ONLINE APPENDIX RESEARCH DESIGN AND METHODS

Study cohorts

Cambridgeshire

A total of 552 patients aged 45–76 years with T2D were randomly selected from general practitioner diabetes registers in Cambridgeshire, UK (1). Presence of T2D was based on clinical criteria; onset of diabetes after the age of 30 years without treatment with insulin in the first year after diagnosis. The controls were recruited at random from the same population sampling frames, and were individually matched to cases for age, sex and GP practice. Diabetes was excluded in controls by medical record search and by a glycated haemoglobin measurement of less than 6%. The study received ethical approval from the Cambridge Local Research Ethics Committee, and participants provided informed consent.

EPIC-Norfolk

The EPIC-Norfolk case-control study is nested within the EPIC-Norfolk Study, a population based cohort study of European men and women aged 40-78 years. Both the case-control (2) and full cohort (3) study have been previously described in detail. Briefly, the case-control study consists of 417 incident type 2 diabetes cases and two sets of 417 controls, each matched in terms of age, sex, general practice, recruitment date, with one set additionally matched for BMI. A case was defined by a physician's diagnosis of type 2 diabetes, with no insulin prescribed within the first year following diagnosis, and/or HbA_{1c} > 7% at the health check. Controls were randomly selected from the EPIC-Norfolk cohort from among those without diabetes, cancer, stroke, or myocardial infarction at baseline and who had not developed diabetes by the time of selection. Potential controls with measured HbA_{1c} levels > 6% were excluded. The EPIC-Norfolk study was approved by the Norfolk Local Research Ethics Committtee.

Exeter

The diabetic subjects from Exeter came from two sources (i) a consecutive-case series of patients with T2D diagnosed before 45 years from North and East Devon (4). The patients were unrelated and recruited through questionnaires distributed though general practitioners (97% agreed to send out questionnaires, >70% return rate and >90% recruitment of those identified through the questionnaires). Validation of the diagnosis of diabetes was based on either current prescribed treatment with sulphonylureas, biguanides and/or insulin, or, in the case of individuals treated with diet alone, historical or contemporary laboratory evidence of hyperglycaemia (as defined by present WHO guidelines). All patients were off insulin for at least 1 year after diagnosis, and patients were excluded if they had pancreatic autoantibodies (GAD), first degree history of type 1 diabetes or clinical features (or DNA test results) suggestive of monogenic diabetes (4). (ii) Probands from a collection of type 2 diabetes families that had either both parents available, or one parent and at least two siblings (5). Only subjects collected in Exeter were used in this study. The sex matched controls are taken from the parents in the Exeter Family Study, a cohort study of newly born babies and both their parents (6). This study recruits from central Exeter so the controls come from a similar geographical region as the cases. Diabetes and hyperglycaemia were excluded by measuring fasting glucose and HbA1c. In total 601 cases and 610 controls were included in this study. Informed consent was obtained from all participants.

ADDITION/Ely

©2009 American Diabetes Association. Published online at http://diabetes.diabetesjournals.org/cgi/content/full/db09-0920/DC1

Cases were participants from the UK Cambridge arm of the ADDITION trial, which aims to evaluate whether screening for prevalent undiagnosed Type 2 diabetes is feasible, and whether subsequent optimised intensive treatment of diabetes is feasible and beneficial (7). All cases were aged 40–69 and screen detected using OGTT and WHO diagnostic criteria. We used participants from the Medical Research Council (MRC) Ely study as controls—a population-based study of white European men and women aged 35 to 79 years and from a similar population sampling frame as the Cambridge arm of the ADDITION study (8). Ely Study participants were defined as cases or controls based on their OGTT (WHO diagnostic criteria). For this analysis, the ADDITION case-control study comprised 926 cases and 1497 controls. The Cambridge Research Ethics Committee approved both studies.

Ashkenazi

Of the cases, 303 are from the multiplex-affected sibships that were ascertained for the genome scan described by Permutt et al, 2001 (9). The cases are of Ashkenazi Jewish origin, defined as having all four grandparents born in Northern or Eastern Europe. Subjects with known or suspected Sephardic Jewish or non-Jewish ancestry were excluded. T2D was initially defined according to World Health Organization criteria (fasting glucose 140 mg/dl on two or more occasions, or random glucose 200 mg/dl on two or more occasions). Their average age at ascertainment was 60 years. Average age at diagnosis was 47 years and average duration of diabetes was 13 years (range 0-47). In this population, the incidence of type 1 diabetes is relatively low therefore anti-GAD or anti-islet cell antibody titers were not routinely measured. The additional 627 cases were ascertained as part of a study with the dual aim of looking for diabetes related genes and for genes related to the risk of developing diabetic complications. This group has an average age at ascertainment of 65.8 years, age of diagnosis of 46.8 years and duration of diabetes of 19.1 years. The Ashkenazi control samples consist of 149 elderly subjects (average age 76 years) with no personal or first-degree family history of T2D. The remaining 312 samples were obtained from The National Laboratory for the Genetics of Israeli Populations at Tel Aviv University, Israel. The institutional review boards of Washington University (St. Louis, MO) and Hadassah University Hospital (Jerusalem, Israel) approved the study.

Västerbotten

Twelve-hundred-ninety-six adults with type 2 diabetes were identified through registries covering the county of Västerbotten in northern Sweden, and 1,412 non-diabetic individuals, group matched on age, sex, examination date and geographic region of residence, were selected from the Västerbotten Intervention Programme (VIP) as controls. Virtually all of these individuals were European whites. Type 2 diabetes was determined using the 1999 diagnostic criteria of the World Health Organization. Participants with fasting capillary glucose concentrations <7.0mmol/l and no document history of diabetes underwent a 75g anhydrous oral glucose tolerance test. Accordingly, control subjects were those without a documented history of diabetes and with glucose concentrations below the thresholds for type 2 diabetes. Type 2 diabetes in the case group was defined by clinical diagnoses. All living participants provided written informed consent. Ethics permission was obtained from the Local Research Ethics Committee of Umeå University and approval for genetics investigations in this material was granted by the Swedish Data Inspection Board. Protocols for clinical measurements used in this study have been described previously (10).

©2009 American Diabetes Association. Published online at http://diabetes.diabetesjournals.org/cgi/content/full/db09-0920/DC1

Selection of conserved regions for sequencing

MultiPIP-maker (http://pipmaker.bx.psu.edu/pipmaker/) and VISTA MLAGAN (http://lagan.stanford.edu/lagan_web/index.shtml) were used to create alignments of human *WFS1* genomic sequence and 5kb flanking regions with chimp, macaque, dog, cow, mouse and rat, and the Dcode ECR browser (http://ecrbrowser.dcode.org/) was used to create alignments with mouse, dog, macaque, opossum, chicken, frog and fugu pufferfish. Conserved sequence >100bp with >80% sequence identity between human and at least one other species were selected for sequencing (two regions in intron 1: NCBI B36 coordinates 6325012-6325193 (181 bp) and 6325875-6326013 (138 bp), and two regions upstream of *WFS1*: 6321756-6321858 (102 bp) and 6322195-6322297 (102 bp)).

PCR and sequencing

PCR was performed on genomic DNA from Cambridgeshire case-control participants, or genomic DNA randomly preamplified in GenomiPhi reactions (GE Healthcare UK, Chalfont St. Giles, UK) from ADDITION and Ely study participants. PCR was performed using standard conditions and products purified using exonuclease I and shrimp alkaline phosphatase (USB Corporation, Cleveland, OH). Bi-directional sequencing was performed using Big Dye Terminator 3.1 kit (Applied Biosystems, Foster City, CA).

Tagging SNP selection

Tagging SNPs were selected to cover variants detected by sequencing of a subset of cases (N=24) and controls (N=68) from the Cambridgeshire case-control study. The tagging SNPs also included four additional variants not covered by sequencing (rs4689391, rs752854, rs6446482, and rs3821943) that were genotyped as part of the original study (11) (62 variants in total). For reasons of power, variants with MAF<0.05 were excluded unless they altered the amino acid sequence, leaving 49 variants. Linkage disequilibrium (LD) was calculated using Haploview v4.0 (http://www.broad.mit.edu/mpg/haploview), and tagging SNPs were selected by Tagger using an r^2 threshold of 0.8, force including non-synonymous variants. Thirty tagging SNPs (including the seven variants genotyped by the original study) were selected to cover the 49 *WFS1* variants (Online Appendix Figure 2). An additional tagging SNP, rs12642481, was required to cover variation in HapMap CEU trios.

Quality control

Of 31 tagging SNPs genotyped in Cambridgeshire and EPIC (7 of which had already been genotyped in these populations (11)), three - rs7655482, rs1046316, and WFS1_K800E - failed during manual confirmation of the genotype clusters. All remaining SNPs were checked for deviation from Hardy-Weinberg equilibrium (P<0.001), low call rates (N<85%) and significant discrepancy in call rate between cases and controls (P<0.001). All failed SNPs except WFS1_K800E and WFS1_V871M were imputed, and WFS1_V871M was tested in Cambridgeshire and ADDITION/Ely samples as part of the low frequency variant analysis. For the low frequency variant analysis, variants and samples with call rate<85% were excluded from analysis (2 synonymous, 1 non-synonymous, and 2 non-coding variants, all with MAF<0.001, were eliminated). All common variants were also tested for deviation from Hardy-Weinberg equilibrium and for statistically significant differences in call rate between cases and controls using the χ^2 statistic (1df) in Stata SE9.

©2009 American Diabetes Association. Published online at http://diabetes.diabetesjournals.org/cgi/content/full/db09-0920/DC1

Imputation

The best guess genotypes of ungenotyped SNPs were imputed in Cambridgeshire and EPIC separately using BIMBAM software (http://stephenslab.uchicago.edu/software.html). SNPs that failed QC were not used to impute untyped SNPs but were instead imputed themselves. Using LD patterns between variants in the 96 sequenced Cambridgeshire case-control samples, we attempted to impute 39 additional variants detected by sequencing. Of these we were able to impute genotypes of 25 additional variants in all Cambridgeshire and EPIC samples. The other fourteen variants were imputed to be monomorphic in Cambridgeshire and EPIC-Norfolk and were therefore excluded from further analysis. This is probably because these fourteen variants were rare (MAF<0.05) and were not correlated with any genotyped SNPs. We also used LD patterns in HapMap CEU trios to attempt to impute 89 HapMap SNPs in the interval between recombination hotspots flanking the association signal. Of these, 73 passed quality control (no deviation from HWE, P>0.001). Thirty-two successfully imputed HapMap SNPs overlapped with variants imputed from sequencing data or those directly genotyped, therefore a total of 66 additional variants were imputed in total. Genotyped and imputed data (89 variants in total) covered 96% of Cambridgeshire variants (all but two low frequency, non-synonymous changes) and 100% of HapMap SNPs.

ONLINE APPENDIX RESULTS

Analysis of low frequency variants

Mutation load

As a small number of individuals carried more than one rare non-synonymous allele, we used logistic regression to assess the trend in odds of disease with increasing mutation load (Online Appendix Table 4) but there was no significant association with risk of T2D (OR = 1.01 ± 0.12, $P = 0.937$). There was also no statistically significant trend in T2D risk with increasing numbers of inferred mutations (OR = 0.98 (0.68-1.41), $P = 0.913$), or synonymous mutations (OR = 1.32 ± 0.22, $P = 0.089$).

PANTHER scores for rare variants

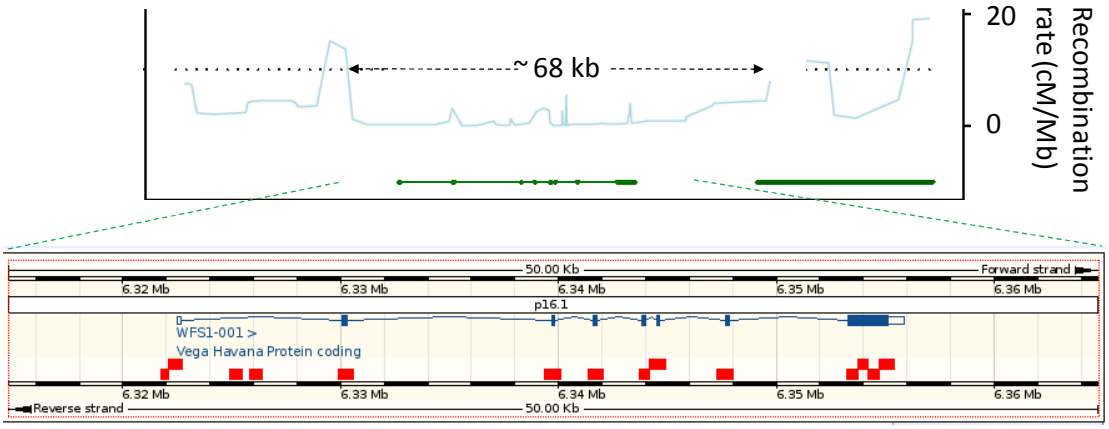
In an exploratory analysis we assessed differences in the load of rare (MAF<0.01) nonsynonymous variants between Cambridgeshire, ADDITION/Ely cases and controls, with mutations weighted by how likely they are to have deleterious effects on protein function. Instead of assigning carriers of mutations a score of 1 (as before), we weighted their score based on the PANTHER pdeleterious score for the mutation(s) they were carrying. In other words, their score was now the sum total of the pdeleterious scores of all the rare (MAF<0.01) non-synonymous alleles they were carrying. However, there was no significant difference in mean score between cases and controls in a two sample T-test (P = 0.5926).

Combined Multivariate and Collapsing (CMC) method

Finally, collapsed rare (MAF<0.01) and intermediate frequency putative functional variants were analysed using the CMC method (12). Hotelling's T^2 test was used as the multivariate test. There was no statistical association between the three marker groups and T2D ($P = 0.286$).

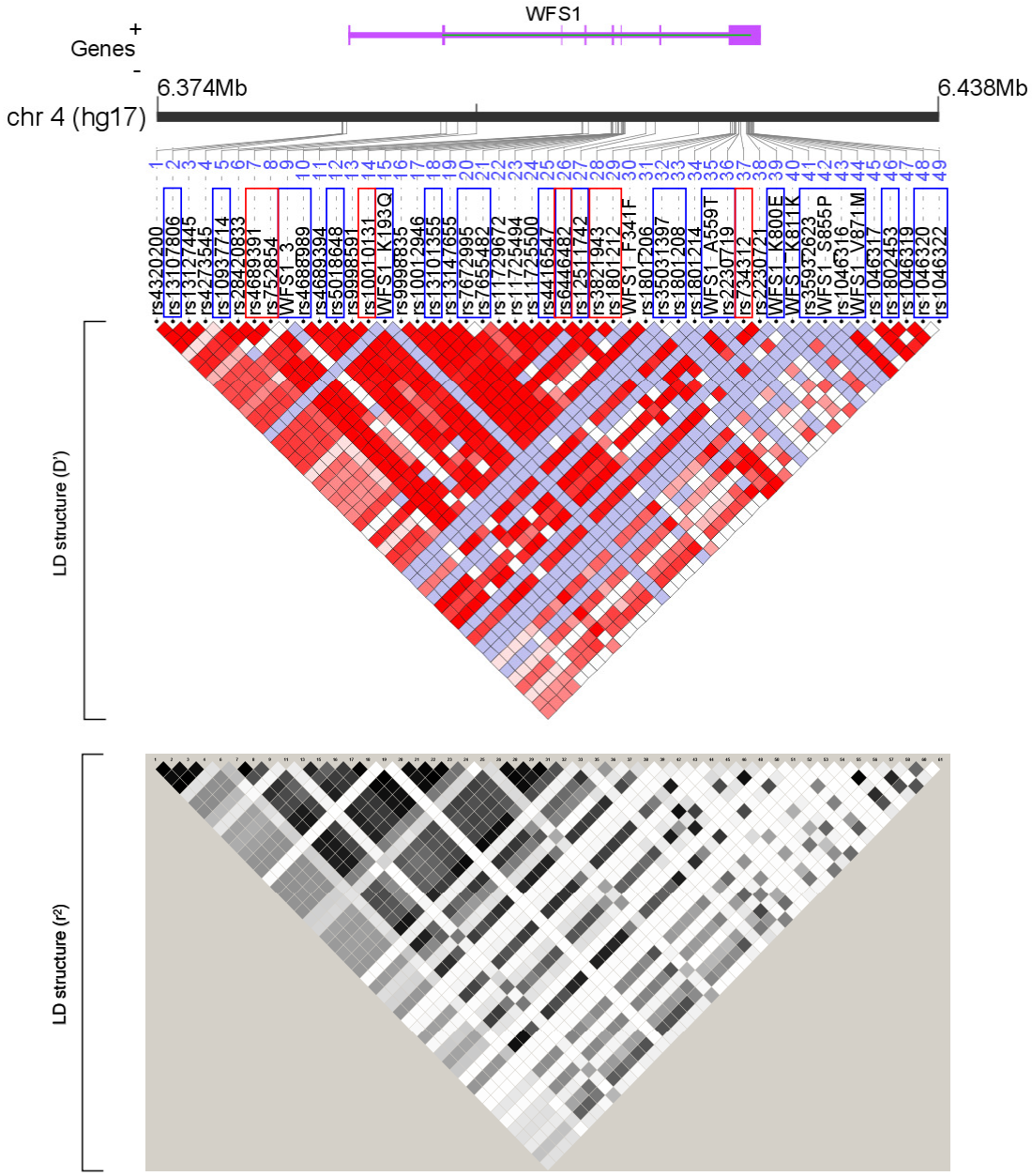
©2009 American Diabetes Association. Published online at http://diabetes.diabetesjournals.org/cgi/content/full/db09-0920/DC1

ONLINE APPENDIX TABLES AND FIGURES



Online Appendix Figure 1 Schematic of the region between recombination hotspots surrounding the WFS1 gene. The blue line in the top graph represents the recombination rate. The distance between recombination hotspots is approximately 68 kb. The bottom panel (from Ensembl) is zoomed in on the WFS1 gene (in blue) with amplicons sequenced as part of this study shown in red. Sequencing covers 9.335 kb of the region.

©2009 American Diabetes Association. Published online at http://diabetes.diabetesjournals.org/cgi/content/full/db09-0920/DC1



Online Appendix Figure 2 Feature map of the *WFS1* gene showing variants discovered during resequencing and tagging variants

The positions of the 49 common and/or non-synonymous variants detected during sequencing of *WFS1* in 96 Cambridgeshire cases and controls (including 4 additional SNPs genotyped during the original association study) are shown relative to the locus (purple) and chromosome 4 (black bar) (see Online Appendix Methods for details). The seven SNPs typed in the original studies are highlighted in red. Newly selected tagging SNPs are highlighted in blue. The bottom of the figure depicts two LD plots for the *WFS1* locus with pairwise LD values presented for SNPs. The upper plot presents LD as D' - see figure key for details. The figure was generated using LocusView (T. Petryshen, A. Kirby, M. Ainscow, unpublished software, available from the

©2009 American Diabetes Association. Published online at http://diabetes.diabetesjournals.org/cgi/content/full/db09-0920/DC1

Broad Institute, Cambridge, MA (http://www.broad.mit.edu/mpg/locusview/). In the lower plot, LD among SNPs is given as r^2. r^2 values of 1.0 are represented by black diamonds, intermediate r^2 values are shown in grey and r^2 values of 0 as white. This plot was generated using Haploview (13), available from the HapMap website (http://www.broad.mit.edu/mpg/haploview/index.php).

Online Appendix Table 1 Correlations among *WFS1* SNPs associated with T2D in the Cambridgeshire and EPIC case-control studies

	rs752854	rs4688989	rs5018648	**rs10010131**	rs13101355	rs7672995	rs6446482
rs752854							
rs4688989	0.71						
rs5018648	0.696	0.988					
rs10010131	0.717	0.963	0.967				
rs13101355	0.702	0.987	0.995	0.962			
rs7672995	0.59	0.7	0.699	0.686	0.7		
rs6446482	0.684	0.923	0.923	0.955	0.919	0.656	
rs1046320	0.655	0.932	0.939	0.92	0.939	0.666	0.883

LD values are r^2, where 1 denotes complete correlation and 0 denotes no correlation. Blue text highlights the SNPs from the original study described in Sandhu et al. Bold text reveals the most significant SNPs in the original and fine-mapping studies.

©2009 American Diabetes Association. Published online at http://diabetes.diabetesjournals.org/cgi/content/full/db09-0920/DC1

Online Appendix Table 2 Log-likelihood ratio tests assessing the association among SNPs in the *WFS1* gene with risk of type 2 diabetes

	rs752854	rs4688989	rs5018648	rs10010131	rs13101355	rs7672995	rs6446482	rs1046320
rs752854		0.69	0.41	0.51	0.31	0.29	0.58	0.41
rs4688989	0.45		0.86	0.81	0.50	0.65	0.89	0.58
rs5018648	0.61	0.68		0.83	0.82	0.62	0.82	0.35
rs10010131	0.64	0.82	0.75		0.36	0.73	0.94	0.49
rs13101355	0.86	0.35	0.96	0.68		0.46	0.96	0.26
rs7672995	0.59	0.41	0.35	0.23	0.46		0.30	0.23
rs6446482	0.52	0.63	0.62	0.71	0.52	0.62		0.25
rs1046320	0.48	0.93	0.78	0.91	0.60	0.78	0.81	

Each number is a P-value (log likelihood ratio test) for addition of the red SNP (log additive) to the model containing the black SNP (2 degrees-of-freedom).

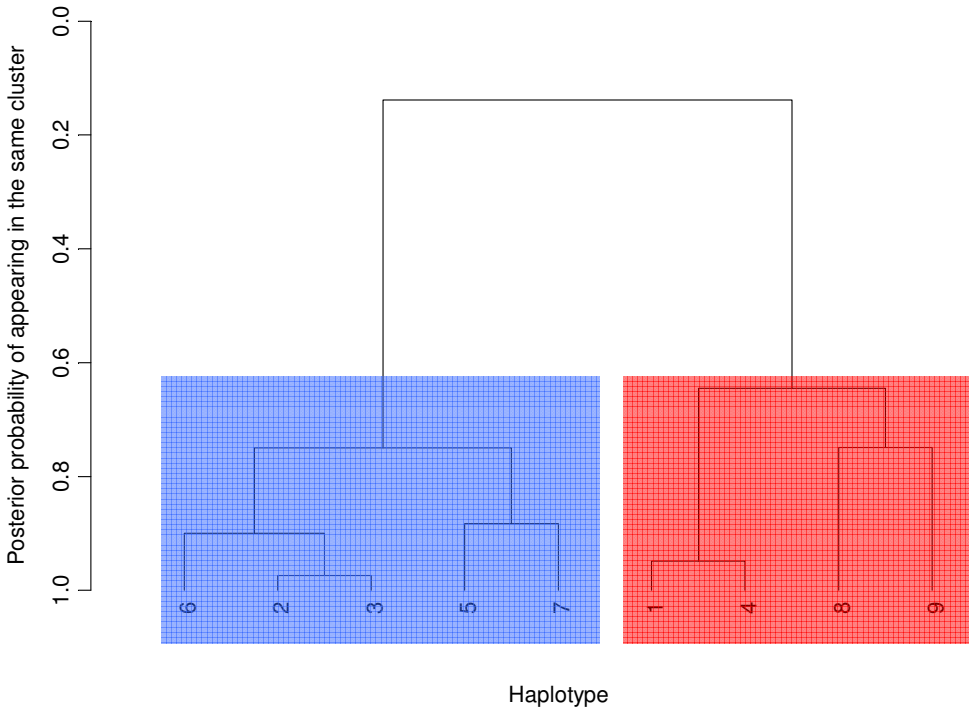
©2009 American Diabetes Association. Published online at http://diabetes.diabetesjournals.org/cgi/content/full/db09-0920/DC1

Online Appendix Table 3 Effect of common *WFS1* haplotypes on risk of type 2 diabetes

Rank	Haplotype	Frequency (%)	Posterior mean (SD) log-OR	Posterior probability in baseline cluster
1	11111111111111111111	43.5	Baseline	1.000
2	21221222222122112121	14.7	-0.112 (0.070)	0.026
3	22221222222122112121	11.6	-0.111 (0.070)	0.026
4	11111111112111111111	6.1	-0.006 (0.033)	0.949
5	21222222212121122222	4.6	-0.095 (0.080)	0.118
6	22211222222121112121	3.3	-0.110 (0.077)	0.048
7	22211222212121122222	2.3	-0.088 (0.085)	0.128
8	11111111111121212112	2.2	0.001 (0.059)	0.745
9	22211111111121212212	1.2	-0.011 (0.079)	0.576

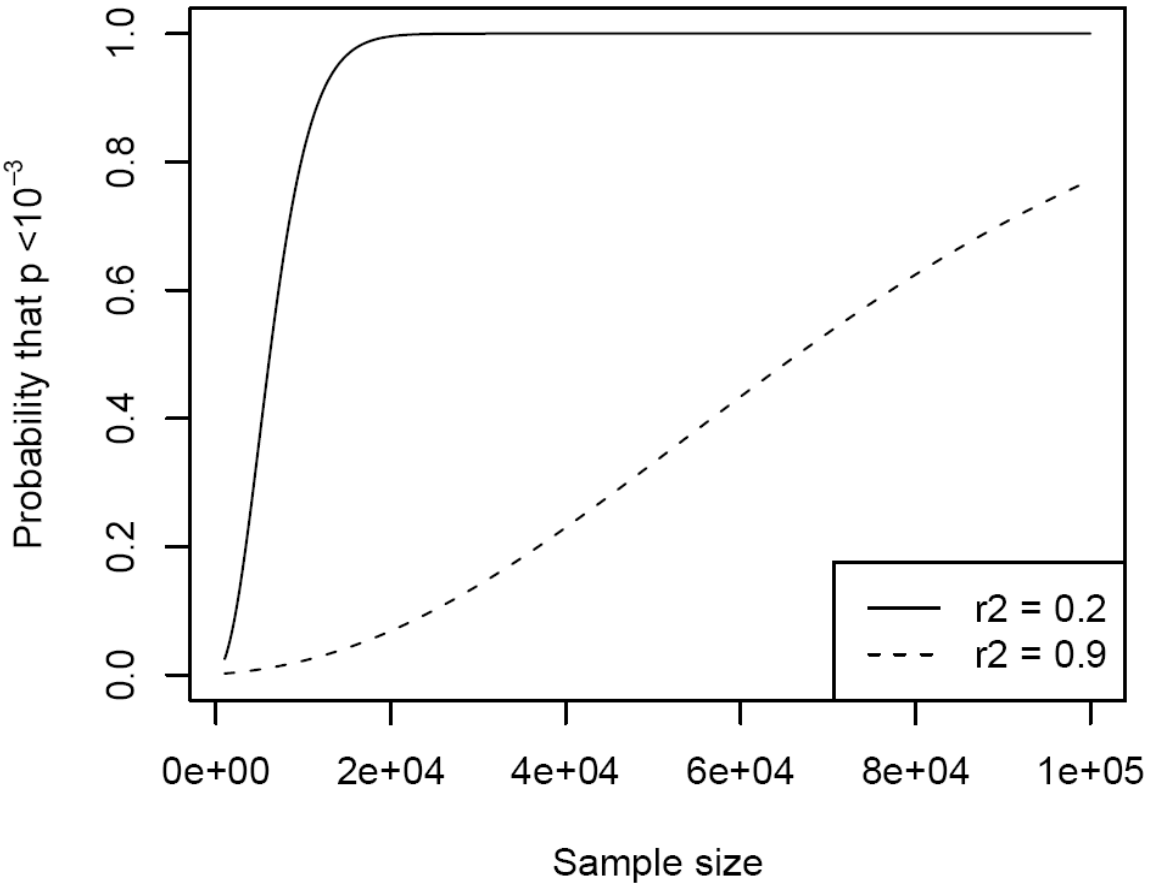
Haplotypes are clustered by allelic make-up and risk of type 2 diabetes. Haplotypes in cells highlighted in red belong to the baseline cluster and those highlighted in blue belong to a second, lower risk cluster. SNPs in white font can entirely separate the two clusters.

©2009 American Diabetes Association. Published online at http://diabetes.diabetesjournals.org/cgi/content/full/db09-0920/DC1



Online Appendix Figure 3 Clustering of WFS1 haplotypes by allelic make-up and effect on type 2 diabetes risk. The nine common *WFS1* haplotypes in our samples fall into one of two clusters (red and blue). The red cluster contains the most common (baseline haplotype) and the blue cluster contains haplotypes most protective against type 2 diabetes.

©2009 American Diabetes Association. Published online at http://diabetes.diabetesjournals.org/cgi/content/full/db09-0920/DC1



Online Appendix Figure 4 The relationship between sample size and power to distinguish statistically between the effects on disease risk of a causative SNP and a SNP in high or low linkage disequilibrium.

This data has been simulated based on two scenarios: an $r^2 = 0.2$ between a disease-causing SNP and a neutral SNP and an $r^2 = 0.9$ between a disease-causing SNP and a neutral SNP. What the simulation demonstrates is that in both scenarios, power to statistically distinguish between the effects of these SNPs on disease risk increases as sample size increases. However, lower sample sizes are required to have >80% power to distinguish between less correlated SNPs (for SNPs in LD $r^2 = 0.2$, >80% power is reached at ~10,000 samples), whereas larger sample sizes are required to have >80% power to distinguish between highly correlated SNPs (for SNPs in LD $r^2 = 0.9$, >80% power is only achieved with sample sizes>100,000). The model assumes the causative SNP affects risk of disease at a magnitude of OR = 0.85 and the threshold for statistical significance is 10^{-3}. This figure was produced by Jason Cooper at the Diabetes and Inflammation Laboratory, Cambridge.

©2009 American Diabetes Association. Published online at http://diabetes.diabetesjournals.org/cgi/content/full/db09-0920/DC1

Online Appendix Table 4 *WFS1* variants detected by sequencing 1235 type 2 diabetes cases and 1668 controls

Genic position	Genomic position	Nucleotide substitution (minor|major)	Protein consequence	MAF	SNP ID
Upstream	6316152	C|T		0.0722	rs11726771
Upstream	6316308	T|C		0.0001	
Upstream	6316341	A|C		0.0001	
Upstream	6316361	T|G		0.0063	
Upstream	6316417	A|T		0.0003	
Upstream	6316465	A|G		0.0001	
Upstream	6316475	T|C		0.0006	
Upstream	6316476	A|G		0.0001	
Upstream	6316614	A|G		0.0001	
Upstream	6316637	C|G		0.001	
Upstream	6321827	T|C		0.0004	
Upstream	6321914	G|T		0.0001	
Upstream	6321944	A|T		0.1856	rs4320200
Upstream	6321972	T|C		0.1835	rs13107806
Upstream	6321981	T|C		0.0001	
Upstream	6322051	C|G		0.1856	rs13127445
Upstream	6322203	G|T		0.0003	
Upstream	6322207	A|C		0.0001	
Upstream	6322317	G|T		0.1781	rs4273545
Upstream	6322384	G|A		0.0001	
Upstream	6322420	T|G		0.0001	
Upstream	6322429	A|C		0.0001	
Upstream	6322436	C|A		0.0001	
5'UTR	6322518	T|C		0.002	
5'UTR	6322519	G|A		0.0004	
5'UTR	6322527	T|C		0.0001	
5'UTR	6322576	G|A		0.0001	
5'UTR	6322580	T|C		0.0001	
5'UTR	6322593	G|C		0.0004	
5'UTR	6322609	A|G		0.0001	
Intron 1	6322673	T|G		0.0001	
Intron 1	6322727	T|C		0.0331	rs6830765
Intron 1	6324924	G|A		0.003	WFS1_1
Intron 1	6324986	A|G		0.0001	
Intron 1	6324997	C|T		0.0001	
Intron 1	6325007	G|A		0.0001	
Intron 1	*6325055*	*A|T*		*0.0003*	
Intron 1	*6325091*	*A|G*		*0.0003*	
Intron 1	6325206	G|T		0.0003	
Intron 1	6325258	G|T		0.0003	
Intron 1	6325330	A|G		0.0001	
Intron 1	6325386	A|G		0.0003	rs7657752

©2009 American Diabetes Association. Published online at http://diabetes.diabetesjournals.org/cgi/content/full/db09-0920/DC1

Genic position	Genomic position	Nucleotide substitution (minor|major)	Protein consequence	MAF	SNP ID
Intron 1	6325387	C|T		0.0024	
Intron 1	6325978	T|G		0.0003	
Intron 1	6326039	C|T		0.0003	
Intron 1	6326040	A|G		0.0001	
Intron 1	6326063	T|G		0.0003	
Intron 1	6326161	C|G		0.0001	
Intron 1	6326178	G|A		0.0003	
Intron 1	6326225	A|G		0.0001	
Intron 1	6326253	A|G		0.0023	
Intron 1	6326269	A|G		0.0001	
Intron 1	6326276	T|C		0.0001	
Intron 1	6326376	A|C		0.0001	
Intron 1	6329948	G|A		0.0087	rs10937714
Intron 1	6330010	A|G		0.0003	
Exon 2	6330104	A|C	P7P	0.0003	
Exon 2	6330124	C|T	Q14R	0.0001	
Exon 2	6330139	A|G	P19L	0.0003	
Exon 2	6330151	A|G	A23V	0.0001	
Exon 2	6330160	T|C	R26Q	0.0001	
Exon 2	6330207	A|G	R42X	0.0001	
Exon 2	6330212	A|T	A43A	0.0001	
Exon 2	6330213	T|G	P44T	0.0003	
Exon 2	6330215	T|G	P44P	0.0003	
Exon 2	6330314	A|G	T77T	0.0001	
Exon 2	6330315	T|C	G78R	0.0001	
Intron 2	6330358	T|C		0.0003	
Intron 2	6330360	G|A		0.0001	
Intron 2	6330363	A|G		0.0001	
Intron 2	6330375	T|C		0.0001	
Intron 2	6330405	C|T		0.3429	rs28420833
Intron 2	6330455	A|G		0.0001	
Intron 2	6330456	T|C		0.0001	
Intron 2	6330540	T|G		0.0001	
Intron 2	6339429	T|C		0.0003	
Intron 2	6339464	T|C		0.0026	
Intron 2	6339486	C|G		0.0003	
Intron 2	6339501	T|C		0.0001	
Intron 2	6339641	A|G		0.0006	
Intron 2	6339648	A|G		0.0001	
Intron 2	6339692	T|G		0.0001	
Intron 3	6339815	T|C		0.0001	
Intron 3	6339847	T|C		0.0001	
Intron 3	6339866	A|G		0.0001	
Intron 3	6339907	T|C		0.0105	
Intron 3	6339923	T|C		0.0001	

©2009 American Diabetes Association. Published online at http://diabetes.diabetesjournals.org/cgi/content/full/db09-0920/DC1

Genic position	Genomic position	Nucleotide substitution (minor|major)	Protein consequence	MAF	SNP ID
Intron 3	6339931	C|G		0.0001	
Intron 3	6339979	A|C		0.0001	
Intron 3	6340039	A|T		0.002	WFS1_2
Intron 3	6341451	C|T		0.0007	
Intron 3	6341495	A|G		0.3016	rs4688989
Intron 3	6341578	A|G		0.0054	rs4688990
Intron 3	6341579	T|C		0.0006	
Exon 4	6341699	T|C	A134T	0.0004	
Exon 4	6341701	T|C	A134A	0.0003	
Intron 4	6341787	A|G		0.0001	rs7688426
Intron 4	6341904	G|C		0.2982	rs4689394
Intron 4	6343810	G|A		0.2139	rs9998519
Intron 4	6343816	T|C		0.1497	rs10010131
Exon 5	6343835	A|G		0.0001	
Exon 5	6343846	T|C	R161Q	0.0001	
Exon 5	6343905	A|G	L181L	0.0001	
Exon 5	6343928	T|G	N188K	0.0001	
Exon 5	6343941	G|T	K193Q	0.0039	
Intron 5	6344085	C|G		0.0001	
Intron 5	6344132	C|T		0.0004	
Intron 5	6344138	G|C		0.3889	rs9998835
Intron 5	6344150	T|C		0.0001	
Intron 5	6344159	G|C		0.0001	
Intron 5	6344161	A|T		0.0001	
Intron 5	6344210	T|C		0.0001	
Intron 5	6344250	A|T		0.001	
Intron 5	6344251	A|G		0.2425	rs10012946
Intron 5	6344302	C|T		0.0001	
Intron 5	6344339	C|G		0.0011	
Intron 5	6344347	A|G		0.1577	rs13101355
Intron 5	6344375	T|C		0.3827	rs13147655
Intron 5	6344439	A|G		0.0001	
Intron 5	6344454	A|G		0.0001	
Intron 5	6344456	C|G		0.0001	
Intron 5	6344495	A|G		0.0001	
Exon 6	6344567	A|G	P218P	0.0001	
Exon 6	6344596	T|C	R228H	0.0009	
Exon 6	6344597	G|C	R228R	0.3081	rs7672995
Exon 6	6344608	T|C	R232H	0.0001	
Intron 6	6344641	T|C		0.0017	
Intron 6	6344703	G|A		0.112	rs7655482
Intron 6	6344717	A|G		0.0001	
Intron 6	6344730	T|C		0.0024	
Intron 6	6344739	T|C		0.0328	rs11729672
Intron 6	6344756	G|A		0.0641	rs11725494

©2009 American Diabetes Association. Published online at http://diabetes.diabetesjournals.org/cgi/content/full/db09-0920/DC1

Genic position	Genomic position	Nucleotide substitution (minor|major)	Protein consequence	MAF	SNP ID
Intron 6	6344806	A|G		0.0001	
Intron 6	6344820	G|C		0.0232	rs11725500
Intron 6	6344868	C|T		0.0082	rs4416547
Intron 6	6347348	A|C		0.0115	rs12511742
Intron 6	6347438	T|C		0.0066	WFS1_7
Intron 6	6347519	T|C		0.0001	
Intron 6	6347522	T|C		0.0004	
Intron 6	6347524	C|G		0.0001	
Intron 6	6347582	T|C		0.007	
Intron 6	6347602	T|C		0.0001	
Intron 6	6347603	A|G		0.0001	
Exon 7	6347682	A|G	I242I	0.0001	
Exon 7	6347684	A|G	A243V	0.0003	
Exon 7	6347685	T|C	A243A	0.0001	
Exon 7	6347755	T|C	D267N	0.0007	
Exon 7	6347810	T|C	R285H	0.0004	
Exon 7	6347813	G|A	L286P	0.0001	
Intron 7	6347845	T|C		0.0003	
Intron 7	6347856	A|G		0.0003	
Exon 8	6353353	G|A	A310A	0.0001	
Exon 8	6353386	C|G	T321T	0.0001	
Exon 8	6353398	T|C	N325N	0.0003	
Exon 8	6353400	T|C	A326V	0.0004	
Exon 8	6353402	T|C	L327F	0.0001	
Exon 8	6353408	A|T	F329I	0.0003	
Exon 8	6353420	G|A	V333I	0.2613	rs1801212
Exon 8	6353431	G|C	L336L	0.0001	
Exon 8	6353446	T|C	F341F	0.0718	
Exon 8	6353447	A|G	A342T	0.0003	
Exon 8	6353487	G|T	I355S	0.0001	
Exon 8	6353502	A|G	C360Y	0.0004	
Exon 8	6353548	T|C	R375R	0.0001	
Exon 8	6353557	A|C	T378T	0.0004	
Exon 8	6353571	A|G	R383H	0.0001	
Exon 8	6353576	A|G	E385K	0.0004	
Exon 8	6353608	C|T	V395V	0.3809	rs1801206
Exon 8	6353614	T|C	F397F	0.0003	
Exon 8	6353669	G|A	I416V	0.0001	
Exon 8	6353688	T|C	A422V	0.0004	
Exon 8	6353698	T|C	D425D	0.0001	
Exon 8	6353700	A|G	C426Y	0.0007	rs35218685
Exon 8	6353717	G|C	L432V	0.005	rs35031397
Exon 8	6353720	A|G	A433T	0.0001	
Exon 8	6353725	A|C	V434V	0.0001	
Exon 8	6353731	T|C	T436T	0.0083	

©2009 American Diabetes Association. Published online at http://diabetes.diabetesjournals.org/cgi/content/full/db09-0920/DC1

Genic position	Genomic position	Nucleotide substitution (minor|major)	Protein consequence	MAF	SNP ID
Exon 8	6353739	G|T	F439C	0.0003	
Exon 8	6353782	T|C	P453P	0.0001	
Exon 8	6353788	A|G	T455T	0.0001	
Exon 8	6353790	A|G	R456H	0.0417	rs1801208
Exon 8	6353794	T|G	R457S	0.0004	
Exon 8	6353795	C|G	A458P	0.0004	
Exon 8	6353815	T|C	T464T	0.0001	
Exon 8	6353822	T|C	L467L	0.0004	
Exon 8	6353893	T|C	T490T	0.0001	
Exon 8	6353903	A|G	G494S	0.0001	
Exon 8	6353918	T|C	L499F	0.0007	
Exon 8	6353923	C|T	N500N	0.3161	rs1801214
Exon 8	6353929	T|C	S502S	0.0001	
Exon 8	6353975	G|A	M518V	0.0001	
Exon 8	6354003	T|C	T527I	0.0001	
Exon 8	6354020	T|C	P533S	0.0014	
Exon 8	6354031	A|G	V536V	0.0001	
Exon 8	6354055	T|C	S544S	0.0001	
Exon 8	6354056	A|G	V545M	0.0001	
Exon 8	6354067	A|G	L548L	0.0001	
Exon 8	6354068	A|C	L549L	0.0001	rs1801211
Exon 8	6354080	A|G	G553S	0.0003	
Exon 8	6354096	A|G	R558H	0.0001	
Exon 8	6354098	A|G	A559T	0.0037	
Exon 8	6354105	G|T	I561S	0.0004	
Exon 8	6354106	T|C	I561I	0.0006	
Exon 8	6354107	A|G	G562S	0.0001	
Exon 8	6354148	T|C	A575A	0.0562	rs2230719
Exon 8	6354149	A|G	G576S	0.0006	rs1805069
Exon 8	6354151	A|C	G576G	0.0001	
Exon 8	6354166	T|C	G581G	0.0001	
Exon 8	6354183	A|G	R587Q	0.0001	
Exon 8	6354199	T|G	L592L	0.0001	
Exon 8	6354220	A|C	V599V	0.0004	
Exon 8	6354223	T|C	T600T	0.0001	
Exon 8	6354228	T|C	A602V	0.0004	rs2230720
Exon 8	6354229	A|G	A602A	0.0003	
Exon 8	6354255	G|A	R611H	0.2863	rs734312
Exon 8	6354262	A|G	W613X	0.0001	
Exon 8	6354306	T|C	T628M	0.0001	
Exon 8	6354308	T|C	R629W	0.0001	
Exon 8	6354346	A|G	T641T	0.0001	
Exon 8	6354435	T|C	A671V	0.0006	
Exon 8	6354436	A|G	A671A	0.0001	
Exon 8	6354442	T|C	C673C	0.0006	

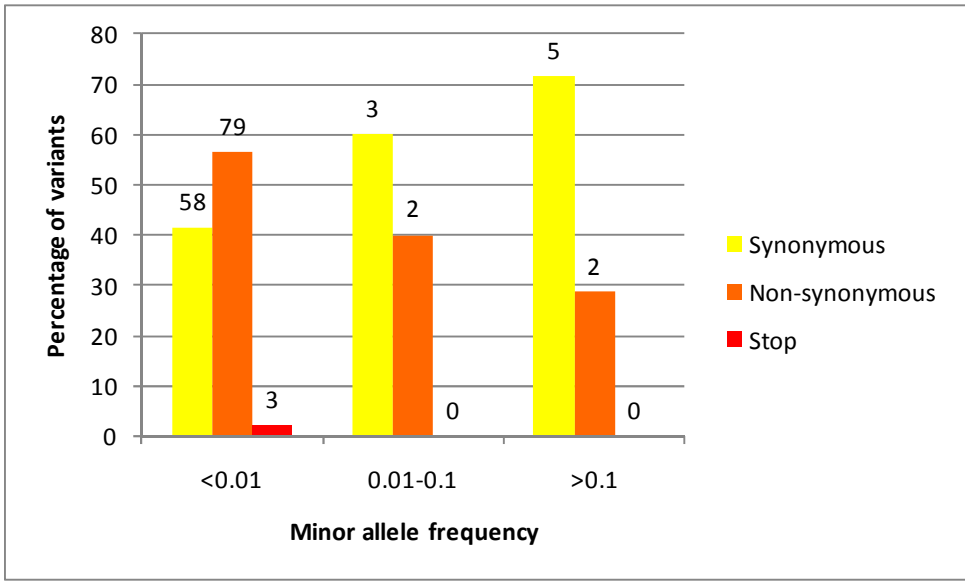
©2009 American Diabetes Association. Published online at http://diabetes.diabetesjournals.org/cgi/content/full/db09-0920/DC1

Genic position	Genomic position	Nucleotide substitution (minor|major)	Protein consequence	MAF	SNP ID
Exon 8	6354443	A|G	G674R	0.0001	
Exon 8	6354449	T|C	R676C	0.0006	
Exon 8	6354450	A|G	R676H	0.0003	
Exon 8	6354451	T|C	R676R	0.0001	
Exon 8	6354475	A|G	A684A	0.0026	
Exon 8	6354476	T|C	R685C	0.0001	
Exon 8	6354477	A|G	R685H	0.0003	
Exon 8	6354522	A|G	W700X	0.0001	
Exon 8	6354530	T|C	R703C	0.0003	
Exon 8	6354545	T|C	R708C	0.0003	
Exon 8	6354547	T|C	R708R	0.0011	
Exon 8	6354560	A|G	D713N	0.0001	
Exon 8	6354571	T|C	A716A	0.0001	
Exon 8	6354572	A|G	E717K	0.0001	
Exon 8	6354580	T|C	A719A	0.0004	
Exon 8	6354581	G|A	I720V	0.0006	rs1805070
Exon 8	6354607	T|C	G728G	0.0006	
Exon 8	6354628	T|C	Y735Y	0.0001	
Exon 8	6354631	T|C	G736G	0.0003	
Exon 8	6354636	A|C	A738D	0.0001	
Exon 8	6354661	A|C	N746K	0.0003	
Exon 8	6354709	A|G	K762K	0.0001	
Exon 8	6354720	C|A	H766P	0.0001	
Exon 8	6354733	T|C	F770F	0.0001	rs34384569
Exon 8	6354737	T|C	R772C	0.0004	
Exon 8	6354745	A|G	K774K	0.0669	rs2230721
Exon 8	6354750	T|A	E776V	0.0046	
Exon 8	6354758	A|G	V779M	0.0006	
Exon 8	6354782	A|G	A787T	0.0001	
Exon 8	6354792	G|C	S790W	0.0001	
Exon 8	6354793	A|G	S790S	0.0003	
Exon 8	6354808	C|G	E795D	0.0001	
Exon 8	6354815	A|G	V798I	0.0001	
Exon 8	6354847	T|C	S808S	0.0003	
Exon 8	6354856	G|A	K811K	0.3936	rs1046314
Exon 8	6354875	T|C	R818C	0.0046	rs35932623
Exon 8	6354889	T|C	L822L	0.0001	
Exon 8	6354892	T|C	I823I	0.0006	rs1801215
Exon 8	6354917	T|C	R832C	0.0003	
Exon 8	6354923	A|G	G834S	0.0003	
Exon 8	6354986	C|T	S855P	0.0001	
Exon 8	6354987	T|C	S855L	0.0003	
Exon 8	6354988	A|G	S855S	0.2484	rs1046316
Exon 8	6355004	A|G	V861M	0.0001	
Exon 8	6355012	T|C	I863I	0.0011	

©2009 American Diabetes Association. Published online at http://diabetes.diabetesjournals.org/cgi/content/full/db09-0920/DC1

Genic position	Genomic position	Nucleotide substitution (minor|major)	Protein consequence	MAF	SNP ID
Exon 8	6355019	A|G	D866N	0.0001	rs3821945
Exon 8	6355034	A|G	V871M	0.0129	
Exon 8	6355045	T|C	A874A	0.0001	
Exon 8	6355046	A|G	V875M	0.0003	
Exon 8	6355054	T|C	F877F	0.0001	
Exon 8	6355061	A|G	D880N	0.0001	
Exon 8	6355077	T|C	P885L	0.0001	
Exon 8	6355089	T|C	A889V	0.0001	
3'UTR	6355100	G|T		0.0001	
3'UTR	*6355126*	*T|C*		*0.0001*	
3'UTR	6355143	T|C		0.3131	rs1046317
3'UTR	6355150	G|A		0.0003	
3'UTR	6355165	T|C		0.0001	
3'UTR	6355178	G|C		0.0004	
3'UTR	6355186	A|G		0.0688	rs1802453
3'UTR	6355187	C|T		0.2659	rs1046319
3'UTR	*6355218*	*A|G*		*0.0001*	
3'UTR	6355227	T|C		0.0199	WFS1_8
3'UTR	6355245	G|A		0.303	rs1046320
3'UTR	*6355321*	*G|A*		*0.0003*	
3'UTR	*6355337*	*C|T*		*0.0006*	
3'UTR	6355349	A|G		0.051	rs1046322
3'UTR	6355370	G|A		0.0526	rs1046325

Genomic coordinates are NCBI build 36 (B36). Intron 1 SNPs in blue italics are conserved and within highly conserved regions. 3'UTR SNPs in red italics are rare variants within predicted miRNA seed sequences



Online Appendix Figure 5 Distribution of types of *WFS1* coding variation discovered during resequencing of cases and controls at different minor allele frequency ranges
Numbers above the bars are the actual numbers found.

©2009 American Diabetes Association. Published online at http://diabetes.diabetesjournals.org/cgi/content/full/db09-0920/DC1

Online Appendix Table 5 Known or inferred functional *WFS1* mutations

Chr:base	Variant	rs ID	Biochemical/genetic evidence	Pdel*	SIFT	PolyPhen	MAF in cases	MAF in controls	Reference
4:6330207	R42X		Novel				0	0.0002998	
4:6353402	L327F		Novel	0.55726	affects protein	possibly damaging	0.0004049	0	
4:6353502	C360Y		Novel	0.77597	affects protein	probably damaging	0.0008097	0	
4:6353739	F439C		Novel	0.59591	affects protein	probably damaging	0.0004049	0.0002998	
4:6353903	G494S		Novel	0.40523	affects protein	possibly damaging	0	0.0002998	
4:6354096	R558H		WS	0.61604	affects protein	possibly damaging	0.0004049	0	(14)
4:6354098	A559T		WS and psychiatric disorders	0.34899	tolerated	benign	0.0048583	0.0029976	(15)
4:6354105	I561S		Novel	0.44338	affects protein	possibly damaging	0	0.0005995	
4:6354306	T628M		Novel	0.74256	affects protein	possibly damaging	0.0004049	0	
4:6354308	R629W		WS & reduces half-life of wolframin	0.87775	affects protein	probably damaging	0.0004049	0	(16; 17)
4:6354435	A671V		WS and psychiatric disorders	0.21811	tolerated	benign	0.0012146	0	(15)
4:6354443	G674R		Polymorphism	0.60981	affects protein	probably damaging	0.0004049	0	
4:6354449	R676C		Novel	0.75663	affects protein	probably damaging	0.0004049	0.0002998	
4:6354476	R685C		Polymorphism	0.84434	affects protein	probably damaging	0.0004049	0	
Chr:base	**Variant**	**rs ID**	**Biochemical/genetic evidence**	**Pdel***	**SIFT**	**PolyPhen**	**MAF in cases**	**MAF in controls**	**Reference**
4:6354545	R708C		WS	0.77798	affects protein	probably damaging	0	0.0005995	(18)
4:6354572	E717K		WS and psychiatric disorders	0.32653	tolerated	benign	0	0.0002998	(19)
4:6354737	R772C		Psychiatric disorders	0.91098	affects protein	probably damaging	0	0.0008993	(19)
4:6354750	E776V		WS	0.49302	affects	probably	0.0040486	0.006295	(15)

©2009 American Diabetes Association. Published online at http://diabetes.diabetesjournals.org/cgi/content/full/db09-0920/DC1

					protein	damaging			
4:6354792	S790W		Novel	0.71718	affects protein	possibly damaging	0	0.0002998	
4:6354875	R818C	rs35932623	WS and psychiatric disorders	0.68043	affects protein	possibly damaging	0.0048583	0.0053957	(20)
4:6354917	R832C		Novel	0.75614	affects protein	probably damaging	0.0004049	0.0002998	
4:6355061	D880N		Novel	0.49211	affects protein	possibly damaging	0	0.0002998	
4:6355077	P885L		WS & reduces half-life of wolframin	0.54691	affects protein	probably damaging	0	0.0002998	(16; 21)

* Pdel score from PANTHER indicates the probability that an amino acid substitution will cause a deleterious effect on protein function based on alignment of evolutionarily related sequences (PANTHER classifies Pdel>0.38 as possibly deleterious). WS = Wolfram Syndrome

Online Appendix Table 6 Number of cases and controls carrying none, one, two, or three non-synonymous, synonymous, or inferred functional mutations with MAF<0.01

	Number of mutations	Case	Control	Total	P-value
Non-synonymous (MAF<0.01)	**0**	1128	1529	2657	
	1	103	130	233	
	2	4	7	11	
	3	0	2	2	
	Total	1235	1668	2903	0.94
Synonymous (MAF<0.01)	**0**	1173	1596	2769	
	1	55	72	127	
	2	6	0	6	
	3	1	0	1	
	Total	1235	1668	2903	0.09
Inferred mutations (MAF<0.01)	**0**	1189	1605	2794	
	1	44	62	106	
	2	2	0	2	
	3	0	1	1	
	Total	1235	1668	2903	0.91

©2009 American Diabetes Association. Published online at http://diabetes.diabetesjournals.org/cgi/content/full/db09-0920/DC1

ONLINE APPENDIX REFERENCES

1. Rathmann W, Haastert B, Icks A, Giani G, Hennings S, Mitchell J, Curran S, Wareham NJ: Low faecal elastase 1 concentrations in type 2 diabetes mellitus. *Scand J Gastroenterol* 36:1056-1061, 2001

2. Harding AH, Day NE, Khaw KT, Bingham S, Luben R, Welsh A, Wareham NJ: Dietary fat and the risk of clinical type 2 diabetes: the European prospective investigation of Cancer-Norfolk study. *Am J Epidemiol* 159:73-82, 2004

3. Day N, Oakes S, Luben R, Khaw KT, Bingham S, Welch A, Wareham N: EPIC-Norfolk: study design and characteristics of the cohort. European Prospective Investigation of Cancer. *Br J Cancer* 80 Suppl 1:95-103, 1999

4. Owen KR, Stride A, Ellard S, Hattersley AT: Etiological investigation of diabetes in young adults presenting with apparent type 2 diabetes. *Diabetes Care* 26:2088-2093, 2003

5. Frayling T, Walker M, McCarthy M, Evans J, Allen L, Lynn S, Ayres S, Millauer B, Turner C, Turner R, Sampson M, Hitman G, Ellard S, Hattersley A: Parent-offspring Trios: a resource to facilitate the identification of Type 2 diabetes genes. *Diabetes* 48:2475-2479, 1999

6. Knight B, Shields BM, Hattersley AT: The Exeter Family Study of Childhood Health (EFSOCH): study protocol and methodology. *Paediatric And Perinatal Epidemiology* 20:172-179, 2006

7. Lauritzen T, Griffin S, Borch-Johnsen K, Wareham NJ, Wolffenbuttel BH, Rutten G: The ADDITION study: proposed trial of the cost-effectiveness of an intensive multifactorial intervention on morbidity and mortality among people with Type 2 diabetes detected by screening. *Int J Obes Relat Metab Disord* 24 Suppl 3:S6-11, 2000

8. Wareham NJ, Hennings SJ, Byrne CD, Hales CN, Prentice AM, Day NE: A quantitative analysis of the relationship between habitual energy expenditure, fitness and the metabolic cardiovascular syndrome. *Br J Nutr* 80:235-241, 1998

9. Permutt MA, Wasson JC, Suarez BK, Lin J, Thomas J, Meyer J, Lewitzky S, Rennich JS, Parker A, DuPrat L, Maruti S, Chayen S, Glaser B: A genome scan for type 2 diabetes susceptibility loci in a genetically isolated population. *Diabetes* 50:681-685, 2001

10. Franks PW, Rolandsson O, Debenham SL, Fawcett KA, Payne F, Dina C, Froguel P, Mohlke KL, Willer C, Olsson T, Wareham NJ, Hallmans G, Barroso I, Sandhu MS: Replication of the association between variants in WFS1 and risk of type 2 diabetes in European populations. *Diabetologia* 51:458-463, 2008

11. Sandhu MS, Weedon MN, Fawcett KA, Wasson J, Debenham SL, Daly A, Lango H, Frayling TM, Neumann RJ, Sherva R, Blech I, Pharoah PD, Palmer CN, Kimber C, Tavendale R, Morris AD, McCarthy MI, Walker M, Hitman G, Glaser B, Permutt MA, Hattersley AT, Wareham NJ, Barroso I: Common variants in WFS1 confer risk of type 2 diabetes. *Nat Genet* 39:951-953, 2007

12. Li B, Leal SM: Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. *Am J Hum Genet* 83:311-321, 2008

13. Barrett JC, Fry B, Maller J, Daly MJ: Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 21:263-265, 2005

14. Colosimo A, Guida V, Rigoli L, Di Bella C, De Luca A, Briuglia S, Stuppia L, Salpietro DC, Dallapiccola B: Molecular detection of novel WFS1 mutations in patients with Wolfram syndrome by a DHPLC-based assay. *Hum Mutat* 21:622-629, 2003

15. Smith CJ, Crock PA, King BR, Meldrum CJ, Scott RJ: Phenotype-genotype correlations in a series of wolfram syndrome families. *Diabetes Care* 27:2003-2009, 2004

©2009 American Diabetes Association. Published online at http://diabetes.diabetesjournals.org/cgi/content/full/db09-0920/DC1

16. Hofmann S, Bauer MF: Wolfram syndrome-associated mutations lead to instability and proteasomal degradation of wolframin. *FEBS Lett* 580:4000-4004, 2006
17. Kadayifci A, Kepekci Y, Coskun Y, Huang Y: Wolfram syndrome in a family with variable expression. *Acta Medica (Hradec Kralove)* 44:115-118, 2001
18. Tessa A, Carbone I, Matteoli MC, Bruno C, Patrono C, Patera IP, De Luca F, Lorini R, Santorelli FM: Identification of novel WFS1 mutations in Italian children with Wolfram syndrome. *Hum Mutat* 17:348-349, 2001
19. Cryns K, Sivakumaran TA, Van den Ouweland JM, Pennings RJ, Cremers CW, Flothmann K, Young TL, Smith RJ, Lesperance MM, Van Camp G: Mutational spectrum of the WFS1 gene in Wolfram syndrome, nonsyndromic hearing impairment, diabetes mellitus, and psychiatric disease. *Hum Mutat* 22:275-287, 2003
20. Gomez-Zaera M, Strom TM, Rodriguez B, Estivill X, Meitinger T, Nunes V: Presence of a major WFS1 mutation in Spanish Wolfram syndrome pedigrees. *Mol Genet Metab* 72:72-81, 2001
21. Hardy C, Khanim F, Torres R, Scott-Brown M, Seller A, Poulton J, Collier D, Kirk J, Polymeropoulos M, Latif F, Barrett T: Clinical and molecular genetic analysis of 19 Wolfram syndrome kindreds demonstrating a wide spectrum of mutations in WFS1. *Am J Hum Genet* 65:1279-1290, 1999

©2009 American Diabetes Association. Published online at http://diabetes.diabetesjournals.org/cgi/content/full/db09-0920/DC1