

## A Genome-Wide Association Study of Type 2 Diabetes in Finns Detects Multiple Susceptibility Variants

Laura J. Scott,<sup>1</sup> Karen L. Mohlke,<sup>2</sup> Lori L. Bonnycastle,<sup>3</sup> Cristen J. Willer,<sup>1</sup> Yun Li,<sup>1</sup> William L. Duren,<sup>1</sup> Michael R. Erdos,<sup>3</sup> Heather M. Stringham,<sup>1</sup> Peter S. Chines,<sup>3</sup> Anne U. Jackson,<sup>1</sup> Ludmila Prokunina-Olsson,<sup>3</sup> Chia-Jen Ding,<sup>1</sup> Amy J. Swift,<sup>3</sup> Narisu Narisu,<sup>3</sup> Tianle Hu,<sup>1</sup> Randall Pruim,<sup>4</sup> Rui Xiao,<sup>1</sup> Xiao-Yi Li,<sup>1</sup> Karen N. Conneely,<sup>1</sup> Nancy L. Riebow,<sup>3</sup> Andrew G. Sprau,<sup>3</sup> Maurine Tong,<sup>3</sup> Peggy P. White,<sup>1</sup> Kurt N. Hetrick,<sup>5</sup> Michael W. Barnhart,<sup>5</sup> Craig W. Bark,<sup>5</sup> Janet L. Goldstein,<sup>5</sup> Lee Watkins,<sup>5</sup> Fang Xiang,<sup>1</sup> Jouko Saramies,<sup>6</sup> Thomas A. Buchanan,<sup>7</sup> Richard M. Watanabe,<sup>8,9</sup> Timo T. Valle,<sup>10</sup> Leena Kinnunen,<sup>10,11</sup> Gonçalo R. Abecasis,<sup>1</sup> Elizabeth W. Pugh,<sup>5</sup> Kimberly F. Doheny,<sup>5</sup> Richard N. Bergman,<sup>9</sup> Jaakko Tuomilehto,<sup>10,11,12</sup> Francis S. Collins,<sup>3\*</sup> Michael Boehnke<sup>1\*</sup>

<sup>1</sup>Department of Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor, MI 48109, USA.

<sup>2</sup>Department of Genetics, University of North Carolina, Chapel Hill, NC 27599, USA. <sup>3</sup>Genome Technology Branch, National Human Genome Research Institute, Bethesda, MD 20892, USA. <sup>4</sup>Department of Mathematics and Statistics, Calvin College, Grand Rapids, MI 49546, USA. <sup>5</sup>Center for Inherited Disease Research, Institute of Genetic Medicine, Johns Hopkins School of Medicine, Baltimore, MD 21224, USA. <sup>6</sup>Savitaipale Health Center, 54800 Savitaipale, Finland. <sup>7</sup>Division of Endocrinology, Keck School of Medicine, University of Southern California, Los Angeles, CA 90033, USA. <sup>8</sup>Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA 90089, USA. <sup>9</sup>Department of Physiology and Biophysics, Keck School of Medicine, University of Southern California, Los Angeles, CA 90033, USA. <sup>10</sup>Diabetes Unit, Department of Epidemiology and Health Promotion, National Public Health Institute, 00300 Helsinki, Finland. <sup>11</sup>Department of Public Health, University of Helsinki, 00014 Helsinki, Finland. <sup>12</sup>South Ostrobothnia Central Hospital, 60220 Seinäjoki, Finland.

\*To whom correspondence should be addressed. E-mail: boehnke@umich.edu (M.B.); francisc@mail.nih.gov (F.S.C.)

**Identifying the genetic variants that increase risk of type 2 diabetes (T2D) has been a formidable challenge. Adopting a genome wide association strategy, we genotyped 1,161 Finnish T2D cases and 1,174 Finnish normal glucose tolerant (NGT) controls with >315,000 SNPs, and imputed genotypes for an additional >2 million autosomal SNPs. We carried out association analysis with these SNPs to identify genetic variants that predispose to T2D, compared our T2D association results with results of two other such studies, and genotyped 80 SNPs in an additional 1,215 Finnish T2D cases and 1,258 Finnish NGT controls. We identify T2D-associated variants in an intergenic region of chromosome 11p12, contribute to the identification of T2D-associated variants near the genes *IGF2BP2*, *CDKAL1*, and *CDKN2A/CDKN2B*, and confirm that variants near *TCF7L2*, *SLC30A8*, *HHEX*, *FTO*, *PPARG*, and *KCNJ11* are associated with T2D risk. This brings to at least ten the number of T2D loci now confidently identified.**

Type 2 diabetes (T2D) is a disease characterized by insulin resistance and impaired pancreatic beta-cell function that affects >170 million people worldwide (1). With first-degree relatives at ~3.5-fold increased risk compared to the general

middle-aged population (2), hereditary factors play an important role in determining T2D risk, together with lifestyle and behavioral factors (3). Intense efforts to identify genetic risk factors in T2D have to date met with only limited success. This report, reports from our collaborators (4–6), and the recently published work of Sladek *et al.* (7), describe results of genome-wide association (GWA) studies that further define the genetic architecture of T2D and identify biological pathways involved in T2D pathogenesis.

We genotyped 1,161 Finnish T2D cases and 1,174 Finnish NGT controls on 317,503 SNPs on the Illumina HumanHap300 BeadChip in stage 1 of a two-stage GWA study of T2D (8). These samples are from the Finland-United States Investigation of NIDDM Genetics (FUSION) (9, 10) and Finrisk 2002 (11) studies (tables S1 and S2A). Among the 317,503 GWA SNPs, 315,635 had  $\geq 10$  copies of the less common allele (minor allele frequency (MAF)  $> .002$ ) and passed quality control criteria (8). We tested these 315,635 SNPs for association with T2D using a model that is additive on the log-odds scale (Table 1 and tables S3 and S4) (8). We observed a modest excess of SNPs with p-values  $< 10^{-4}$  (41 versus 31.6 expected,  $p = .19$ ) (fig. S1). These results argue against the existence of multiple common SNPs with large impact on T2D disease risk, but are consistent with multiple

SNPs that each confer modest risk. They also suggest that matching of cases and controls on birth province, sex, and age (8) has been successful; in support of this conclusion is a genomic control (12) correction value of 1.026.

Analysis of our Illumina HumanHap300 data allowed us to query much of the known SNP variation in the genome. To increase this proportion, we developed an imputation method (8, 13) which uses genotype data and linkage disequilibrium (LD) information from the HapMap CEU samples to predict genotypes of autosomal SNPs not genotyped in our subjects. A total of 2.09 million HapMap CEU SNPs (14) had imputed MAF > 1% in FUSION and passed our imputation quality control criteria. In the HapMap CEU sample, imputed SNPs passing these criteria increased coverage of SNPs with MAF > 1% from 71.9% to 89.1% at  $r^2$  threshold .8.

To increase power to detect T2D predisposing variants, we compared our stage 1 results to GWA results from the Diabetes Genetics Initiative (DGI) and the Wellcome Trust Case Control Consortium (WTCCC). We selected 82 SNPs for FUSION stage 2 follow-up genotyping based on evidence from: (a) FUSION genotyped and imputed SNPs; (b) FUSION-DGI-WTCCC GWA results comparison; and (c) prior T2D association results. For (a) and (b), we used a prioritization algorithm that advantaged SNPs based on genome annotation (8) (table S7) and gave preference to genotyped SNPs over nearby imputed SNPs. We successfully genotyped 80 of the 82 SNPs in our stage 2 sample of 1,215 Finnish T2D cases and 1,258 Finnish NGT controls (8) (table S2B) and carried out joint analysis of the combined FUSION stage 1+2 sample (table S5). DGI (4) and UK T2D Genetics Consortium (UKT2D) (5) investigators also followed up DGI and WTCCC GWAs by genotyping replication samples.

We confirmed well-established T2D associations with *TCF7L2*, *PPARG*, and *KCNJ11* (Table 1) (15–18). SNPs in *TCF7L2* reached genome-wide significance in the FUSION stage 1+2 sample (OR = 1.34,  $p = 1.3 \times 10^{-8}$ ) and in the FUSION-DGI-WTCCC/UKT2D all-data (all GWA and follow-up samples) meta-analysis (OR = 1.37,  $p = 1.0 \times 10^{-48}$ ) (Table 1 and table S5). *PPARG* Pro12Ala (rs1801282) and *KCNJ11* Glu23Lys (rs5219) were not genotyped in the FUSION GWA, but nearby SNPs showed some evidence for T2D association, as did the imputed genotypes for the coding variants. All-data meta-analysis resulted in genome-wide significant T2D association with *KCNJ11* Glu23Lys (OR = 1.14,  $p = 6.7 \times 10^{-11}$ ) and strong evidence for *PPARG* Pro12Ala (OR = 1.14,  $p = 1.7 \times 10^{-6}$ ). The *PPARG* and *KCNJ11* results emphasize the value of combining data across studies, and suggest other T2D-associated loci remain to be found.

The combined samples from the three studies provide evidence for seven additional T2D loci. For the first three of these we had strong evidence in the FUSION stage 1 GWA

data, and for the latter four our FUSION stage 1 evidence was more modest.

A cluster of variants in the *IGF2BP2* (insulin-like growth factor 2 mRNA binding protein 2) region was associated with T2D in our stage 1 sample (e.g. rs1470579 with OR = 1.27,  $p = 1.6 \times 10^{-4}$ ) (Fig. 1A). Combining results for rs4402960 with the DGI and the WTCCC/UKT2D resulted in genome-wide significance (OR = 1.14,  $p = 8.9 \times 10^{-16}$ ) in the all-data meta-analysis. Including rs4402960 genotype as a covariate essentially eliminates evidence for T2D association for other variants in the cluster (Fig. 1A), consistent with all SNPs representing the same T2D-predisposing variant(s). *IGF2BP2* is a paralog of *IGF2BP1*, which binds to the 5' UTR of the insulin-like growth factor 2 (*IGF2*) mRNA and regulates *IGF2* translation (19). *IGF2* is a member of the insulin family of polypeptide growth factors involved in development, growth, and stimulation of insulin action. The most strongly associated *IGF2BP2* SNPs are located in a 50 kb region within intron 2 (Fig. 1A); diabetes-predisposing variants may therefore affect regulation of *IGF2BP2* expression.

SNP rs13266634, a non-synonymous Arg325Trp variant in the pancreatic beta-cell specific zinc-transporter *SLC30A8* (20), showed evidence for T2D association in stage 1 (Table 1 and fig. S2) using our annotation-based algorithm. Modest evidence in stage 2 resulted in stronger evidence in our stage 1+2 sample (OR = 1.18,  $p = 7.0 \times 10^{-5}$ ) (Table 1 and table S5). Subsequent DGI and UKT2D genotyping resulted in strong evidence in the combined samples (OR = 1.12,  $p = 5.3 \times 10^{-8}$ ). Sladek *et al.* (7) recently reported independent T2D association evidence with the same allele in two French samples ( $p = 1.8 \times 10^{-5}$  and  $p = 5.0 \times 10^{-7}$ ). *SLC30A8* transports zinc from the cytoplasm into insulin secretory vesicles (20, 21), where insulin is stored as a hexamer bound with two  $Zn^{2+}$  ions prior to secretion (22). Variation in *SLC30A8* may affect zinc accumulation in insulin granules, affecting insulin stability, storage, or secretion. In high glucose conditions, overexpression of *SLC30A8* in INS-1E cells enhanced glucose-induced insulin secretion (21).

SNP rs9300039 in an intergenic region on chromosome 11 showed evidence for T2D association in stage 1 (Table 1 and Fig. 1B); genotyping stage 2 resulted in near genome-wide significance in our stage 1+2 sample (OR = 1.48,  $p = 5.7 \times 10^{-8}$ ) (Table 1 and tables S3 and S5). In the WTCCC and DGI scans, the nearby SNP rs1514823 ( $r^2 = .97$  with rs9300039) provided weak evidence for T2D association with the appropriate allele; combining results across all three studies gave OR = 1.25 and  $p = 4.3 \times 10^{-7}$ . Fifty-six imputed and two more genotyped SNPs spanning 219 kb are in LD with rs9300039 and show substantial evidence for T2D association ( $p < 10^{-4}$ ) in our stage 1 (table S3 and Fig. 1B). Including genotype for rs9300039 as a covariate essentially eliminates evidence for T2D association with the remaining SNPs (Fig.

1B). This region includes three sets of spliced ESTs but no annotated genes. Identification of a T2D-associated variant >1 Mb from the nearest annotated gene highlights the value of a genome-wide approach. Interestingly, Sladek *et al.* (7) reported strongly associated SNPs in two nearby regions on chromosome 11. SNP rs7480010 near hypothetical gene *LOC387761* is 331 kb centromeric to rs9300039. LD between rs9300039 and rs7480010 is essentially zero ( $r^2 = .00063$ ,  $D' = .036$ ), and rs7480010 showed little evidence for association in our stage 1+2 sample (OR = 1.03,  $p = .54$ ). Sladek *et al.* (7) also reported T2D association with three intronic variants of *EXT2*, located ~2.4 Mb centromeric of rs9300039; we had no evidence for association with *EXT2* SNPs.

SNP rs4712523, located within intron 5 of *CDKAL1*, showed modest evidence for T2D association in our FUSION stage 1 sample, which strengthened slightly in our combined stage 1+2 sample (OR = 1.12,  $p = .0095$ ) (Table 1 and table S5). Nearby SNPs in strong LD with rs4712523 showed modest evidence for T2D association in the DGI scan and considerably stronger evidence in the WTCCC scan. Including strong DGI and UKT2D replication data resulted in genome-wide significance (OR = 1.12,  $p = 4.1 \times 10^{-11}$ ) in the all-data meta-analysis. *CDKAL1* (CDK5 regulatory subunit associated protein 1-like 1) shares protein domain similarity with *CDK5RAP1*, which specifically inhibits activation of *CDK5* by *CDK5R1* (23). Using quantitative RT-PCR analysis of a panel of RNA samples from human tissues and cells, we detected the highest expression of *CDKAL1* in skeletal muscle and brain, and in 293T and HepG2 cells (fig. S3A). The associated SNPs within intron 5, or SNPs in LD with them, may regulate expression of *CDKAL1*, and so affect expression of *CDK5*. *CDK5/CDK5R1* activity is influenced by glucose and may influence beta-cell processes (24, 25); over-activity of *CDK5* in the pancreas may lead to beta-cell degeneration, especially under glucotoxic conditions (26).

SNP rs10811661 near *CDKN2A* and *CDKN2B* showed modest evidence for T2D association in our stage 1+2 sample (OR = 1.20,  $p = .0022$ ) (Table 1 and table S5) and showed genome-wide significance in the all-data meta-analysis (OR = 1.20,  $p = 7.8 \times 10^{-15}$ ). rs10811661 is located upstream of cyclin-dependent kinase inhibitors *CDKN2A* and *CDKN2B*, may have a long-range effect on one of these genes, or may influence a gene not yet annotated. *CDKN2A* and *CDKN2B* inhibit the activity of cyclin-dependent protein kinases *CDK4* and *CDK6*. *Cdk4* activity has been shown to influence beta-cell proliferation and mass in mice, with loss of *Cdk4* leading to diabetes (27, 28). We find *CDKN2A* to be expressed at high levels in islets, adipocytes, brain, and pancreas, and even higher levels in 293T, HeLa, and HepG2 cells (fig. S3B); *CDKN2B* is expressed in islets and adipocytes, and to a lesser degree in small intestine, colon, 293T, and HepG2 cells (fig.

S3C). *CDKN2A* and *CDKN2B* are also tumor suppressor genes, and may play a role in aging (29).

SNPs rs1111875 and rs7923837 showed modest evidence of T2D association in the FUSION and DGI scans, much stronger evidence in the WTCCC scan, and genome-wide significant evidence (OR = 1.13,  $p = 5.7 \times 10^{-10}$ ) in the all-data meta-analysis. These SNPs are in LD ( $r^2 = .70$ ) in a region that includes *HHEX* (hematopoietically expressed homeobox), critical for development of the ventral pancreas (30), the insulin degrading enzyme gene *IDE*, and the kinesin-interacting factor 11 gene *KIF11*. Sladek *et al.* (7) recently reported independent genome-wide significant evidence for T2D association with these SNPs.

The WTCCC/UKT2D group identified evidence for T2D and BMI associations with a set of SNPs including rs8050136 in the *FTO* region; the T2D association appears to be mediated through a primary effect on adiposity (5, 6, 31). We observed modest evidence for association with T2D in the combined FUSION 1+2 sample (OR = 1.11,  $p = .016$ ) (Table 1 and table S5).

T2D can be a component of a larger syndrome of metabolic abnormalities, and we were interested to assess the effects of T2D-related traits on our association results. We repeated our T2D association analysis for the ten SNPs in Table 1 with one of several variables included as an additional covariate. Adjustment for BMI strengthened T2D association with *TCF7L2* and *SLC30A8*, weakened association with rs9300039 and *FTO*, and had little effect on the other loci. The effect of waist was similar to that of BMI; blood pressure variables had essentially no effect.

We previously carried out T2D linkage analysis in the families of many of our stage 1 cases (10). None of the ten Table 1 loci had large T2D LOD scores, although those for *FTO* and *TCF7L2* were 0.63 and 0.60, and so nominally significant. Interestingly, LOD scores for six of the ten loci were > 0.2, compared to 2.2 expected for random genome locations, suggesting enrichment for T2D-associated loci in regions with modest evidence of T2D linkage ( $p = .01$ ), but that the power of the linkage approach was insufficient to distinguish these signals from background.

The ability to construct a list of ten robust and replicated T2D-associated loci (Table 1) represents a landmark in efforts to identify genetic variants that predispose to complex human diseases, although the specific predisposing variants and even the relevant genes remain to be defined. We examined the combined risk of T2D based on these ten loci in our stage 1+2 sample by constructing a logistic regression model and predicting T2D risk for each person (8). Our model distinguished groups of individuals with up to four-times different T2D risk, of potential interest for a personalized preventive medicine program (Fig. 2). However, these predictions from our data may be biased compared to the

general population owing to likely overestimation of ORs due to the “winner’s curse”, enrichment for familial T2D cases, and exclusion of individuals with impaired glucose tolerance or impaired fasting glucose.

Thirty years ago, James V. Neel labeled type 2 diabetes as “the geneticist’s nightmare” (32), predicting that discovery of genetic factors in T2D would be profoundly challenging. Until recently, his prediction has proven true. While large samples and collaboration between three groups were required, we can confidently state that new diabetes risk factors have been identified. Each gene discovery points to a pathway that contributes to pathogenesis, and all of these proteins and their relevant pathways represent potential drug targets for the prevention or treatment of diabetes. Based on the number of other interesting results observed in these studies, it is likely that there are additional T2D-predisposing loci to be found. While much remains to be done, we are at last awakening from Jim Neel’s nightmare.

## References and Notes

1. S. Wild, G. Roglic, A. Green, R. Sicree, H. King, *Diabetes Care* **27**, 1047 (2004).
2. S. S. Rich, *Diabetes* **39**, 1315 (1990).
3. J. Kaprio *et al.*, *Diabetologia* **35**, 1060 (1992).
4. Diabetics Genetics Initiative, *Science*, 26 April 2007 (10.1126/science.1142358).
5. E. Zeggini *et al.*, *Science*, 26 April 2007 (10.1126/science.1142364).
6. P. Donnelly and the WTCCC, personal communication. Data from the Wellcome Trust Case Control Consortium scan.
7. R. Sladek *et al.*, *Nature* **445**, 881 (2007).
8. Materials and methods are available as supporting material on *Science Online*.
9. T. Valle *et al.*, *Diabetes Care* **21**, 949 (1998).
10. K. Silander *et al.*, *Diabetes* **53**, 821 (2004).
11. T. Saaristo *et al.*, *Diab. Vasc. Dis. Res.* **2**, 67 (2005).
12. B. Devlin, K. Roeder, *Biometrics* **55**, 997 (1999).
13. Y. Li, P. Scheet, J. Ding, G. R. Abecasis, (Submitted for publication; manuscript available from GRA).
14. International HapMap Consortium, *Nature* **437**, 1299 (2005).
15. S. F. Grant *et al.*, *Nat. Genet.* **38**, 320 (2006).
16. S. S. Deeb *et al.*, *Nat. Genet.* **20**, 284 (1998).
17. D. Altshuler *et al.*, *Nat. Genet.* **26**, 76 (2000).
18. A. L. Gloyn *et al.*, *Diabetes* **52**, 568 (2003).
19. J. Nielsen *et al.*, *Mol. Cell Biol.* **19**, 1262 (1999).
20. F. Chimienti, S. Devergnas, A. Favier, M. Seve, *Diabetes* **53**, 2330 (2004).
21. F. Chimienti *et al.*, *J. Cell Sci.* **119**, 4199 (2006).
22. M. F. Dunn, *Biometals* **18**, 295 (2005).
23. Y. P. Ching, A. S. Pang, W. H. Lam, R. Z. Qi, J. H. Wang, *J. Biol. Chem.* **277**, 15237 (2002).
24. M. Ubeda, D. M. Kemp, J. F. Habener, *Endocrinology* **145**, 3023 (2004).
25. F. Y. Wei *et al.*, *Nat. Med.* **11**, 1104 (2005).
26. M. Ubeda, J. M. Rukstalis, J. F. Habener, *J. Biol. Chem.* **281**, 28858 (2006).
27. S. G. Rane *et al.*, *Nat. Genet.* **22**, 44 (1999).
28. T. Tsutsui *et al.*, *Mol. Cell Biol.* **19**, 7011 (1999).
29. W. Y. Kim, N. E. Sharpless, *Cell* **127**, 265 (2006).
30. R. Bort, J. P. Martinez-Barbera, R. S. Beddington, K. S. Zaret, *Development* **131**, 797 (2004).
31. T. M. Frayling *et al.*, *Science*, Published online 12 April 2007; 10.1126/science.1141634
32. J. V. Neel, in *The Genetics of Diabetes Mellitus*, W. Creutzfeldt, J. Köbberling, J. V. Neel, Eds. (Springer-Verlag, Berlin; New York, 1976), pp. 1–11.
33. We thank the Finnish citizens who generously participated in this study and our colleagues from the Diabetes Genetics Initiative, the Wellcome Trust Case Control Consortium, and the UK Type 2 Diabetes Genetics Consortium for sharing pre-publication data from their studies. We thank S. Enloe of FUSION, and E. Kwasnik, J. Gearhart, J. Romm, M. Zilka, C. Ongaco, A. Robinson, R. King, B. Craig, and E. Hsu of the Center for Inherited Disease Research (CIDR) for expert technical work, and D. Leja of NHGRI for expert assistance with a figure. Support for this research was provided by NIH grants DK062370 (MB), DK072193 (KLM), HL084729 (GRA), HG002651 (GRA), and U54 DA021519, National Human Genome Research Institute intramural project number 1 Z01 HG000024 (FSC), a postdoctoral fellowship award from the American Diabetes Association (CJW), a Wenner-Gren Fellowship (LPO), and a Calvin Research Fellowship (RP). 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.

## Supporting Online Material

[www.sciencemag.org/cgi/content/full/1142382/DC1](http://www.sciencemag.org/cgi/content/full/1142382/DC1)

Author Contributions

Materials and Methods

Figures S1 to S3

Tables S1 to S7

References

12 March 2007; accepted 20 April 2007

Published online 26 April 2007; 10.1126/science.1142382

Include this information when citing this paper.

**Fig. 1.** Plots of T2D association and LD in FUSION stage 1 samples for regions surrounding (A) *IGF2BP2* and (B)

rs9300039. The top panel contains RefSeq genes; there are none in the rs9300039 region. The second panel shows the T2D association  $-\log_{10}$  p-values in FUSION stage 1 samples for SNPs genotyped in the GWA panel (•) or imputed (o). The third panel shows T2D association  $-\log_{10}$  p-values for each SNP in a logistic regression model correcting for the reference SNP (•, red dot): (A) rs4402960 and (B) rs9300039. SNP rs7480010, reported by Sladek *et al.* (7), is also labeled in the rs9300039 plot (B) (•, green dot). A decrease in the  $-\log_{10}$  p-value from the second to the third panel indicates that the association signal of the tested SNPs can be explained, at least in part, by the reference SNP. In both regions, the reference SNP was chosen for convenience; choice of another strongly associated SNP nearby would have resulted in a similar picture. The fourth panel shows recombination rate in cM per Mb for the HapMap CEU sample (14). The fifth and sixth panels show linkage disequilibrium  $r^2$  and  $D'$  based on FUSION stage 1 genotyped and imputed data.

**Fig. 2.** Prediction of T2D risk in the FUSION sample using ten T2D susceptibility variants. T2D cases and NGT controls with complete genotype data were included in the analysis. To obtain a sample with T2D prevalence ~10%, 2,176 NGT controls were included nine times each and 2,102 T2D cases once each. The predicted risk for each individual was estimated from a logistic regression model containing the ten risk variants listed in Table 1. The proportion of T2D cases is shown for twenty equal intervals of predicted T2D risk. 95% CIs for the proportion of T2D cases in each interval were constructed using the original sample of 2,102 cases and 2,176 controls. The constructed sample T2D prevalence (.096) is shown as a horizontal line. The proportion of T2D cases increases from ~5% in the lowest to 20% in the highest predicted risk categories.

Table 1. Confirmed T2D susceptibility loci based on all available data from the FUSION, DGI, and UK samples

FUSION	Chr	Position (bp)	Genes	Risk allele/ non-risk allele	FUSION control risk allele freq	FUSION Stage 1			FUSION Stage 2			FUSION Stage 1 + 2			DGI All Samples			UK All Samples			FUSION DGI UK All Samples			Sample size for 80% power
						OR	95% CI	p-value	OR	95% CI	p-value	OR	95% CI	p-value	OR	95% CI	p-value	OR	95% CI	p-value	OR	95% CI	p-value	
<b>New T2D Loci</b>																								
rs4402960	3	186,994,389	<i>IGF2BP2</i>	T/G	.30	1.28	1.13-1.45	1.2 x 10 <sup>-4</sup>	1.08	0.96-1.22	.22	1.18	1.08-1.28	2.1 x 10 <sup>-4</sup>	1.17	1.11-1.23	1.7 x 10 <sup>-9</sup>	1.11	1.05-1.16	1.6 x 10 <sup>-4</sup>	1.14	1.11-1.18	8.9 x 10 <sup>-16</sup>	~4,300
rs7754840 <sup>c</sup>	6	20,769,229	<i>CDKAL1</i>	C/G	.36	1.16	1.02-1.30	.021	1.08	0.96-1.22	.20	1.12	1.03-1.22	.0095	1.08	1.03-1.14	2.4 x 10 <sup>-3</sup>	1.16	1.10-1.22	1.3 x 10 <sup>-8</sup>	1.12	1.08-1.16	4.1 x 10 <sup>-11</sup>	~5,300
rs10811661	9	22,124,094	<i>CDKN2A/B</i>	T/C	.85	1.17	0.98-1.39	.082	1.22	1.04-1.44	.015	1.20	1.07-1.36	.0022	1.20	1.12-1.28	5.4 x 10 <sup>-8</sup>	1.19	1.11-1.28	4.9 x 10 <sup>-7</sup>	1.20	1.14-1.25	7.8 x 10 <sup>-15</sup>	~3,900
rs9300039 <sup>d</sup>	11	41,871,942		C/A	.89	1.52	1.24-1.87	6.0 x 10 <sup>-5</sup>	1.45	1.19-1.77	2.7 x 10 <sup>-4</sup>	1.48	1.28-1.71	5.7 x 10 <sup>-8</sup>	1.16 <sup>b</sup>	0.95-1.42	.12	1.13 <sup>i</sup>	0.99-1.29	.068	1.25	1.15-1.37	4.3 x 10 <sup>-7</sup>	~3,400
rs8050136	16	52,373,776	<i>FTO</i>	A/C	.38	1.03	0.92-1.16	.58	1.18	1.05-1.33	.0063	1.11	1.02-1.20	.016	1.03 <sup>b</sup>	0.91-1.17	.25	1.23	1.18-1.32	7.3 x 10 <sup>-14</sup>	1.17	1.12-1.22	1.3 x 10 <sup>-12</sup>	~2,700
<b>Previously published T2D association</b>																								
rs1801282	3	12,368,125	<i>PPARG</i>	C/G	.82	1.30	1.11-1.53	.0011	1.08	0.93-1.26	.33	1.20	1.07-1.33	.0014	1.09	1.01-1.16	.019	1.23 <sup>i</sup>	1.09-1.41	.0013	1.14	1.08-1.20	1.7 x 10 <sup>-6</sup>	~6,400
rs13266634	8	118,253,964	<i>SLC30A8</i>	C/T	.61	1.22	1.08-1.38	.0010	1.14	1.02-1.28	.026	1.18	1.09-1.29	7.0 x 10 <sup>-5</sup>	1.07	1.0-1.16	.047	1.12	1.05-1.18	7.0 x 10 <sup>-5</sup>	1.12	1.07-1.16	5.3 x 10 <sup>-8</sup>	~5,100
rs1111875 <sup>c</sup>	10	94,452,862	<i>HHEX</i>	C/T	.52	1.13	1.01-1.27	.039	1.06	0.94-1.19	.34	1.10	1.01-1.19	.026	1.14	1.06-1.22	1.7 x 10 <sup>-4</sup>	1.13	1.07-1.19	4.6 x 10 <sup>-6</sup>	1.13	1.09-1.17	5.7 x 10 <sup>-10</sup>	~4,200
rs7903146 <sup>f</sup>	10	114,748,339	<i>TCF7L2</i>	T/C	.18	1.39	1.20-1.61	1.2 x 10 <sup>-5</sup>	1.30	1.12-1.50	3.5 x 10 <sup>-4</sup>	1.34	1.21-1.49	1.3 x 10 <sup>-8</sup>	1.38	1.31-1.46	2.3 x 10 <sup>-31</sup>	1.37 <sup>i</sup>	1.25-1.49	6.7 x 10 <sup>-13</sup>	1.37	1.31-1.43	1.0 x 10 <sup>-48</sup>	~1,000
rs5219 <sup>e</sup>	11	17,366,148	<i>KCNJ11</i>	T/C	.46	1.20	1.07-1.36	.0022	1.04	0.92-1.16	.55	1.11	1.02-1.21	.013	1.15	1.09-1.21	1.0 x 10 <sup>-7</sup>	1.15 <sup>i</sup>	1.05-1.25	.0013	1.14	1.10-1.19	6.7 x 10 <sup>-11</sup>	~3,700
Total sample size						2,335			2,473			4,808			13,781			13,965			32,544			
N cases/controls						1,161/1,174			1,215/1,258			2,376/2,432			6,529/7,252			5,681/8,284			14,586/17,968			

<sup>a</sup>Stage 1 + 2 risk allele frequency

<sup>b</sup>Approximate sample size for power 80% to detect T2D-SNP association at significance level .05 based on FUSION control risk allele frequency and the risk ratio calculated from FUSION-DGI-WTCCC/UK all samples OR assuming .10 T2D prevalence. Note the sample sizes vary slightly from those of (4) because study-specific allele frequencies were used in the calculations.

<sup>c</sup>rs10946398 UK r<sup>2</sup> = 1

<sup>d</sup>Multi-marker tag for rs9300039 DGI, rs1514823 UK r<sup>2</sup> = .965

<sup>e</sup>rs5015480 WTCCC GWA only r<sup>2</sup>=1

<sup>f</sup>rs7901695 UK r<sup>2</sup> = .849

<sup>g</sup>rs5215 UK r<sup>2</sup> = .995

<sup>h</sup>DGI GWA samples

<sup>i</sup>WTCCC GWA samples



