nature genetics

Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes

Eleftheria Zeggini[1,10], Laura J Scott[2,10], Richa Saxena[3–8,10] & Benjamin F Voight[3–5,7,10], for the Diabetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium[9]

Genome-wide association (GWA) studies have identified multiple loci at which common variants modestly but reproducibly influence risk of type 2 diabetes (T2D)[1–11]. Established associations to common and rare variants explain only a small proportion of the heritability of T2D. As previously published analyses had limited power to identify variants with modest effects, we carried out meta-analysis of three T2D GWA scans comprising 10,128 individuals of European descent and ~2.2 million SNPs (directly genotyped and imputed), followed by replication testing in an independent sample with an effective sample size of up to 53,975. We detected at least six previously unknown loci with robust evidence for association, including the *JAZF1* ($P = 5.0 \times 10^{-14}$), *CDC123-CAMK1D* ($P = 1.2 \times 10^{-10}$), *TSPAN8-LGR5* ($P = 1.1 \times 10^{-9}$), *THADA* ($P = 1.1 \times 10^{-9}$), *ADAMTS9* ($P = 1.2 \times 10^{-8}$) and *NOTCH2* ($P = 4.1 \times 10^{-8}$) gene regions. Our results illustrate the value of large discovery and follow-up samples for gaining further insights into the inherited basis of T2D.

GWA studies are unbiased by previous hypotheses concerning candidate genes and pathways, but they are limited by the modest effect sizes of individual common susceptibility variants and the need for stringent statistical thresholds. For example, the largest allelic odds ratio (OR) of any established common variant for T2D is ~1.35 (*TCF7L2*), and the nine other validated associations to common variants (excluding *FTO*, which has its primary effect through obesity) have allelic ORs between 1.1 and 1.2 (refs. 1–6,11,12). To augment power to detect additional loci of similar or smaller effect, we increased sample size by combining three previously published GWA studies (Diabetes Genetics Initiative (DGI), Finland–United States Investigation of NIDDM Genetics (FUSION) and Wellcome Trust Case Control Consortium (WTCCC))[1–4], and extended SNP coverage by imputing untyped SNPs on the basis of patterns of haplotype variation from HapMap[13] (**Table 1**).

We started with a set of genotyped autosomal SNPs that passed quality control filters in each study: in WTCCC, 393,143 SNPs from the Affymetrix 500K chip (minor allele frequency (MAF) > 0.01; 1,924 cases and 2,938 population-based controls[3,4]); in DGI, 378,860 SNPs from the Affymetrix 500K chip (MAF > 0.01; Swedish and Finnish sample of 1,464 T2D cases and 1,467 normoglycemic controls, including 326 discordant sibships[1]); and in FUSION, 306,222 SNPs from the Illumina 317K chip (MAF > 0.01, 1,161 T2D cases and 1,174 normal glucose-tolerant controls from Finland[2]) (**Supplementary Table 1** online). 44,750 SNPs (MAF > 0.01) were directly genotyped in all three studies across the two platforms. We used data from the GWA studies and phased chromosomes from the HapMap CEU sample to impute autosomal SNPs with MAF > 0.01 (ref. 14; see also URLs section in Methods). We based our further analyses on 2,202,892 SNPs that met imputation and genotyping quality control criteria across all studies (**Supplementary Methods** online).

Using these directly measured and imputed genotypes, we tested for association of each SNP with T2D in each study separately, corrected each study for residual population stratification, cryptic relatedness or technical artifacts using genomic control, and then combined these results in a genome-wide meta-analysis across a total of 10,128 samples (4,549 cases and 5,579 controls; **Supplementary Methods**). We calculated that this sample size provides reasonable power to detect additional variants with properties similar to those previously identified through less formal data combination efforts[1,2,4] (**Supplementary Table 2** online). Unless otherwise indicated, results presented are derived from individually genomic control–adjusted stage 1 results. We obtained meta-analysis OR and confidence intervals

[1]Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK. [2]Department of Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor, Michigan 48109, USA. [3]Broad Institute of Harvard and Massachusetts Institute of Technology (MIT), Cambridge, Massachusetts 02142, USA. [4]Center for Human Genetic Research, [5]Department of Medicine and [6]Department of Molecular Biology, Massachusetts General Hospital, Boston, Massachusetts 02114, USA. [7]Department of Medicine and [8]Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA. [9]The full list of authors and affiliations appears at the end of this paper. [10]These authors contributed equally to this work. Correspondence should be addressed to M.I.M. (mark.mccarthy@drl.ox.ac.uk), M.B. (boehnke@umich.edu) or D.A. (altshuler@molbio.mgh.harvard.edu).

Received 18 December 2007; accepted 12 February 2008; published online 30 March 2008; doi:10.1038/ng.120

Table 1 Overview of study design

Study	Cases (*n*)[a]	Controls (*n*)[a]	Effective sample size[a]	Number of directly genotyped SNPs[b]	Number of imputed SNPs[b]
Stage 1					
DGI	1,464	1,467	2,521	378,860	1,888,145
WTCCC	1,924	2,938	4,706	393,143	1,915,393
FUSION	1,161	1,174	2,335	306,222	2,110,199
Stage 2					
DGI stage 2	5,065	5,785	9,874	63	–
FUSION stage 2	1,215	1,258	2,473	59	–
UK stage 2	3,757	5,346	9,114	66	–
Stage 3					
deCODE	1,520 (1,422)	25,235 (3,455)	4,280 (3,130)	11	–
KORA	1,241	1,458	2,684	6	–
Danish	4,089	5,043	8,690	11	–
HUNT	1,004	1,503	2,412	11	–
NHS	1,506	2,014	3,468	10	–
CCC	547	533	1,070	11	–
EPIC	388	774	1,036	10	–
ADDITION/Ely	892	1,610	2,288	11	–
Norfolk	2,311	2,400	4,450	11	–
METSIM	659	2,639	2,136	11	–

[a]Sample sizes presented here are the maximum available for each study. For the deCODE stage 3 study, we used genotype data from the Icelandic GWA scan[5] for rs2641348, rs7578597 and rs9472138, and a perfect proxy (rs2793831, based on HapMap) for rs10923931. The remaining SNPs had not been directly typed as part of this scan and were therefore genotyped separately, in a subset of the GWA scan samples (numbers indicated in parentheses) (**Supplementary Methods**). [b]Autosomal SNPs passing quality control, as defined for directly genotyped and imputed SNPs in each study (quality control criteria: SNPTEST information measure ≥ 0.5; r^2hat ≥ 0.3; MAF > 0.01). For the stage 1 meta-analysis, we combined results for 2,202,892 directly genotyped and imputed SNPs passing quality control in all three studies (**Supplementary Methods**).

from a fixed-effects model, and *P* values from a weighted *z* statistic–based meta-analysis (**Supplementary Methods**). As expected, the most significant result was obtained for rs7903146 in *TCF7L2*. We also observed evidence for association ($P < 10^{-3}$) at eight of the ten established T2D loci (as well as at the *FTO* obesity locus)[12] (**Supplementary Table 3** online). This was unsurprising, as these same data supported the identification of many of these loci. As our goal was to identify previously unknown loci, we excluded 1,981 SNPs in the immediate vicinity of these T2D susceptibility loci from further analysis (with the exception of a signal near *PPARG*, which was followed up), and examined the remainder of the autosomal genome (**Supplementary Methods**). Even after excluding known loci, we saw a strong enrichment of highly associated variants: 426 with *P* values $<10^{-4}$, compared to 217 under the null.

Before proceeding to follow-up, we explored the individual studies and the combined data for potential errors and biases. We found a genomic control λ value of 1.04 for the combined results (based on 10,128 samples), which, given the relationship between λ and sample size[15], suggests little residual confounding (**Supplementary Fig. 1** and **Supplementary Note** online). We also used genome-wide genotype data to estimate the principal components of the identity-by-state relationships in each stage 1 sample. For the SNPs presented in **Table 2**, adjustment for principal components in stage 1 T2D association analysis did not diminish the association in the WTCCC (two principal components), FUSION (ten principal components) or DGI (ten principal components) samples (**Supplementary Note**). Additionally, we did not find any evidence for association between UK population ancestry informative markers[3] and disease status in the UK replication sets (**Supplementary Note**). To ensure that the observed stage 1 associations taken forward to follow-up were not due to imputation errors, we directly genotyped originally imputed variants in the stage 1 samples (**Supplementary Methods**). We found strong agreement between the genotype-based and imputed *P* values: in 38 of 43 cases where a direct genotype-based result was obtained, the *P* value was within one order of magnitude of that derived from imputation, and in the remaining five cases, *P* values were less than two orders of magnitude different (**Supplementary Table 4** online).

We selected SNPs for replication principally on the basis of the statistical evidence for association in stage 1, excluding SNPs with evidence for heterogeneity of ORs ($P < 10^{-4}$) across studies (**Supplementary Methods**). We took 69 SNPs forward to an initial round of replication (stage 2) in up to 22,426 additional samples of European descent (**Table 1** and **Supplementary Table 1**). The distribution of association *P* values in stage 2 was highly inconsistent with a null distribution. Of the 69 signals selected for follow up, 65 were successfully genotyped in stage 2, and represented loci that were independent of each other and of previously established susceptibility loci. Nine of these had a *P* value ≤ 0.01 with association in the same direction as the original signal, far in excess of the 0.33 expected under the null ($P = 1.4 \times 10^{-12}$, binomial test; **Supplementary Methods**), and two SNPs had $P < 10^{-4}$ as compared to an expectation of 0.0033 ($P = 5.2 \times 10^{-6}$) (**Supplementary Methods** and **Supplementary Table 5** online).

We identified 11 SNPs (ten separate signals, nine of which represent previously unknown loci) with $P < 0.005$ in stage 2 for which the combined stage 1 and stage 2 data (based on direct genotyping of stage

Table 2 Eleven T2D-associated SNPs taken forward to stages 2 and 3

							Stage 1 (DGI, FUSION, WTCCC)		Stage 2 (DGI, FUSION, UKT2D)		Stage 3 (deCODE, KORA, Steno, HUNT, NHS, CCC, EPIC, ADDITION/Ely, Norfolk, METSIM)		All data				
SNP	Chr	Position NCBI35 (bp)	Nonrisk allele[a]	Risk allele[a]	Risk allele frequency[a]	Nearest gene(s)	OR (95% CI)	*P* value	OR (95% CI)	*P* value	OR (95% CI)	*P* value	n_{eff}	OR (95% CI)	*P* value	P_{het}	Number of samples for 80% power[b]
rs864745	7	27,953,796	C	**T**	0.501	*JAZF1*	1.14 (1.07–1.20)	1.5E–04	1.08 (1.04–1.12)	8.1E–05	1.10 (1.06–1.15)	1.3E–07	59,617	1.10 (1.07–1.13)	5.0E–14	0.70	10,610
rs12779790	10	12,368,016	**A**	G	0.183	*CDC123, CAMK1D*	1.15 (1.06–1.24)	4.2E–04	1.11 (1.06–1.16)	5.4E–05	1.09 (1.04–1.14)	1.5E–04	62,366	1.11 (1.07–1.14)	1.2E–10	0.67	9,334
rs7961581	12	69,949,369	**T**	C	0.269	*TSPAN8, LGR5*	1.18 (1.10–1.26)	1.8E–05	1.06 (1.02–1.11)	9.8E–03	1.09 (1.04–1.13)	4.3E–05	62,301	1.09 (1.06–1.12)	1.1E–09	0.20	23,206
rs7578597	2	43,644,474	C	**T**	0.902	*THADA*	1.25 (1.12–1.40)	1.8E–04	1.15 (1.07–1.22)	1.6E–03	1.12 (1.05–1.20)	9.2E–05	60,832	1.15 (1.10–1.20)	1.1E–09	0.008	9,624
rs4607103	3	64,686,944	T	**C**	0.761	*ADAMTS9*	1.13 (1.06–1.22)	5.4E–04	1.10 (1.05–1.15)	1.0E–04	1.06 (1.01–1.11)	3.5E–03	62,387	1.09 (1.06–1.12)	1.2E–08	0.17	9,748
rs10923931[c]	1	120,230,001	**G**	T	0.106	*NOTCH2*	1.30 (1.17–1.43)	1.1E–04	1.09 (1.03–1.16)	2.9E–03	1.11 (1.05–1.18)	1.9E–03	58,667	1.13 (1.08–1.17)	4.1E–08	0.004	21,568
rs1153188	12	53,385,263	T	**A**	0.733	*DCD*	1.15 (1.08–1.23)	3.2E–05	1.07 (1.03–1.12)	3.1E–03	1.06 (1.02–1.10)	8.8E–03	62,301	1.08 (1.05–1.11)	1.8E–07	0.79	17,808
rs17036101[d]	3	12,252,845	**A**	G	0.927	*SYN2,* PPARG	1.33 (1.18–1.50)	1.0E–05	1.13 (1.04–1.22)	4.5E–03	1.11 (1.02–1.20)	1.2E–02	59,682	1.15 (1.10–1.21)	2.0E–07	0.19	16,370
rs2641348[c]	1	120,149,926	**A**	G	0.107	*ADAM30*	1.14 (1.05–1.25)	1.4E–03	1.10 (1.03–1.17)	1.2E–03	1.09 (1.03–1.16)	7.8E–03	60,048	1.10 (1.06–1.15)	4.0E–07	0.08	17,428
rs9472138	6	43,919,740	**C**	T	0.282	*VEGFA*	1.13 (1.06–1.21)	5.4E–05	1.07 (1.02–1.12)	1.5E–03	1.03 (1.00–1.07)	9.5E–02	63,537	1.06 (1.04–1.09)	4.0E–06	0.43	16,696
rs10490072	2	60,581,582	C	**T**	0.724	*BCL11A*	1.17 (1.10–1.26)	3.4E–05	1.08 (1.03–1.13)	1.4E–03	1.00 (0.97–1.04)	6.5E–01	59,682	1.05 (1.03–1.08)	1.0E–04	0.0035	13,502
Maximum available effective sample size							9,562		21,461		32,514						

Results from the analysis of directly genotyped data only, except for FUSION stage 1 results for rs7961581 (**Supplementary Methods**). Combined estimates of ORs were calculated using a fixed effects, inverse variance meta-analysis; DGI discordant sibling pairs were not included in OR estimates. *P* values were combined using a weighted *z* score–based meta-analysis including DGI sibships; *P* values for the three stage 1 studies were individually corrected by genomic control before meta analysis.

[a]Ancestral allele is denoted in bold, based on Entrez SNP and derived by comparison against chimpanzee sequence. The risk allele frequencies presented are sample size-weighted risk allele frequencies across the stage 2 studies. [b]Sample size (sum of case and control samples) required for 80% power (to achieve nominal replication at $\alpha = 0.05$) is calculated on the basis of the stage 2 OR estimate, sample size–weighted risk allele frequency across the stage 2 studies and assuming an equal number of cases and controls (**Supplementary Methods**). [c]SNPs rs10923931 and rs2641348 appear to represent the same signal ($r^2 = 0.92$ in HapMap CEU)[13]. Results for rs2934381 and rs2793831 (perfect proxies for rs10923931) are presented for UK (stage 1,2) and deCODE (stage 3) respectively. [d]The signal at SNP rs17036101 is indistinguishable from that at rs1801282, the established P12A variant in *PPARG*.

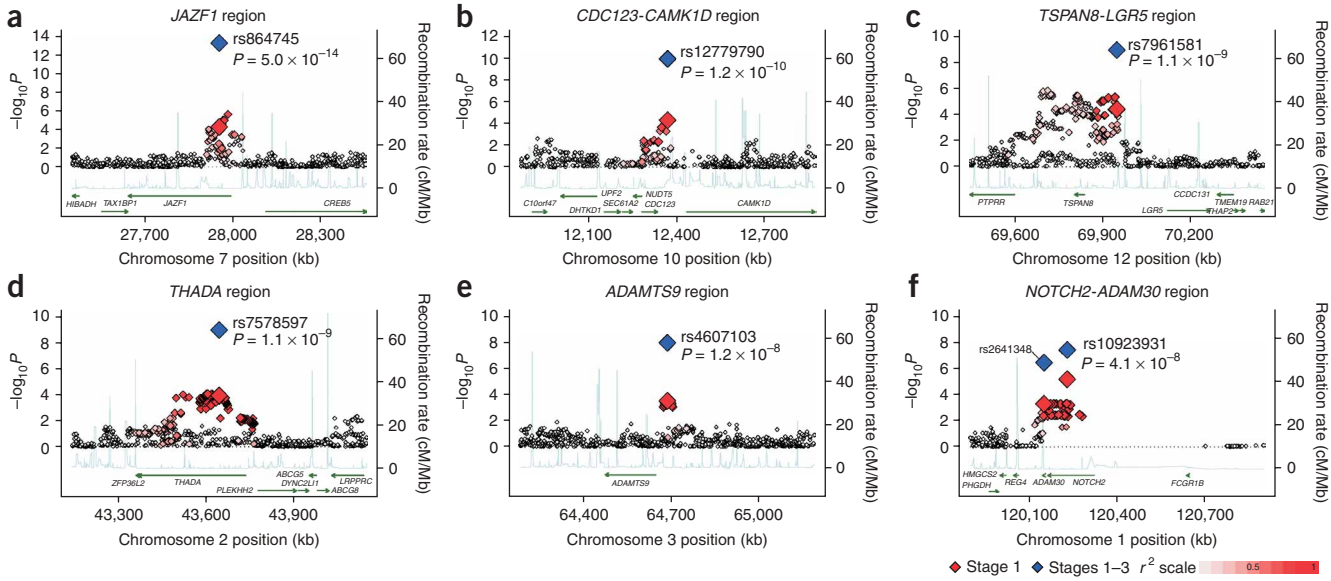

Figure 1 Regional plots of six confirmed associations. (**a**–**f**) For each of the *JAZF1* (**a**), *CDC123-CAMK1D* (**b**), *TSPAN8-LGR5* (**c**), *THADA* (**d**), *ADAMTS9* (**e**) and *NOTCH2-ADAM30* (**f**) regions, genotyped and imputed SNPs passing quality control across all three stage 1 studies are plotted with their meta-analysis *P* values (as $-\log_{10}$ values) as a function of genomic position (NCBI Build 35). In each panel, the SNP taken forward to stages 2 and 3 is represented by a blue diamond (meta-analysis *P* value across stages 1–3), and its initial *P* value in stage 1 data is denoted by a red diamond. Estimated recombination rates (taken from HapMap)[13] are plotted to reflect the local LD structure around the associated SNPs and their correlated proxies (according to a white to red scale from $r^2 = 0$ to $r^2 = 1$; based on pairwise r^2 values from HapMap CEU)[13]. Gene annotations were taken from the University of California Santa Cruz Genome Browser.

1 samples, where previously imputed) generated $P < 10^{-5}$. We further genotyped these 11 SNPs in up to 57,366 additional samples (14,157 cases and 43,209 controls) of European descent in stage 3 (**Table 1**, **Supplementary Table 1** and **Supplementary Methods**). The distribution of *P* values for these 11 SNPs was again inconsistent with a null distribution: all nine newly identified and independent SNPs had effects in the same direction as in the stage 1 + 2 meta-analysis ($P = 0.002$), and seven had $P < 0.05$ in the direction of the original association ($P = 2.1 \times 10^{-10}$) (**Table 2**).

On the basis of the combined stage 1–3 analyses, we found that six signals reached compelling levels of evidence ($P = 5.0 \times 10^{-8}$ or better) for association with T2D (**Table 2**). As in all linkage disequilibrium (LD)-mapping approaches, characterization of the causal variants responsible, their effect sizes and the genes through which they act will require extensive resequencing and fine-mapping. However, on the basis of current evidence, we found that the most associated variants in each of these signals map to intron 1 of *JAZF1*, between *CDC123* and *CAMK1D*, between *TSPAN8* and *LGR5*, in exon 24 of *THADA*, near *ADAMTS9* and in intron 5 of *NOTCH2*.

The strongest statistical evidence for a new association signal was for rs864745 in intron 1 of *JAZF1* (**Fig. 1**), one of a cluster of associated SNPs with strong evidence for association in the stage 1 meta-analysis and across each replication sample (**Table 2** and **Supplementary Table 6** online). The overall estimate of effect was an OR of 1.10 (95% CI = 1.07–1.13; $P = 5.0 \times 10^{-14}$ under an additive model), based on 68,042 individuals. *JAZF1* (juxtaposed with another zinc finger gene 1) encodes a transcriptional repressor of NR2C2 (nuclear receptor subfamily 2, group C, member 2)[16]. Mice deficient in *Nr2c2* show growth retardation, low IGF1 serum concentrations and perinatal and early postnatal hypoglycaemia[17]. Very recently, a SNP in *JAZF1* was identified as associated with prostate cancer[18]; this is particularly interesting given the recent finding that SNPs in *HNF1B* are also associated both with T2D and prostate cancer[19,20].

The second strongest statistical evidence for a new signal was for rs12779790 (combined OR = 1.11, 95% CI = 1.07–1.14, $P = 1.2 \times 10^{-10}$), which lies in an intergenic region ~90 kb from *CDC123* (cell division cycle 123 homolog (*S. cerevisiae*)) and ~63.5 kb from *CAMK1D* (calcium/calmodulin-dependent protein kinase ID) (**Fig. 1**, **Table 2** and **Supplementary Table 6**). CDC123 is regulated by nutrient availability in *S. cerevisiae* and has a role in cell cycle regulation[21]. Evidence from previous GWA studies implicating variants in *CDKAL1* and near *CDKN2A/B* in T2D predisposition suggests that cell cycle dysregulation may be a common pathogenetic mechanism in T2D[1,2,4].

The third strongest statistical signal was found for rs7961581, which resides upstream of *TSPAN8* (tetraspanin 8; combined OR = 1.09, 95% CI = 1.06–1.12, $P = 1.1 \times 10^{-9}$) (**Fig. 1**, **Table 2** and **Supplementary Table 6**). Tetraspanin 8 is a cell-surface glycoprotein expressed in carcinomas of the colon, liver and pancreas.

The fourth strongest new association signal was found for rs7578597, a nonsynonymous SNP (T1187A; combined OR = 1.15, 95% CI = 1.10–1.20, $P = 1.1 \times 10^{-9}$) that resides in exon 24 of the widely expressed *THADA* (thyroid adenoma associated) gene (**Fig. 1**, **Table 2** and **Supplementary Table 6**). Disruption of *THADA* by chromosomal rearrangements (including fusion with intronic sequence from *PPARG*) is observed in thyroid adenomas[22]. The function of THADA has not been well characterized, but there is some evidence to suggest it may be involved in the death receptor pathway and apoptosis[23].

rs4607103 (combined OR = 1.09, 95% CI = 1.06–1.12, $P = 1.2 \times 10^{-8}$), representing a cluster of associated SNPs, resides ~38 kb upstream of *ADAMTS9* (ADAM metallopeptidase with thrombospondin

type 1 motif, 9), and is the SNP with the fifth strongest signal (**Fig. 1**, **Table 2** and **Supplementary Table 6**). ADAMTS9 is a secreted metalloprotease that cleaves the proteoglycans versican and aggrecan, and it is expressed widely, including in skeletal muscle and pancreas.

The sixth strongest signal is marked by rs10923931, which resides within intron 5 of *NOTCH2* (Notch homolog 2 (*Drosophila*); combined OR = 1.13, 95% CI = 1.08–1.17, $P = 4.1 \times 10^{-8}$) (**Fig. 1**, **Table 2** and **Supplementary Table 6**). We also followed up on rs2641348, a nonsynonymous SNP (L359P) within the neighboring gene *ADAM30* (ADAM metallopeptidase domain 30) that represents the same signal ($r^2 = 0.92$ based on HapMap CEU data), but we found that its overall signal (combined OR = 1.10, 95% CI = 1.06–1.15, $P = 4.0 \times 10^{-7}$; **Table 2**) was slightly weaker. NOTCH2 is a type 1 transmembrane receptor; in mice, Notch2 is expressed in embryonic ductal cells of branching pancreatic buds during pancreatic organogenesis, the likely source of endocrine and exocrine stem cells[24].

The strength of the association evidence for the remaining four variants taken into stage 3 did not meet our prespecified threshold of $P \leq 5.0 \times 10^{-8}$. However, it is likely (based on individual significance values and their overall distribution) that several of these variants also represent genuine association signals. In all, three of these additional SNPs showed *P* values $<10^{-5}$ across the combined data (**Table 2**), and two had $P < 0.05$ in stage 3 in the same direction as in stages 1 and 2. Variants near *DCD* (dermcidin) showed evidence for association (rs1153188; overall $P = 1.8 \times 10^{-7}$) (**Supplementary Fig. 2** online). A signal in *VEGFA* had previously been noted in the WTCCC GWA scan[4], but it showed inconsistent evidence for replication: further studies will be required to establish its status. We also found association at rs17036101, ~44 kb downstream of *SYN2* (synapsin II) and 115.3 kb upstream of the established T2D susceptibility variant rs1801282 (P12A) in *PPARG* ($r^2 = 0.54$ in HapMap CEU) (**Supplementary Fig. 3** online). Conditional analyses in stage 1 + 2 samples could not differentiate between the effect of these two SNPs (**Supplementary Note** and **Supplementary Table 7** online).

None of the 11 SNPs (**Table 2**) were convincingly associated with body mass index (BMI) (**Supplementary Table 8** online) or other T2D-related traits (with $P < 10^{-3}$) (**Supplementary Table 9** online). The largest fold-change in T2D association *P* values before and after adjusting for BMI was for rs17036101 ($P = 8.1 \times 10^{-8}$ before adjustment and $P = 7.5 \times 10^{-6}$ after adjustment for BMI; **Supplementary Table 10** online). Conditioning on the associated SNP that was taken forward to stages 2 and 3 in each region showed no additional independent association signals ($P < 10^{-4}$) in stage 1 data (**Supplementary Note** and **Supplementary Fig. 4** online).

By combining three GWA scans involving 10,128 samples (enhanced through imputation approaches) and undertaking large-scale replication in up to 79,792 additional samples, we identified six additional loci that apparently harbor common genetic variants influencing susceptibility to T2D. These findings are consistent with a model in which the preponderance of loci detectable through the GWA approach (using current arrays and indirect LD mapping) have modest effects (ORs between 1.1 and 1.2). Given such a model, our study (in which we followed up only 69 signals out of over 2 million meta-analysed SNPs) would be expected to recover only a subset of the loci with similar characteristics (that is, those that managed to reach our stage 1 selection criteria). Further efforts to expand GWA meta-analyses and to extend the number of SNPs taken forward to large-scale replication should confirm additional genomic loci, as should targeted analysis of copy number variation. However, the present data provide only crude estimates of the overall effect on susceptibility attributable to variants at these loci. The effect of the actual common causal variant responsible for the index association (once identified) will typically be larger, and many of these loci are likely to carry additional causal variants, including, on occasion, low-frequency variants of larger effect: three genes with common variants that influence risk of T2D were first identified on the basis of rare mendelian mutations (in *KCNJ11*, *WFS1* and *HNF1B*). Regardless of effect size, these loci provide important clues to the processes involved in the maintenance of normal glucose homeostasis and in the pathogenesis of T2D.

METHODS

Stage 1 samples, genome-wide genotyping and quality control. An expanded description of these methods is provided in **Supplementary Methods**.

The WTCCC stage 1 sample consists of 1,924 T2D cases and 2,938 population controls from the UK[3,4]. These samples were genotyped on the Affymetrix GeneChip Human Mapping 500K Array Set. The call frequency of included samples was >0.97. In total, 393,143 autosomal SNPs passed quality control criteria (Hardy-Weinberg equilibrium (HWE) $P > 10^{-4}$ in T2D cases and controls; call frequency >0.95, MAF > 0.01 and good clustering[3,4]).

The DGI stage 1 Swedish and Finnish sample consists of 1,464 T2D cases and 1,467 normoglycemic controls. Of these, 2,097 are population-based T2D cases and controls matched for body mass index (BMI), gender and geographic origin, and 834 are T2D cases and controls in 326 sibships discordant for T2D[1]. These samples were genotyped on the Affymetrix GeneChip Human Mapping 500K Array Set, and all included samples had a genotype call rate >0.95. In total, 378,860 autosomal SNPs passed quality control criteria (call frequency >0.95, HWE $P > 10^{-6}$ in controls and MAF > 0.01 in both population and familial components)[1].

The FUSION stage 1 sample consists of 1,161 Finnish T2D cases and 1,174 Finnish normal glucose-tolerant controls[2]. In addition, we included 122 FUSION offspring with genotyped parents for quality control purposes and quantitative trait analysis. Samples were genotyped with the Illumina HumanHap300 BeadChip (v1.1). All samples included had a call frequency >0.975. In sum, 306,222 autosomal SNPs passed quality control (HWE $P \geq 10^{-6}$ in the total sample, ≤3 combined duplicate or nonmendelian inheritance errors (out of 79 duplicate samples and 122 parent–offspring sets), call frequency ≥0.90 and MAF > 0.01) (ref. 2).

Analysis of stage 1 genotype data. In combining data across the three studies, we did not attempt, given differences in study design and implementation, to harmonize every aspect of individual study analysis and quality control. For the UK, DGI and FUSION studies, respectively, 393,143, 378,860 and 306,222 SNPs were analyzed under an additive model. The genomic control values for these directly genotyped SNPs were 1.08 (UK), 1.06 (DGI) and 1.03 (FUSION) (**Supplementary Methods**).

Stage 1 imputation and T2D analysis. For each stage 1 sample set, we imputed genotypes for autosomal SNPs that were present in HapMap Phase II but that were not present in the genome-wide chip or that did not pass direct genotyping quality control. In each sample, genotypes were imputed using the genotype data from the GWA chips and phased HapMap II genotype data from the 60 CEU HapMap founders. We retained SNPs that had an estimated MAF > 0.01 in the control or total sample. Imputed SNPs were then tested for T2D association. The genomic control values for these imputed SNPs were 1.08 (UK), 1.07 (DGI) and 1.04 (FUSION) (**Supplementary Methods**).

Stage 1 meta-analysis. An expanded description of these methods is provided in **Supplementary Methods**. We used meta-analysis to combine the T2D association results for the stage 1 WTCCC, DGI and FUSION samples. The combined stage 1 data are comprised of 10,128 samples: 4,549 T2D cases and 5,579 controls. We used association results from directly genotyped SNPs, where available, and imputed genotype association results at all other positions. 2,202,892 genotyped and imputed autosomal SNPs passed quality control and had MAF > 0.01 in each of the three samples (44,750 were genotyped in all three samples, 308,685 were genotyped in two samples, 250,280 were genotyped in one sample, and 1,599,177 were imputed in all samples). All association

results were expressed relative to the forward strand of the reference genome based on dbSNP125. In our initial analysis, which was used to select signals for stage 2 genotyping, for each SNP we combined the ORs for a given reference allele weighted by the confidence intervals using a fixed effects model. We investigated evidence for heterogeneity of ORs using two commonly used statistics: Cochrans's Q statistic and I^2 (ref. 25).

We repeated the meta-analysis, combining evidence for association solely on the basis of the P values. Specifically, for each study, we converted the two-sided P value to a z statistic that was signed to reflect the direction of the association given the reference allele. Each z score was then weighted; the squared weights were chosen to sum to 1, and each sample-specific weight was proportional to the square root of the effective number of individuals in the sample. We summed the weighted z statistics across studies and converted the summary z score to a two-sided P value.

SNP prioritization for stage 2 genotyping. We prioritized 69 SNPs for replication in stage 2 on the basis of the results from the three-study stage 1 meta-analysis, using a set of criteria we developed as part of a heuristic approach to the prioritization of loci for follow-up (**Supplementary Methods**). We considered SNPs with a meta-analysis P value $<10^{-4}$ and a meta-analysis heterogeneity P value $>10^{-4}$. These selections were largely made using the initial OR-based version of the meta-analysis. We allowed some exceptions to the above follow-up criteria.

Five SNPs were selected for replication genotyping on the basis of their strong association with T2D in the DGI GWA study (two SNPs), association with T2D and with insulinogenic index in the DGI study (one SNP), and overlap with FUSION or WTCCC ($P < 0.05$ in DGI and one or both studies; two SNPs). For known T2D loci (*TCF7L2*, *CDKAL1*, *IGF2BP2*, *KCNJ11*, *HHEX/IDE*, *SLC30A8*, *CDKN2A/B* region, *WFS1*, *HNF1B* and *FTO*), we excluded from follow-up all SNPs that resided within the surrounding region, with region boundaries defined by the furthest neighboring SNPs with P values remaining $\sim$0.01 ($n = 1{,}981$). For the *PPARG* region, we identified a SNP, rs17036101, with a P value two orders of magnitude lower than the established P12A susceptibility variant, rs1801282, and we took this signal forward to replication. In total, we took 69 SNPs forward to stage 2 genotyping.

Stage 2 samples, genotyping and analysis. We genotyped the prioritized SNPs in cases and controls from three UK replication sets (RS1, RS2 and RS3, described in ref. 4; **Supplementary Table 1** and **Supplementary Methods**). Genotyping of prioritized SNPs in RS1, RS2 and RS3 was done by KBiosciences. All assays were validated prior to use, using a standard 96-well validation plate (KBiosciences) and up to 296 samples from the WTCCC study (**Supplementary Methods**). Concordance rates between the Affymetrix and KASPar/TaqMan genotypes (based on up to 296 replicate stage 1 samples) were 97.5% on average. All genotyped SNPs had genotype call frequency rates >94% in the replication sets, and no SNPs had HWE $P < 0.001$ in cases or controls. We tested for association with T2D using the Cochran-Armitage test for trend. Results from the three replication sets were combined in a Cochran-Mantel-Haenszel meta-analysis framework.

For DGI, we genotyped the prioritized SNPs in three stage 2 case-control samples[1] (**Supplementary Table 1** and **Supplementary Methods**). The prioritized SNPs were genotyped in all DGI stage 1 and 2 samples using the iPLEX Sequenom MassARRAY platform. We used 63 SNPs passing quality control (>94% call rate, MAF > 0.01 and HWE P value >0.001) for association testing. We tested for T2D association in each DGI stage 2 case-control set using a χ^2 analysis (assuming an additive genetic model). Results from the three DGI stage 2 samples were combined using Cochran-Mantel-Haenszel meta-analysis.

For FUSION, we genotyped the prioritized SNPs in a Finnish case-control sample (**Supplementary Table 1** and **Supplementary Methods**) using the Sequenom Homogeneous Mass EXTEND or iPLEX Gold SBE assays, carried out at the National Human Genome Research Institute (NHGRI). In sum, 59 SNPs had genotype call frequency >94% and HWE P value >0.001. The genotype consistency rate among 56 duplicate samples was 100%, and the average call frequency of successfully genotyped SNPs was 97.3%. SNPs were analyzed using logistic regression with adjustment for sex, 5-year age category and birth province and an additive model for the genetic effect.

Comparison of genotypes from imputation and direct genotyping. We genotyped a proportion of the prioritized imputed signals in the stage 1 samples of the three studies, and calculated respective concordance rates (**Supplementary Methods** and **Supplementary Table 4**). All results presented in the main manuscript text are based on directly typed stage 1 data, except rs7961581 in FUSION stage 1.

Combined meta-analysis for stages 1 and 2. We combined stage 1 and stage 2 data using both the OR-based and the weighted z score–based meta-analysis approaches described above. We also assessed our results using random effects meta-analysis to better account for any heterogeneity between the studies (**Supplementary Table 6**). Locus-specific and combined sibling relative risk estimates were calculated using sample size–weighted estimates of the effect size and risk-allele frequency derived from stage 2 replication samples only, and under the assumption of allelic and locus independence, as described[26,27].

Stage 3 sample, genotyping and association analysis. We followed up 11 SNPs (rs2641348, rs10490072, rs7578597, rs17036101, rs4607103, rs9472138, rs864745, rs12779790, rs1153188, rs10923931 and rs7961581) in stage 3 samples from the deCODE, KORA, Danish, HUNT, NHS, GEM Consortium (CCC, EPIC, ADDITION/Ely, Norfolk) and METSIM studies (**Supplementary Table 1** and **Supplementary Methods**).

Combined meta-analysis for stages 1, 2 and 3. We combined stage 1, 2 and 3 data using both meta-analysis approaches (fixed-effects model to combine ORs and weighted P value–based z statistic combination across all sample sets) described above. We also assessed our results using random effects meta-analysis (**Supplementary Table 6**). We observed some evidence for heterogeneity across studies (the I^2 statistic ranged from 0 to 57.8% depending on the SNP), with rs7578597 and rs10923931 showing the largest fold differences in association P value between the fixed- and random-effects model analyses. Differences in strength of association across studies (leading to evidence for heterogeneity) could reflect interesting biological associations that vary from study to study depending on subject ascertainment scheme.

Genomic control. An expanded description of these methods is described in **Supplementary Methods**. We adopted two strategies in reporting the findings from this study. In the first, we performed GC-correction of data from DGI, FUSION and WTCCC before stage 1 meta-analysis. We corrected each individual study for the GC inflation observed (directly genotyped and imputed data separately), and combined results across studies. We present the genome-wide distribution of association statistics in **Supplementary Figure 1**. We note that, after study-specific genomic control adjustment, the estimated inflation factor for the stage 1 meta-analysis test statistic was 1.04.

In the second strategy, we combined GC-uncorrected data from DGI, FUSION and WTCCC for stage 1 meta-analysis and did not correct the meta-analysis test statistics for the overall GC (to guard against over-conservativeness in the estimate of strength of association for interesting signals). We also present the genome-wide distribution of these statistics in **Supplementary Figure 1**.

For the combination of data across stages 1, 2 and 3, we also adopted these two strategies (of using GC-corrected and GC-uncorrected stage 1 data). In the first, we performed individual GC-correction of DGI, FUSION and WTCCC stage 1 data before meta-analysis with stage 2 and stage 3 data (an approach which may be over-conservative where, as was the case here, none of the T2D-associated SNPs had particular hallmarks of stratification) (**Supplementary Note**). In the second, we combined only uncorrected data (except for the deCODE data, for which we applied GC correction, given a more marked genomic control inflation (GC $\sim$1.3) in that sample). We present the resulting data from both approaches (of using GC-corrected and GC-uncorrected stage 1 data for stage 1–3 meta-analysis) in **Supplementary Table 6** and a comparison of results (showing very small differences) in the **Supplementary Note**. All data presented elsewhere in the manuscript reflect the GC-corrected analysis strategy outcome.

Conditional analysis of T2D signals. For each SNP in **Table 2**, we assessed the additive SNP association in the stage 1 and 2 samples before and after including

BMI in the logistic regression model. For each genotyped and imputed SNP surrounding a specific T2D signal, we assessed the additive SNP association in the stage 1 sample before and after including the **Table 2** SNP from the same region in the model. We analyzed the data and adjusted for covariates for the stage 1 and stage 2 analysis of each sample. Data were combined across studies as described above. The ORs and CIs were calculated using a fixed-effects model, and *P* values were calculated using the weighted *z* score method. For the WTCCC stage 1 samples, we did not have BMI information available for ~1,500 of the population-based controls. We therefore carried out the conditional BMI analyses by using all T2D cases and only those controls for whom BMI data were available.

Quantitative trait analyses. Quantitative trait analyses were carried out in the UK, DGI and FUSION samples for the 11 SNPs taken forward to stage 3. We tested BMI, quantitative glycemic traits (fasting and 2-h levels of glucose and insulin, HOMA-IR (homeostasis model assessment of insulin resistance)), lipid traits (total, HDL and LDL cholesterol, and serum triglycerides) and blood pressure (systolic and diastolic), where available, for association using an additive genetic model (**Supplementary Methods**).

URLs. MACH, http://www.sph.umich.edu/csg/abecasis/MaCH/download.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

UK: Collection of the UK type 2 diabetes cases was supported by Diabetes UK, BDA Research and the UK Medical Research Council (Biomedical Collections Strategic Grant G0000649). The UK Type 2 Diabetes Genetics Consortium collection was supported by the Wellcome Trust (Biomedical Collections Grant GR072960). The GWA genotyping was supported by the Wellcome Trust (076113), and the replication genotyping was supported by the European Commission (EURODIA LSHG-CT-2004- 518153), MRC (Project Grant G0601261), Wellcome Trust, Peninsula Medical School and Diabetes UK. E.Z. is a Wellcome Trust Research Career Development Fellow. We acknowledge the contribution of M. Sampson and our team of research nurses. We acknowledge the efforts of J. Collier, P. Robinson, S. Asquith and others at KBiosciences for their rapid and accurate large-scale genotyping.

DGI: We thank the study participants who made this research possible. We thank colleagues in the Broad Genetic Analysis and Biological Samples Platforms for their expertise and contributions to genotyping, data and sample management, and analysis. The initial GWAS genotyping was supported by Novartis (to D.A.); support for additional analysis and genotyping in this report was provided by funding from the Broad Institute of Harvard and MIT, by the Richard and Susan Smith Family Foundation/American Diabetes Association Pinnacle Program Project Award (to D.A.), and by a Freedom to Discovery award of the Foundation of Bristol Myers Squibb (to D.A.). P.I.W.dB., M.J.D. and D.A. acknowledge support from US National Institutes of Health/National Heart, Lung, and Blood Institute grant (U01 HG004171). D.A. was a Burroughs Wellcome Fund Clinical Scholar in Translational Research and is a Distinguished Clinical Scholar of the Doris Duke Charitable Foundation. L.G., T.T., B.I. and M.R.T. and the Botnia Study are principally supported by the Sigrid Juselius Foundation, the Finnish Diabetes Research Foundation, The Folkhalsan Research Foundation and Clinical Research Institute HUCH Ltd; work in Malmö, Sweden was also funded by a Linné grant from the Swedish Research Council (349-2006-237). We thank the Botnia and Skara research teams for clinical contributions, and colleagues at MGH, Harvard, Broad, Novartis and Lund for helpful discussions throughout the course of this work.

FUSION: We thank the Finnish citizens who generously participated in this study and R.Welch for bioinformatics support. Support for this research was provided by US National Institutes of Health grants DK062370 (M.B.), DK072193 (K.L.M.), HL084729 (G.R.A.), HG002651 (G.R.A.) and U54 DA021519; National Human Genome Research Institute intramural project number 1 Z01 HG000024 (F.S.C.); and a postdoctoral fellowship award from the American Diabetes Association (C.J.W.). Genome-wide genotyping was performed by the Johns Hopkins University Genetic Resources Core Facility (GRCF) SNP Center at the Center for Inherited Disease Research (CIDR) with support from CIDR NIH Contract Number N01-HG-65403 and the GRCF SNP Center.

deCODE: We thank the Icelandic study participants whose contribution made this work possible. We also thank the nurses at Noatun (deCODE's sample recruitment center) and personnel at the deCODE core facilities.

KORA study: We thank C. Gieger and G. Fischer for expert data handling. The MONICA/KORA Augsburg studies were financed by the GSF-National Research Center for Environment and Health, Neuherberg, Germany and supported by grants from the German Federal Ministry of Education and Research (BMBF). Part of this work was financed by the German National Genome Research Network (NGFN). Our research was also supported within the Munich Center of Health Sciences (MC Health) as part of LMUinnovativ. We thank all members of field staffs who were involved in the planning and conduct of the MONICA/KORA Augsburg studies.

Danish study: This work was supported by the European Union (EUGENE2, grant no. LSHM-CT-2004-512013), Lundbeck Foundation centre of Applied Medical Genomics in Personalized Disease Prediction, Prevention and Care and The Danish Medical Research Council.

HUNT: The Nord-Trøndelag Health Study (The HUNT Study) is a collaboration between The HUNT Research Centre, Faculty of Medicine, Norwegian University of Science and Technology (NTNU), The National Institute of Public Health, The National Screening Service of Norway and The Nord-Trøndelag County Council.

NHS: The Nurses' Health Study is funded by National Cancer Institute grant CA87969. L.Q. is supported by an American Heart Association Scientist Development Grant. F.B.H. is supported by NIH grants DK58845 and U01 HG004399.

GEM Consortium: We thank all study participants. The work on the Cambridgeshire case-control, Ely, ADDITION and EPIC-Norfolk studies was funded by support from the Wellcome Trust and MRC. The Norfolk Diabetes study is funded by the MRC with support from NHS Research & Development and the Wellcome Trust. We are grateful to S. Griffin, MRC Epidemiology Unit, for assistance with the ADDITION study and M. Sampson and E. Young for help with the Norfolk Diabetes Study. We thank S. Bumpstead, W.E. Bottomley and A. Chaney for rapid and accurate genotyping and J. Ghori for assay design and informatics support. We are grateful to P. Deloukas for overall genotyping support. F.P. and I.B. are funded by the Wellcome Trust.

METSIM: The METSIM study has received grant support from the Academy of Finland (no. 124243).

AUTHOR CONTRIBUTIONS

Writing team and project management: L.J.S., E.Z., R.S., B.F.V., D.A., M.B. & M.I.M. **Study design:** R.S., B.F.V., E.Z., L.J.S., T.E.H., F.B.H., J.J.R., H.C., K.S., O.P., T.I., K.H., M.L., A.T.H., I.B., N.J.W., F.S.C., L.G., D.A., M.I.M. & M.B. **Analysis:** K.S.E., R.M.F., H.L., C.M.L., J.R.B.P., I.P., N.W.R., N.J.T., M.N.W., J.L.M. & E.Z. (UK), P.S.C., C.-J.D., W.L.D., T. Hu, A.U.J., Y.L., H.M.S., C.J.W., G.R.A. & L.J.S. (FUSION), R.S., B.F.V., P.I.W.dB., F.G.K., P.A. & M.J.D. (DGI), U.T. & A.K. (deCODE), N.G., G.A., T.H. & O.P. (Danish), K.M. (HUNT), L. Qi (NHS), C.L. (GEM Consortium), M.L. (Metsim). **Clinical samples and genotyping:** WTCCC, A.S.F.D., T.M.F., C.J.G., G.A.H., K.R.O., C.N.A.P., B.S., M.W., A.D.M., A.T.H. & M.I.M. (UK), L.L.B., P.D., M.R.E., K.K., M.A.M., N.N., M.R., A.J.S., R.N.B., K.L.M., J.T., A.F.M., L. Qin & R.M.W. (FUSION), K.A., K.B.B., N.P.B., L. Gianniny, C.G., B.I., V.L., P.N., M.S., T.T. & L. Groop (DGI), V.S., G.T. & K.S. (deCODE), H.G., C.H., C.M. & T.I. (KORA), G.A., N.G., T.H., T.J., T.L., A.S., K.B.-J. & O.P. (Danish), K.M., E.P., C.P. & K.H. (HUNT), F.B.H. (NHS), F.P., I.B. & N.J.W. (GEM Consortium), J.K. (METSIM).

COMPETING INTERESTS STATEMENT

The authors declare competing financial interests: details accompany the full-text HTML version of the paper at http://www.nature.com/naturegenetics/.

Published online at http://www.nature.com/naturegenetics
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions

1. Diabetes Genetics Initiative.. Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science* **316**, 1331–1336 (2007).
2. Scott, L.J. *et al.* A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science* **316**, 1341–1345 (2007).
3. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**, 661–678 (2007).
4. Zeggini, E. *et al.* Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. *Science* **316**, 1336–1341 (2007).
5. Steinthorsdottir, V. *et al.* A variant in *CDKAL1* influences insulin response and risk of type 2 diabetes. *Nat. Genet.* **39**, 770–775 (2007).

6. Sladek, R. *et al.* A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature* **445**, 881–885 (2007).
7. Florez, J.C. *et al.* A 100K genome-wide association scan for diabetes and related traits in the Framingham Heart Study: replication and integration with other genome-wide datasets. *Diabetes* **56**, 3063–3074 (2007).
8. Rampersaud, E. *et al.* Identification of novel candidate genes for type 2 diabetes from a genome-wide association scan in the Old Order Amish: evidence for replication from diabetes-related quantitative traits and from independent populations. *Diabetes* **56**, 3053–3062 (2007).
9. Hanson, R.L. *et al.* A search for variants associated with young-onset type 2 diabetes in American Indians in a 100K genotyping array. *Diabetes* **56**, 3045–3052 (2007).
10. Hayes, M.G. *et al.* Identification of type 2 diabetes genes in Mexican Americans through genome-wide association studies. *Diabetes* **56**, 3033–3044 (2007).
11. Salonen, J. *et al.* Type 2 diabetes whole-genome association study in four populations: the DiaGen consortium. *Am. J. Hum. Genet.* **81**, 338–345 (2007).
12. McCarthy, M.I. & Zeggini, E. Genome-wide association scans for Type 2 diabetes: new insights into biology and therapy. *Trends Pharmacol. Sci.* **28**, 598–601 (2007).
13. International HapMap Consortium. A second generation human haplotype map of over 3.1 million SNPs. *Nature* **449**, 851–861 (2007).
14. Marchini, J., Howie, B., Myers, S., McVean, G. & Donnelly, P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat. Genet.* **39**, 906–913 (2007).
15. Freedman, M.L. *et al.* Assessing the impact of population stratification on genetic association studies. *Nat. Genet.* **36**, 388–393 (2004).
16. Nakajima, T., Fujino, S., Nakanishi, G., Kim, Y.S. & Jetten, A.M. TIP27: a novel repressor of the nuclear orphan receptor TAK1/TR4. *Nucleic Acids Res.* **32**, 4194–4204 (2004).
17. Collins, L.L. *et al.* Growth retardation and abnormal maternal behavior in mice lacking testicular orphan nuclear receptor 4. *Proc. Natl. Acad. Sci. USA* **101**, 15058–15063 (2004).
18. Thomas, G. *et al.* Multiple loci identified in a genome-wide association study of prostate cancer. *Nat. Genet.* **40**, 310–315 (2008).
19. Gudmundsson, J. *et al.* Two variants on chromosome 17 confer prostate cancer risk, and the one in *TCF2* protects against type 2 diabetes. *Nat. Genet.* **39**, 977–983 (2007).
20. Winckler, W. *et al.* Evaluation of common variants in the six known maturity-onset diabetes of the young (MODY) genes for association with type 2 diabetes. *Diabetes* **56**, 685–693 (2007).
21. Bieganowski, P., Shilinski, K., Tsichlis, P.N. & Brenner, C. Cdc123 and checkpoint forkhead associated with RING proteins control the cell cycle by controlling eIF2γ abundance. *J. Biol. Chem.* **273**, 44656–44666 (2004).
22. Drieschner, N. *et al.* Evidence for a 3p25 breakpoint hot spot region in thyroid tumors of follicular origin. *Thyroid* **16**, 1091–1096 (2006).
23. Drieschner, N. *et al.* A domain of the thyroid adenoma associated gene (THADA) conserved in vertebrates becomes destroyed by chromosomal rearrangements observed in thyroid adenomas. *Gene* **403**, 110–117 (2007).
24. Lammert, E., Brown, J. & Melton, D.A. Notch gene expression during pancreatic organogenesis. *Mech. Dev.* **94**, 199–203 (2000).
25. Higgins, J.P., Thompson, S.G., Deeks, J.J. & Altman, D.G. Measuring inconsistency in meta-analyses. *Br. Med. J.* **327**, 557–560 (2003).
26. Risch, N. & Merikangas, K. The future of genetic studies of complex human diseases. *Science* **273**, 1516–1517 (1996).
27. Lin, S., Chakravarti, A. & Cutler, D.J. Exhaustive allelic transmission disequilibrium tests as a new approach to genome-wide association studies. *Nat. Genet.* **36**, 1181–1188 (2004).

The complete list of authors is as follows:

Eleftheria Zeggini[1,10], Laura J Scott[2,10], Richa Saxena[3–8,10], Benjamin F Voight[3–5,7,10], Jonathan L Marchini[11], Tianle Hu[2], Paul IW de Bakker[3,7,12], Gonçalo R Abecasis[2], Peter Almgren[13], Gitte Andersen[14], Kristin Ardlie[3], Kristina Bengtsson Boström[15], Richard N Bergman[16], Lori L Bonnycastle[17], Knut Borch-Johnsen[14,18], Noël P Burtt[3], Hong Chen[19], Peter S Chines[17], Mark J Daly[3–5,7], Parimal Deodhar[17], Chia-Jen Ding[2], Alex S F Doney[20], William L Duren[2], Katherine S Elliott[1], Michael R Erdos[17], Timothy M Frayling[21,22], Rachel M Freathy[21,22], Lauren Gianniny[3], Harald Grallert[23], Niels Grarup[14], Christopher J Groves[24], Candace Guiducci[3], Torben Hansen[14], Christian Herder[25], Graham A Hitman[26], Thomas E Hughes[19], Bo Isomaa[27,28], Anne U Jackson[2], Torben Jørgensen[29], Augustine Kong[30], Kari Kubalanza[17], Finny G Kuruvilla[3,4,6], Johanna Kuusisto[31], Claudia Langenberg[32], Hana Lango[21,22], Torsten Lauritzen[33], Yun Li[2], Cecilia M Lindgren[1,24], Valeriya Lyssenko[13], Amanda F Marvelle[34], Christa Meisinger[23], Kristian Midthjell[35], Karen L Mohlke[34], Mario A Morken[17], Andrew D Morris[20], Narisu Narisu[17], Peter Nilsson[13], Katharine R Owen[24], Colin NA Palmer[36], Felicity Payne[37], John R B Perry[21,22], Elin Pettersen[38], Carl Platou[35], Inga Prokopenko[1,24], Lu Qi[39,40], Li Qin[34], Nigel W Rayner[1,24], Matthew Rees[17], Jeffrey J Roix[19], Anelli Sandbæk[18], Beverley Shields[22], Marketa Sjögren[13], Valgerdur Steinthorsdottir[30], Heather M Stringham[2], Amy J Swift[17], Gudmar Thorleifsson[30], Unnur Thorsteinsdottir[30], Nicholas J Timpson[1,41], Tiinamaija Tuomi[28,42], Jaakko Tuomilehto[43–45], Mark Walker[46], Richard M Watanabe[47], Michael N Weedon[21,22], Cristen J Willer[2], Wellcome Trust Case Control Consortium[49], Thomas Illig[23], Kristian Hveem[35], Frank B Hu[39,40], Markku Laakso[31], Kari Stefansson[30], Oluf Pedersen[14,18], Nicholas J Wareham[32], Inês Barroso[37], Andrew T Hattersley[21,22], Francis S Collins[17], Leif Groop[13,42], Mark I McCarthy[1,24,50], Michael Boehnke[2,50] & David Altshuler[3,4,6–8,48,50]

[11]Department of Statistics, University of Oxford, Oxford, OX1 3TG, UK. [12]Division of Genetics, Brigham and Women's Hospital, Harvard-Partners Center for Genetics and Genomics, Boston, Massachusetts 02115, USA. [13]Department of Clinical Sciences, Diabetes and Endocrinology Research Unit, University Hospital Malmö, Lund University, S-205 02 Malmö, Sweden. [14]Steno Diabetes Center, Copenhagen, DK-2820, Denmark. [15]Skaraborg Institute, S-541 30 Skövde, Sweden. [16]Department of Physiology and Biophysics, Keck School of Medicine, University of Southern California, Los Angeles, California 90033, USA. [17]Genome Technology Branch, National Human Genome Research Institute, Bethesda, Maryland 20892, USA. [18]Faculty of Health Science, University of Aarhus, Aarhus, DK-8000, Denmark. [19]Diabetes and Metabolism Disease Area, Novartis Institutes for BioMedical Research, 100 Technology Square, Cambridge, Massachusetts 02139, USA. [20]Diabetes Research Group, Division of Medicine and Therapeutics, Ninewells Hospital and Medical School, Dundee, DD1 9SY, UK. [21]Genetics of Complex Traits, Institute of Biomedical and Clinical Science, Peninsula Medical School, Magdalen Road, Exeter, EX1 2LU, UK. [22]Diabetes Genetics, Institute of Biomedical and Clinical Science, Peninsula Medical School, Barrack Road, Exeter, EX2 5DW, UK. [23]Gesellschaft für Strahlenforschung-National Research Center for Environment and Health, Institute of Epidemiology, D-85764 Neuherberg, Germany. [24]Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, Oxford, OX3 7LJ, UK. [25]Institute for Clinical Diabetes Research, German Diabetes Center, Leibniz Institute at Heinrich Heine University, D-40225 Düsseldorf, Germany. [26]Centre for Diabetes and Metabolic Medicine, Barts and The London, Royal London Hospital, Whitechapel, London, E1 1BB, UK. [27]Malmska Municipal Health Center and Hospital, FIN-68601 Jakobstad, Finland. [28]Folkhälsan Research Center, FIN-00014 Helsinki, Finland. [29]Research Centre for Prevention and Health, Glostrup University Hospital, DK-2600 Glostrup, Denmark. [30]deCODE genetics, Sturlugata 8, IS-101 Reykjavik, Iceland. [31]Department of Medicine, University of Kuopio and Kuopio University Hospital, 70210, Kuopio, Finland. [32]MRC Epidemiology Unit, Institute of Metabolic Science, Addenbrooke's Hospital, Cambridge CB2 0QQ, UK. [33]Department of General Practice, University of Aarhus, DK-8000 Aarhus, Denmark. [34]Department of Genetics, University of North Carolina, Chapel Hill, North Carolina 27599, USA. [35]HUNT Research Centre, Department of Public Health and General Practice, Faculty of Medicine, Norwegian University of Science and Technology (NTNU), 7650 Verdal, Norway. [36]Population Pharmacogenetics Group, Biomedical Research Centre, Ninewells Hospital and Medical School, Dundee, DD1 9SY, UK. [37]Metabolic Disease Group, Wellcome Trust Sanger Institute, Hinxton, Cambridge, CB10 1SA, UK. [38]Department of Cancer Research and Molecular Medicine, Norwegian University of Science and Technology (NTNU), N-7600 Levanger, Norway. [39]Departments of Nutrition and Epidemiology, Harvard School of Public Health, Boston, Massachusetts 02115, USA. [40]Channing Laboratory, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts 02115, USA. [41]The MRC Centre for Causal Analyses in Translational Epidemiology, Bristol University, Canynge Hall, Whiteladies Road, Bristol, BS2 8PR, UK. [42]Department of Medicine, Helsinki University Hospital, University of Helsinki, FIN-00300 Helsinki, Finland. [43]Diabetes Unit, Department of Epidemiology and Health Promotion, National Public Health Institute, 00300 Helsinki, Finland. [44]Department of Public Health, University of Helsinki, 00014 Helsinki, Finland. [45]South Ostrobothnia Central Hospital, 60220 Seinäjoki, Finland. [46]Diabetes Research Group, School of Clinical Medical Sciences, Newcastle University, Framlington Place, Newcastle upon Tyne, NE2 4HH, UK. [47]Department of Preventative Medicine, Keck School of Medicine, University of Southern California, Los Angeles, California 90033, USA. [48]Diabetes Unit, Massachusetts General Hospital, Boston, Massachusetts 02114, USA. [49]Membership of the Wellcome Trust Case Control Consortium is provided in the **Supplementary Note**. [50]These authors contributed equally to this work.