Biobank, Biobank Japan, MedStar, PennCath, Malmo Cancer and Diet study, and Pakistan Genomic Resource. Participants were excluded if they failed to meet case or control definitions.
Data exclusions	Data were excluded if they did not pass our pre-established quality control metrics, or if they did not fall within the main ancestries used for analysis. Patient were excluded if they had diagnosis codes for type 1 diabetes or secondary diabetes mellitus.
Replication	Replication is not applicable, as all available data was used for discovery analysis.
Randomization	Randomization is not applicable, as this is a population based case-control analysis of prevalent data.
Blinding	Blinding is not applicable, as this is a population based case-control analysis of prevalent data.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☐	☒ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Human research participants

Policy information about studies involving human research participants

Population characteristics	MVP participants (n= 273,409) are comprised predominantly of male subjects (91.6%) and were classified as Europeans (72.1%), African Americans (19.5%), Hispanics (7.5%), and Asians (0.9%)(Supplemental Table 2). The average age at study enrollment ranged from 56.1 for Asian to 68.2 for European participants (Supplementary Table 3). Average body mass index (BMI) ranged from 28.5 for Asians to 30.8 for African Americans. The proportion of males ranged from 87.2% for African Americans to 93.9% for Asians. The prevalence of T2D was 35.5% for Europeans, 36.4% for Asians, 42.1% for Hispanics, and 43.6% for African Americans.
Recruitment	Individuals aged 19 to 104 years have been recruited voluntarily from more than 50 VA Medical Centers nationwide for participation in the Million Veteran Program biobank study. Recruitment is currently occurring in person at selected sites in the VHA health care system. Every Veteran is assigned a study ID number, which is used to track them throughout the entire process of recruitment, enrollment, sample collection and use; this approach also provides a level of protection for personal identifiers from the outset. Given that study enrollment is voluntary, biases of this study are similar to those of any mega-biobank with voluntary enrollment, including survivorship bias. A complete description of the entire MVP Biobank study including recruitment can be found at PMID: 26441289. The recruitment criteria for the meta-analyzed studies can be obtained through the following PMIDS: DIAMANTE Consortium (30297969), Penn Medicine Biobank (30571185), Pakistan Genomic Resource (28869590), Biobank Japan (30718926), Malmo Diet and Cancer Study (8429286), Medstar (21239051) and PennCath (21239051).
Ethics oversight	The Million Veteran Program received ethical and study protocol approval from the VA Central Institutional Review Board (IRB) in accordance with the principles outlined in the Declaration of Helsinki.

Note that full information on the approval of the study protocol must also be provided in the manuscript.