LETTER

doi:10.1038/nature12828

Sequence variants in *SLC16A11* are a common risk factor for type 2 diabetes in Mexico

The SIGMA Type 2 Diabetes Consortium*

Performing genetic studies in multiple human populations can identify disease risk alleles that are common in one population but rare in others[1], with the potential to illuminate pathophysiology, health disparities, and the population genetic origins of disease alleles. Here we analysed 9.2 million single nucleotide polymorphisms (SNPs) in each of 8,214 Mexicans and other Latin Americans: 3,848 with type 2 diabetes and 4,366 non-diabetic controls. In addition to replicating previous findings[2–4], we identified a novel locus associated with type 2 diabetes at genome-wide significance spanning the solute carriers *SLC16A11* and *SLC16A13* ($P = 3.9 \times 10^{-13}$; odds ratio (OR) = 1.29). The association was stronger in younger, leaner people with type 2 diabetes, and replicated in independent samples ($P = 1.1 \times 10^{-4}$; OR = 1.20). The risk haplotype carries four amino acid substitutions, all in SLC16A11; it is present at ~50% frequency in Native American samples and ~10% in east Asian, but is rare in European and African samples. Analysis of an archaic genome sequence indicated that the risk haplotype introgressed into modern humans via admixture with Neanderthals. The *SLC16A11* messenger RNA is expressed in liver, and V5-tagged SLC16A11 protein localizes to the endoplasmic reticulum. Expression of SLC16A11 in heterologous cells alters lipid metabolism, most notably causing an increase in intracellular triacylglycerol levels. Despite type 2 diabetes having been well studied by genome-wide association studies in other populations, analysis in Mexican and Latin American individuals identified *SLC16A11* as a novel candidate gene for type 2 diabetes with a possible role in triacylglycerol metabolism.

The Slim Initiative in Genomic Medicine for the Americas (SIGMA) Type 2 Diabetes Consortium set out to characterize the genetic basis of type 2 diabetes in Mexican and other Latin American populations, where the prevalence is roughly twice that of US non-Hispanic whites[5] (see also http://www.cdc.gov/diabetes/pubs/factsheet11.htm). This report considers 3,848 type 2 diabetes cases and 4,366 controls (Table 1) genotyped using the Illumina OMNI 2.5 array that were unrelated to other samples, and that fall on a cline of Native American and European ancestry[6] (Extended Data Fig. 1). Association analysis included 9.2 million variants that were imputed[7,8] from the 1000 Genomes Project Phase I release[9] based on 1.38 million SNPs directly genotyped at high quality with minor allele frequency (MAF) >1%.

The association of SNP genotype with type 2 diabetes was evaluated using LTSOFT[10], a method that increases power by jointly modelling case–control status with non-genetic risk factors. Our analysis used body mass index (BMI) and age to construct liability scores and also included adjustment for sex and ancestry via principal components[6]. The quantile–quantile (QQ) plot is well calibrated under the null ($\lambda_{GC} = 1.05$; Fig. 1a, red), indicating adequate control for confounders, with substantial excess signal at $P < 10^{-4}$.

We first examined SNPs previously reported to be associated to risk of type 2 diabetes. Two such variants reached genome-wide significance: *TCF7L2* (rs7903146; $P = 2.5 \times 10^{-17}$; OR = 1.41 (95% confidence interval 1.30–1.53)) and *KCNQ1* (rs2237897; $P = 4.9 \times 10^{-16}$; OR = 0.74 (0.69–0.80)) (Extended Data Figs 2, 3a), with effect sizes and frequencies consistent with previous studies[3,4,11]. At *KCNQ1*, we identified a signal[3] of association that shows limited linkage disequilibrium both to rs2237897 ($r^2 = 0.056$) and to rs231362 ($r^2 = 0.028$) (previously seen in Europeans[11]), suggesting a third allele at this locus (rs139647931; after conditioning, $P = 5.3 \times 10^{-8}$; OR = 0.78 (0.70–0.86); Extended Data Fig. 3b and Supplementary Note).

More generally, of SNPs previously associated with type 2 diabetes at genome-wide significance, 56 of 68 are directionally consistent with the initial report ($P = 3.1 \times 10^{-8}$; Supplementary Table 1). Nonetheless, a QQ plot excluding all SNPs within 1 megabase (Mb) of the 68 type 2 diabetes associations remains strikingly non-null (Fig. 1a, blue).

This excess signal of association is entirely attributable to two regions of the genome: chromosome 11p15.5 and 17p13.1 (Fig. 1a, black). The genome-wide significant association at 11p15.5 spans insulin, *IGF2* and other genes (Extended Data Fig. 3a): the SNP with the strongest association lies in the 3′ untranslated region (UTR) of *IGF2* and the non-coding *INS–IGF2* transcript (rs11564732, $P = 2.6 \times 10^{-8}$; OR = 0.77 (0.70–0.84); Supplementary Table 2). The associated SNPs are ~700 kilobases (kb) from the genome-wide significant signal in *KCNQ1* (above), and analysis conditional on the two significant *KCNQ1* SNPs reduced the *INS–IGF2* association signal to just below genome-wide significance ($P = 7.5 \times 10^{-7}$, Extended Data Fig. 3c). Conditioning on the two *KCNQ1* SNPs and the *INS–IGF2* SNP reduces the signal to background (Extended Data Fig. 3d). Further analysis is needed to determine whether the *INS–IGF2* signal is reproducible and independent of that at *KCNQ1*.

Table 1 | Study cohorts comprising the SIGMA type 2 diabetes project data set

Study	Sample location	Study design		*n* (before quality control)	Per cent male	Age (years)	Age-of-onset (years)	BMI (kg m^{-2})	Fasting plasma glucose (mmol l^{-1})
UNAM/INCMNSZ Diabetes Study (UIDS)	Mexico City, Mexico	Prospective cohort	Controls	1,138 (1,195)	41.1	55.3 ± 9.4	–	28.1 ± 4.0	4.8 ± 0.5
			T2D cases	815 (872)	40.9	56.2 ± 12.3	44.2 ± 11.3	28.4 ± 4.5	–
Diabetes in Mexico Study (DMS)	Mexico City, Mexico	Prospective cohort	Controls	472 (505)	25.8	52.5 ± 7.7	–	28.0 ± 4.4	5.0 ± 0.4
			T2D cases	690 (762)	33.0	55.8 ± 11.1	47.8 ± 10.6	29.0 ± 5.4	–
Mexico City Diabetes Study (MCDS)	Mexico City, Mexico	Prospective cohort	Controls	613 (790)	39.3	62.5 ± 7.7	–	29.4 ± 4.8	5.0 ± 0.5
			T2D cases	287 (358)	41.1	64.2 ± 7.5	55.1 ± 9.7	29.9 ± 5.4	–
Multiethnic Cohort (MEC)	Los Angeles, California, USA	Case–control	Controls	2,143 (2,464)	48.3	59.3 ± 7.0	–	26.6 ± 3.9	N/A
			T2D cases	2,056 (2,279)	47.9	59.2 ± 6.9	N/A	30.0 ± 5.4	–

The table shows sample location, study design, numbers of cases and controls (including numbers before quality control checks), per cent male participants, age ± standard deviation (s.d.), age-of-onset in cases ± s.d., body mass index ± s.d., and fasting plasma glucose in controls ± s.d. N/A, not applicable; T2D, type 2 diabetes.

*Lists of participants and their affiliations appear at the end of the paper.

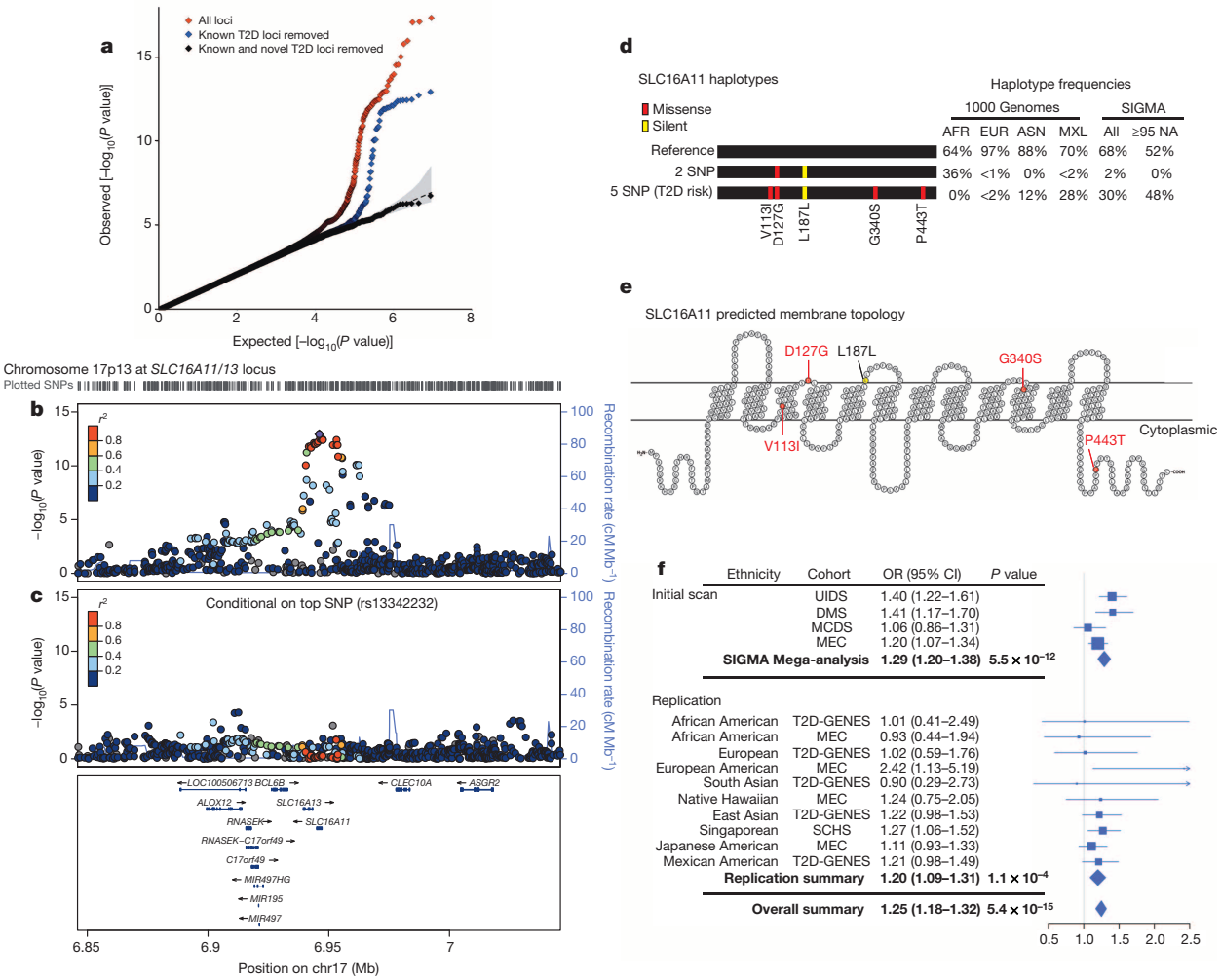


Figure 1 | Identification of a novel type 2 diabetes risk haplotype carrying 5 SNPs in *SLC16A11*. **a**, QQ plot of association statistics in genome-wide scan of $n = 8{,}214$ samples shows calibration under the null and enrichment in the tail for all SNPs (red), and after removing SNPs within 1 Mb of previously published type 2 diabetes associations (blue). Removal of sites within 1 Mb of 68 known loci and two novel loci results in a null distribution (black). Association with liability threshold quantitative traits tested via linear regression. T2D, type 2 diabetes. **b**, Regional plot of association at 17p13.1 that spans *SLC16A11* and *SLC16A13*. **c**, Analysis conditional on genotype at rs13342232 (the top associated variant) reduces signal to far below genome-wide significance across the surrounding region. Colour indicates r^2 to the most strongly associated site; recombination rate is shown, each based on the 1000 Genomes ASN population. **d**, Graphical depictions of SLC16A11 haplotypes constructed from the synonymous and four missense SNPs associated to type 2 diabetes, with haplotype frequencies derived from the 1000 Genomes Project and SIGMA samples. AFR, African ($n = 185$); ASN, east Asian ($n = 286$); EUR, European ($n = 379$); MXL, Mexican samples from Los Angeles ($n = 66$). Frequencies from SIGMA samples are calculated from genotypes and represent either the entire data set (All) or only samples estimated to have ≥95% Native American ancestry (≥95 NA, $n = 290$; Supplementary Methods). Haplotypes with population frequency <1% are not depicted. **e**, Predicted membrane topology of human SLC16A11 generated using TMHMM 2.0 and visualized with TeXtopo. Locations of SNPs carried by the type-2-diabetes-associated haplotype are indicated. **f**, Forest plot depicting odds ratio estimates at rs75493593 from the four SIGMA cohorts, the SIGMA pooled mega-analysis, the replication cohorts, replication-only meta-analysis based on inverse standard error weighting of effect sizes, and the overall meta-analysis (including all replication cohorts and the SIGMA mega-analysis). Accompanying table lists ethnicity, cohort names, estimated odds ratio (OR) and 95% confidence interval (95% CI). Replication cohorts are the Type 2 Diabetes Genetic Exploration by Next-generation sequencing in multi-Ethnic Samples (T2D-GENES), Multiethnic Cohort (MEC), and Singapore Chinese Health Study (SCHS). Further details including sample sizes are provided in Supplementary Table 8.

The strongest novel association is at 17p13.1 spanning *SLC16A11* and *SLC16A13* (Fig. 1b), both poorly characterized members of the monocarboxylic acid transporter family of solute carriers[12]. The strongest signal of association includes a silent mutation as well as four missense SNPs, all in *SLC16A11* (Fig. 1d, e). These five variants are (1) in strong linkage disequilibrium ($r^2 \geq 0.85$ in 1000 Genomes samples from the Americas) and co-segregate on a single haplotype; (2) common in samples of Latin American ancestry; and (3) show equivalent levels of association to type 2 diabetes ($P = 2.4 \times 10^{-12}$ to $P = 3.9 \times 10^{-13}$; OR = 1.29 (1.20–1.38); Supplementary Tables 3–5). Analysis conditional on any of these variants leaves no genome-wide significant signal (Fig. 1c and Extended Data Fig. 4). Computational prediction with SIFT[13] (which considers each site independently) labels one of the missense SNPs (rs13342692, D127G) as damaging and the other three 'tolerated' (Supplementary Table 6).

Individuals that carry the risk haplotype develop type 2 diabetes 2.1 years earlier ($P = 3.1 \times 10^{-4}$), and at 0.9 kg m^{-2} lower BMI ($P = 5.2 \times 10^{-4}$) than non-carriers (Extended Data Fig. 5). The odds ratio for the risk haplotype estimated using young cases (≤45 years) was higher than in older cases (OR = 1.48 versus 1.11; $P_{heterogeneity} = 1.7 \times 10^{-3}$). We tested the haplotype for association with related metabolic quantitative traits in the fasting state in a subset of SIGMA participants ($n =$ 1,505–3,855). No associations surpass nominal significance ($P < 0.05$; Supplementary Table 7).

Given that large genome-wide association studies (GWAS) have been performed for type 2 diabetes in samples of European and Asian ancestry, it may seem surprising that associated variants at *SLC16A11/13* were not previously identified. Using data generated by the 1000 Genomes Project and the current study, we observed that the risk haplotype (hereafter referred to as '5 SNP' haplotype) is rare or absent in samples from Europe and Africa, has intermediate frequency (~10%) in samples from east Asia, and up to ~50% frequency in samples from the Americas (Fig. 1d and Extended Data Fig. 6a). A second haplotype carrying one of the four missense SNPs (D127G) and the synonymous variant (termed the '2 SNP' haplotype) is very common in samples from Africa but rare elsewhere, including in the Americas (Fig. 1d). The low frequency of the 5 SNP haplotype in Africa and Europe may explain why this association was not found in previous studies.

We attempted to replicate this association in ~22,000 samples from a variety of ancestry groups. A proxy for the 5 SNP haplotype of *SLC16A11* showed strong association with type 2 diabetes ($P_{replication} = 1.1 \times 10^{-4}$; $OR_{replication} = 1.20$ (1.09–1.31); $P_{combined} = 5.4 \times 10^{-15}$; $OR_{combined} = 1.25$ (1.18–1.32); Fig. 1f and Supplementary Table 8). The association was clearly observed in east Asian samples, a population that lacks admixture of Native American and European populations and shows little genetic substructure. This result argues against population stratification as an explanation for the finding in Latin American populations.

We estimated the difference in disease prevalence attributable to a risk factor with OR = 1.20 (1.09–1.31), 26% frequency in Mexican Americans (as in the SIGMA control samples) and 2% in European Americans. Approximately 20% (9.2–29%) of the difference in prevalence could be explained by such a risk factor (Supplementary Methods).

Two population genetic features of the 5 SNP haplotype struck us as discordant. The haplotype sequence is highly divergent, with an estimated time to most recent common ancestor (TMRCA) of 799,000 years to a European haplotype (Supplementary Table 9 and Supplementary Note). This long precedes the 'out of Africa' bottleneck. And yet, the haplotype is not observed in Africa and is rare throughout Europe (Fig. 1d).

This combination of age and geographical distribution could be consistent with admixture from Neanderthals into modern humans. Neither the published Neanderthal genome[14] nor the Denisova genome[15] contained the variants observed on the 5 SNP haplotype. However, an unpublished genome of a Neanderthal from Denisova Cave[16,17] is homozygous across 5 kb for the 5 SNP haplotype at *SLC16A11*, including all four missense SNPs. Over a span of 73 kb this Neanderthal sequence is nearly identical to that of individuals from the 1000 Genomes Project who are homozygous for the 5 SNP haplotype (Supplementary Note).

Two lines of evidence indicate that the 5 SNP haplotype entered modern humans through archaic admixture. First, the Neanderthal sequence is more closely related to the extended 73 kb 5 SNP haplotype than to random non-risk haplotypes (mean TMRCA = 250,000 years versus 677,000 years; Supplementary Tables 10 and 11 and Supplementary Note), forming a clade with the risk haplotype (Extended Data Fig. 6b) with a coalescence time that post-dates the range of estimated split times between modern humans and Neanderthals[15,18]. Second, the genetic length of the 73-kb haplotype is longer than would be expected if it had undergone recombination for ~9,000 generations since the split with Neanderthals ($P = 3.9 \times 10^{-5}$; Supplementary Note). These two features indicate that the 5 SNP haplotype is not only similar to the Neanderthal sequence, but was probably introduced into modern humans relatively recently through archaic admixture. We note that whereas this particular Neanderthal-derived haplotype is common in the Americas, Latin Americans have the same proportion of Neanderthal ancestry genome-wide as other Eurasian populations (~2%)[15].

With an absence of multiple independently segregating functional mutations in the same gene, we lack formal genetic proof that *SLC16A11* is the gene responsible for association to type 2 diabetes at 17p13.1. Nonetheless, as the associated haplotype encodes four missense SNPs in a single gene (Supplementary Table 12), we set out to begin characterizing the function of SLC16A11.

We examined the tissue distribution of *SLC16A11* mRNA expression using Nanostring and ~55,000 curated microarray samples. In both data sets, we observed *SLC16A11* expression in liver, salivary gland and thyroid (Extended Data Figs 7 and 8). We used immunofluorescence to determine the subcellular localization of V5-tagged SLC16A11 introduced into HeLa cells. SLC16A11–V5 co-localizes with the endoplasmic reticulum membrane protein calnexin, but shows minimal overlap with plasma membrane, Golgi apparatus and mitochondria (Fig. 2a). Distinct patterns were seen for other SLC16 family members, which are known to have diverse cellular functions[19]: SLC16A13–V5 localizes to the Golgi apparatus and SLC16A1–V5 appears at the plasma membrane[20] (Extended Data Fig. 9 and data not shown).

As SLC16 family members are solute carriers, we expressed SLC16A11 (or control proteins) in HeLa cells (which do not express SLC16A11 at

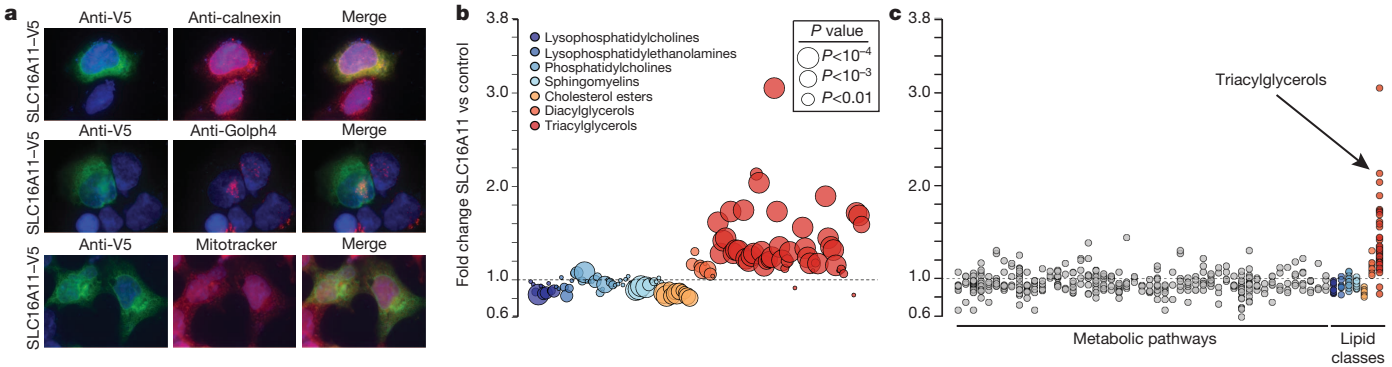


Figure 2 | SLC16A11 localizes to the endoplasmic reticulum and alters lipid metabolism in HeLa cells. **a**, Localization of SLC16A11 to the endoplasmic reticulum. HeLa cells expressing C terminus, V5-tagged SLC16A11 were immunostained for SLC16 expression (anti-V5) along with markers for the endoplasmic reticulum (anti-calnexin), cis-Golgi apparatus (anti-Golph4), or mitochondria (MitoTracker). Imaging of each protein was optimized for clarity of localization rather than comparison of expression level across proteins. Representative images from multiple independent transfections are shown. **b**, Changes in intracellular lipid metabolites after expression of SLC16A11–V5 in HeLa cells. The fold change in cells expressing SLC16A11 relative to cells expressing control proteins is plotted for individual lipid metabolites, with lipid classes indicated by point colour and *P* values (of the Wilcoxon rank-sum test) by point size. **c**, Fold change plotted for both polar and lipid metabolites, grouped according to metabolic pathway or class. Pathways shown include all KEGG pathways from the human reference set for which metabolites were measured as well as eight additional classes of metabolites covering carnitines and lipid subtypes. Each point within a pathway or class shows the fold change of a single metabolite within that pathway or class. Pathway names and statistical analyses are shown in Extended Data Fig. 10 and Supplementary Table 14. Metabolite data shown are the combined results from three independent experiments, each of which included 12 biological replicates each for SLC16A11 and control.

appreciable levels) and profiled ~300 polar and lipid metabolites. Expression of SLC16A11 resulted in substantial increases in triacylglycerol (TAG) levels ($P = 7.6 \times 10^{-12}$), with smaller increases in intracellular diacylglycerols ($P = 7.8 \times 10^{-3}$) and decreases in lysophosphatidylcholine ($P = 2.0 \times 10^{-3}$), cholesterol ester ($P = 9.8 \times 10^{-4}$) and sphingomyelin ($P = 3.9 \times 10^{-3}$) lipids (Fig. 2b, c and Supplementary Tables 13 and 14). As TAG synthesis takes place in the endoplasmic reticulum in the liver[21], these results indicate that SLC16A11 may have a role in hepatic lipid metabolism. We note that serum levels of specific TAGs have been prospectively associated with future risk of type 2 diabetes[22] and accumulation of intracellular lipids has been implicated in insulin resistance in human populations[23,24].

In summary, GWAS in Mexican and other Latin American samples identified a haplotype containing four missense SNPs, all in *SLC16A11*, that is much more common in individuals with Native American ancestry than in other populations. Each haplotype copy is associated with a ~20% increased risk of type 2 diabetes. With these properties, the haplotype would be expected to contribute to the higher burden of type 2 diabetes in Mexican and Latin American populations[25]. The haplotype derives from Neanderthal introgression, providing an example of Neanderthal admixture affecting physiology and disease susceptibility today. Our data suggest the hypothesis for future studies that *SLC16A11* may influence diabetes risk through effects on lipid metabolism in the liver. Our results also indicate that genetic mapping in understudied populations can identify previously undiscovered aspects of disease pathophysiology[1].

Note added in proof: While this paper was in final revision, Hara *et al.* reported[29] a SNP in *SLC16A13* (rs312457) as associated with risk of T2D in an east Asian population with OR = 1.20, $P = 10^{-12}$.

METHODS SUMMARY

DNA samples were prepared using strict quality control procedures and genotyped using the Illumina HumanOmni2.5 array. Stringent sample and SNP quality (including ancestry) filters were applied on the resulting genotypes. After imputation[7,8], SNPs were quality filtered (MAF ≥1% and info score ≥0.6) and association testing was performed via LTSOFT[10] with type 2 diabetes status, BMI, and age modelling liability and adjusting for sex and top two principal components as fixed effect covariates. *P* values were corrected for genomic control ($\lambda_{GC} = 1.046$). Odds ratios (ORs) are from logistic regression in PLINK[26] using BMI, age, sex, and top 2 principal components as covariates. Proportion of Native American ancestry was estimated using ADMIXTURE[27] ($K = 3$) run including unadmixed individuals from several populations.

Odds ratios for young (≤45 years) and older age of onset cases were calculated using logistic regression in each group compared to two randomly selected non-overlapping sets of controls. Significance testing used a *Z*-score calculated from these odds ratios.

Population prevalence was modelled using odds ratio to approximate relative risk in a log-additive effect model[28]. Relative change in population prevalences is reported based on removing a locus with relative risk of 1.20 and the indicated frequency.

Gene expression analyses were performed on data collected using Nanostring and a compendium of publicly available Affymetrix U133 Plus 2.0 microarrays. The subcellular localization of SLC16A11–V5 and metabolic profiling studies were performed after expression of carboxy-terminus, V5-tagged SLC16A11 in HeLa cells. Metabolite values were normalized to the total metabolite signal obtained for each sample. Measurements were obtained in replicate from each of three independent experiments, with data combined after subtracting the mean of the log-transformed values. The Wilcoxon rank sum test was used to test for differences in individual metabolite levels in cells expressing SLC16A11 compared to controls; the Wilcoxon signed rank test was used to assess differences in lipid classes.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 30 November 2012; accepted 4 November 2013.
Published online 25 December 2013.

1. Rosenberg, N. A. *et al.* Genome-wide association studies in diverse populations. *Nature Rev. Genet.* **11,** 356–366 (2010).
2. Grant, S. F. A. *et al.* Variant of transcription factor 7-like 2 (*TCF7L2*) gene confers risk of type 2 diabetes. *Nature Genet.* **38,** 320–323 (2006).
3. Unoki, H. *et al.* SNPs in *KCNQ1* are associated with susceptibility to type 2 diabetes in East Asian and European populations. *Nature Genet.* **40,** 1098–1102 (2008).
4. Yasuda, K. *et al.* Variants in *KCNQ1* are associated with susceptibility to type 2 diabetes mellitus. *Nature Genet.* **40,** 1092–1097 (2008).
5. Villalpando, S. *et al.* Prevalence and distribution of type 2 diabetes mellitus in Mexican adult population: a probabilistic survey. *Salud Publica Mex.* **52,** S19–S26 (2010).
6. Patterson, N., Price, A. L. & Reich, D. Population structure and eigenanalysis. *PLoS Genet.* **2,** e190 (2006).
7. 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).
8. Williams, A. L., Patterson, N., Glessner, J., Hakonarson, H. & Reich, D. Phasing of many thousands of genotyped samples. *Am. J. Hum. Genet.* **91,** 238–251 (2012).
9. The 1000 Genomes Project Consortium. An integrated map of genetic variation from 1,092 human genomes. *Nature* **491,** 56–65 (2012).
10. Zaitlen, N. *et al.* Informed conditioning on clinical covariates increases power in case-control aAssociation Studies. *PLoS Genet.* **8,** e1003032 http://dx.doi.org/10.1371/journal.pgen.1003032 (2012).
11. Voight, B. F. *et al.* Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nature Genet.* **42,** 579–589 (2010).
12. Halestrap, A. P. The monocarboxylate transporter family—Structure and functional characterization. *IUBMB Life* **64,** 1–9 (2012).
13. Ng, P. C. & Henikoff, S. SIFT: predicting amino acid changes that affect protein function. *Nucleic Acids Res.* **31,** 3812–3814 (2003).
14. Green, R. E. *et al.* A draft sequence of the Neandertal genome. *Science* **328,** 710–722 (2010).
15. Meyer, M. *et al.* A high-coverage genome sequence from an Archaic Denisovan individual. *Science* **338,** 222–226 (2012).
16. Mednikova, M. B. A proximal pedal phalanx of a Paleolithic hominin from denisova cave, Altai. *Archaeol. Ethnol. Anthropol. Eurasia* **39,** 129–138 (2011).
17. Max Planck Institute for Evolutionary Anthropology. A high-quality Neandertal genome sequence. http://www.eva.mpg.de/neandertal/ (2013).
18. Hublin, J. J. The origin of Neandertals. *Proc. Natl Acad. Sci. USA* **106,** 16022–16027 (2009).
19. Halestrap, A. P. & Wilson, M. C. The monocarboxylate transporter family—Role and regulation. *IUBMB Life* **64,** 109–119 (2012).
20. Garcia, C. K., Goldstein, J. L., Pathak, R. K., Anderson, R. G. W. & Brown, M. S. Molecular characterization of a membrane transporter for lactate, pyruvate, and other monocarboxylates: Implications for the Cori cycle. *Cell* **76,** 865–873 (1994).
21. Fu, S., Watkins, S. M. & Hotamisligil, G. S. The role of endoplasmic reticulum in hepatic lipid homeostasis and stress signaling. *Cell Metab.* **15,** 623–634 (2012).
22. Rhee, E. P. *et al.* Lipid profiling identifies a triacylglycerol signature of insulin resistance and improves diabetes prediction in humans. *J. Clin. Invest.* **121,** 1402–1411 (2011).
23. Savage, D. B. & Semple, R. K. Recent insights into fatty liver, metabolic dyslipidaemia and their links to insulin resistance. *Curr. Opin. Lipidol.* **21,** 329–336 (2010).
24. Samuel, V. T. & Shulman, G. I. Mechanisms for insulin resistance: Common threads and missing links. *Cell* **148,** 852–871 (2012).
25. Florez, J. *et al.* Strong association of socioeconomic status with genetic ancestry in Latinos: implications for admixture studies of type 2 diabetes. *Diabetologia* **52,** 1528–1536 (2009).
26. 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).
27. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19,** 1655–1664 (2009).
28. Risch, N. & Merikangas, K. The future of genetic studies of complex human diseases. *Science.* **273,** 1516–1517 (1996).
29. Hara, K. *et al.* Genome-wide association study identifies three novel loci for type 2 diabetes. *Hum. Mol. Genet.*. http://dx.doi.org/10.1093/hmg/ddt399 (14 August 2013).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank M. Daly, V. Mootha, E. Lander and K. Estrada for comments on the manuscript, B. Voight, A. Segre, J. Pickrell and the Scientific Advisory Board of the SIGMA Project (especially C. Bustamante) for useful discussions, and A. Subramanian and V. Rusu for assistance with expression analyses. This work was conducted as part of the Slim Initiative for Genomic Medicine, a joint US–Mexico project funded by the Carlos Slim Health Institute. The UNAM/INCMNSZ Diabetes Study was supported by Consejo Nacional de Ciencia y Tecnología grants 138826, 128877, CONACyT- SALUD 2009-01-115250, and a grant from Dirección General de Asuntos del Personal Académico, UNAM, IT 214711. The Diabetes in Mexico Study was supported by Consejo Nacional de Ciencia y Tecnología grant 86867 and by Instituto Carlos Slim de la Salud, A.C. The Mexico City Diabetes Study was supported by National Institutes of Health (NIH) grant R01HL24799 and by the Consejo Nacional de Ciencia y Tenologia grants 2092, M9303, F677-M9407, 251M and 2005-C01-14502, SALUD 2010-2-151165. The Multiethnic Cohort was supported by NIH grants CA164973, CA054281 and CA063464. The Singapore Chinese Health Study was funded by the National Medical Research Council of Singapore under its individual research grant scheme and by NIH grants R01 CA55069, R35 CA53890, R01 CA80205 and R01 CA144034. The Type 2 Diabetes Genetic Exploration by Next-generation sequencing in multi-Ethnic Samples (T2D-GENES) project was supported by NIH grant U01DK085526. The San Antonio Mexican American Family Studies (SAMAFS) were supported by R01 DK042273, R01 DK047482, R01 DK053889, R01 DK057295, P01 HL045522 and a Veterans Administration Epidemiologic grant to R.A.D. A.L.W. was

supported by National Institutes of Health Ruth L. Kirschstein National Research Service Award number F32 HG005944.

Author Contributions See the author list for details of author contributions.

Author Information Genotype data have been deposited in dbGaP under accession number phs000683.v1.p1. Microarray data used in the '55k screen' is publicly available through the NCBI Gene Expression Omnibus and the Cancer Cell Line Encyclopedia. A list of sample identities and accession numbers are available in the Supplementary Information. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.A. (altshuler@molbio.mgh.harvard.edu) or T.T.L. (mttusie@gmail.com).

The SIGMA Type 2 Diabetes Genetics Consortium

Writing team: Amy L. Williams[1,2], Suzanne B. R. Jacobs[1], Hortensia Moreno-Macías[3], Alicia Huerta-Chagoya[4,5], Claire Churchhouse[1], Carla Márquez-Luna[6], Humberto García-Ortíz[6], María José Gómez-Vázquez[4,7], Noël P. Burtt[1], Carlos A. Aguilar-Salinas[4], Clicerio González-Villalpando[8], Jose C. Florez[1,9,10], Lorena Orozco[6], Christopher A. Haiman[11], Teresa Tusié-Luna[4,5], David Altshuler[1,2,9,10,12,13,14]

Analysis team: Amy L. Williams[1,2], Carla Márquez-Luna[6], Alicia Huerta-Chagoya[4,5], Stephan Ripke[1,15], María José Gómez-Vázquez[4,7], Alisa K. Manning[1], Hortensia Moreno-Macías[3], Humberto García-Ortíz[6], Benjamin Neale[1,15], Noël P. Burtt[1], Carlos A. Aguilar-Salinas[4], David Reich[1,2], Daniel O. Stram[11], Juan Carlos Fernández-López[6], Sandra Romero-Hidalgo[6], David Altshuler[1,2,9,10,12,13,14], Jose C. Florez[1,9,10], Teresa Tusié-Luna[4,5], Nick Patterson[1], Christopher A. Haiman[11]

Clinical research, study design and metabolic phenotyping: Diabetes in Mexico Study Irma Aguilar-Delfín[6], Angélica Martínez-Hernández[6], Federico Centeno-Cruz[6], Elvia Mendoza-Caamal[6], Cristina Revilla-Monsalve[16], Sergio Islas-Andrade[16], Emilio Córdova[6], Eunice Rodríguez-Arellano[17], Xavier Soberón[6], Lorena Orozco[6]; **Massachusetts General Hospital** Jose C. Florez[1,9,10]; **Mexico City Diabetes Study** Clicerio González-Villalpando[8], María Elena González-Villalpando[8]; **Multiethnic Cohort** Christopher A. Haiman[11], Brian E. Henderson[11], Kristine Monroe[11], Lynne Wilkens[18], Laurence N. Kolonel[18], Loic Le Marchand[18]; **UNAM/INCMNSZ Diabetes Study** Laura Riba[5], María Luisa Ordóñez-Sánchez[4], Rosario Rodríguez-Guillén[4], Ivette Cruz-Bautista[4], Maribel Rodríguez-Torres[4], Linda Liliana Muñoz-Hernández[4], Tamara Sáenz[4], Donají Gómez[4], Ulices Alvirde[4]

Sample quality control and whole-genome genotyping: Noël P. Burtt[1], Robert C. Onofrio[19], Wendy M. Brodeur[19], Diane Gage[19], Jacquelyn Murphy[1], Jennifer Franklin[19], Scott Mahan[19], Kristin Ardlie[19], Andrew T. Crenshaw[19], Wendy Winckler[19]

Neanderthal analysis team: Kay Prüfer[20], Michael V. Shunkov[21], Susanna Sawyer[20], Udo Stenzel[20], Janet Kelso[20], Monkol Lek[1,15], Sriram Sankararaman[1,2], Amy L. Williams[1,2], Nick Patterson[1], Daniel G. MacArthur[1,15], David Reich[1,2], Anatoli P. Derevianko[21], Svante Pääbo[20]

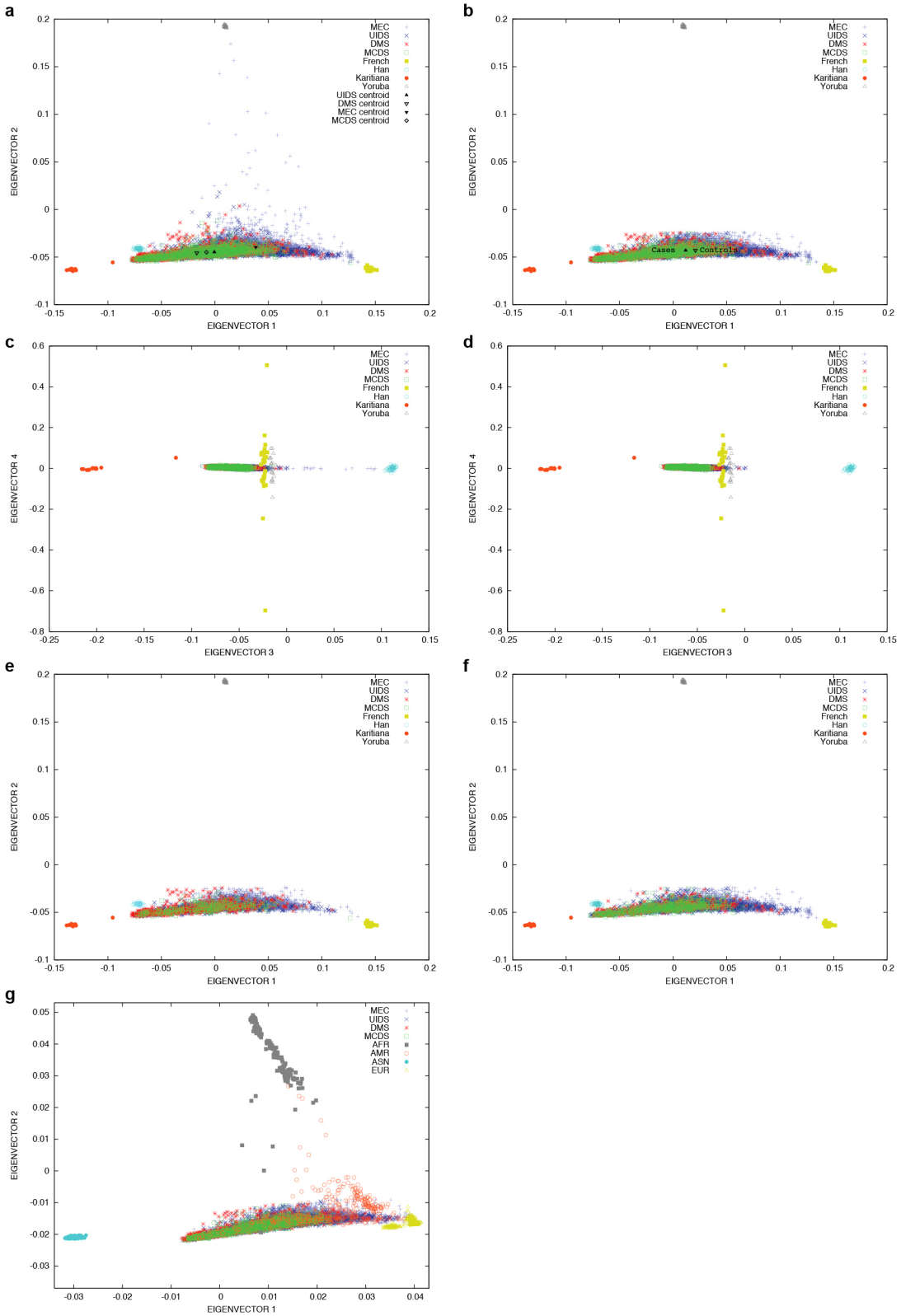
Functional analysis and metabolite profiling: Suzanne B. R. Jacobs[1], Claire Churchhouse[1], Shuba Gopal[22], James A. Grammatikos[22], Ian C. Smith[23], Kevin H. Bullock[22], Amy A. Deik[22], Amanda L. Souza[22], Kerry A. Pierce[22], Clary B. Clish[22], David Altshuler[1,2,9,10,12,13,14]

Replication genotyping and analysis: Broad Institute of Harvard and MIT Timothy Fennell[19], Yossi Farjoun[19], Broad Genomics Platform*, Stacey Gabriel[19]; **Singapore Chinese Health Study** Daniel O. Stram[11], Myron D. Gross[24], Mark A. Pereira[24], Mark Seielstad[25], Woon-Puay Koh[26,27], E-Shyong Tai[26,27,28]; **T2D-GENES Consortium** Jason Flannick[1,9], Pierre Fontanillas[1], Andrew Morris[29], Tanya M. Teslovich[30], Noël P. Burtt[1], Gil Atzmon[31], John Blangero[32], Donald W. Bowden[33], John Chambers[34,35,36], Yoon Shin Cho[37], Ravindranath Duggirala[32], Benjamin Glaser[38,39], Craig Hanis[40], Jaspal Kooner[35,36,41], Markku Laakso[42], Jong-Young Lee[43], E-Shyong Tai[26,27,28], Yik Ying Teo[44,45,46,47,48], James G. Wilson[49], the T2D-GENES Consortium*; **Multiethnic Cohort** Christopher A. Haiman[11], Brian E. Henderson[11], Kristine Monroe[11], Lynne Wilkens[18], Laurence N. Kolonel[18], Loic Le Marchand[18]; **Texas Biomedical Research Institute and University of Texas Health Science Center at San Antonio** Sobha Puppala[32], Vidya S. Farook[32], Farook Thameem[50], Hanna E. Abboud[50], Ralph A. DeFronzo[51], Christopher P. Jenkinson[51], Donna M. Lehman[52], Joanne E. Curran[32], John Blangero[32], Ravindranath Duggirala[32]

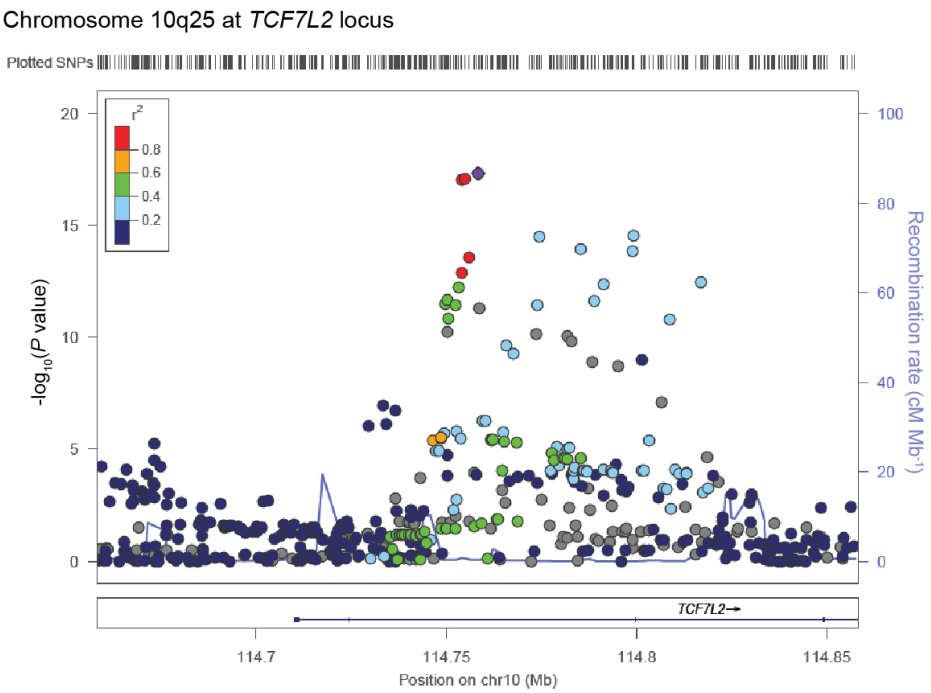
Scientific and project management: Noël P. Burtt[1], Maria L. Cortes[53]

Steering committee: David Altshuler[1,2,9,10,12,13,14], Jose C. Florez[1,9,10], Christopher A. Haiman[11], Brian E. Henderson[11], Carlos A. Aguilar-Salinas[4], Clicerio González-Villalpando[8], Lorena Orozco[6] & Teresa Tusié-Luna[4,5]

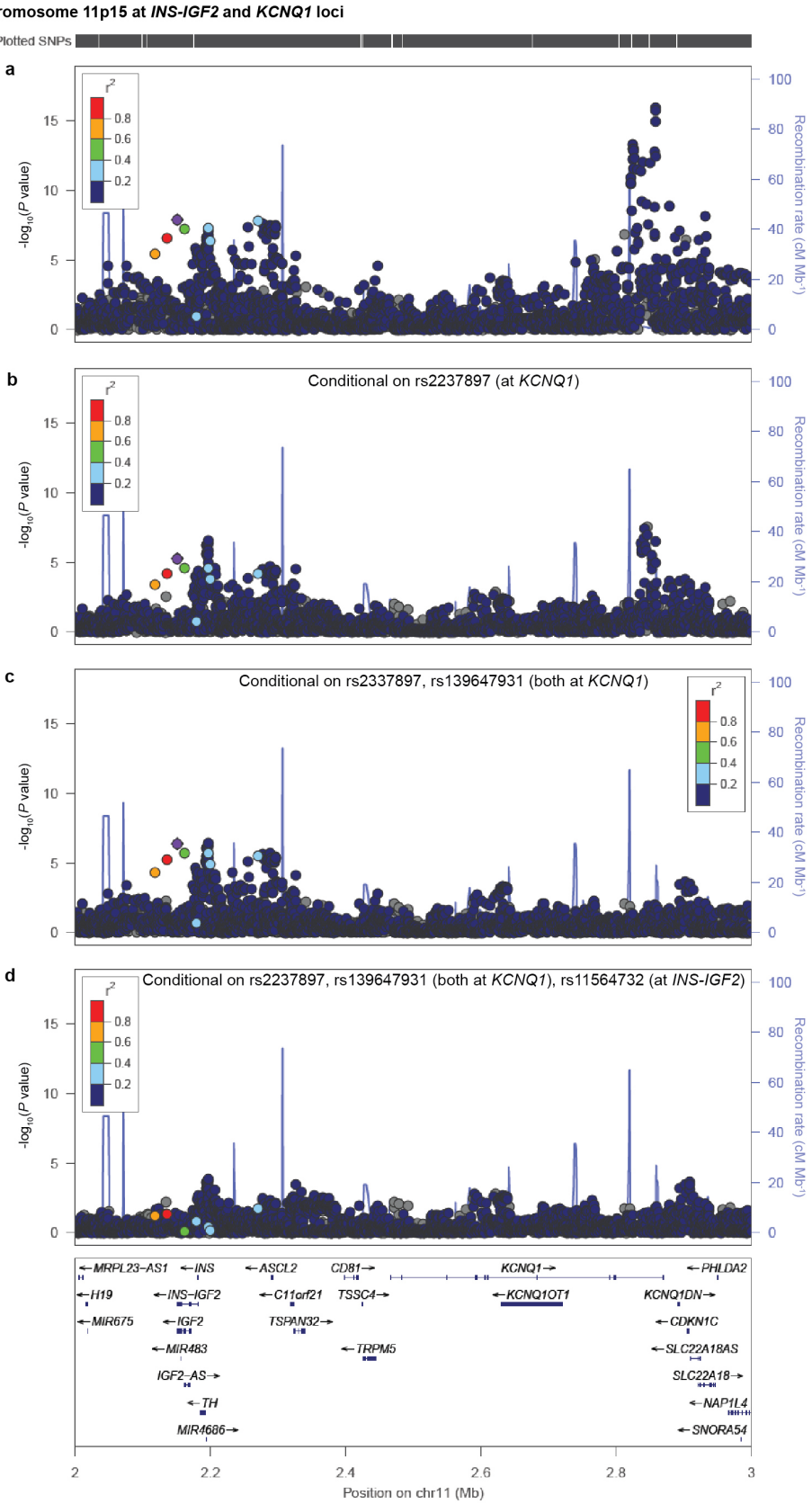
[1]Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, Massachusetts 02142, USA. [2]Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA. [3]Universidad Autonoma Metropolitana, Tlalpan 14387, Mexico City, Mexico. [4]Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán, Sección XVI, Tlalpan, 14000 Mexico City, Mexico. [5]Instituto de Investigaciones Biomédicas, UNAM. Unidad de Biología Molecular y Medicina Genómica, UNAM/INCMNSZ, Coyoacán, 04510 Mexico City, Mexico. [6]Instituto Nacional de Medicina Genómica, Tlalpan, 14610 Mexico City, Mexico. [7]Universidad Autónoma de Nuevo León, San Nicolás de los Garza, Nuevo León 66451, México. [8]Centro de Estudios en Diabetes, Unidad de Investigacion en Diabetes y Riesgo Cardiovascular, Centro de Investigacion en Salud Poblacional, Instituto Nacional de Salud Publica, 01120 Mexico City, Mexico. [9]Center for Human Genetic Research and Diabetes Research Center (Diabetes Unit), Massachusetts General Hospital, Boston 02114, Massachusetts, USA. [10]Department of Medicine, Harvard Medical School, Boston, Massachusetts 02115, USA. [11]Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, California 90089, USA. [12]Center for Human Genetic Research, Massachusetts General Hospital, Boston, Massachusetts 02114, USA. [13]Department of Molecular Biology, Harvard Medical School, Boston, Massachusetts 02114, USA. [14]Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. [15]Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, Massachusetts 02114, USA. [16]Unidad de Investigación Médica en Enfermedades Metabólicas, Instituto Mexicano del Seguro Social SXXI, Cuauhtémoc, 06720 Mexico City, Mexico. [17]Instituto de Seguridad y Servicios Sociales para los Trabajadores del Estado, Álvaro Obregón, 01030 Mexico City, Mexico. [18]Epidemiology Program, University of Hawaii Cancer Center, Honolulu, Hawaii 96813, USA. [19]The Genomics Platform, The Broad Institute of Harvard and MIT, Cambridge, Massachusetts 02142, USA. [20]Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, D-04103 Leipzig, Germany. [21]Palaeolithic Department, Institute of Archaeology and Ethnography, Russian Academy of Sciences, Siberian Branch, 630090 Novosibirsk, Russia. [22]The Metabolite Profiling Platform, The Broad Institute of Harvard and MIT, Cambridge, Massachusetts 02142, USA. [23]Cancer Biology Program, The Broad Institute of Harvard and MIT, Cambridge, Massachusetts 02142, USA. [24]University of Minnesota, Minneapolis, Minnesota 55455, USA. [25]University of California San Francisco, San Francisco, California 94143, USA. [26]Duke National University of Singapore Graduate Medical School, Singapore 169857, Singapore. [27]Saw Swee Hock School of Public Health, National University of Singapore, Singapore 117597, Singapore. [28]Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore, Singapore 117597, Singapore. [29]Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, UK. [30]Department of Biostatistics, Center for Statistical Genetics, University of Michigan, Ann Arbor, Michigan 48109, USA. [31]Department of Medicine, Department of Genetics, Albert Einstein College of Medicine, Bronx, New York 10461, USA. [32]Department of Genetics, Texas Biomedical Research Institute, San Antonio, Texas 78227, USA. [33]Center for Genomics and Personalized Medicine Research, Center for Diabetes Research, Department of Biochemistry, Department of Internal Medicine, Wake Forest School of Medicine, Winston-Salem, North Carolina 27157, USA. [34]Department of Epidemiology and Biostatistics, Imperial College London, London SW7 2AZ, UK. [35]Imperial College Healthcare NHS Trust, London W2 1NY, UK. [36]Ealing Hospital National Health Service (NHS) Trust, Middlesex UB1 3HW, UK. [37]Department of Biomedical Science, Hallym University, Chuncheon, Gangwon-do, 200-702 South Korea. [38]Endocrinology and Metabolism Service, Hadassah-Hebrew University Medical School, Jerusalem 91120, Israel. [39]Israel Diabetes Research Group (IDRG), Diabetes Unit, The E. Wolfson Medical Center, Holon 58100, Israel. [40]Human Genetics Center, University of Texas Health Science Center at Houston, Houston, Texas 77030, USA. [41]National Heart and Lung Institute (NHLI), Imperial College London, Hammersmith Hospital, London W12 0HS, UK. [42]Department of Medicine, University of Eastern Finland, Kuopio Campus and Kuopio University Hospital, FI-70211 Kuopio, Finland. [43]Center for Genome Science, Korea National Institute of Health, Osong Health Technology Administration Complex, Chungcheongbuk-do 363-951, South Korea. [44]Department of Epidemiology and Public Health, National University of Singapore, Singapore 117597, Singapore. [45]Centre for Molecular Epidemiology, National University of Singapore, Singapore 117456, Singapore. [46]Genome Institute of Singapore, Agency for Science, Technology and Research, Singapore 138672, Singapore. [47]Graduate School for Integrative Science and Engineering, National University of Singapore, Singapore 117456, Singapore. [48]Department of Statistics and Applied Probability, National University of Singapore, Singapore 117546, Singapore. [49]Department of Physiology and Biophysics, University of Mississippi Medical Center, Jackson, Mississippi 39216, USA. [50]Division of Nephrology, Department of Medicine, University of Texas Health Science Center at San Antonio, San Antonio, Texas 78229, USA. [51]Division of Diabetes, Department of Medicine, University of Texas Health Science Center at San Antonio, San Antonio, Texas 78229, USA. [52]Division of Clinical Epidemiology, Department of Medicine, University of Texas Health Science Center at San Antonio, San Antonio, Texas 78229, USA. [53]Broad Institute of Harvard and MIT, Cambridge, Massachusetts 02142, USA. *Lists of participants and their affiliations appear in the Supplementary Information.



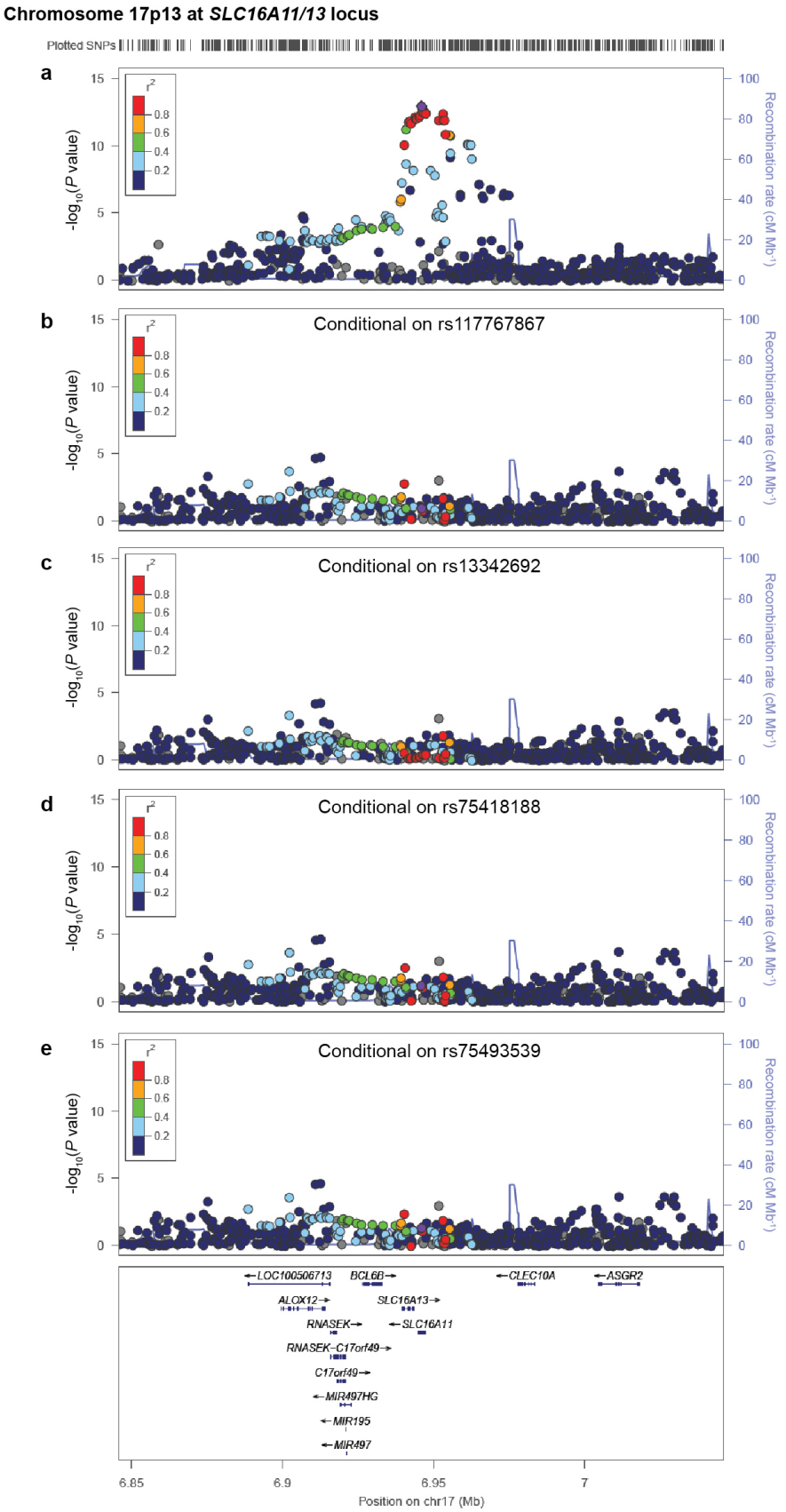
Extended Data Figure 1 | Principal component analysis (PCA) projection of SIGMA samples onto principal components calculated using data from samples collected by the Human Genome Diversity Project (HGDP) and 1000 Genomes Project. **a**, **b**, PCA projection of SIGMA onto HGDP Yoruba, French, Karitiana and Han (Chinese) populations before ancestry quality control filters were applied (**a**), with cohort centroids as indicated, and after all quality control filters were applied (**b**), with case and control centroids as indicated. **c**, **d**, Principal components 3 and 4 before filtering samples on ancestry (a small number of samples in the MEC show East Asian admixture) (**c**), and after all quality control filters were applied (**d**). **e**, **f**, Additional plots as in **b** but separating cases (**e**) and controls (**f**). **g**, SIGMA samples projected onto the 1000 Genomes Project Omni2.5 genotype data. 1000 Genomes samples are labelled by their continental ancestry group: AFR, African; AMR, Native American descent; ASN, east Asian; EUR, European.



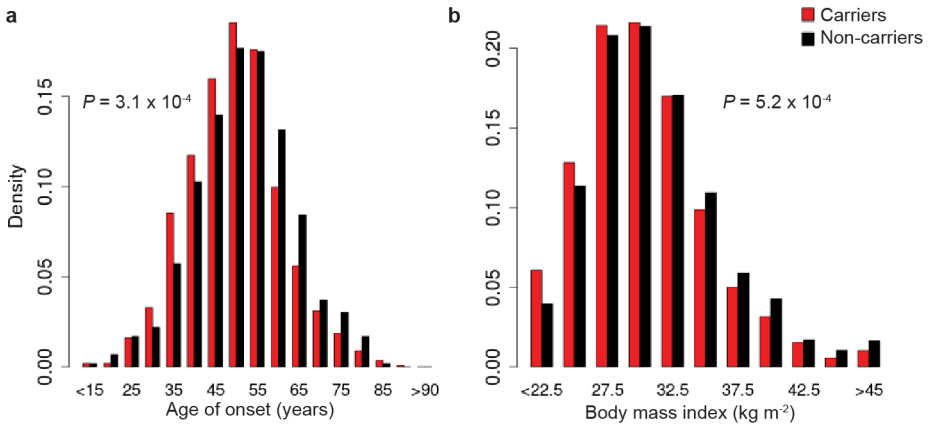
Extended Data Figure 2 | Regional plot for signal at *TCF7L2*. Point colour indicates r^2 to the most strongly associated site (rs7903146) and recombination rate is also shown, both based on the 1000 Genomes ASN population.



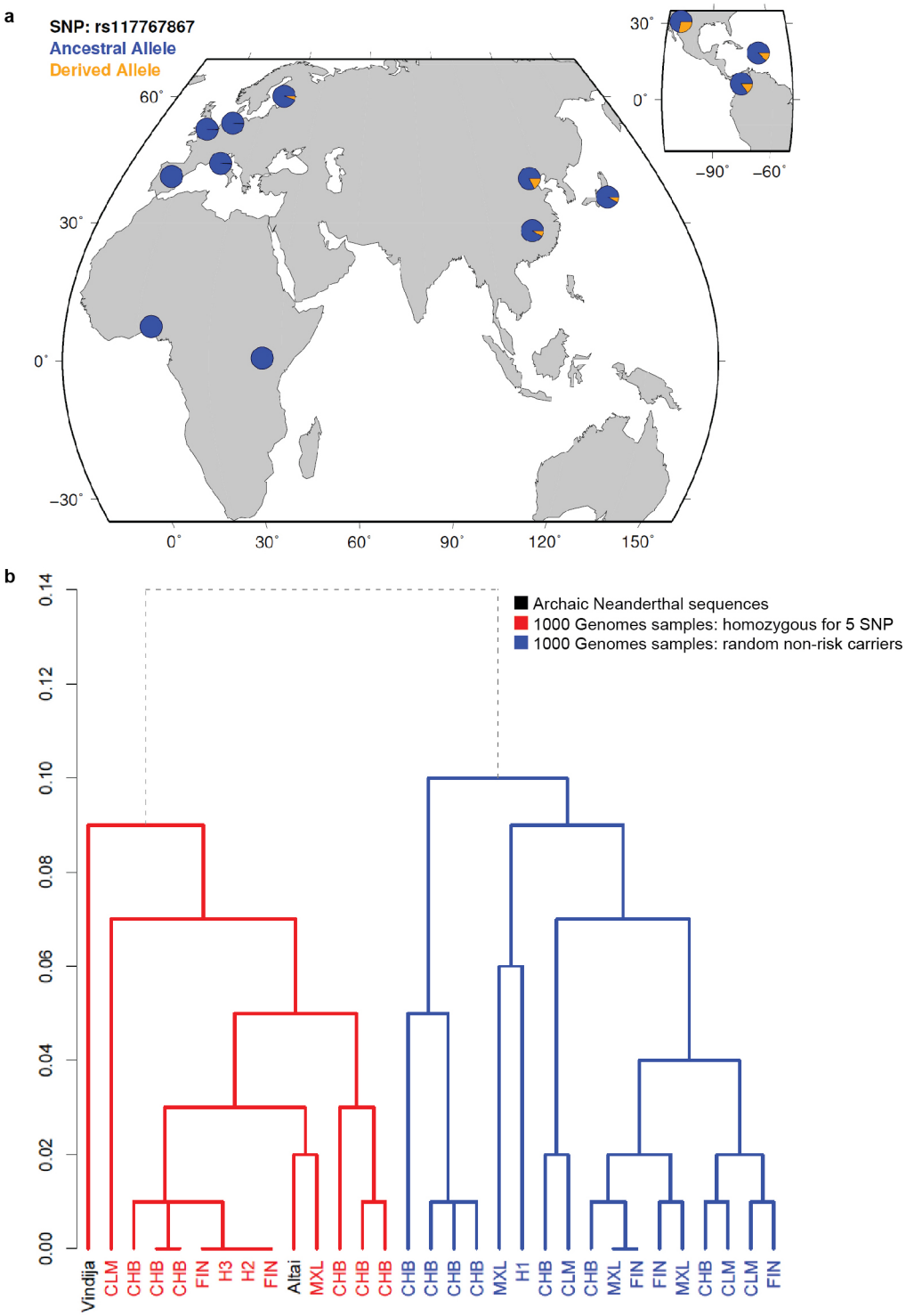
Extended Data Figure 3 | Conditional analyses reveal multiple independent signals at *INS–IGF2* and *KCNQ1*. **a**–**d**, Regional plots are shown for the interval spanning *INS–IGF2* and *KCNQ1* without conditioning (**a**), conditional on rs2237897 at *KCNQ1* (**b**), conditional on rs2237897 and rs139647931 (both at *KCNQ1*) (**c**), and conditional on rs2237897 and rs139647931 (both at *KCNQ1*), and rs11564732 (the top associated variant in the *INS–IGF2–TH* region) (**d**). The top SNPs in 11p15.5 and *KCNQ1* are ~700 kb away from each other, but despite this proximity, there is a strong residual signal of association at *INS–IGF2* after analysis conditional on genotype at *KCNQ1*. Point colour indicates r^2 to rs11564732 and recombination rate is also shown, both based on the 1000 Genomes ASN population.



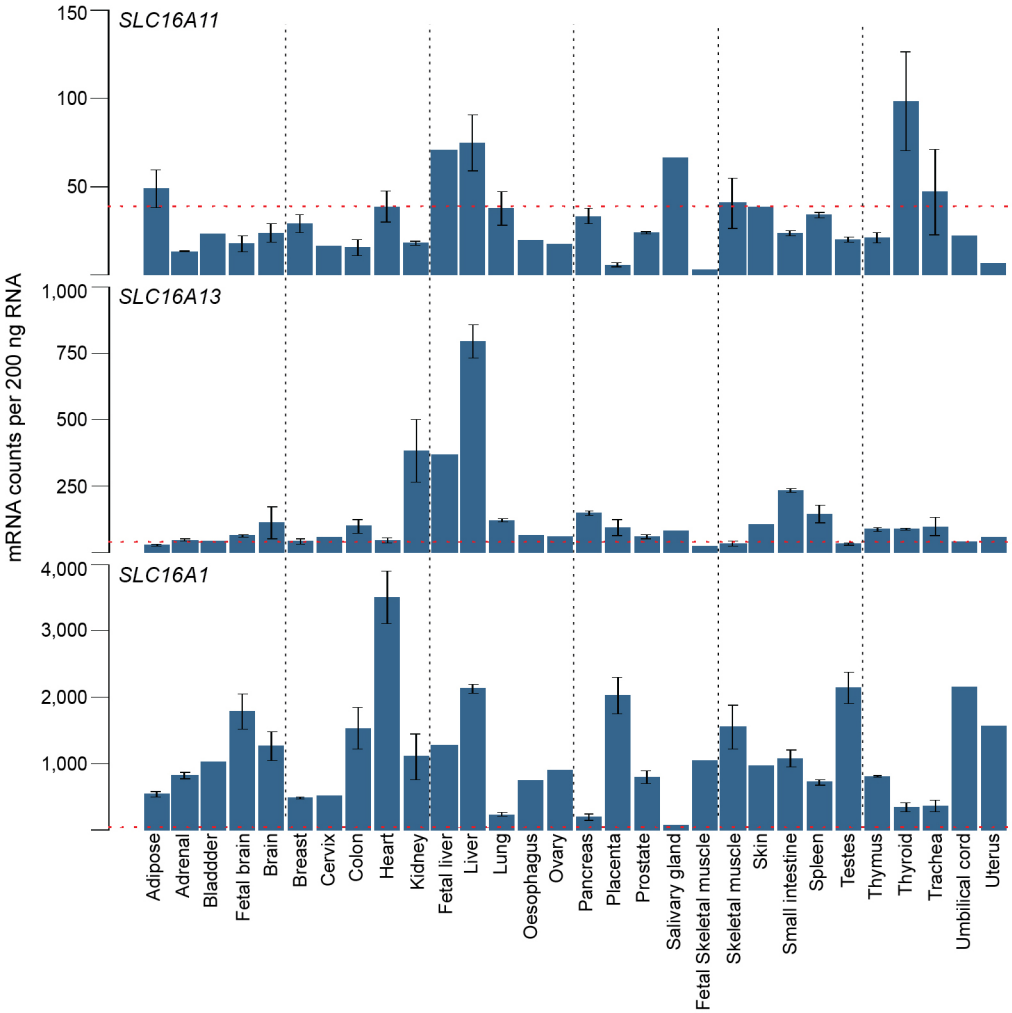
Extended Data Figure 4 | Regional plots for *SLC16A11* conditional on associated missense variants of that gene. **a**–**e**, Association signal at chromosome 17p13 without conditioning (**a**), or conditional on the four missense SNPs in SLC16A11: rs117767867 (**b**), rs13342692 (**c**), rs75418188 (**d**) and rs75493593 (**e**). Point colour indicates r^2 to the most strongly associated SNP (rs13342232) and recombination rate is also shown, both based on the 1000 Genomes ASN population.



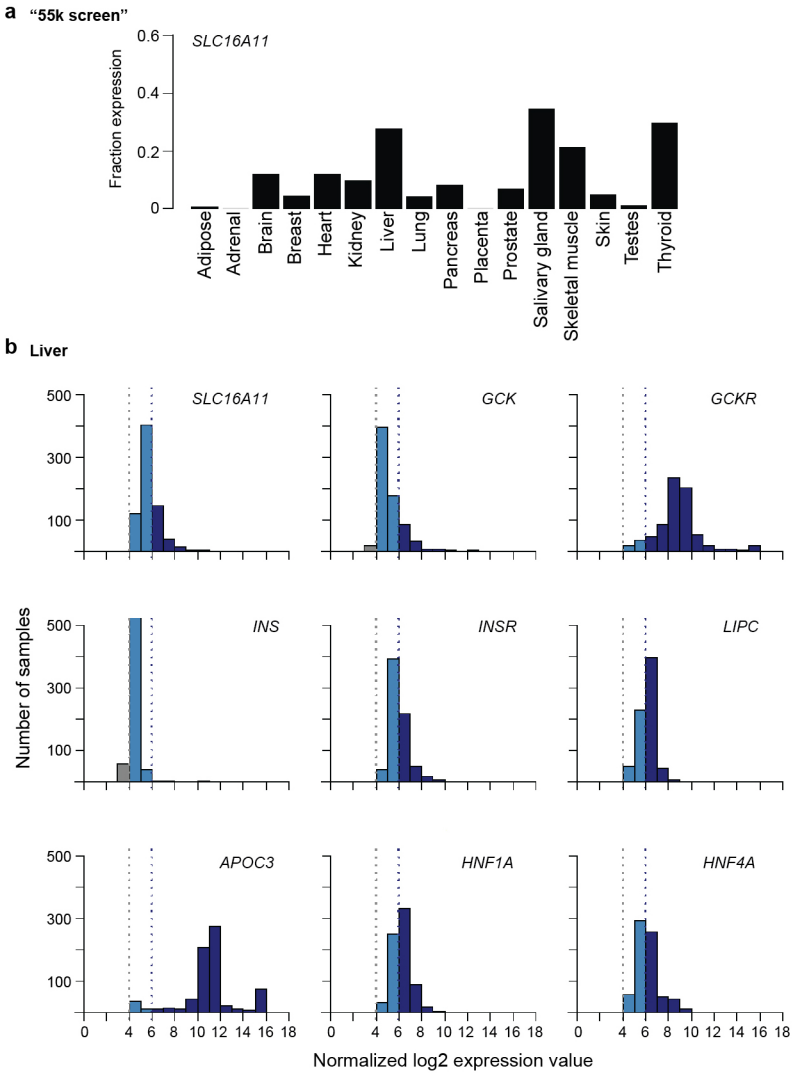
Extended Data Figure 5 | Cases with risk haplotype develop type 2 diabetes younger and at a lower BMI than non-carriers. **a**, Distribution of age-of-onset in type 2 diabetes cases based on genotype at rs13342232, binned every 5 years with upper bounds indicated (carriers $n = 1{,}126$; non-carriers $n = 594$). **b**, Distribution of BMI in type 2 diabetes cases for carriers and non-carriers of rs13342232, binned every 2.5 kg m^{-2} with upper bounds indicated (carriers $n = 2{,}161$; non-carriers $n = 1{,}647$). *P* values from two-sample *t*-test between type 2 diabetes risk haplotype carriers and type 2 diabetes non-carriers.



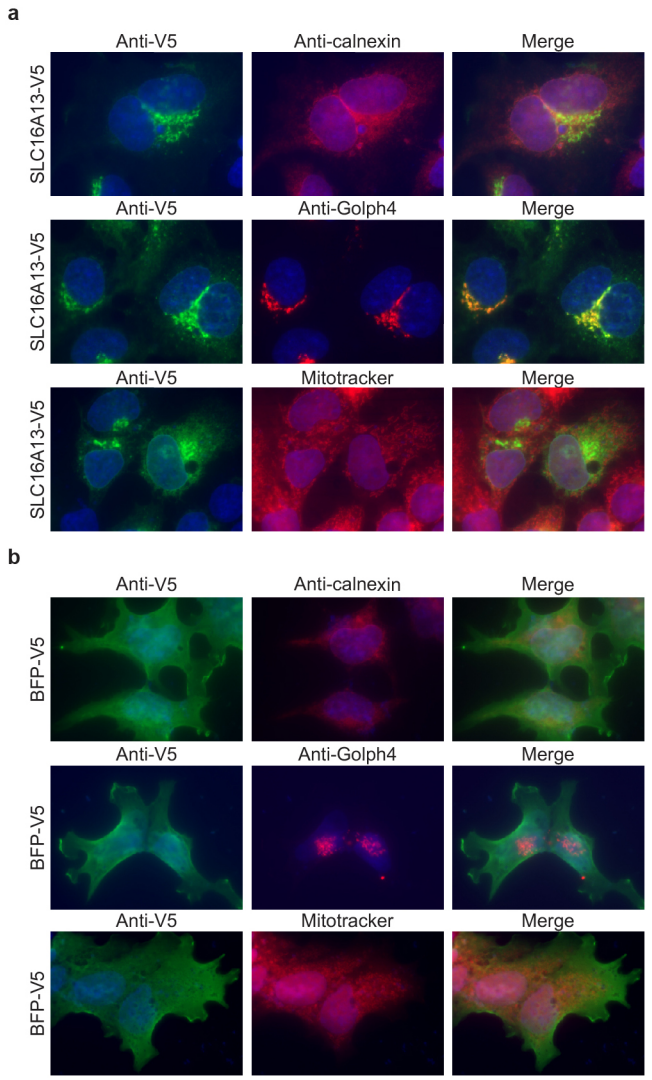
Extended Data Figure 6 | Frequency distribution of the risk haplotype and dendrogram depicting clustering with Neanderthal haplotypes. **a**, Allele frequency of missense SNP rs117767867 (tag for risk haplotype) in the 1000 Genomes Phase I data set. **b**, Dendrogram generated from haplotypes across the 73-kb Neanderthal introgressed region. Nodes for modern human haplotypes are labelled in red or blue with the 1000 Genomes population in which the corresponding haplotype resides. Archaic Neanderthal sequences are labelled in black and include the low-coverage Neanderthal sequence[14] (labelled Vindija), and the unpublished Neanderthal sequence that is homozygous for the 5 SNP risk haplotype[17] (Altai). H1 includes haplotypes from MXL and FIN, and H2 and H3 both include haplotypes from CLM, MXL, CHB and ASW. Modern human sequences included are all 1000 Genomes Phase I samples that are homozygous for the 5 SNP risk haplotype ($n = 15$), and 16 non-risk haplotypes—four haplotypes (from two randomly selected individuals) from each of the CLM (Colombian in Medellin, Colombia), MXL (Mexican Ancestry in Los Angeles, California), CHB (Han Chinese in Beijing, China) and FIN (Finnish in Finland) 1000 Genomes populations (the populations with carriers of the 5 SNP haplotype). The red subtree depicts the Neanderthal clade, with all risk haplotypes clustering with the Altai and Vindija sequences. In blue are all other modern human haplotypes. The dendrogram was generated by the R function hclust using a complete linkage clustering algorithm on a distance matrix measuring the fraction of SNPs called in the 1000 Genomes project at which a pair of haplotypes differs (the *y* axis represents this distance). Because haplotypes are unavailable for the archaic samples, we picked a random allele to compute the distance matrix.



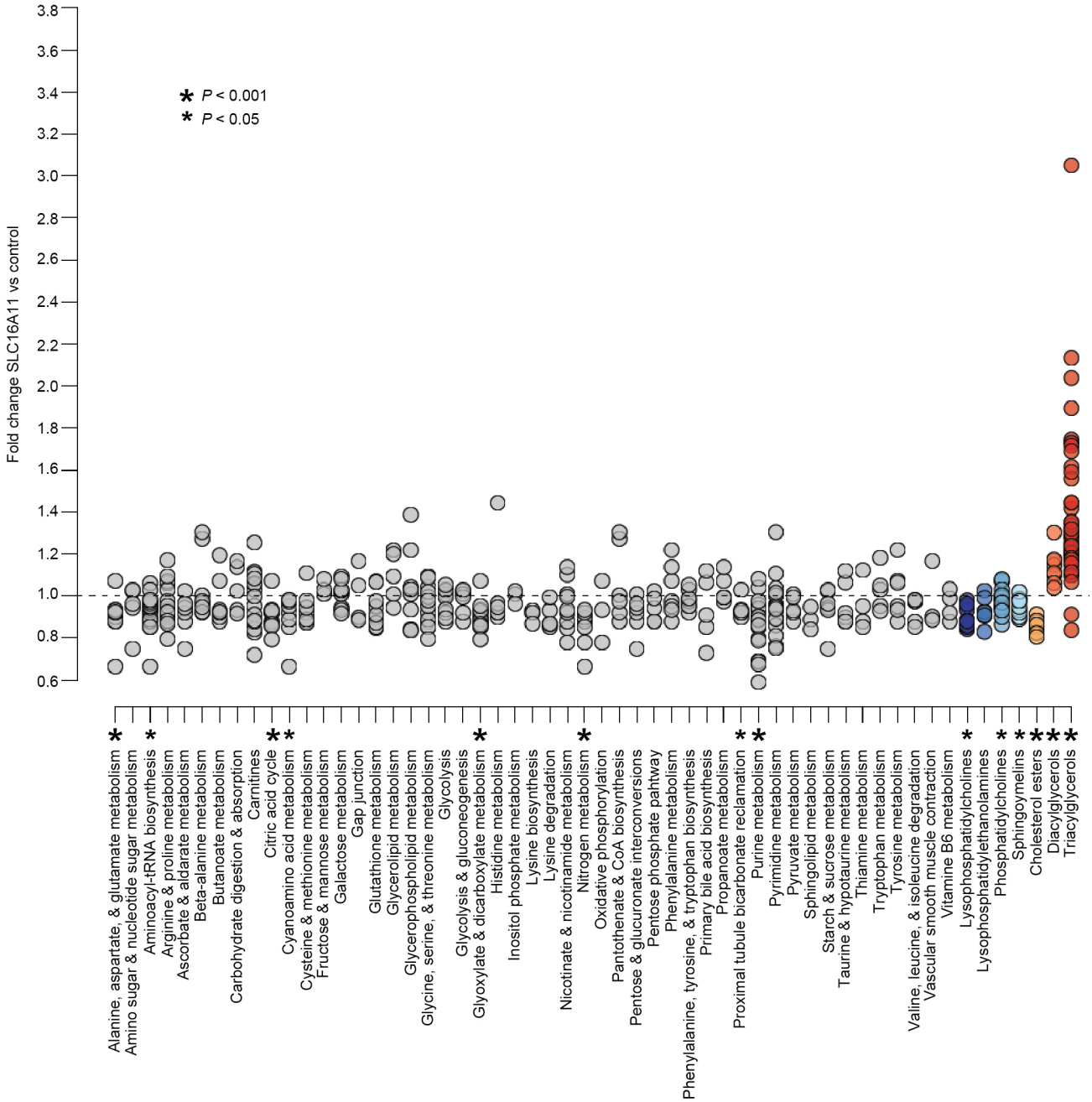
Extended Data Figure 7 | Analysis of gene expression for *SLC16A11*, *SLC16A13* and *SLC16A1* in 30 human tissues. Data measured using nCounter are shown as mean, normalized mRNA counts per 200 ng RNA ± s.e.m. Threshold for background (nonspecific) binding is indicated by the red line. Sample size for each tissue (*n*): pancreas (5); adipose, brain, colon, liver, skeletal muscle and thyroid (3); adrenal, fetal brain, breast, heart, kidney, lung, placenta, prostate, small intestine, spleen, testes, thymus and trachea (2); bladder, cervix, oesophagus, fetal liver, ovary, salivary gland, fetal skeletal muscle, skin, umbilical cord and uterus (1).



Extended Data Figure 8 | Microarray-based analysis of *SLC16A11* expression in human tissues. **a**, Results from the '55k screen', a survey of gene expression in 55,269 samples profiled on the Affymetrix U133 plus 2.0 array, are shown as the fraction of samples of a given tissue in which *SLC16A11* is expressed. Sample size for each tissue (*n*): adipose (394), adrenal (69), brain (1,990), breast (4,104), heart (178), kidney (675), liver (721), lung (1,442), pancreas (150), placenta (107), prostate (578), salivary gland (26), skeletal muscle (793), skin (947), testis (102), thyroid (108). **b**, Histograms show the expression level distribution of *SLC16A11* and other well-studied liver genes in 721 liver samples from the '55k screen'. *INS* is shown as reference for a gene not expressed in liver. On the basis of negative controls, a normalized $\log_2$ expression of 4 is considered baseline and $\log_2$ expression values greater than 6 are considered expressed.



Extended Data Figure 9 | SLC16A13 localizes to Golgi apparatus. a, **b**, HeLa cells transiently expressing C terminus, V5-tagged SLC16A13 (**a**) or BFP (**b**) were immunostained for SLC16A13 or BFP expression (anti-V5) along with specific markers for the endoplasmic reticulum (anti-calnexin), cis-Golgi apparatus (anti-Golph4) and mitochondria (MitoTracker). Representative images from multiple independent transfections are shown. Owing to heterogeneity in expression levels of overexpressed proteins and endogenous organelle markers, imaging of each protein was optimized for clarity of localization and varied across images; therefore, images are not representative of relative expression levels of each protein as compared to the other proteins.



Extended Data Figure 10 | Pathway and class-based metabolic changes induced by SLC16A11 expression. Changes in metabolite levels in HeLa cells expressing SLC16A11–V5 compared to control-transfected cells are plotted in groups according to metabolic pathway or class. Data shown are the combined results from three independent experiments, each of which included 12 biological replicates each for SLC16A11 and control. Pathways shown include all KEGG pathways from the human reference set for which metabolites were measured as well as eight additional classes of metabolites covering carnitines and lipid subtypes. Each point within a pathway or class shows the fold change of a single metabolite within that pathway or class. For each pathway or class with at least six measured metabolites, enrichment was computed as described in Supplementary Methods. Asterisks indicate pathways with $P \leq 0.05$ and FDR ≤ 0.25. Supplementary Table 14 shows additional details from the enrichment analysis.