nature genetics

Identification of new susceptibility loci for type 2 diabetes and shared etiological pathways with coronary heart disease

To evaluate the shared genetic etiology of type 2 diabetes (T2D) and coronary heart disease (CHD), we conducted a genome-wide, multi-ancestry study of genetic variation for both diseases in up to 265,678 subjects for T2D and 260,365 subjects for CHD. We identify 16 previously unreported loci for T2D and 1 locus for CHD, including a new T2D association at a missense variant in *HLA-DRB5* (odds ratio (OR) = 1.29). We show that genetically mediated increase in T2D risk also confers higher CHD risk. Joint T2D–CHD analysis identified eight variants—two of which are coding—where T2D and CHD associations appear to colocalize, including a new joint T2D–CHD association at the *CCDC92* locus that also replicated for T2D. The variants associated with both outcomes implicate new pathways as well as targets of existing drugs, including icosapent ethyl and adipocyte fatty-acid-binding protein.

The global epidemic of T2D is expected to worsen over the coming decades, and the number of people with T2D is projected to reach ~592 million by 2035 (ref. 1). T2D is also a major risk factor for CHD, which is the leading cause of death worldwide[2]. Patients with T2D are also at a twofold higher risk of mortality due to CHD as compared to individuals who do not have T2D[3], although the mechanisms that link T2D with increased risk of CHD remain inadequately understood. Recently, a coding variant in the *GLP1R* gene encoding the glucagon-like peptide-1 receptor was reported[4] that was associated with lower fasting glucose levels, lower T2D susceptibility, and, more modestly, with reduced risk for CHD, a result consistent with existing therapeutic perturbation of this gene. This type of result is noteworthy and motivates the search for additional loci with this type of genetic support: associations with protective effects for both T2D and CHD in humans. Such targets would merit detailed molecular, functional, and therapeutic experimentation, but these candidate loci need to be first identified from existing and newly generated data sets.

Genome-wide association studies (GWAS) have advanced understanding of the genetic architecture individually for each disease, yielding discovery of several dozen loci for T2D and CHD[5,6]. Previous work has also demonstrated a genetic correlation between these endpoints[7,8], although no study has directly compared individual variants beyond established sites across the genome or examined the pathways that are shared by the two outcomes. Regional association for multiple SNPs for both endpoints at a locus has been observed (for example, *CDKN2A–CDKN2B* or *APOE*)[5,6]. These initial observations indicate that the genetic pathways that connect T2D and CHD may have a modest impact on disease risk, hence requiring large sample sizes to enable robust discovery.

We therefore assembled a discovery association set for T2D comprising 73,337 T2D cases and 192,341 controls to first enable discovery of new loci for T2D. Second, we used additional genetic data on 90,831 CHD cases and 169,534 controls to identify genetic pathways connected with both outcomes.

RESULTS

Genome-wide association and replication testing for T2D

We used genetic data from 48,437 individuals (13,525 T2D cases and 34,912 controls) of South Asian (n = 28,139; 9,654 T2D cases and 18,485 controls) and European (n = 20,298; 3,871 T2D cases and 16,427 controls) descent. We used non-overlapping data for T2D from the DIAGRAM Consortium[5] and conducted combined discovery analysis on 198,258 participants (48,365 T2D cases and 149,893 controls). Characteristics of the participants and information on genotyping quality control are summarized in **Supplementary Tables 1–3** and **Supplementary Figure 1**. After removing known loci, we advanced into replication 21 new loci with suggestive association with T2D ($P \leq 5 \times 10^{-6}$). We performed further testing of these SNPs in additional samples of up to 67,420 individuals (24,972 cases and 42,448 controls) of South Asian (n = 13,960; 4,587 T2D cases and 9,373 controls), European (n = 2,479; 387 T2D cases and 2,092 controls), and East-Asian (n = 50,981; 19,998 T2D cases and 30,983 controls) descent. Our combined discovery and replication analyses included 265,678 participants (73,337 T2D cases and 192,341 controls) (**Supplementary Fig. 1a**). In the combined analysis across both stages, 14 SNPs at previously unreported loci for T2D obtained genome-wide significance (fixed-effects meta-analysis $P < 5 \times 10^{-8}$; **Table 1**). A previous report found one of our variants (rs10507349) strongly associated with T2D[9], but we report genome-wide significance for this variant here for the first time. Population-specific analyses (Europeans only or South Asians only) identified one additional locus where a sentinel variant obtained genome-wide significance in only European participants (**Table 1**, **Fig. 1**, **Supplementary Figs. 2** and **3**, and **Supplementary Tables 4** and **5**). Aside from this case, there

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

Received 30 March; accepted 3 August; published online 4 September 2017; doi:10.1038/ng.3943

Table 1 Sixteen new loci associated with T2D

Lead variant	Closest gene	Chr.	Position (hg19)	EA	NEA	EAF	OR	95% CI	*P* value	I^2	P_{het}
Loci associated with T2D in the combined analysis of Europeans, South Asians, and East Asians at $P < 5 \times 10^{-8}$											
rs2867125	*TMEM18*	2	622,827	C	T	0.83	1.06	1.04–1.08	2.18×10^{-9}	18	2.3×10^{-1}
rs11123406	*BCL2L11*	2	111,950,541	T	C	0.36	1.04	1.03–1.06	8.57×10^{-9}	2	4.4×10^{-1}
rs2706785	*TMEM155*	4	122,660,250	G	A	0.05	1.13	1.08–1.17	2.40×10^{-8}	0	9.0×10^{-1}
rs329122	*PHF15*	5	133,864,599	A	G	0.43	1.04	1.03–1.06	3.02×10^{-9}	0	5.1×10^{-1}
rs622217	*SLC22A3*	6	160,766,770	T	C	0.52	1.05	1.03–1.07	2.47×10^{-10}	0	7.0×10^{-1}
rs9648716	*BRAF*	7	140,612,163	T	A	0.15	1.06	1.04–1.09	1.16×10^{-9}	0	4.8×10^{-1}
rs12681990	*KCNU1*	8	36,859,186	C	T	0.15	1.05	1.04–1.07	3.07×10^{-9}	0	6.4×10^{-1}
rs10507349	*RNF6*	13	26,781,528	G	A	0.78	1.05	1.04–1.07	9.69×10^{-10}	0	6.1×10^{-1}
rs576674	*KL*	13	33,554,302	G	A	0.16	1.07	1.05–1.10	9.27×10^{-13}	4	4.0×10^{-1}
rs7985179	*MIR17HG*	13	91,940,169	T	A	0.72	1.07	1.05–1.10	4.16×10^{-9}	0	6.2×10^{-1}
rs9940149	*ITFG3*	16	300,641	G	A	0.83	1.05	1.04–1.07	1.09×10^{-9}	0	9.2×10^{-1}
rs2050188	*HLA-DRB5*[a]	6	32,339,897	T	C	0.67	1.06	1.04–1.08	7.56×10^{-9}	17	5.8×10^{-1}
rs2421016	*PLEKHA1*	10	124,167,512	C	T	0.53	1.05	1.03–1.06	3.68×10^{-11}	17	2.3×10^{-1}
rs2925979	*CMIP*	16	81,534,790	T	C	0.29	1.05	1.03–1.07	2.41×10^{-8}	6	3.8×10^{-1}
rs825476[b]	*CCDC92*	12	123,568,456	T	C	0.57	1.04	1.03–1.06	4.3×10^{-8}	0	6.2×10^{-1}
Locus associated with T2D in Europeans at $P < 5 \times 10^{-8}$											
rs7674212	*CISD2*[a]	4	103,988,899	G	T	0.58	1.07	1.04–1.09	6.85×10^{-9}	0	7.2×10^{-1}

Chr., chromosome; EA, effect allele; NEA, non-effect allele; EAF, risk allele frequency in Europeans (allele frequencies by ancestry are reported in **Supplementary Table 2**); OR, odds ratio; CI, confidence interval; I^2, heterogeneity inconsistency index; P_{het}, *P* value for heterogeneity across the meta-analysis data sets. Position is under hg19.
[a]Candidate gene based on ExomeChip lookup or Mendelian subform. [b]Variant discovered from the bivariate scan with additional support from replication.

was little evidence of heterogeneity of effect between the ancestry groups in either our primary genetic analyses or across the two stages (**Supplementary Fig. 2**). We replicated previously reported associations with T2D[5,9] at 60 loci at genome-wide significance; a further 25 known loci were associated with T2D at $P < 0.05$ (**Fig. 1** and **Supplementary Table 4**). We did not observe association at three loci (rs76895963, rs7330796, and rs4523957) in our overall meta-analyses, owing to their previous discovery in subjects of East Asian ancestry (**Supplementary Table 4**). To nominate candidate genes and pathways, we obtained expression quantitative trait locus (eQTL) data from the MuTHER Consortium and the Genotype-Tissue Expression (GTEx) Project (v6; **Supplementary Table 6**)[10,11]. These data suggest a candidate gene at two loci (*ITFG3* and *PLEKHA1*) where the lead eQTL association strongly tagged the T2D association ($r^2 = 1.0$).

Coding variants at new genetic loci

To identify coding variants that may influence protein structure at unreported T2D loci, we obtained data on up to 31,207 individuals (9,500 T2D cases, 21,707 controls) of South Asian (7,832 T2D cases, 16,703 controls) and European origin (1,668 T2D cases, 5,004 controls) (Online Methods) genotyped on the ExomeChip[12]. We investigated 505 variants captured by the ExomeChip within ±250-kb regions flanking the sentinel SNPs. We identified one missense variant in *HLA-DRB5* (rs701884) that was associated with T2D risk at close to exome-wide significance (fixed-effects meta-analysis $P < 2 \times 10^{-7}$) (**Supplementary Table 7**). The missense variant was found to have a relatively stronger impact on disease risk than the lead noncoding variant. At *HLA-DRB5*, the OR for T2D for the missense variant (rs701884) was 1.29 (95% confidence interval (CI) = 1.23–1.35; $P = 4.8 \times 10^{-7}$) as compared to the T2D OR of 1.06 (95% CI = 1.04–1.08; $P = 7.56 \times 10^{-9}$) for the noncoding variant (rs2050188). Existing knowledge on gene function is summarized in **Supplementary Table 8**.

Variant association with traits and circulating biomarkers

To help understand the underlying biological mechanisms, we examined the association of genetic variation at newly discovered loci with a range of phenotypes and biomarkers ($n = 70$ traits) (**Supplementary Table 9**). We also used an association screen against a panel of 105 phenotypic traits measured in the PROMIS study in up to 17,542 participants (**Supplementary Table 10**)[13]. For the new loci, we conducted 2,800 variant–phenotype analyses using linear regression resulting in a Bonferroni-adjusted significance threshold of $P = 1.8 \times 10^{-5}$. Allelic variation that increased T2D risk was associated at *TMEM18* with increased body mass index (BMI; $P = 4.39 \times 10^{-52}$), BMI in childhood ($P = 7.95 \times 10^{-12}$), obesity ($P = 2.50 \times 10^{-25}$) and obesity in childhood ($P = 2.85 \times 10^{-20}$); at *KL* with increased fasting glucose ($P = 2.26 \times 10^{-8}$); at *PLEKHA1* with increased risk of neovascular disease ($P = 2.71 \times 10^{-94}$); at *SLC22A1* with increased lipoprotein(a) levels ($P = 5.10 \times 10^{-6}$); and at *CMIP* with decreased HDL cholesterol (HDL-C) ($P = 1.32 \times 10^{-19}$), increased triglycerides ($P = 2.14 \times 10^{-7}$), and decreased adiponectin ($P = 1.87 \times 10^{-18}$).

Genetic risk for T2D and CHD shared at established loci

We next examined the relationships of sentinel T2D SNPs with the risk of CHD at all T2D loci (**Supplementary Table 11**). For analyses in relation to CHD, we used data on up to 260,365 participants (90,831 CHD cases and 169,534 controls) (Online Methods). We found allelic variation at 17 T2D loci to be nominally associated with CHD risk at $P < 0.01$, which was more than expected (17 of 106 T2D SNPs; binomial test $P = 5.9 \times 10^{-13}$). In one case, we found that the T2D sentinel SNP rs7578326 (the *IRS1* locus) was associated with both T2D and CHD at genome-wide levels of significance (**Supplementary Table 11** and **Supplementary Fig. 4**). To the best of our knowledge, this is the first definitive report of genetic variation at *IRS1* associated with CHD (**Supplementary Note**)[6]. We further investigated the relationship between these two endpoints in more detail.

Genetically elevated T2D risk overall increases CHD risk

First, we examined whether elevated T2D risk conferred a higher risk of CHD using the framework of Mendelian randomization (MR)[14–16] and examined whether all genetic T2D risk pathways influence CHD susceptibility in a similar way. We calculated genetic

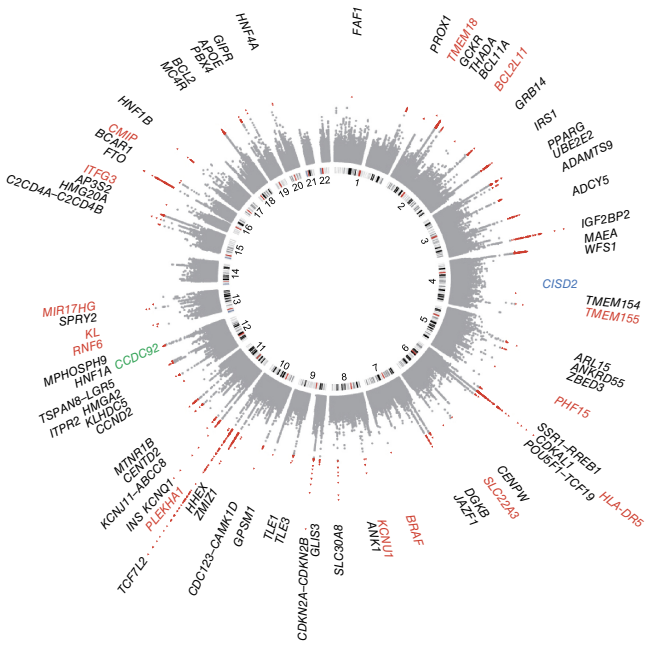

Figure 1 A circular Manhattan plot summarizing the association results for the T2D scan. Black, previously established T2D loci; red, previously unreported T2D loci from trans-ancestry meta-analysis; blue, previously unreported T2D loci from European-only meta-analysis; green, previously unreported T2D locus discovered from the bivariate scan with additional support from replication.

risk scores comprising collections of SNPs associated with T2D and potentially a range of cardiometabolic traits (Online Methods). These analyses underscore three key findings. First, a genotype risk score based on variants exclusively associated with increased risk of T2D (Online Methods and **Supplementary Tables 12** and **13**) was significantly associated with increased CHD risk (OR = 1.26; Wald test $P = 3.3 \times 10^{-8}$), supporting a causal role for T2D in CHD etiology in a directionally consistent manner (**Supplementary Table 14**). Second, T2D risk scores that involved variants based on their association with established risk factors for CHD (blood pressure, BMI, lipids, and anthropometric traits; **Supplementary Table 13**) showed significant differences in their estimated effects in relation to CHD (OR = 1.07–1.43; Cochran's heterogeneity test $P = 1.4 \times 10^{-5}$), indicating that the genetic mechanisms and underlying pathways that increase risk of T2D do not uniformly influence CHD risk in the same manner (**Supplementary Table 14**). Finally, in contrast to these scores, variants associated with T2D and glucose or insulin-related traits, but not other traits, were not associated with CHD (OR = 1.07; Wald test $P = 0.06$) (**Supplementary Table 14**); however, this could be due to reduced power of this instrument relative to others, as has been observed previously[15]. These analyses indicate that pathways segregating genetic susceptibility for T2D may not have an equivalent impact on CHD risk.

Genetic risk for T2D and CHD shared across the genome

We next looked for enrichment in the consistency of the risk allele associated with both T2D and CHD across the genome. In our meta-analyses, of the 1,260 variants associated with T2D at $P < 5 \times 10^{-8}$, we found that 76.1% of the T2D risk alleles were associated with higher risk of CHD as well, in comparison with an expectation of 50% under the null hypothesis (binomial test $P = 2.6 \times 10^{-33}$; **Table 2**). In contrast, variants associated with CHD at $P < 5 \times 10^{-8}$ were not enriched for directional consistency in allelic associations with T2D (48.2 versus the 50% expected; binomial test $P = 0.79$). Among the loci nominally associated with T2D and CHD ($P < 0.05$ but excluding associations above $P < 5 \times 10^{-8}$), 81.8% of the allelic variation associated with both the outcomes in a directionally consistent manner (binomial test $P < 10^{-100}$). Furthermore, of the allelic variation that was not associated with both T2D and CHD at $P > 0.05$, only 50.6% of the allelic variation (as compared to the 50% expected under the null hypothesis) was associated with the two outcomes in a directionally consistent manner (**Table 2**). To rule out any biases introduced as a result of allelic variations at a limited set of loci associated with both CHD and T2D, we conducted sensitivity analyses in PROMIS using genome-wide variants pruned for linkage disequilibrium (LD) and found results consistent for an overall enrichment of loci associated with both T2D and CHD in a directionally consistent manner (**Supplementary Table 15**).

Joint test identifies an additional locus for T2D and CHD

Motivated by the enrichment of directionally consistent associations of allelic variation between T2D and CHD SNPs, we performed a genome-wide association scan that modeled the joint distribution of association with both T2D and CHD (T2D–CHD; Online Methods), a test to help improve power for discovery of new loci that are associated with both the outcomes (**Supplementary Fig. 5**). After verifying that our test statistic was calibrated (**Supplementary Fig. 6**), we used this approach to identify a set of loci that were associated with both T2D and CHD (trait-specific fixed-effects meta-analysis $P < 10^{-3}$) and that were overall associated at genome-wide levels of significance (bivariate $P < 5 \times 10^{-8}$; **Supplementary Table 16**). Nineteen loci met these criteria, which included many established loci for T2D or CHD.

We identified one association near *CCDC92* (bivariate $P = 2.7 \times 10^{-9}$; **Supplementary Fig. 7a**). The sentinel variant (rs825476) was associated with both T2D (fixed-effects meta-analysis $P = 2.2 \times 10^{-6}$) and CHD (fixed-effects meta-analysis $P = 2.9 \times 10^{-7}$) (**Supplementary Fig. 7b**,**c**); rs825476[T] at this locus increased risk for both outcomes. To demonstrate conclusive association of rs825476 with T2D, we sought replication data from eight additional cohorts, comprising 21,560 T2D cases and 42,814 controls. We observed marginal replication for this variant in those data alone (OR = 1.04, 95% CI = 1.01–1.07; fixed-effects meta-analysis $P = 5.5 \times 10^{-3}$) and obtained genome-wide significance when combined with the previous data (OR = 1.04, 95% CI = 1.03–1.06; fixed-effects meta-analysis $P = 4.3 \times 10^{-8}$; **Supplementary Fig. 7b**). Analyses conditioned on the lead SNP accounted for all the residual joint T2D–CHD association in the region (Online Methods), indicating that the underlying genetic associations for both endpoints colocalize to a shared genetic risk factor potentially tagged by the sentinel SNP (**Supplementary Fig. 8**). The rs825476[T] allele also increased the expression of *CCDC92* in subcutaneous adipose tissue (**Supplementary Table 6**) in eQTL analyses conducted in the MuTHER Consortium and GTEx[10,11], suggesting a possible candidate gene for the association.

We sought to reduce our list to a subset of loci that colocalized the T2D and CHD associations to a single underlying genetic risk variant by conducting formal colocalization analyses (Online Methods). Eight of these 19 loci met this criterion, and at 7 of those 8 loci the risk allele for T2D also increased the risk for CHD (**Table 3**). This included loci with known associations with T2D (*TCF7L2*, *HNF1A*, *CTRB1*, and *CTRB2*) as well as previously unreported T2D loci reported here (*MIR17HG* and *CCDC92*) or known association with

Table 2 Enrichment in directional consistency for all SNPs in the T2D and CHD association scan

P-value cutoff			T2D and CHD associations from trait-specific meta-analyses		
T2D	CHD	No. of SNPs in total	No. of SNPs T2D/CHD consistent	% of SNPs T2D/CHD consistent	Adjusted $-\log_{10}(P)$[a]
$(0, 5 \times 10^{-8}]$	–[b]	1,260	959	76.11	76.966
–[b]	$(0, 5 \times 10^{-8}]$	595	287	48.24	0.062
(0.5, 1]	(0.5, 1]	1,874,138	948,292	**50.60**	–
$(5 \times 10^{-8}, 0.05]$	$(5 \times 10^{-8}, 0.05]$	36,242	29,634	81.77	3319.168

[a]*P* values from the binomial sign test are reported. The probability used to estimate the *P* values in the binomial sign test is the percentage highlighted in bold. [b]All variants, irrespective of association *P* value.

CHD (*MRAS* and *ZC3HC1*). Interestingly, this set of directionally consistent loci included coding variants in two genes encoding transcription factors: the missense variant(s) p.Ile27Leu in *HNF1A* and p.Arg326His in *ZC3HC1*. At the *APOE* locus, where the effect of association for T2D and CHD risk was opposite, localization was observed at rs4420638, but the tagging among the lead sentinel SNPs was incomplete, making it challenging to distinguish between multiple conditionally independent variant associations with both traits versus partial tagging of a single, common association. At the *IRS1* locus, while we found rs7578326 to be associated with both T2D and CHD ($P < 5 \times 10^{-8}$), formal colocalization analyses failed to identify a single underlying genetic risk factor for the two outcomes at this locus.

Next, we used biomarker data to help understand the mechanisms linking T2D and CHD at two new loci discovered through bivariate scan, *MIR17HG* and *CCDC92*. The region around *CCDC92* segregates numerous cardiometabolic trait associations, including T2D (rs1727313)[9], HDL-C (rs4759375 and rs838880), triglycerides (rs4765127)[17], and waist–hip ratio adjusted for BMI (rs4765219)[18]. However, variants in these previous reports were not strongly linked to our sentinel SNP ($r^2 < 0.02$ in all cases). The risk variant for T2D–CHD at the *CCDC92* locus also decreased HDL-C levels (fixed-effects meta-analysis $P = 2.2 \times 10^{-9}$) in analyses by the Global Lipids Genetics Consortium (GLGC)[17]; this variant was in partial linkage with a variant (rs10773049; $r^2 = 0.6$ and 0.3 in Europe and South Asia, respectively) previously known for association with BMI. *MIR17HG* appeared to harbor only modest associations with HDL (fixed-effects meta-analysis $P = 1.3 \times 10^{-4}$), fasting insulin levels ($P = 6.4 \times 10^{-4}$), and homeostatic model assessment of insulin resistance (HOMA-IR; $P = 7.9 \times 10^{-4}$).

We also examined association at *APOE* where the T2D risk allele was associated with decreased CHD risk. The T2D risk allele was also found associated with increased HDL-C (fixed-effects meta-analysis $P = 1.72 \times 10^{-21}$), decreased LDL cholesterol (LDL-C; $P = 1.51 \times 10^{-178}$), decreased total cholesterol ($P = 1.14 \times 10^{-149}$), decreased triglycerides ($P = 1.55 \times 10^{-14}$), reduced LDL particle size ($P = 3.80 \times 10^{-11}$), increased waist–hip ratio (BMI adjusted; $P = 1.80 \times 10^{-6}$), and increased risk of neovascular disease ($P = 2.78 \times 10^{-8}$).

Joint T2D–CHD associations highlight new pathways

We next aimed to identify a subset of highly connected loci that indicate unidentified pathways that were jointly related to T2D and CHD. To achieve this, we used results from our bivariate T2D–CHD association scan and pruned SNPs for LD to obtain a set of unlinked regions across the genome ($r^2 < 0.05$). From this list, we selected 299 LD-independent SNPs that were found to be associated with T2D and CHD in our bivariate scan ($P < 0.001$; **Supplementary Tables 17** and **18**) and sought to prioritize candidate genes implicated in the association using the text-mining approach GRAIL[19]. Seventy-nine of 299 regions were found to have prioritized specific genes in associated intervals (GRAIL $P < 0.05$), significantly more overall than expected (26.4%; binomial test $P < 1 \times 10^{-34}$). Next, protein–protein interaction connectivity analysis among these 79 genes[20] demonstrated more direct and indirect connections than expected (permuted $P < 1 \times 10^{-4}$; Online Methods), thus motivating us to focus on this subset for further analysis. Several plausible candidates from this list emerged, including the hepatic glucose transporter gene *SLC2A2*, the adipocyte fatty-acid-binding protein gene *FABP4* (aP2), *LPIN1* (lipin-1), *PPARGC1B* (PGC-1β), and the free fatty acid receptor 1 gene *FFAR1*, among others (**Supplementary Table 18**).

Table 3 Genome-wide significant loci by bivariate scan at sentinel SNPs that are associated with both T2D and CHD ($P < 1 \times 10^{-3}$) where leading associations colocalize

						T2D			CHD			BVN
Gene	Lead variant	Chr.	Position (hg19)	EA	NEA	OR	95% CI	*P* value	OR	95% CI	*P* value	*P* value
Established loci with T2D–CHD risk allele agreement and colocalization ($r^2 > 0.7$ between T2D and CHD associations and coloc Prob. > 0.5)												
TCF7L2	rs7903146	10	114,758,349	T	C	1.35	1.33–1.38	1.3×10^{-219}	1.04	1.02–1.05	2.9×10^{-5}	2.6×10^{-212}
HNF1A (I27L)	rs1169288	12	12,146,650	A	C	1.06	1.04–1.08	9.3×10^{-10}	1.04	1.03–1.06	3.9×10^{-7}	2.0×10^{-12}
CTRB1/2	rs7202877	16	75,247,245	T	G	1.06	1.03–1.08	4.0×10^{-6}	1.06	1.04–1.09	2.9×10^{-6}	1.0×10^{-8}
MRAS	rs2306374	3	138,119,952	C	T	1.05	1.02–1.07	6.5×10^{-4}	1.06	1.04–1.08	2.3×10^{-8}	9.8×10^{-9}
ZC3HC1 (R342H)	rs11556924	7	129,663,496	C	T	1.03	1.01–1.05	4.9×10^{-4}	1.08	1.06–1.10	3.3×10^{-20}	1.4×10^{-19}
New loci with T2D–CHD risk allele agreement and colocalization ($r^2 > 0.7$ between T2D and CHD associations and coloc Prob. > 0.5)												
MIR17HG	rs7985179	13	91,940,169	A	T	1.07	1.05–1.10	3.7×10^{-9}	1.05	1.02–1.08	6.4×10^{-4}	1.5×10^{-9}
CCDC92	rs825476	12	124,568,456	T	C	1.04	1.03–1.06	2.2×10^{-6}	1.03	1.02–1.05	3.0×10^{-7}	2.7×10^{-9}
Opposite risk allele for T2D and CHD with colocalization ($r^2 > 0.7$ between T2D and CHD associations and coloc Prob > 0.5)												
APOE	rs4420638	19	45,422,946	A	G	1.08	1.05–1.11	8.8×10^{-8}	0.89	0.85–0.93	1.8×10^{-6}	2.6×10^{-13}

Chr., chromosome; EA, effect allele; NEA, non-effect allele; OR, odds ratio; CI, confidence interval; BVN, bivariate normal distribution of T2D and CHD statistics. Position is under hg19.

We next performed ontology analysis on the set of 79 genes that emerged from the T2D–CHD bivariate scan for connectivity[21]. To compare our findings, we also conducted similar ontological analysis on loci identified for T2D or CHD in previous GWAS for each of these traits. As expected, ontological analysis of established T2D loci alone indicated robust enrichment of diabetes, hyperglycemia, and insulin resistance disease annotations (enrichment test $P < 1 \times 10^{-55}$), as well as enrichment for pathways related to insulin secretion and transport, glucose homeostasis, and pancreas development (all $P < 1 \times 10^{-9}$). Also as expected, ontological analysis of CHD loci alone demonstrated robust enrichment of disease annotations related to coronary disease, myocardial infarction, and arteriosclerosis ($P < 1 \times 10^{-36}$), as well as enrichment for pathways related to lipid homeostasis and cholesterol transport ($P < 1 \times 10^{-8}$). As expected, analysis of the 79 gene intervals associated with T2D and CHD identified loci that were also modestly enriched for disease ontologies related to vascular resistance ($P < 1 \times 10^{-12}$), T2D, cardiovascular disease, fatty liver, obesity, gestational hypertension, and pre-eclampsia (all $P < 1 \times 10^{-5}$), as well as cancer ($P < 1 \times 10^{-9}$). In contrast to the pathways described above, we also observed enrichment for additional pathways related to cardiovascular system development, cell signaling, signal transduction, regulation of phosphorylation, and transmembrane receptor protein kinase signaling among the categories (adjusted $P < 1 \times 10^{-7}$) (**Supplementary Fig. 9**).

DISCUSSION

We report the discovery of 16 new loci for T2D using discovery and replication studies in 265,678 participants. Using ExomeChip data, we were able to identify a coding variant that was more strongly associated with T2D risk than the corresponding noncoding variant. Using additional data on 260,365 participants, we report a new locus for CHD and identify genetic loci that are shared by T2D and CHD, of which a subset colocalized to the same genetic variant (for example, *CCDC92*, *MIR17HG*, *HNF1A*, *ZC3HC1*, and *APOE*; **Table 3**). Finally, using a bivariate scan for T2D and CHD together, genetic association data pointed to new pathways that are implicated in the etiology of both the disease outcomes.

Many of the loci discovered in the current meta-analyses suggest new T2D biology or confirm pathways previously implicated in T2D. For instance, *MIR17HG*, *KL*, and *BCL2L11* have been shown to be involved in cell survival, apoptosis, and cellular aging, respectively[22–24]. Genetic variants near *KL* have also been shown to be associated with fasting glucose levels as well[25]. *TMEM18* is involved in cellular migration; *HLA-DR5* and *CMIP* have crucial roles in immune-mediated responses and have been implicated in various immunological disorders[26,27]. Genetic variation at the HLA locus (rs9272346) has previously been implicated in type 1 diabetes (T1D); however, we did not find any evidence of association of rs9272346 with T2D in our meta-analyses. Additionally, rs9272346 was not in LD with the T2D sentinel SNP (rs2050188) at this locus ($r^2 = 0.06$ in Europe and $r^2 = 0.01$ in South Asian). However, rs7111341 has previously been reported as a risk factor for T1D[28]. We found that rs7111341[T] is associated with increased risk for T2D but decreased risk for T1D, a similar pattern of association to a previously established association (rs7202877, near *CTRB1*).

The contrasting associations of *APOE* with T2D and CHD were challenging to interpret. *APOE* encodes apolipoprotein E found in the chylomicron and intermediate-density lipoproteins (IDLs). Genetic variation at the *APOE* locus is associated with major lipids and CHD[6,17]. Here the T2D risk variant was associated with decreased CHD risk and LDL-C and with reduced LDL particle size. These observations are consistent with recent studies indicating that reduction in levels of LDL-C, a major CHD risk factor, may confer a higher, but modest risk of T2D. Evidence from a meta-analysis of randomized controlled trials has shown that reduction of LDL-C by statin treatment, as compared to placebo, led to a higher but a very small absolute risk of T2D[29]. Moreover, genetic variants associated with reduced expression of HMG–CoA reductase, the target of statins, and reduced LDL-C levels have been shown to be associated with increased risk of T2D[30]. Also, two MR studies concluded that genetically mediated decreases in LDL-C associated with a higher risk of T2D[31,32]. Furthermore, it has been shown that genetic variants in the *PCSK9* gene that lower LDL-C levels are associated with a higher risk of T2D, fasting glucose concentration, body weight, and waist–hip ratio[33]. In contrast to the findings from our overall meta-analyses, these results suggest that LDL-C may represent one of a small subset of discrete pathways that display opposing associations for the two outcomes. These findings underscore how human genetics can help focus future investigations on T2D therapeutics that have either neutral or beneficial effects on vascular outcomes.

The collection of 79 regions identified through our joint T2D–CHD bivariate scan involves targets of existing drugs. These includes icosapent, a polyunsaturated fatty acid found in fish oil that is an *FFAR1* and *PPARG* agonist and a *COX1*/*COX2* inhibitor[34]. The ANCHOR trial showed that icosapent ethyl, marketed as the drug Vascepa, has efficacy in lowering triglycerides in patients with high triglyceride levels[35] as well as non-HDL-C and HDL-C[36]. A second plausible candidate gene is *FABP4* (encoding the adipocyte fatty-acid-binding protein; also known as aP2). Mouse models deficient in aP2 display protection against atherosclerosis and antidiabetic phenotypes[37–39]. Moreover, small-molecule inhibition of aP2 has been shown to reduce atherosclerosis, glucose and insulin levels, and triglycerides in a mouse model[40]; inhibition of this pathway through a monoclonal antibody also appears to be efficacious in mouse models[41].

Careful evaluation of the pathways or biological processes where T2D, CHD, and related traits overlap could help to highlight new avenues for therapeutic targeting. First, using gene discovery and biomarker studies, we have identified new pathways, outside of the established glucose and cholesterol homeostatic networks, that could be investigated in more detail. Second, we have found that some genetic variants associated with T2D singly or in aggregate are enriched for associations with CHD. With one exception (pathways involving LDL-C), genetic pathways that increase T2D risk tend to overall increase CHD risk. Hence, existing or future therapeutic programs designed for the prevention of T2D could be better guided by evidence from genetic studies, to prioritize targets that have either neutral or directionally consistent effects on vascular outcomes. Overall, identification of genetic loci associated with both T2D and CHD risk in a directionally consistent manner could provide therapeutic opportunities to lower the risk of both outcomes.

Note added in proof: A report[42] was published while our article was under review that mapped T2D associations to three regions reported here at genome-wide significance. These regions include (i) rs2925979; (ii) rs2292626, which is perfectly linked ($r^2 = 1.0$) to rs2421016 reported here; and (iii) rs9271774, which maps to nearby rs2050188 reported here but is not strongly linked ($r^2 = 0.14$).

URLs. CARDIoGRAMC4D Consortium, http://www.cardiogramplusc4d.org/; DIAGRAM Consortium, http://diagram-consortium.org/; coloc tool, https://github.com/chr1swallace/coloc; bivariate scan analysis code, https://github.com/WWinstonZ/bivariate_scan/; WebGestalt,

http://www.webgestalt.org/option.php; GRAIL, http://software.broadinstitute.org/mpg/grail/; DAPPLE, http://archive.broadinstitute.org/mpg/dapple/dapple.php; SNPTEST, https://mathgen.stats.ox.ac.uk/genetics_software/snptest/snptest.html; LocusZoom, http://locuszoom.org/; METAL, http://genome.sph.umich.edu/wiki/METAL; GTEx, https://gtexportal.org/home/; optiCall, https://opticall.bitbucket.io/; zCall, https://github.com/jigold/zCall; RAREMETALWORKER, http://genome.sph.umich.edu/wiki/RAREMETALWORKER; 1000 Genomes, http://www.internationalgenome.org/; PLINK, https://www.cog-genomics.org/plink2; GCTA, http://cnsgenomics.com/software/gcta/; summary data availability, http://www.med.upenn.edu/ccebfiles//t2d_meta_cleaned.zip and http://www.med.upenn.edu/ccebfiles/chd_t2d_af_gwas12_cleaned_combined_1000gRefAlt_added_pvalRescaled_varSetID_added.zip.

METHODS

Methods, including statements of data availability and any associated accession codes and references, are available in the online version of the paper.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

D.S. has received support from NHLBI, NINDS, Pfizer, Regeneron Pharmaceuticals, Genentech, and Eli Lilly. Genotyping in PROMIS was funded by the Wellcome Trust, UK, and Pfizer. Biomarker assays in PROMIS have been funded through grants awarded by the NIH (RC2HL101834 and RC1TW008485) and Fogarty International (RC1TW008485). The RACE study has been funded by NINDS (R21NS064908), Fogarty International (R21NS064908), and the Center for Non-Communicable Diseases (Karachi, Pakistan). B.F.V. was supported by funding from the American Heart Association (13SDG14330006), the W.W. Smith Charitable Trust (H1201), and the NIH/NIDDK (R01DK101478). J.D. is a British Heart Foundation Professor, European Research Council Senior Investigator, and NIHR Senior Investigator. V.S. was supported by the Finnish Foundation for Cardiovascular Research. S. Ripatti was supported by the Academy of Finland (251217 and 255847), the Center of Excellence in Complex Disease Genetics, the European Union's Seventh Framework Programme projects ENGAGE (201413) and BioSHaRE (261433), the Finnish Foundation for Cardiovascular Research, Biocentrum Helsinki, and the Sigrid Juselius Foundation. The Mount Sinai IPM Biobank Program is supported by the Andrea and Charles Bronfman Philanthropies. S. Anand is supported by grants from the Canada Research Chair in Ethnic Diversity and CVD and from the Heart and Stroke Michael G. DeGroote Chair in Population Health, McMaster University. Data contributed by Biobank Japan were partly supported by a grant from the Leading Project of the Ministry of Education, Culture, Sports, Science and Technology, Japan. We thank the participants and staff of the Copenhagen Ischemic Heart Disease Study and the Copenhagen General Population Study for their important contributions.
The CHD Exome+ Consortium was funded by the UK Medical Research Council (G0800270), the British Heart Foundation (SP/09/002), the UK NIHR Cambridge Biomedical Research Centre, the European Research Council (268834), the European Commission's Framework Programme 7 (HEALTH-F2-2012-279233), Merck, and Pfizer. PROSPER has received funding from the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement HEALTH-F2-2009-223004.

AUTHOR CONTRIBUTIONS

A.R., B.F.V., B.G.N., D.J.R., D.S., D.S.A., J.I.R., M.R., O.M., P.F., R.C., R.J.F.L., S. Anand, S.E., S.M., S. Ripatti, T.-D.W., W.H.-H.S., and W. Zhao conceived of and designed the experiments. A.I., A.M., A.R., A.S.B., B.F.V., B.G.N., B.R.S., C.A.H., D.J.R., D.K.S., D.S., D.S.A., E.P.B., E.S.T., E.T., F.M., F.-u.-R.M., G.P., I.-T.L., I.H.Q., J.-J.L., J.C.C., J.I.R., J.M.M.H., J.S.K., J.W.J., J.Z.K., K.-W.L., K.D.T., K.M., K.T., M.B., N.M., M.O.-M., N.A., N.H.M., S.N.H.R., N.Q., N. Sattar, O.M., P.C., P.F., P.S., R.-H.C., R.C., R.F.-S., R.J.F.L., S. Abbas, S. Anand, S.E., S.J., S.M., S.N.H.R., S. Ralhan, S. Ripatti, S.Z.R., T.-D.W., T.-u.S., T.K., T.L.A., T.S., T.Y., U.M., W.H.-H.S., W.I., W. Zhang, W. Zhao, X.G., Y.-D.I.C., Y.-J.H., Y.L., Y.Y.T., and Z.Y. performed the experiments. A.R., A.S.B., B.F.V., C.A.H., D.S., E.D.A., E.P.B., E.T., I.-T.L., J.-J.L., J.-M.J.J., J.C.C., J.D., J.M.M.H., J.S.K., J.W.J., J.Z.K., K.-W.L., K.D.T., M.B., M.I., M.O.-M., M.R., N.K.M., N. Sattar, N. Shah, O.M., P.C., P.F., P.R.K., P.S., R.-H.C., R.C., R.F.-S., R.J.F.L., R.S., R.Y., S. Anand, S. Asma, S.D., S.F.N., S.M., S. Ralhan, T.-D.W., N.M., T.L.A., T.Q., T.S., T.Y., U.M., V.S., W.-J.L., W.H.-H.S., W. Zhang, W. Zhao, X.G., Y.-D.I.C., Y.-J.H., Y.L., and Y.Y.T. performed statistical analyses. A.I., A.R., A.S., A.S.B., A.T.-H., B.F.V., B.R.S., C.-C.H., C.A.H., D.K.S., D.S., E.D.A., E.P.B., E.S.T., E.T., F.M., G.P., I.-T.L., J.-J.L., J.-M.J.J., J.C.C., J.D., J.I.R., J.S.K., J.Z.K., K.-W.L., K.D.T., K.T., M.I., M.O.-M., M.R., N.H.M., N.Q., N. Sattar, N. Shah, O.M., P.C., P.R.K., P.S., R.-H.C., R.F.-S., R.J.F.L., R.S., R.Y., S. Abbas, S. Asma, S.D., S.J., S. Ralhan, T.K., T.L.A., T.Q., T.S., T.Y., U.M., V.S., W.-J.L., W. Zhao, X.G., Y.-D.I.C., Y.-J.H., Y.L., Y.Y.T., and Z.Y. analyzed the data. A.M., A.S., A.T.-H., B.F.V., B.G.N., D.J.R., D.S., F.u.R.M., G.P., I.H.Q., J.-M.J.J., J.D., J.W.J., N.M., K.M., M.B., M.R., N.A., P.R.K., R.S., S.E., S.N.H.R., S. Ripatti, S.Z.R., T.-u.-S., V.S., W.I., and W. Zhao contributed reagents, materials, and/or analysis tools. A.R., A.S.B., B.F.V., D.J.R., D.K.S., D.S., E.T., G.P., J.-J.L., J.D., J.I.R., J.M.M.H., M.R., N. Sattar, V.S., and W. Zhao wrote the manuscript. D.S. and B.F.V. led the writing group. W. Zhao, A.R., E.T., B.F.V., and D.S. were equal contributors. B.F.V. and D.S. jointly supervised all aspects of the work.

COMPETING FINANCIAL INTERESTS

The authors declare competing financial interests: details are available in the online version of the paper.

1. Guariguata, L. *et al.* Global estimates of diabetes prevalence for 2013 and projections for 2035. *Diabetes Res. Clin. Pract.* **103**, 137–149 (2014).
2. Xu, J., Murphy, S.L., Kochanek, K.D. & Bastian, B.A. Deaths: final data for 2013. *Natl. Vital Stat. Rep.* **64**, 1–119 (2016).
3. Rao Kondapally Seshasai, S. *et al.* Diabetes mellitus, fasting glucose, and risk of cause-specific death. *N. Engl. J. Med.* **364**, 829–841 (2011).
4. Scott, R.A. *et al.* A genomic approach to therapeutic target validation identifies a glucose-lowering *GLP1R* variant protective for coronary heart disease. *Sci. Transl. Med.* **8**, 341ra76 (2016).
5. Morris, A.P. *et al.* Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat. Genet.* **44**, 981–990 (2012).
6. Nikpay, M. *et al.* A comprehensive 1000 Genomes–based genome-wide association meta-analysis of coronary artery disease. *Nat. Genet.* **47**, 1121–1130 (2015).
7. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47**, 1236–1241 (2015).
8. Jansen, H. *et al.* Genetic variants primarily associated with type 2 diabetes are related to coronary artery disease risk. *Atherosclerosis* **241**, 419–426 (2015).
9. Mahajan, A. *et al.* Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. *Nat. Genet.* **46**, 234–244 (2014).
10. GTEx Consortium. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348**, 648–660 (2015).
11. Grundberg, E. *et al.* Mapping *cis*- and *trans*-regulatory effects across multiple tissues in twins. *Nat. Genet.* **44**, 1084–1089 (2012).
12. Huyghe, J.R. *et al.* Exome array analysis identifies new loci and low-frequency variants influencing insulin processing and secretion. *Nat. Genet.* **45**, 197–201 (2013).
13. Saleheen, D. *et al.* The Pakistan Risk of Myocardial Infarction Study: a resource for the study of genetic, lifestyle and other determinants of myocardial infarction in South Asia. *Eur. J. Epidemiol.* **24**, 329–338 (2009).
14. Smith, G.D. & Ebrahim, S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int. J. Epidemiol.* **32**, 1–22 (2003).
15. Ross, S. *et al.* Mendelian randomization analysis supports the causal role of dysglycaemia and diabetes in the risk of coronary artery disease. *Eur. Heart J.* **36**, 1454–1462 (2015).
16. Ahmad, O.S. *et al.* A Mendelian randomization study of the effect of type-2 diabetes on coronary heart disease. *Nat. Commun.* **6**, 7060 (2015).
17. Willer, C.J. *et al.* Discovery and refinement of loci associated with lipid levels. *Nat. Genet.* **45**, 1274–1283 (2013).
18. Shungin, D. *et al.* New genetic loci link adipose and insulin biology to body fat distribution. *Nature* **518**, 187–196 (2015).
19. Raychaudhuri, S. *et al.* Identifying relationships among genomic disease regions: predicting genes at pathogenic SNP associations and rare deletions. *PLoS Genet.* **5**, e1000534 (2009).
20. Rossin, E.J. *et al.* Proteins encoded in genomic regions associated with immune-mediated disease physically interact and suggest underlying biology. *PLoS Genet.* **7**, e1001273 (2011).
21. Wang, J., Duncan, D., Shi, Z. & Zhang, B. WEB-based GEne SeT AnaLysis Toolkit (WebGestalt): update 2013. *Nucleic Acids Res.* **41**, W77–W83 (2013).
22. de Pontual, L. *et al.* Germline deletion of the miR-17~92 cluster causes skeletal and growth defects in humans. *Nat. Genet.* **43**, 1026–1030 (2011).
23. Kuro-o, M. *et al.* Mutation of the mouse *klotho* gene leads to a syndrome resembling ageing. *Nature* **390**, 45–51 (1997).
24. O'Connor, L. *et al.* Bim: a novel member of the Bcl-2 family that promotes apoptosis. *EMBO J.* **17**, 384–395 (1998).

25. Suzuki, M. *et al.* Plasma FGF21 concentrations, adipose fibroblast growth factor receptor-1 and β-klotho expression decrease with fasting in northern elephant seals. *Gen. Comp. Endocrinol.* **216**, 86–89 (2015).
26. Grimbert, P. *et al.* Truncation of C-mip (Tc-mip), a new proximal signaling protein, induces c-maf Th2 transcription factor and cytoskeleton reorganization. *J. Exp. Med.* **198**, 797–807 (2003).
27. Madsen, L.S. *et al.* A humanized model for multiple sclerosis using HLA-DR2 and a human T-cell receptor. *Nat. Genet.* **23**, 343–347 (1999).
28. Barrett, J.C. *et al.* Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. *Nat. Genet.* **41**, 703–707 (2009).
29. Sattar, N. *et al.* Statins and risk of incident diabetes: a collaborative meta-analysis of randomised statin trials. *Lancet* **375**, 735–742 (2010).
30. Swerdlow, D.I. *et al.* HMG–coenzyme A reductase inhibition, type 2 diabetes, and bodyweight: evidence from genetic analysis and randomised trials. *Lancet* **385**, 351–361 (2015).
31. White, J. *et al.* Association of lipid fractions with risks for coronary artery disease and diabetes. *JAMA Cardiol.* **1**, 692–699 (2016).
32. Fall, T. *et al.* Using genetic variants to assess the relationship between circulating lipids and type 2 diabetes. *Diabetes* **64**, 2676–2684 (2015).
33. Schmidt, A.F. *et al. PCSK9* genetic variants and risk of type 2 diabetes: a mendelian randomisation study. *Lancet Diabetes Endocrinol.* **5**, 97–105 (2017).
34. Law, V. *et al.* DrugBank 4.0: shedding new light on drug metabolism. *Nucleic Acids Res.* **42**, D1091–D1097 (2014).
35. Ballantyne, C.M. *et al.* Efficacy and safety of eicosapentaenoic acid ethyl ester (AMR101) therapy in statin-treated patients with persistent high triglycerides (from the ANCHOR study). *Am. J. Cardiol.* **110**, 984–992 (2012).
36. Ballantyne, C.M. *et al.* Effects of icosapent ethyl on lipoprotein particle concentration and size in statin-treated patients with persistent high triglycerides (the ANCHOR Study). *J. Clin. Lipidol.* **9**, 377–383 (2015).
37. Boord, J.B. *et al.* Adipocyte fatty acid–binding protein, aP2, alters late atherosclerotic lesion formation in severe hypercholesterolemia. *Arterioscler. Thromb. Vasc. Biol.* **22**, 1686–1691 (2002).
38. Hotamisligil, G.S. *et al.* Uncoupling of obesity from insulin resistance through a targeted mutation in aP2, the adipocyte fatty acid binding protein. *Science* **274**, 1377–1379 (1996).
39. Makowski, L. *et al.* Lack of macrophage fatty-acid-binding protein aP2 protects mice deficient in apolipoprotein E against atherosclerosis. *Nat. Med.* **7**, 699–705 (2001).
40. Furuhashi, M. *et al.* Treatment of diabetes and atherosclerosis by inhibiting fatty-acid-binding protein aP2. *Nature* **447**, 959–965 (2007).
41. Burak, M.F. *et al.* Development of a therapeutic monoclonal antibody that targets secreted fatty acid-binding protein aP2 to treat type 2 diabetes. *Sci. Transl. Med.* **7**, 319ra205 (2015).
42. Scott, R.A. *et al.* An expanded genome-wide association study of type 2 diabetes in Europeans. *Diabetes* http://dx.doi.org/10.2337/db16-1253 (2017).

Wei Zhao[1,2,72], Asif Rasheed[3,72], Emmi Tikkanen[4,5,72], Jung-Jin Lee[2], Adam S Butterworth[6], Joanna M M Howson[6], Themistocles L Assimes[7], Rajiv Chowdhury[6], Marju Orho-Melander[8], Scott Damrauer[9], Aeron Small[10], Senay Asma[11], Minako Imamura[12–14], Toshimasa Yamauch[15], John C Chambers[16–18], Peng Chen[19], Bishwa R Sapkota[20], Nabi Shah[3,21], Sehrish Jabeen[3], Praveen Surendran[6], Yingchang Lu[22–24], Weihua Zhang[16,17], Atif Imran[3], Shahid Abbas[25], Faisal Majeed[3], Kevin Trindade[2], Nadeem Qamar[26], Nadeem Hayyat Mallick[27], Zia Yaqoob[26], Tahir Saghir[26], Syed Nadeem Hasan Rizvi[26], Anis Memon[26], Syed Zahed Rasheed[28], Fazal-ur-Rehman Memon[29], Khalid Mehmood[30], Naveeduddin Ahmed[31], Irshad Hussain Qureshi[32], Tanveer-us-Salam[33], Wasim Iqbal[33], Uzma Malik[32], Narinder Mehra[34], Jane Z Kuo[35,36], Wayne H-H Sheu[37,38], Xiuqing Guo[36], Chao A Hsiung[39], Jyh-Ming J Juang[40], Kent D Taylor[36], Yi-Jen Hung[41], Wen-Jane Lee[42], Thomas Quertermous[7], I-Te Lee[37,38], Chih-Cheng Hsu[39], Erwin P Bottinger[22], Sarju Ralhan[43], Yik Ying Teo[19,44–47], Tzung-Dau Wang[40], Dewan S Alam[48], Emanuele Di Angelantonio[6], Steve Epstein[49], Sune F Nielsen[50], Børge G Nordestgaard[51,52], Anne Tybjaerg-Hansen[51,53], Robin Young[6], CHD Exome+ Consortium, Marianne Benn[54], Ruth Frikke-Schmidt[53], Pia R Kamstrup[51], EPIC-CVD Consortium, EPIC-Interact Consortium, Michigan Biobank, J Wouter Jukema[54,55], Naveed Sattar[56], Roelof Smit[57], Ren-Hua Chung[39], Kae-Woei Liang[58], Sonia Anand[59], Dharambir K Sanghera[20,60,61], Samuli Ripatti[4,5,62,63], Ruth J F Loos[22–24], Jaspal S Kooner[16,18], E Shyong Tai[64], Jerome I Rotter[36], Yii-Der Ida Chen[36], Philippe Frossard[3], Shiro Maeda[12–14], Takashi Kadowaki[15], Muredach Reilly[65], Guillaume Pare[66], Olle Melander[8,67], Veikko Salomaa[62], Daniel J Rader[2,10], John Danesh[6,68,69], Benjamin F Voight[10,70,71,73] & Danish Saleheen[1–3,73]

[1]Department of Biostatistics and Epidemiology, University of Pennsylvania, Philadelphia, Pennsylvania, USA. [2]Translational Medicine and Human Genetics, Department of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA. [3]Center for Non-Communicable Diseases, Karachi, Pakistan. [4]Department of Public Health, University of Helsinki, Helsinki, Finland. [5]Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland. [6]MRC/BHF Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK. [7]Department of Medicine, Stanford University School of Medicine, Stanford, California, USA. [8]Department of Clinical Sciences, Lund University, Malmö, Sweden. [9]Division of Vascular Surgery and Endovascular Therapy, Department of Surgery, University of Pennsylvania, Philadelphia, Pennsylvania, USA. [10]Department of Genetics, University of Pennsylvania, Philadelphia, Pennsylvania, USA. [11]Department of Mathematics and Statistics, McMaster University, Hamilton, Ontario, Canada. [12]Laboratory for Endocrinology, Metabolism, and Kidney Diseases, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan. [13]Department of Advanced Genomic and Laboratory Medicine, Graduate School of Medicine, University of the Ryukyus, Nishihara, Japan. [14]Division of Clinical Laboratory and Blood Transfusion, University of the Ryukyus Hospital, Nishihara, Japan. [15]Department of Diabetes and Metabolic Diseases, Graduate School of Medicine, University of Tokyo, Tokyo, Japan. [16]Department of Epidemiology and Biostatistics, Imperial College London, London, UK. [17]Ealing Hospital NHS Trust, Southall, UK. [18]Imperial College Healthcare NHS Trust, London, UK. [19]Saw Swee Hock School of Public Health, National University of Singapore and National University Health System, Singapore. [20]Department of Pediatrics, College of Medicine, University of Oklahoma Health Sciences Center, Oklahoma City, Oklahoma, USA. [21]Department of Pharmacy, COMSATS Institute of Information Technology, Abbottabad, Pakistan. [22]Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, New York, USA. [23]Genetics of Obesity and Related Metabolic Traits Program, Icahn School of Medicine at Mount Sinai, New York, New York, USA. [24]Division of Epidemiology, Department of Medicine, Vanderbilt-Ingram Cancer Center, Vanderbilt Epidemiology Center, Vanderbilt University School of Medicine, Nashville, Tennessee, USA. [25]Faisalabad Institute of Cardiology, Faisalabad, Pakistan. [26]National Institute of Cardiovascular Disorders, Karachi, Pakistan. [27]Punjab Institute of Cardiology, Lahore, Pakistan. [28]Karachi Institute of Heart Diseases, Karachi, Pakistan. [29]Red Crescent Institute of Cardiology, Hyderabad, Pakistan. [30]Department of Medicine, Civil Hospital, Karachi, Pakistan. [31]Department of Neurology, Liaquat National Hospital, Karachi, Pakistan. [32]Department of Medicine, Mayo Hospital and King Edward Medical University, Lahore, Pakistan. [33]Department of Neurology, Lahore General Hospital, Lahore, Pakistan. [34]All-India Institute of Medical Sciences, New Delhi, India. [35]Clinical and Medical Affairs, CardioDx, Inc., Redwood City, California, USA. [36]Institute for Translational Genomics and Population Sciences, Los Angeles Biomedical Research Institute and Department of Pediatrics, Harbor-UCLA Medical Center, Torrance, California, USA.

[37]Division of Endocrine and Metabolism, Department of Internal Medicine, Taichung Veterans General Hospital, Taichung, Taiwan. [38]School of Medicine, National Yang-Ming University, Taipei, Taiwan, and Chung Shan Medical University, Taichung, Taiwan. [39]Institute of Population Health Sciences, National Health Research Institutes, Miaoli, Taiwan. [40]Cardiovascular Center and Division of Cardiology, Department of Internal Medicine, National Taiwan University Hospital, National Taiwan University College of Medicine, Taipei, Taiwan. [41]Division of Endocrinology and Metabolism, Tri-Service General Hospital, National Defense Medical Center, Taipei, Taiwan. [42]Department of Medical Research, Taichung Veterans General Hospital, Taichung, Taiwan. [43]Hero DMC Heart Institute Ludhiana, Punjab, India. [44]Genome Institute of Singapore, A*STAR, Singapore. [45]Singapore Eye Research Institute, Singapore National Eye Centre, Singapore. [46]National University of Singapore Graduate School for Integrative Science and Engineering, National University of Singapore, Singapore. [47]Life Sciences Institute, National University of Singapore, Singapore. [48]International Centre for Diarrhoeal Disease Research, Dhaka, Bangladesh. [49]MedStar Heart and Vascular Institute, MedStar Washington Hospital Center, Washington, DC, USA. [50]Copenhagen City Heart Study, Frederiksberg Hospital, Frederiksberg, Denmark. [51]Department of Clinical Biochemistry, Herlev and Gentofte Hospital, Copenhagen University Hospital, Herlev, Denmark. [52]Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. [53]Department of Clinical Biochemistry, Rigshospitalet, Copenhagen University Hospital, Copenhagen, Denmark. [54]Department of Cardiology, Leiden University Medical Center, Leiden, the Netherlands. [55]Interuniversity Cardiology Institute of the Netherlands, Utrecht, the Netherlands. [56]Institute of Cardiovascular and Medical Sciences, University of Glasgow, Glasgow, UK. [57]Department of Gerontology and Geriatrics, Leiden University Medical Center, Leiden, the Netherlands. [58]Cardiovascular Center, Taichung Veterans General Hospital, Taichung, Taiwan. [59]Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario, Canada. [60]Department of Pharmaceutical Sciences, College of Pharmacy, University of Oklahoma Health Sciences Center, Oklahoma City, Oklahoma, USA. [61]Oklahoma Center for Neuroscience, Oklahoma City, Oklahoma, USA. [62]National Institute for Health and Welfare, Helsinki, Finland. [63]Human Genetics, Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, UK. [64]Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore, Singapore. [65]Cardiology Division, Department of Medicine and the Irving Institute for Clinical and Translational Research, Columbia University, New York, New York, USA. [66]Department of Pathology and Molecular Medicine, McMaster University, Hamilton, Ontario, Canada. [67]Hypertension and Cardiovascular Disease, Department of Clinical Sciences, Lund University, Malmö, Sweden. [68]Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, UK. [69]NIHR Blood and Transplant Research Unit in Donor Health and Genomics, University of Cambridge, Cambridge, UK. [70]Department of Systems Pharmacology and Translational Therapeutics, University of Pennsylvania, Philadelphia, Pennsylvania, USA. [71]Institute for Translational Medicine and Therapeutics, University of Pennsylvania, Philadelphia, Pennsylvania, USA. [72]These authors contributed equally to this work. [73]These authors jointly directed this work. Correspondence should be addressed to D.S. (saleheen@mail.med.upenn.edu) or B.F.V. (bvoight@upenn.edu).

ONLINE METHODS

Study subjects. In the discovery phase, we performed meta-analysis on data from eight different studies; four studies (PROMIS, RACE, BRAVE, and EPIDREAM) include participants of South Asian origin living in Pakistan, Bangladesh, and Canada and four studies (FINRISK, MedStar, MDC, and PennCATH) include subjects of European origin (**Supplementary Table 1** and **Supplementary Note**). GWAS/Metabochip data and information on T2D risk were available on 48,437 individuals (13,525 T2D cases and 34,912 controls) from these eight studies. We further used published data from the DIAGRAM Consortium and conducted combined discovery analysis on 198,258 participants (48,365 T2D cases and 149,893 controls). Characteristics of the participants and information on genotyping arrays and imputation are summarized in **Supplementary Tables 1–3**. Replication studies were completed in participants enrolled in the LOLIPOP, SINDI, SDS, MSSE, TAICHI, and BBJ studies (**Supplementary Table 1**), collectively composed of 67,420 individuals (24,972 cases and 42,448 controls) who were of South Asian (n = 13,960; 4,587 T2D cases and 9,373 controls), European (n = 2,479; 387 T2D cases and 2,092 controls), and East Asian (n = 50,981; 19,998 T2D cases and 30,983 controls) descent. Hence, our combined discovery and replication analyses included 265,678 participants (73,337 T2D cases and 192,341 controls). Further details of the contributing cohorts and characteristics of the participants are provided in **Supplementary Table 1** and the **Supplementary Note**.

Institutional review boards and informed consent. All participating studies were approved by the relevant local institutional review boards. All participants enrolled in each of the participating studies provided informed consent.

Genotyping and quality control in the discovery stage. All studies used a high-density genotyping array (GWAS/Metabochip) (**Supplementary Table 2** and **Supplementary Note**). Quality control procedures were performed for each individual study. Details on study-specific quality control are provided in **Supplementary Table 2**. Each study individually assessed and controlled for any population stratification using principal-component analysis.

Imputation. In all studies, the genomic locations of all variants were first harmonized using NCBI Build 37/UCSC hg19 coordinates. Only studies that contributed GWAS data underwent imputation. Imputation of genotypes across the genome was computed using data from the 1000 Genomes Project (phase 1 integrated release 3, March 2012)[43]. Imputed SNPs were removed if they had (i) a minor allele frequency (MAF) of <0.01; (ii) an info score of <0.90; or (iii) an average maximum posterior call <0.90. **Supplementary Table 2** provides further details on the imputation protocol used by each of the participating studies.

Statistical analysis in the discovery stage. To test for an association between each SNP and risk of T2D, a logistic regression model was computed with adjustment for age, sex, and the first study-specific principal components using SNPTEST[44]. SNPs were modeled under an additive genetic model, and imputation uncertainty was accounted for under an allele dosage approach. Inflation of association statistics was assessed within each study by the genomic control method (**Supplementary Fig. 1** and **Supplementary Table 3**). Variants that were retained in at least two studies were subjected to meta-analysis using the weighted z-score method implemented in METAL[45]. Heterogeneity was assessed by Cochran's Q statistic and the I^2 heterogeneity index. Pairwise LD between SNPs was assessed and visualized using the 1000 Genomes European reference panel[43]. Regional association plots were visualized using LocusZoom software. After removing regions harboring known loci, the top associated SNP and one or more SNPs based on LD with the lead variant found in association with any of the above phenotypes ($P < 5 \times 10^{-6}$) were selected for the replication studies.

Analyses in the replication stage. Studies that participated in the replication stage had conducted genotyping on GWAS or Metabochip arrays. The association of SNPs with T2D was calculated separately using a trend test, with heterogeneity between studies assessed using Cochran's Q statistic. Meta-analysis was then conducted using the weighted z-score method implemented in METAL to combine the results across all replication studies and with the discovery stage. For the combined analysis of discovery and replication data, genome-wide significance was inferred at $P < 5 \times 10^{-8}$.

eQTL and functional prioritization. To determine whether the identified risk variants influence expression of any nearby genes, we accessed a variety of sources, including (i) GTEx *cis*-eQTL data in all available tissues, including liver, brain, endothelial cells, and whole blood[10], and (ii) *cis*-eQTL data for adipose, lymphoblastoid cell lines, and skin from the MuTHER Consortium[11].

ExomeChip analysis. To assess whether there are coding variants associated with T2D in the proximity of the newly discovered sentinel T2D SNPs, we performed an ExomeChip-based meta-analysis in four studies (PROMIS, BRAVE, CIHDS-CGPS, and PROSPER) in the ±500-kb regions centered on the sentinel T2D SNPs. For all the studies, genotyping and quality control were carried out centrally in Cambridge, UK. In each study, samples with extreme intensity values and outlying plates or arrays were removed before genotype calling. Genotype calling was initially performed with optiCall. Samples with a call rate (CR) less than (mean CR – 3 s.d.) were removed before postprocessing optiCall calls with zCall. Scanner-specific Z-values (calculated using the 1,000 samples with the highest optiCall CR) were adopted as they gave the best global concordance within each batch. Rare variants (optiCall, MAF < 0.05) were then postprocessed with zCall using the scanner-specific Z-values. Within each genotyping batch, variants were removed if variant CR was <0.97 or if Hardy–Weinberg equilibrium P value was $<1 \times 10^{-6}$ for common variants or $<1 \times 10^{-15}$ for rare variants (MAF < 0.05). Variants within each genotyping batch were aligned to the human genome reference sequence plus strand, and the standardized files were used for sample quality control. Samples were excluded from each batch or study if sample heterozygosity was >±3 s.d. from the mean heterogeneity or sample call rate was >3 s.d. from the mean call rate. Variants were further selected on the basis of stringent quality control thresholds (CR < 0.99, Hardy–Weinberg equilibrium $P < 1 \times 10^{-4}$, MAF > 0.05) and LD pruned ($r^2 < 0.2$) for principal-component analysis and kinship calculations. Duplicates within each collection (kinship coefficient > 0.45) and ancestral outliers identified by principal-component analysis were removed. Samples and variants that failed quality control were removed from individual batches. Where studies were analyzed in multiple batches, the batches were combined and any single-nucleotide variants (SNVs) out of Hardy–Weinberg equilibrium across the study were removed.

Study-specific analyses were conducted using RAREMETALWORKER[46,47] incorporating the kinship matrix and adjusting for age and sex. In each study, variants with minor allele count (MAC) <10 were removed before meta-analysis. Meta-analysis was performed in METAL. In the meta-analysis, the sample-size-weighted approach was used to estimate P values and an inverse-variance-weighted approach was used to calculate pooled effect estimates and corresponding standard errors. Study-specific information is provided in **Supplementary Table 19**.

Phenome/biomarker scan analyses. We downloaded online-available GWAS data from 12 consortia for 70 traits (**Supplementary Table 9**) and harmonized genome positions to Build37/hg19. We then performed a lookup for the newly discovered T2D SNPs using these harmonized data sets. We also performed a phenotypic scan for the same new T2D SNPs across 105 biomarkers measured in the PROMIS participants using a linear regression model adjusted for the first five principal components (**Supplementary Table 10**). We used a Bonferroni-adjusted P-value cutoff of 1.8×10^{-5} (= 0.05/175 traits/16 SNPs) to declare statistical significance.

Coronary heart disease meta-analysis. We assembled 56,354 samples of European, East Asian, and South Asian ancestry genotyped on the CardioMetabochip to identify genetic determinants of CHD. These results were combined with those reported by CARDIoGRAMplusC4D to yield analyses comprising 260,365 subjects (90,831 CHD cases) for CHD. Additional new CHD studies comprised 16,093 CHD cases and 16,616 unaffected individuals: EPIC-CVD study, a case–cohort study recruited across ten European countries; the Copenhagen City Heart Study (CCHS), the Copenhagen Ischemic Heart Disease Study (CIHDS), and the Copenhagen General Population Study (CGPS) all recruited within Copenhagen, Denmark; the South Asian studies

comprised up to 7,654 CHD cases and 7,014 controls from the Pakistan Risk of Myocardial Infarction Study (PROMIS), a case–control study that recruited samples from nine sites in Pakistan, and the Bangladesh Risk of Acute Vascular Events (BRAVE) study based in Dhaka, Bangladesh; the East Asian studies comprised 4,129 CHD cases and 6,369 controls recruited from seven studies across Taiwan that collectively make up the TAIwan metaboCHIp (TAICHI) Consortium. Samples from EPIC-CVD, CCHS, CIHDS, CGPS, BRAVE, and PROMIS were all genotyped on a customized version of the Illumina CardioMetabochip, referred to as Metabochip+, in two Illumina-certified laboratories located in Cambridge, UK, and Copenhagen, Denmark. TAICHI samples were genotyped using the latest version of the CardioMetabochip. For each study, samples were removed if they had CR <0.97, had average heterozygosity >±3 s.d. from the overall mean heterozygosity, or the genotypic sex did not match the reported sex. One of each pair of duplicate samples and first-degree relatives (identified by a kinship coefficient >0.2) was removed. CardioMetabochip data were also obtained from the Women's Health Initiative Study and the ARIC study; the two studies underwent the same quality control as described for the TAICHI study. Across all studies, SNP exclusions were based on MAF < 0.01, $P < 1 \times 10^{-6}$ for Hardy–Weinberg equilibrium, or CR < 0.97. CARDIoGRAMplusC4D Consortium data were obtained online (see URLs). Only non-overlapping samples were used for meta-analyses. Fixed-effects inverse-variance-weighted meta-analysis was used to combine the effects across studies in METAL[45].

Genetic risk score analysis. We used a two-sample MR method[48] to estimate effects for a multi-SNP genetic instrument by using summary statistics. This method has previously been validated to infer causal effects (odds ratios) and associated standard error[49]. Briefly, association data for both T2D and CHD were obtained using data from two separate genome-wide meta-analyses. For T2D, we used the data from the current meta-analyses, whereas we used data from the most recent CHD meta-analyses as described in the **Supplementary Note**. Using sentinel SNPs for all established T2D associations, we identified a set of variants ($n = 16$) exclusively associated with T2D by screening against the GWAS catalog of publicly available data[50] for anthropometric traits (BMI, waist–hip ratio, waist circumference, waist–hip ratio adjusted for BMI, waist circumference adjusted for BMI, and hip), glucose/insulin or MAGIC traits (fasting glucose, 2-h glucose, fasting insulin, and proinsulin levels), blood lipids (HDL-C, LDL-C, and triglycerides), and blood pressure (systolic and diastolic). We next attempted to group the remaining pleiotropic T2D SNPs into different categories on the basis of their observed associations for various cardiometabolic intermediate traits ($P < 0.01$). These groupings included (i) variants associated with glucose/insulin traits only ($n = 13$), (ii) variants associated with triglycerides/HDL-C and waist circumference/waist–hip ratio but not glucose/insulin, blood pressure, LDL-C, or BMI ($n = 6$), (iii) variants associated with triglycerides/HDL-C and obesity/anthropometric traits but not glucose/insulin, blood pressure, or LDL-C ($n = 6$), (iv) variants associated with triglycerides/HDL-C, blood pressure, and BMI but not glucose/insulin or LDL-C ($n = 8$), and (v) variants associated with triglycerides/HDL-C, blood pressure, BMI, and glucose/insulin but not blood pressure or LDL-C ($n = 24$) (**Supplementary Table 12**). Established T2D SNPs that did not fall into any of these categories were excluded. Heterogeneity in odds ratios was assessed via Cochran's Q test for heterogeneity.

T2D and CHD enrichment analysis. We used a binomial distribution with baseline enrichment probability P_b to derive the density for the test statistic $E \sim \text{Binomial}(n, P_b)$, where n is the number of SNPs in a variant set. E is the number of SNPs with a directionally consistent effect on T2D and CHD (the allele that increases the risk for T2D also increases the risk for CHD). Using SNPs that were not associated with T2D or CHD ($P \geq 0.05$ for T2D and CHD), we calculated the percentage of SNPs with a directionally consistent effect in T2D and CHD and used this as an estimate for P_b. We then performed the enrichment analysis in two variant sets: (i) the variant set with all variants available and (ii) the variant set with LD-clumped variants. The results are shown in **Supplementary Table 15**. In the LD clumping procedure, the SNPs with more significant T2D P values were retained as seeds and the other SNPs that were in LD ($r^2 > 0.1$ based on data from the 1000 Genomes Project (phase 3, v5 variant set)) with the seed SNPs were removed[43].

Estimating the T2D–CHD bivariate normal density. To establish the T2D–CHD bivariate normal density, we used all variants that we identified in our analyses on T2D and CHD; we further pruned them for LD with 1000 Genomes Project data (Phase 3, v5 variant set) to $r^2 \leq 0.1$ using PLINK[51,52]. The reference and alternate alleles of the variants that survived LD pruning were retrieved from the same 1000 Genomes VCF file used for pruning, and the variants' effects on CHD and T2D were aligned to their reference alleles. The statistics used to estimate the bivariate normal density were produced using the following formula:

$$Z_{\text{CHD}} = \Phi^{-1}\left(\frac{P_{\text{CHD}}}{2}\right) \times \frac{\beta_{\text{CHD}}}{|\beta_{\text{CHD}}|}, Z_{\text{T2D}} = \Phi^{-1}\left(\frac{P_{\text{T2D}}}{2}\right) \times \frac{\beta_{\text{T2D}}}{|\beta_{\text{T2D}}|} \quad (1)$$

where Φ^{-1} is the inverse-cumulative distribution function of the standard normal distribution, P_{CHD} and P_{T2D} are the P values for CHD and T2D, respectively, and β_{CHD} and β_{T2D} are the effect sizes of the reference allele on CHD and T2D, respectively. Because a successful estimation of the bivariate distribution depends on both positive and negative Z-scores, we used the signs of the corresponding effect estimates ($\beta/|\beta|$) to determine the signs of Z_{CHD} and Z_{T2D}. The distributions of Z_{CHD} and Z_{T2D} are shown in **Supplementary Figure 5**. Parameters for the bivariate normal density were estimated using the mvn.ub() function in the R package miscF. The estimated bivariate normal density has the following parameter values.

$$\begin{pmatrix}\mu_{\text{CHD}}\\ \mu_{\text{CHD}}\end{pmatrix} = \begin{pmatrix}-0.00676\\ -0.00731\end{pmatrix}, \Sigma = \begin{pmatrix}1.0361 & 0.1435\\ 0.1435 & 1.0265\end{pmatrix} \quad (2)$$

Two-degree-of-freedom test under the bivariate normal density. Assuming that Z is distributed as a bivariate normal, then

$$Y = \Sigma^{-\frac{1}{2}}(Z-\mu) \sim \text{N}_2(0, I) \quad \text{and}$$

$$(Z-\mu)'\Sigma^{-1}(Z-\mu) = \sum_{i=1}^{2} Y_i^2 \sim \chi_2^2 \quad (3)$$

where $\text{N}_2(\mu, \Sigma)$ denotes a bivariate normal distribution with a vector of means μ and variance–covariance matrix Σ, and χ_2^2 is the chi-squared distribution with 2 degrees of freedom. Using Y as the test statistic, we performed a two-degree-of-freedom test on $Z = (Z_{\text{CHD}}, Z_{\text{T2D}})$, and our null hypothesis was that a SNP was not associated with either of the two traits. **Supplementary Figure 5** depicts the rejection region of the two-degree-of-freedom test.

Conditional analysis for *CCDC92*. We performed approximate conditional analysis for the *CCDC92* locus using the software package GCTA[53]. We used the summary meta-analysis data for our primary T2D and CHD scans (before replication) as data input from each continental group (European, South Asian, and East Asian). As the reference input, we used population data from the 1000 Genomes Project (version 3) matching the continental ancestry for the respective conditional analysis. We then conditioned on rs825476—the lead SNP associated with CHD and T2D—for each continental group. We combined the summary results from each continental group via inverse-weighted fixed-effects meta-analysis. LocusZoom plots for these conditional meta-analysis association results are presented in **Supplementary Figure 8**.

Colocalization analysis. To determine whether the T2D and CHD association signals colocalized to the same genetic variant, we used the R package coloc. For each of the 19 loci that met our T2D–CHD association criteria, we obtained association data from all SNPs within 500 kb of the sentinel bivariate associated SNP (**Supplementary Table 16**). From there, we used the coloc.abf() function to calculate the probability that both traits are associated and share a single causal variant (H_4), using the P values from the overall inverse-variance fixed-effects meta-analysis for T2D (without replication) and CHD, the overall case–control sample sizes for both scans, and the allele frequencies for the variant based on all 1000 Genomes data (version 3). We call variants colocalized if the H_4 colocalization probability was greater than 0.5.

doi:10.1038/ng.3943

Selection of loci for connectivity and ontology analyses. For T2D, we used the previously reported loci[5] ($n = 88$) and the loci discovered in this report ($n = 16$). For CHD, we used the previously reported loci described in the most recent report published by the CARDIoGRAMplusC4D Consortium ($n = 58$)[6]. Prioritization of genes from this list of established loci for T2D and CHD (**Supplementary Table 17**) was based on evidence from monogenic association with disease[53], coding mutations in nearby genes, functional evidence implicating genes, or the gene nearest to the sentinel SNP. For T2D–CHD associations arising from our bivariate scan, we first pruned the data set for LD ($r^2 < 0.1$). We further selected 299 LD-independent SNPs that were found to be associated with T2D and CHD with $P < 0.001$ in our bivariate scan and used them to identify underlying candidate genes using GRAIL[19]. For protein–protein interaction connectivity analysis, we used DAPPLE[20] on the 79 loci that were found to be significant in GRAIL[19]. Empirical significance for excess connectivity in protein–protein interactions was assessed by 10,000 permutations.

Ontology analysis and drug target annotations. We used the online tool WebGestalt[21] to perform ontology enrichment analysis. For analysis of the query loci, we nominated genes ($n = 79$) that were prioritized from text mining (GRAIL $P < 0.05$). We also performed ontology analyses using separate gene lists for T2D ($n = 104$) and CHD ($n = 58$) loci separately. The hypergeometric distribution was used to assess significance, and adjustment for multiple testing was controlled for using the Benjamini–Hochberg procedure[54] implemented in WebGestalt[21].

Data availability. Summary GWAS estimates for the T2D meta-analysis and bivariate summary data, respectively, are publicly available in the following files: http://www.med.upenn.edu/ccebfiles//t2d_meta_cleaned.zip, and http://www.med.upenn.edu/ccebfiles/chd_t2d_af_gwas12_cleaned_combined_1000gRefAlt_added_pvalRescaled_varSetID_added.zip. A **Life Sciences Reporting Summary** is available.

43. Auton, A. *et al.* A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
44. Howie, B.N., Donnelly, P. & Marchini, J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet.* **5**, e1000529 (2009).
45. Willer, C.J., Li, Y. & Abecasis, G.R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
46. Feng, S., Liu, D., Zhan, X., Wing, M.K. & Abecasis, G.R. RAREMETAL: fast and powerful meta-analysis for rare variants. *Bioinformatics* **30**, 2828–2829 (2014).
47. Liu, D.J. *et al.* Meta-analysis of gene-level tests for rare variant association. *Nat. Genet.* **46**, 200–204 (2014).
48. Dastani, Z. *et al.* Novel loci for adiponectin levels and their influence on type 2 diabetes and metabolic traits: a multi-ethnic meta-analysis of 45,891 individuals. *PLoS Genet.* **8**, e1002607 (2012).
49. Evans, D.M. & Davey Smith, G. Mendelian randomization: new applications in the coming age of hypothesis-free causality. *Annu. Rev. Genomics Hum. Genet.* **16**, 327–350 (2015).
50. Welter, D. *et al.* The NHGRI GWAS Catalog, a curated resource of SNP–trait associations. *Nucleic Acids Res.* **42**, D1001–D1006 (2014).
51. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
52. Yang, J. *et al.* Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat. Genet.* **44**, 369–375 (2012).
53. Doria, A., Patti, M.E. & Kahn, C.R. The emerging genetic architecture of type 2 diabetes. *Cell Metab.* **8**, 186–200 (2008).
54. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* **57**, 289–300 (1995).

nature research

Corresponding Author: Danish Saleheen

Date: Jul 26, 2017

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work we publish. This form is published with all life science papers and is intended to promote consistency and transparency in reporting. All life sciences submissions use this form; while some list items might not apply to an individual manuscript, all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

Experimental design

1. Sample size

   Describe how sample size was determined.

   For our meta-analysis, we aimed to conduct the largest meta-analyses by generation new data and assembling publicly available information.

2. Data exclusions

   Describe any data exclusions.

   A brief description on all participating studies has been provided in the supplementary note. If participants were excluded by any particular study, details have been provided in the supplementary note. No animal studies have been conducted in the current analyses.

3. Replication

   Describe whether the experimental findings were reliably reproduced.

   We replicated our findings by assembling datasets independent of our discovery studies; only those genetic variants which were successfully replicated were declared to be novel in association with type-2 diabetes

4. Randomization

   Describe how samples/organisms/participants were allocated into experimental groups.

   N/A

5. Blinding

   Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

   N/A

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

6. Statistical parameters

   For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or the Methods section if additional space is needed).

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
☐	☒	A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly.
☐	☒	A statement indicating how many times each experiment was replicated
☐	☒	The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)
☐	☒	A description of any assumptions or corrections, such as an adjustment for multiple comparisons
☐	☒	The test results (e.g. *p* values) given as exact values whenever possible and with confidence intervals noted
☐	☒	A summary of the descriptive statistics, including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
☐	☒	Clearly defined error bars

See the web collection on statistics for biologists for further resources and guidance.

Nature Genetics: doi:10.1038/ng.3943

Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this study.	All analyses were conducted in SNPTEST, PLINK, R and STATA which are available to the wider scientific community. Methods to perform the bivariate scan are available through a public github repository. All other tools used in the manuscript derive from computational tools that are publicly available.

For all studies, we encourage code deposition in a community repository (e.g. GitHub). Authors must make computer code available to editors and reviewers upon request. The *Nature Methods* guidance for providing algorithms and software for publication may be useful for any submission.

Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.	N/A

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).	N/A

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.	N/A
b. Describe the method of cell line authentication used.	N/A
c. Report whether the cell lines were tested for mycoplasma contamination.	N/A
d. If any of the cell lines used in the paper are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.	N/A

Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.	N/A

Policy information about studies involving human research participants

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.	A brief description on all participating studies has been provided in the supplementary note.

Nature Genetics: doi:10.1038/ng.3943