ARTICLES

A genome-wide association study identifies novel risk loci for type 2 diabetes

Robert Sladek[1,2,4], Ghislain Rocheleau[1]*, Johan Rung[4]*, Christian Dina[5]*, Lishuang Shen[1], David Serre[1], Philippe Boutin[5], Daniel Vincent[4], Alexandre Belisle[4], Samy Hadjadj[6], Beverley Balkau[7], Barbara Heude[7], Guillaume Charpentier[8], Thomas J. Hudson[4,9], Alexandre Montpetit[4], Alexey V. Pshezhetsky[10], Marc Prentki[10,11], Barry I. Posner[2,12], David J. Balding[13], David Meyre[5], Constantin Polychronakos[1,3] & Philippe Froguel[5,14]

Type 2 diabetes mellitus results from the interaction of environmental factors with a combination of genetic variants, most of which were hitherto unknown. A systematic search for these variants was recently made possible by the development of high-density arrays that permit the genotyping of hundreds of thousands of polymorphisms. We tested 392,935 single-nucleotide polymorphisms in a French case–control cohort. Markers with the most significant difference in genotype frequencies between cases of type 2 diabetes and controls were fast-tracked for testing in a second cohort. This identified four loci containing variants that confer type 2 diabetes risk, in addition to confirming the known association with the *TCF7L2* gene. These loci include a non-synonymous polymorphism in the zinc transporter *SLC30A8*, which is expressed exclusively in insulin-producing β-cells, and two linkage disequilibrium blocks that contain genes potentially involved in β-cell development or function (*IDE–KIF11–HHEX* and *EXT2–ALX4*). These associations explain a substantial portion of disease risk and constitute proof of principle for the genome-wide approach to the elucidation of complex genetic traits.

The rapidly increasing prevalence of type 2 diabetes mellitus (T2DM) is thought to be due to environmental factors, such as increased availability of food and decreased opportunity and motivation for physical activity, acting on genetically susceptible individuals. The heritability of T2DM is one of the best established among common diseases and, consequently, genetic risk factors for T2DM have been the subject of intense research[1]. Although the genetic causes of many monogenic forms of diabetes (maturity onset diabetes in the young, neonatal mitochondrial and other syndromic types of diabetes mellitus) have been elucidated, few variants leading to common T2DM have been clearly identified and individually confer only a small risk (odds ratio ≈ 1.1–1.25) of developing T2DM[1]. Linkage studies have reported many T2DM-linked chromosomal regions and have identified putative, causative genetic variants in *CAPN10* (ref. 2), *ENPP1* (ref. 3), *HNF4A* (refs 4, 5) and *ACDC* (also called *ADIPOQ*)[6]. In parallel, candidate-gene studies have reported many T2DM-associated loci, with coding variants in the nuclear receptor *PPARG* (P12A)[7] and the potassium channel *KCNJ11* (E23K)[8] being among the very few that have been convincingly replicated. The strongest known (odds ratio ≈ 1.7) T2DM association[9] was recently mapped to the transcription factor *TCF7L2* and has been consistently replicated in multiple populations[10–20].

Subjects and study design

The recent availability of high-density genotyping arrays, which combine the power of association studies with the systematic nature of a genome-wide search, led us to undertake a two-stage, genome-wide association study to identify additional T2DM susceptibility loci (Supplementary Fig. 1). In the first stage of this study, we obtained genotypes for 392,935 single-nucleotide polymorphisms (SNPs) in 1,363 T2DM cases and controls (Supplementary Table 1). In order to enrich for risk alleles[21], the diabetic subjects studied in stage 1 were selected to have at least one affected first degree relative and age at onset under 45 yr (excluding patients with maturity onset diabetes in the young). Furthermore, in order to decrease phenotypic heterogeneity and to enrich for variants determining insulin resistance and β-cell dysfunction through mechanisms other than severe obesity, we initially studied diabetic patients with a body mass index (BMI) $<30\,kg\,m^{-2}$. Control subjects were selected to have fasting blood glucose $<5.7\,mmol\,l^{-1}$ in DESIR, a large prospective cohort for the study of insulin resistance in French subjects[22].

Genotypes for each study subject were obtained using two platforms: Illumina Infinium Human1 BeadArrays, which assay 109,365 SNPs chosen using a gene-centred design; and Human Hap300 BeadArrays, which assay 317,503 SNPs chosen to tag haplotype blocks identified by the Phase I HapMap[23]. Of the 409,927 markers that passed quality control (Supplementary Tables 2 and 3), genotypes were obtained for an average of 99.2% (Human1) and 99.4% (Hap300) of markers for each subject with a reproducibility of >99.9% (both platforms). Forty-three subjects were removed from analysis because of evidence of intercontinental admixture (Supplementary Fig. 3) and an additional four because their genotype-determined gender disagreed with clinical records. In total, T2DM association was tested for 100,764 (Human1) and 309,163 (Hap300) SNPs representing 392,935 unique loci (Fig. 1). Because of unequal male/female ratios in our cases and controls, we analysed the 12,666 sex-chromosome SNPs separately for each gender.

[1]Departments of Human Genetics, [2]Medicine and [3]Pediatrics, Faculty of Medicine, McGill University, Montreal H3H 1P3, Canada. [4]McGill University and Genome Quebec Innovation Centre, Montreal H3A 1A4, Canada. [5]CNRS 8090-Institute of Biology, Pasteur Institute, Lille 59019 Cedex, France. [6]Endocrinology and Diabetology, University Hospital, Poitiers 86021 Cedex, France. [7]INSERM U780-IFR69, Villejuif 94807, France. [8]Endocrinology-Diabetology Unit, Corbeil-Essonnes Hospital, Corbeil-Essonnes 91100, France. [9]Ontario Institute for Cancer Research, Toronto M5G 1L7, Canada. [10]Montreal Diabetes Research Center, Montreal H2L 4M1, Canada. [11]Molecular Nutrition Unit and the Department of Nutrition, University of Montreal and the Centre Hospitalier de l'Université de Montréal, Montreal H3C 3J7, Canada. [12]Polypeptide Hormone Laboratory and Department of Anatomy and Cell Biology, Montreal H3A 2B2, Canada. [13]Department of Epidemiology & Public Health, Imperial College, St Mary's Campus, Norfolk Place, London W2 1PG, UK. [14]Section of Genomic Medicine, Imperial College London W12 0NN, and Hammersmith Hospital, Du Cane Road, London W12 0HS, UK.
*These authors contributed equally to this work.

Analysis

Markers were selected for assessment in a second cohort using significance thresholds on the basis of the divergence between the observed and expected *P*-values (Supplementary Figs 4 and 5). These included 28 autosomal SNPs from the Human1 chip ($P < 1 \times 10^{-4}$; Supplementary Table 4) and 43 autosomal SNPs from the Hap300 chip ($P < 5 \times 10^{-5}$; Supplementary Table 5) for a total of 66 unique SNPs representing 44 unique loci. No X-linked marker attained significance, a result that may be due to reduced power of the gender-specific analysis. *P*-values calculated using 10,000 permutations of the disease state labels identified the same significant associations (Supplementary Tables 4 and 5). Our stage 1 results included the known T2DM association[9] with the *TCF7L2* SNP rs7903146 ($P = 3.2 \times 10^{-17}$). Several other SNPs at that locus also attained genome-wide significance after correcting for 392,935 tests. In contrast, none of the other previously identified T2DM genes did so (for example, *PPARG*), which is not surprising because our stage 1 had limited power to detect their modest effect and also the arrays did not include the best-associated variants at these loci. However, SNPs tagging four out of seven of these loci attained significance at $P < 0.05$ (Supplementary Table 6).

Because one of the loci showing the strongest T2DM association (rs932206) maps 200 kilobases (kb) telomeric to the lactase gene on 2q21, a region displaying recent positive selection and a north-to-south minor allele frequency (MAF) gradient in Europe, we suspected a spurious association due to population stratification[24]. This was tested with principal component analysis using 20,323 markers with MAF ≥ 0.1, perfectly genotyped in all samples on the Human Hap300 chip, showing no T2DM association in stage 1 ($P > 0.01$) and separated by at least 100 kb. Using the first principal component as a covariate for ancestry differences between cases and controls, we tested for association between rs932206 and disease status. Our result suggests that this apparent association is largely attributable to ancestry differences ($P = 0.0016$ after adjusting for stratification). However, the selection responsible for the European gradient may be related to metabolic fitness and T2DM risk, and therefore the observed association may not be spurious. Similar testing of the other significant loci did not reveal evidence of stratification and the correction did not affect the statistical significance of their association with T2DM.

We thus prioritized 59 SNPs showing significant association in stage 1, including one of the eight significant *TCF7L2* markers, for rapid confirmation on a larger cohort, using the Sequenom iPlex assay (Supplementary Fig. 1). We successfully obtained genotypes from 2,617 T2DM cases and 2,894 controls for 57 SNPs (see Supplementary Information). Unlike the stage 1 sample, the affected individuals used in stage 2 were not required to have a family history of T2DM or to be lean (however, severely obese subjects were excluded by requiring BMI $< 35\,\mathrm{kg\,m^{-2}}$). We also relaxed the inclusion criteria for control subjects to include individuals with normal fasting glucose levels according to 1997 American Diabetes Association (ADA) criteria (<6.1 mM). The SNPs selected for rapid validation were analysed analogously to stage 1 (Supplementary Table 7). In total, eight SNPs representing five unique loci showed significant association after Bonferroni correction was applied for the 57 SNPs tested, based on *P*-values calculated using 10,000,000 permutations of the disease state labels (Table 1). Because the validation stage samples were selected on the basis of more relaxed inclusion criteria than the stage 1 samples, we used a logistic regression model to investigate the effect of phenotypic variables on T2DM association. The Wald test was used to assess effects of age, sex and BMI on the association between marker and disease, as it is asymptotically equivalent to the Armitage trend test used to detect association in stages 1 and 2. None of the associations (Supplementary Table 7) was substantially changed by considering the effects of these covariates.

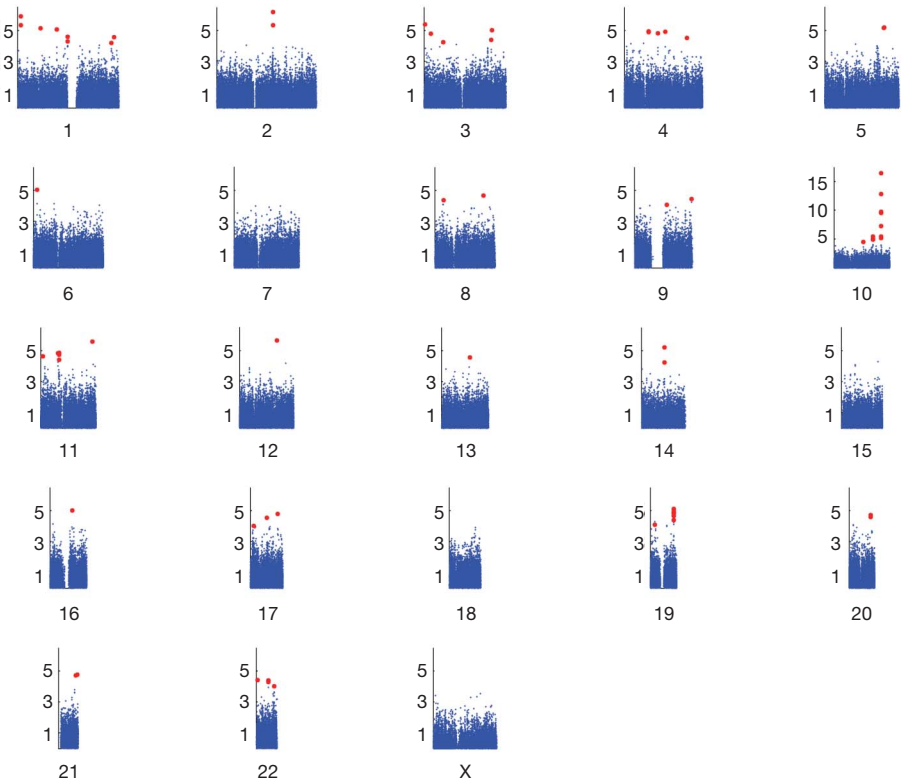

Figure 1 | Graphical summary of stage 1 association results. T2DM association was determined for SNPs on the Human1 and Hap300 chips. The *x* axis represents the chromosome position from pter; the *y* axis shows $-\log_{10}$[pMAX], the *P*-value obtained by the MAX statistic, for each SNP (Note the different scale on the *y* axis of the chromosome 10 plot.). SNPs that passed the cutoff for a fast-tracked second stage are highlighted in red.

Table 1 | Confirmed association results

SNP	Chromosome	Position (nucleotides)	Risk allele	Major allele	MAF (case)	MAF (ctrl)	Odds ratio (het)	Odds ratio (hom)	PAR	λ_s	Stage 2 pMAX	Stage 2 pMAX (perm)	Stage 1 pMAX	Stage 1 pMAX (perm)	Nearest gene
rs7903146	10	114,748,339	T	C	0.406	0.293	1.65 ± 0.19	2.77 ± 0.50	0.28	1.0546	1.5×10^{-34}	$<1.0\times10^{-7}$	3.2×10^{-17}	$<3.3\times10^{-10}$	*TCF7L2*
rs13266634	8	118,253,964	C	C	0.254	0.301	1.18 ± 0.25	1.53 ± 0.31	0.24	1.0089	6.1×10^{-8}	5.0×10^{-7}	2.1×10^{-5}	1.8×10^{-5}	*SLC30A8*
rs1111875	10	94,452,862	G	G	0.358	0.402	1.19 ± 0.19	1.44 ± 0.24	0.19	1.0069	3.0×10^{-6}	7.4×10^{-6}	9.1×10^{-6}	7.3×10^{-6}	*HHEX*
rs7923837	10	94,471,897	G	G	0.335	0.377	1.22 ± 0.21	1.45 ± 0.25	0.20	1.0065	7.5×10^{-6}	2.2×10^{-5}	3.4×10^{-6}	2.5×10^{-6}	*HHEX*
rs7480010	11	42,203,294	G	A	0.336	0.301	1.14 ± 0.13	1.40 ± 0.25	0.08	1.0041	1.1×10^{-4}	2.9×10^{-4}	1.5×10^{-5}	1.2×10^{-5}	LOC387761
rs3740878	11	44,214,378	A	A	0.240	0.272	1.26 ± 0.29	1.46 ± 0.33	0.24	1.0046	1.2×10^{-4}	2.8×10^{-4}	1.8×10^{-5}	1.3×10^{-5}	*EXT2*
rs11037909	11	44,212,190	T	T	0.240	0.271	1.27 ± 0.30	1.47 ± 0.33	0.25	1.0045	1.8×10^{-4}	4.5×10^{-4}	1.8×10^{-5}	1.3×10^{-5}	*EXT2*
rs1113132	11	44,209,979	C	C	0.237	0.267	1.15 ± 0.27	1.36 ± 0.31	0.19	1.0044	3.3×10^{-4}	8.1×10^{-4}	3.7×10^{-5}	2.9×10^{-5}	*EXT2*

Significant T2DM associations were confirmed for eight SNPs in five loci. Allele frequencies, odds ratios (with 95% confidence intervals) and PAR were calculated using only the stage 2 data. Allele frequencies in the controls were very close to those reported for the CEU set (European subjects genotyped in the HapMap project). Induced sibling recurrent risk ratios (λ_s) were estimated using stage 2 genotype counts for the control subjects and assuming a T2DM prevalence of 7% in the French population. hom, homozygous; het, heterozygous; major allele, the allele with the higher frequency in controls; pMAX, *P*-value of the MAX statistic from the χ^2 distribution; pMAX (perm), *P*-value of the MAX statistic from the permutation-derived empirical distribution (pMAX and pMAX (perm) are adjusted for variance inflation); risk allele, the allele with higher frequency in cases compared with controls.

Identification of four novel T2DM loci

Our fast-track stage 2 genotyping confirmed the reported association for rs7903146 (*TCF7L2*) on chromosome 10, and in addition identified significant associations for seven SNPs representing four new T2DM loci (Table 1). In all cases, the strongest association for the MAX statistic (see Methods) was obtained with the additive model. The most significant of these corresponds to rs13266634, a non-synonymous SNP (R325W) in *SLC30A8*, located in a 33-kb linkage disequilibrium block on chromosome 8, containing only the 3′ end of this gene (Fig. 2a). *SLC30A8* encodes a zinc transporter expressed solely in the secretory vesicles of β-cells and is thus implicated in the final stages of insulin biosynthesis, which involve co-crystallization

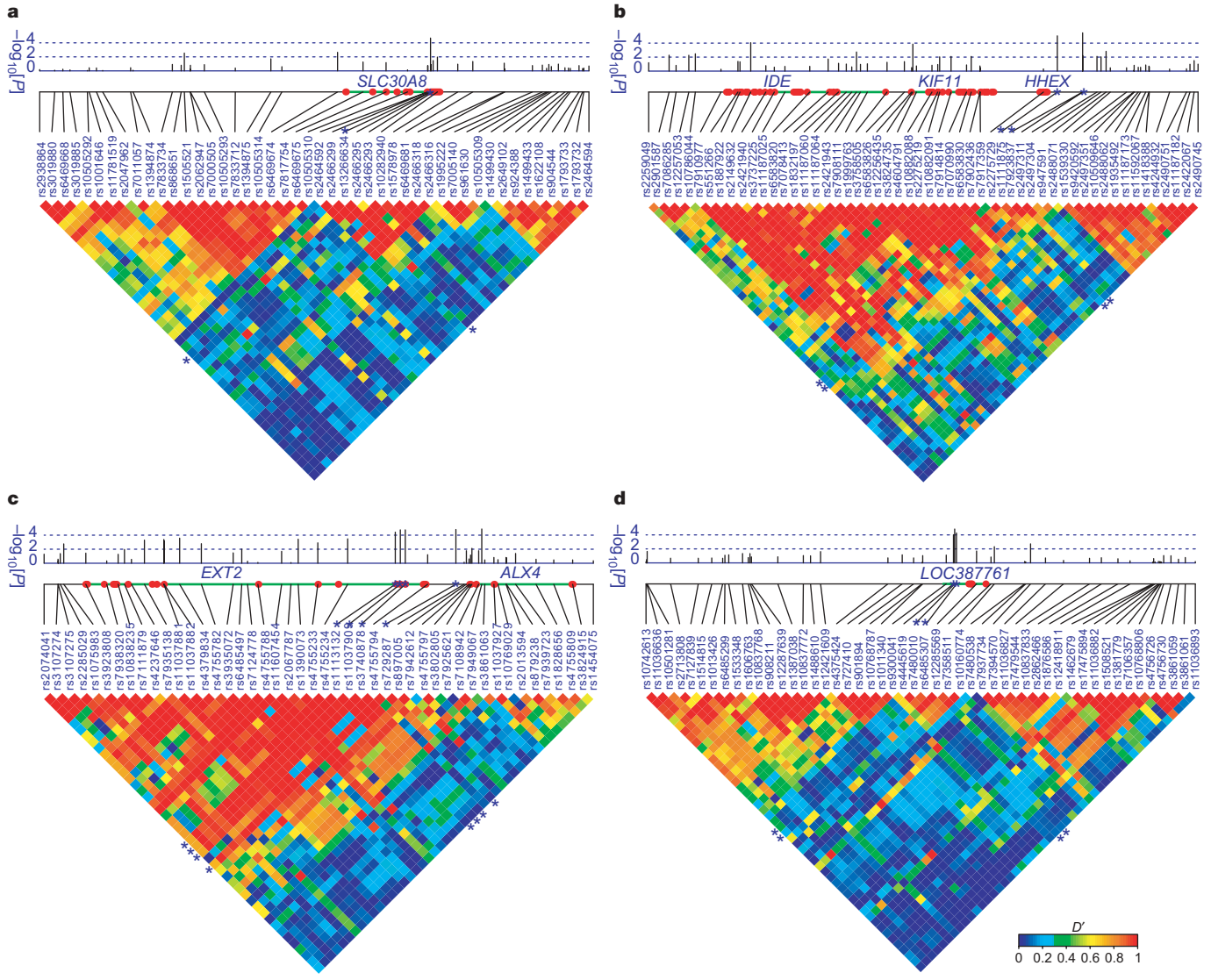

Figure 2 | Pairwise linkage disequilibrium diagrams for four novel T2DM-associated loci. *D′* was calculated from the stage 1 genotyping data as a fraction of observed linkage disequilibrium over the maximal possible. The bar graph indicates the negative logarithm of the stage 1 *P*-value for each SNP. Transcriptional units are indicated by green lines, with exons highlighted in orange. Blue asterisks mark the SNPs chosen for confirmatory studies. **a**, *SLC30A8*; **b**, *IDE–KIF11–HHEX*; **c**, *EXT2–ALX4*; **d**, LOC387761.

with zinc. Notably, overexpression of *SLC30A8* in insulinoma cells increases glucose-stimulated insulin secretion[25]. This finding suggests possible dietary implications and therapeutic approaches with zinc supplementation or, more plausibly, pharmacological manipulation of its transport.

SNPs rs1111875 and rs7923837 are located near the telomeric end of a 270-kb linkage disequilibrium block on chromosome 10 (Fig. 2b), the only one of the novel loci that maps to an interval confirmed in more than one linkage study[26–29]. The linkage disequilibrium block contains two genes of known biological significance—the insulin-degrading enzyme (*IDE*) and the homeodomain protein *HHEX*—as well as kinesin-interacting factor 11 (*KIF11*). *HHEX* is essential for hepatic and pancreatic development[30,31] and is a target of the Wnt signalling pathway[32], as is *TCF7L2*. Reduction of *IDE* activity by a pharmacological inhibitor increases islet amyloid polypeptide (amylin) accumulation and amylin-mediated cytotoxicity in cultured β-cells[33], whereas *IDE* ablation causes glucose intolerance in knockout mice[34]. Although *IDE* showed weak T2DM association[35,36], these findings were not confirmed in a third well-powered study[37]. Fine mapping of the *IDE–KIF11–HHEX* locus in different populations and, ultimately, biological studies will be required to identify the causative variant.

The statistical significance of the top three loci is robust enough to withstand variance-inflation correction for factors of the order of magnitude we observed in stage 1. Two more loci pass Bonferroni correction for 57 markers, but much closer to the cutoff. The first of these involves three SNPs located in introns of exostosin 2 (*EXT2*), at the telomeric end of a 169-kb linkage disequilibrium block on chromosome 11q (Fig. 2c). *EXT2* modulates hedgehog signalling, a pathway involved in early pancreatic development[38] and the regulation of insulin synthesis[39]. This block also contains *ALX4*, a homeodomain protein with possible involvement in the Wnt pathway[40]. Finally, one additional T2DM-associated SNP in chromosome 11 maps to a linkage disequilibrium block that contains the hypothetical gene LOC387761 (Fig. 2d).

To quantify the contribution of these loci to T2DM risk, we calculated the population attributable risk (PAR) for each marker (Table 1). Stepwise logistic regression showed that one SNP per locus explains the entire locus effect and that there was no significant epistasis between loci. Thus, the PAR for the four novel loci together with *TCF7L2* is 70% (see Supplementary Material). In this context, it is worth noting that for three of the four novel loci, the risk allele is the major allele (Table 1). Thus, although our findings can be the source of valuable physiological insights, their contribution to the familial clustering and individual risk prediction of T2DM is relatively small. Of note, for seven of the eight SNPs in Table 1, the risk allele is the ancestral allele, which may be consistent with the hypothesis[41] that the ancestral alleles were adapted to the environment of ancient human populations but today, in a different environment, they increase disease risk. Further population genetics analyses of these loci will allow direct testing of this hypothesis.

Discussion

Our findings permit a number of preliminary insights into the allelic architecture of T2DM susceptibility. We have demonstrated that five relatively common variants (MAF > 0.2) with modest effects (heterozygous relative risk = 1.15–1.65) contribute a significant part towards T2DM risk. Furthermore, we expect that our full stage 2 study will reveal more such loci. Thus, the contribution to T2DM risk of loci with substantial allelic heterogeneity[42] does not seem to be large. Our results also need to be interpreted in the context of the different selection criteria for the cohorts used in the two stages. We sought to increase effect sizes in stage 1 by excluding obese patients, thus diminishing phenotypic heterogeneity. Consequently, loci conferring risk through effects on insulin secretion or insulin response only in the presence of obesity, if they exist, might not have been detected in our study. In addition, selection for positive family history probably enriched our stage 1 cohort for individuals carrying risk alleles at a smaller number of loci with stronger effects; this might have compromised our power to detect loci with weak effects. These issues should be addressed in future studies using stratified analysis of larger and more diverse case–control samples.

As might be expected from the clinical characteristics of the stage 1 cohort, the T2DM risk loci we identified seem to involve genes implicated in pancreatic development and the control of insulin secretion. These associations were confirmed in the stage 2 cohort, which is more representative of the general French T2DM population, and potentially highlight the importance of impaired β-cell adaptation to increased metabolic demands in the pathogenesis of T2DM. In addition, these loci may also affect the peripheral response to insulin. For example, *TCF7L2* variants may alter insulin sensitivity in addition to regulating insulin secretion[11,19]. *HHEX* regulates cell proliferation and tissue specification underlying vascular and hepatic differentiation[31,43] and *EXT2* is implicated in bone cell proliferation[44], suggesting a more complex scenario of pleiotropic effects. We anticipate that identification of the causal variants at these genetic loci and their functional consequences will reveal unexpected players in T2DM pathogenesis, and will point to novel mechanisms and targeted therapeutics.

METHODS

Subjects. Detailed characteristics of subjects used in each of the two stages are described in Supplementary Information. Briefly, the stage 1 cases were non-obese (BMI $< 30\,kg\,m^{-2}$) individuals diagnosed with T2DM according to the 1997 criteria of the ADA, who had at least one first degree relative with T2DM. Stage 1 control subjects were selected to have normal fasting plasma glucose and a BMI $< 27\,kg\,m^{-2}$. Stage 2 cases required a diagnosis of T2DM by the ADA criteria and BMI $< 35\,kg\,m^{-2}$. Stage 2 controls all had normal fasting plasma glucose and BMI $< 35\,kg\,m^{-2}$.

Stage 1 whole-genome scan and quality control. Genotyping was performed by labelling 750 ng of genomic DNA and hybridizing it to the Illumina Infinium Human1 and Hap300 BeadArrays, which interrogated 109,365 and 317,503 SNPs, respectively (see Supplementary Information). No significant difference in call rates between cases and controls was seen. Samples successfully genotyped in less than 95% of markers on either array were excluded from analysis, as were subjects whose genotype-inferred gender disagreed with clinical records. Markers were excluded if they deviated significantly from Hardy–Weinberg equilibrium ($P < 0.001$ in the control samples), if they had low MAF (<0.01 in both the case and control samples), or if they had a call rate $<95\%$ in the case and control samples combined (Supplementary Fig. 2 and Supplementary Tables 2 and 3).

Statistical analysis of stage 1. To identify and correct for possible population stratification, case and control genotypes were analysed using STRUCTURE[45]. For this analysis, our data set was 'spiked' with genotypes of unrelated individuals from the four HapMap populations (see Supplementary Information). T2DM association was tested with additive (Armitage trend test), dominant and recessive models for autosomal SNPs[46], and the largest test statistic obtained from the three models was chosen (MAX statistic). To compensate accurately for testing three models, significance was also estimated against the empirical distribution of the MAX statistic after performing 10,000 permutations of the case and control labels for each marker. To correct for variance inflation owing to systematic genotyping errors or subtle subpopulation structure[47–50], the observed χ^2 statistic was adjusted using the Genomic Control method[48] for each of the three genetic models[47].

Genotyping and analysis of stage 2. Genotypes for markers selected for fast-track confirmation were obtained using the Sequenom iPLEX assay (Sequenom). Allele detection was performed using matrix-assisted laser desorption/ionization–time-of-flight mass spectrometry. Quality control criteria for markers were the same as in stage 1. Association testing was performed using the MAX statistic and 10,000,000 permutations of the disease state labels. To be considered significant, an association had to involve the same risk allele in both stages. Using the permutation *P*-values, Bonferroni correction over the 57 SNPs tested gives a significance threshold of $P = 8.8 \times 10^{-4}$.

Received 11 November 2006; accepted 23 January 2007.
Published online 11 February 2007.

1. Permutt, M. A., Wasson, J. & Cox, N. Genetic epidemiology of diabetes. *J. Clin. Invest.* 115, 1431–1439 (2005).

2. Horikawa, Y. *et al.* Genetic variation in the gene encoding calpain-10 is associated with type 2 diabetes mellitus. *Nature Genet.* **26,** 163–175 (2000).
3. Meyre, D. *et al.* Variants of *ENPP1* are associated with childhood and adult obesity and increase the risk of glucose intolerance and type 2 diabetes. *Nature Genet.* **37,** 863–867 (2005).
4. Love-Gregory, L. D. *et al.* A common polymorphism in the upstream promoter region of the hepatocyte nuclear factor-4α gene on chromosome 20q is associated with type 2 diabetes and appears to contribute to the evidence for linkage in an ashkenazi jewish population. *Diabetes* **53,** 1134–1140 (2004).
5. Silander, K. *et al.* Genetic variation near the hepatocyte nuclear factor-4α gene predicts susceptibility to type 2 diabetes. *Diabetes* **53,** 1141–1149 (2004).
6. Vasseur, F. *et al.* Single-nucleotide polymorphism haplotypes in the both proximal promoter and exon 3 of the *APM1* gene modulate adipocyte-secreted adiponectin hormone levels and contribute to the genetic risk for type 2 diabetes in French Caucasians. *Hum. Mol. Genet.* **11,** 2607–2614 (2002).
7. Altshuler, D. *et al.* The common PPARγ Pro12Ala polymorphism is associated with decreased risk of type 2 diabetes. *Nature Genet.* **26,** 76–80 (2000).
8. Gloyn, A. L. *et al.* Large-scale association studies of variants in genes encoding the pancreatic β-cell K_{ATP} channel subunits Kir6.2 (*KCNJ11*) and SUR1 (*ABCC8*) confirm that the *KCNJ11* E23K variant is associated with type 2 diabetes. *Diabetes* **52,** 568–572 (2003).
9. Grant, S. F. *et al.* Variant of transcription factor 7-like 2 (*TCF7L2*) gene confers risk of type 2 diabetes. *Nature Genet.* **38,** 320–323 (2006).
10. Zhang, C. *et al.* Variant of transcription factor 7-like 2 (*TCF7L2*) gene and the risk of type 2 diabetes in large cohorts of U.S. women and men. *Diabetes* **55,** 2645–2648 (2006).
11. Damcott, C. M. *et al.* Polymorphisms in the transcription factor 7-like 2 (*TCF7L2*) gene are associated with type 2 diabetes in the Amish: replication and evidence for a role in both insulin secretion and insulin resistance. *Diabetes* **55,** 2654–2659 (2006).
12. Scott, L. J. *et al.* Association of transcription factor 7-like 2 (*TCF7L2*) variants with type 2 diabetes in a Finnish sample. *Diabetes* **55,** 2649–2653 (2006).
13. Groves, C. J. *et al.* Association analysis of 6,736 U.K. subjects provides replication and confirms *TCF7L2* as a type 2 diabetes susceptibility gene with a substantial effect on individual risk. *Diabetes* **55,** 2640–2644 (2006).
14. Cauchi, S. *et al. TCF7L2* variation predicts hyperglycemia incidence in a French general population: the data from an epidemiological study on the Insulin Resistance Syndrome (DESIR) study. *Diabetes* **55,** 3189–3192 (2006).
15. Chandak, G. R. *et al.* Common variants in the *TCF7L2* gene are strongly associated with type 2 diabetes mellitus in the Indian population. *Diabetologia* **50**, 63–67 (2007).
16. Florez, J. C. *et al. TCF7L2* polymorphisms and progression to diabetes in the Diabetes Prevention Program. *N. Engl. J. Med.* **355,** 241–250 (2006).
17. Humphries, S. E. *et al.* Common variants in the *TCF7L2* gene and predisposition to type 2 diabetes in UK European whites, Indian Asians and Afro-Caribbean men and women. *J. Mol. Med.* **84,** 1–10 (2006).
18. Parton, L. E. *et al.* Limited role for SREBP-1c in defective glucose-induced insulin secretion from Zucker diabetic fatty rat islets: a functional and gene profiling analysis. *Am. J. Physiol. Endocrinol. Metab.* **291,** E982–E994 (2006).
19. Saxena, R. *et al.* Common single nucleotide polymorphisms in *TCF7L2* are reproducibly associated with type 2 diabetes and reduce the insulin response to glucose in nondiabetic individuals. *Diabetes* **55,** 2890–2895 (2006).
20. Weedon, M. N. *et al.* Combining information from common type 2 diabetes risk polymorphisms improves disease prediction. *PLoS Med.* **3,** e374 (2006).
21. Fingerlin, T. E., Boehnke, M. & Abecasis, G. R. Increasing the power and efficiency of disease-marker case-control association studies through use of allele-sharing information. *Am. J. Hum. Genet.* **74,** 432–443 (2004).
22. Balkau, B. An epidemiologic survey from a network of French Health Examination Centres, (D.E.S.I.R.): epidemiologic data on the insulin resistance syndrome. *Rev. Epidemiol. Sante Publique* **44,** 373–375 (1996).
23. International HapMap Consortium. A haplotype map of the human genome. *Nature* **437,** 1299–1320 (2005).
24. Campbell, C. D. *et al.* Demonstrating stratification in a European American population. *Nature Genet.* **37,** 868–872 (2005).
25. Chimienti, F. *et al. In vivo* expression and functional characterization of the zinc transporter ZnT8 in glucose-induced insulin secretion. *J. Cell Sci.* **119,** 4199–4206 (2006).
26. Duggirala, R. *et al.* Linkage of type 2 diabetes mellitus and of age at onset to a genetic location on chromosome 10q in Mexican Americans. *Am. J. Hum. Genet.* **64,** 1127–1140 (1999).
27. Ghosh, S. *et al.* The Finland-United States investigation of non-insulin-dependent diabetes mellitus genetics (FUSION) study. I. An autosomal genome scan for genes that predispose to type 2 diabetes. *Am. J. Hum. Genet.* **67,** 1174–1185 (2000).
28. Meigs, J. B., Panhuysen, C. I., Myers, R. H., Wilson, P. W. & Cupples, L. A. A genome-wide scan for loci linked to plasma levels of glucose and HbA(1c) in a community-based sample of Caucasian pedigrees: The Framingham Offspring Study. *Diabetes* **51,** 833–840 (2002).
29. Wiltshire, S. *et al.* A genomewide scan for loci predisposing to type 2 diabetes in a U.K. population (the Diabetes UK Warren 2 Repository): analysis of 573 pedigrees provides independent replication of a susceptibility locus on chromosome 1q. *Am. J. Hum. Genet.* **69,** 553–569 (2001).
30. Bort, R., Martinez-Barbera, J. P., Beddington, R. S. & Zaret, K. S. *Hex* homeobox gene-dependent tissue positioning is required for organogenesis of the ventral pancreas. *Development* **131,** 797–806 (2004).
31. Bort, R., Signore, M., Tremblay, K., Martinez Barbera, J. P. & Zaret, K. S. *Hex* homeobox gene controls the transition of the endoderm to a pseudostratified, cell emergent epithelium for liver bud development. *Dev. Biol.* **290,** 44–56 (2006).
32. Foley, A. C. & Mercola, M. Heart induction by Wnt antagonists depends on the homeodomain transcription factor Hex. *Genes Dev.* **19,** 387–396 (2005).
33. Bennett, R. G., Hamel, F. G. & Duckworth, W. C. An insulin-degrading enzyme inhibitor decreases amylin degradation, increases amylin-induced cytotoxicity, and increases amyloid formation in insulinoma cell cultures. *Diabetes* **52,** 2315–2320 (2003).
34. Farris, W. *et al.* Insulin-degrading enzyme regulates the levels of insulin, amyloid β-protein, and the β-amyloid precursor protein intracellular domain *in vivo*. *Proc. Natl Acad. Sci. USA* **100,** 4162–4167 (2003).
35. Groves, C. J. *et al.* Association and haplotype analysis of the insulin-degrading enzyme (IDE) gene, a strong positional and biological candidate for type 2 diabetes susceptibility. *Diabetes* **52,** 1300–1305 (2003).
36. Karamohamed, S. *et al.* Polymorphisms in the insulin-degrading enzyme gene are associated with type 2 diabetes in men from the NHLBI Framingham Heart Study. *Diabetes* **52,** 1562–1567 (2003).
37. Florez, J. C. *et al.* High-density haplotype structure and association testing of the insulin-degrading enzyme (IDE) gene with type 2 diabetes in 4,206 people. *Diabetes* **55,** 128–135 (2006).
38. Apelqvist, A., Ahlgren, U. & Edlund, H. Sonic hedgehog directs specialised mesoderm differentiation in the intestine and pancreas. *Curr. Biol.* **7,** 801–804 (1997).
39. Thomas, M. K., Rastalsky, N., Lee, J. H. & Habener, J. F. Hedgehog signaling regulation of insulin production by pancreatic β-cells. *Diabetes* **49,** 2039–2047 (2000).
40. Boras-Granic, K., Grosschedl, R. & Hamel, P. A. Genetic interaction between *Lef1* and *Alx4* is required for early embryonic development. *Int. J. Dev. Biol.* **50,** 601–610 (2006).
41. Di Rienzo, A. & Hudson, R. R. An evolutionary framework for common diseases: the ancestral-susceptibility model. *Trends Genet.* **21,** 596–601 (2005).
42. Pritchard, J. K. Are rare variants responsible for susceptibility to complex diseases? *Am. J. Hum. Genet.* **69,** 124–137 (2001).
43. Hallaq, H. *et al.* A null mutation of *Hhex* results in abnormal cardiac development, defective vasculogenesis and elevated Vegfa levels. *Development* **131,** 5197–5209 (2004).
44. Stickens, D. *et al.* The *EXT2* multiple exostoses gene defines a family of putative tumour suppressor genes. *Nature Genet.* **14,** 25–32 (1996).
45. Pritchard, J. K., Stephens, M. & Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics* **155,** 945–959 (2000).
46. Sasieni, P. D. From genotypes to genes: doubling the sample size. *Biometrics* **53,** 1253–1261 (1997).
47. Clayton, D. G. *et al.* Population structure, differential bias and genomic control in a large-scale, case-control association study. *Nature Genet.* **37,** 1243–1246 (2005).
48. Devlin, B. & Roeder, K. Genomic control for association studies. *Biometrics* **55,** 997–1004 (1999).
49. Reich, D. E. & Goldstein, D. B. Detecting association in a case-control study while correcting for population stratification. *Genet. Epidemiol.* **20,** 4–16 (2001).
50. Kohler, K. & Bickeboller, H. Case-control association tests correcting for population stratification. *Ann. Hum. Genet.* **70,** 98–115 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was funded by Genome Canada, Génome Québec, and the Canada Foundation for Innovation. Cohort recruitment was supported by the Association Française des Diabétiques, INSERM, CNAMTS, Centre Hospitalier Universitaire Poitiers, La Fondation de France and industrial partners. We thank all individuals who participated as cases or controls in this study. C. Petit, J-P. Riveline and S. Franc were instrumental in recruitment and S. Brunet, F. Bacot, R. Frechette, V. Catudal, M. Deweirder, F. Allegaert, P. Laflamme, P. Lepage, W. Astle, M. Leboeuf and S. Leroux provided technical assistance. K. Shazand and N. Foisset provided organizational guidance. Large-scale computations were made possible by the CLUMEQ supercomputer facility.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to C.P. (Constantin.Polychronakos@McGill.ca).