Detailed Investigation of the Role of Common and Low-Frequency *WFS1* Variants in Type 2 Diabetes Risk

Katherine A. Fawcett,[1] Eleanor Wheeler,[1] Andrew P. Morris,[2] Sally L. Ricketts,[3] Göran Hallmans,[4] Olov Rolandsson,[5] Allan Daly,[1] Jon Wasson,[6] Alan Permutt,[7] Andrew T. Hattersley,[8] Benjamin Glaser,[9] Paul W. Franks,[10,11] Mark I. McCarthy,[12,13,14] Nicholas J. Wareham,[15] Manjinder S. Sandhu,[3,15] and Inês Barroso[1]

OBJECTIVE—Wolfram syndrome 1 (*WFS1*) single nucleotide polymorphisms (SNPs) are associated with risk of type 2 diabetes. In this study we aimed to refine this association and investigate the role of low-frequency *WFS1* variants in type 2 diabetes risk.

RESEARCH DESIGN AND METHODS—For fine-mapping, we sequenced *WFS1* exons, splice junctions, and conserved noncoding sequences in samples from 24 type 2 diabetic case and 68 control subjects, selected tagging SNPs, and genotyped these in 959 U.K. type 2 diabetic case and 1,386 control subjects. The same genomic regions were sequenced in samples from 1,235 type 2 diabetic case and 1,668 control subjects to compare the frequency of rarer variants between case and control subjects.

RESULTS—Of 31 tagging SNPs, the strongest associated was the previously untested 3′ untranslated region rs1046320 ($P = 0.008$); odds ratio 0.84 and $P = 6.59 \times 10^{-7}$ on further replication in 3,753 case and 4,198 control subjects. High correlation between rs1046320 and the original strongest SNP (rs10010131) ($r^2 = 0.92$) meant that we could not differentiate between their effects in our samples. There was no difference in the cumulative frequency of 82 rare (minor allele frequency [MAF] <0.01) nonsynonymous variants between type 2 diabetic case and control subjects ($P = 0.79$). Two intermediate frequency (MAF 0.01–0.05) nonsynonymous changes also showed no statistical association with type 2 diabetes.

CONCLUSIONS—We identified six highly correlated SNPs that show strong and comparable associations with risk of type 2 diabetes, but further refinement of these associations will require large sample sizes ($>100,000$) or studies in ethnically diverse populations. Low frequency variants in *WFS1* are unlikely to have a large impact on type 2 diabetes risk in white U.K. populations, highlighting the complexities of undertaking association studies with low-frequency variants identified by resequencing. *Diabetes* **59:741–746, 2010**

The post genome-wide association study era presents several challenges. These include fine-mapping association signals to genes and/or variants within the genomic regions of interest, assessing the impact of low frequency variants (not tagged in previous association studies) on diseases/traits, and understanding the functional mechanisms behind genetic associations.

WFS1 encodes wolframin (1,2), an endoplasmic reticulum (ER) membrane protein with a role in ER calcium homeostasis (3–5) and in the ER stress response (6,7). Loss-of-function mutations in *WFS1* cause Wolfram syndrome (MIM 222300), which includes young onset nonautoimmune insulin-dependent diabetes (8). Common single nucleotide polymorphisms (SNPs) at *WFS1* have recently been shown to be reproducibly associated with type 2 diabetes risk in white European populations (9–11). However, the strongest associated SNP, rs10010131, is intronic and is not associated with *WFS1* expression in HapMap lymphoblastoid cell lines (12), suggesting that it is tagging a causal variant(s).

Given the impact of rare and common *WFS1* variants on Mendelian and common forms of diabetes, respectively, *WFS1* is an excellent candidate gene in which to look for low frequency variants with intermediate effects on diabetes risk. Furthermore, anecdotal evidence suggests increased type 2 diabetes susceptibility in obligate carriers of Wolfram syndrome mutations (13).

We aimed to refine the association between *WFS1* common variants and type 2 diabetes by sequencing exons, splice junctions, and conserved intragenic and upstream noncoding regions in a subset of case ($n = 24$) and control ($n = 68$) subjects from the Cambridgeshire case-control study. We used these data to select tagging SNPs to capture common (minor allele frequency [MAF] >0.05) and nonsynonymous variants and genotyped these tagging SNPs in two U.K. case-control studies (total 959 case and 1,386 control subjects). Replication studies were

From the [1]Metabolic Disease Group, Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire, U.K.; the [2]Genetic and Genomic Epidemiology Unit, Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, U.K.; the [3]Department of Public Health and Primary Care, Strangeways Research Laboratory, University of Cambridge, Cambridge, U.K.; the [4]Section for Nutritional Research, Department of Public Health & Clinical Medicine, Umeå University Hospital, Umeå, Sweden; the [5]Section for Family Medicine, Department of Public Health & Clinical Medicine, Umeå University Hospital, Umeå, Sweden; the [6]Department of Medicine, Metabolism, Diabetes and Lipid Research Division, Washington University School of Medicine, Saint Louis, Missouri; the [7]Department of Internal Medicine, Metabolism, Diabetes and Lipid Research Division, Washington University School of Medicine, Saint Louis, Missouri; the [8]Institute of Biomedical and Clinical Science, Peninsula Medical School, Exeter, U.K.; the [9]Endocrine and Metabolism Service, The Hadassah Hebrew University Medical Center, Jerusalem, Israel; the [10]Section for Medicine, Department of Public Health & Clinical Medicine, Genetic Epidemiology & Clinical Research Group, Umeå University Hospital, Umeå, Sweden; the [11]Section for Nutritional Research, Department of Public Health & Clinical Medicine, Umeå University Hospital, Umeå, Sweden; the [12]Oxford Centre for Diabetes Endocrinology & Metabolism, Churchill Hospital, University of Oxford, Oxford, U.K.; the [13]Oxford National Institute for Health Research Biomedical Research Centre, Churchill Hospital, Headington, Oxford, U.K.; the [14]Wellcome Trust Centre for Human Genetics, University of Oxford, Headington, Oxford, U.K.; and the [15]Medical Research Council Epidemiology Unit, Institute of Metabolic Science, Cambridge, U.K.

Corresponding author: Inês Barroso, ib1@sanger.ac.uk.

Received 22 June 2009 and accepted 10 December 2009. Published ahead of print at http://diabetes.diabetesjournals.org on 12 December 2009. DOI: 10.2337/db09-0920.

conducted in four additional studies: two U.K., one Swedish, and one Ashkenazi (total 3,753 type 2 diabetic case and 4,198 control subjects). We also aimed to test for the presence of independent type 2 diabetes association signals from low-frequency (MAF <0.05) putative functional *WFS1* variants by sequencing 1,235 type 2 diabetic case and 1,668 control subjects from two U.K. case-control studies.

RESEARCH DESIGN AND METHODS

The Cambridgeshire (14) (552 type 2 diabetic case and 552 control subjects), European Prospective Investigation into Cancer and Nutrition (EPIC)-Norfolk (15) (417 case and 834 control subjects), Anglo-Danish-Dutch Study of Intensive Treatment in People With Screen-Detected Diabetes in Primary Care (ADDITION)/Ely (16,17) (926 case and 1,497 control subjects), and Exeter (18–20) (601 case and 610 control subjects) studies comprise white U.K. participants. The Ashkenazi study comprises 930 type 2 diabetic case and 461 control subjects of Ashkenazi Jewish origin (21). The Västerbotten study comprises predominantly northern Swedish whites (1,296 type 2 diabetic case and 1,412 control subjects) (11). The online appendix (available at http:// diabetes.diabetesjournals.org/cgi/content/full/db09-0920/DC1) describes the cohorts in detail.

PCR and sequencing. PCR was performed on genomic DNA from Cambridgeshire case-control participants or whole-genome amplified DNA from ADDITION and Ely study participants. Fourteen primer pairs (sequences and cycling conditions available upon request), designed using Primer3 software (http://frodo.wi.mit.edu/primer3/), were required to amplify the eight *WFS1* exons, including splice junctions, untranslated regions (UTRs), and selected conserved regions. Coverage is shown in supplemental Fig. 1 in the online appendix. PCR and bi-directional sequencing were performed using standard conditions and following manufacturers' protocols (supplemental Methods). Sequencing reactions were run on ABI3730 capillary machines (Applied Biosystems) and analyzed using an automatic SNP caller, ExoTrace (S. Leonard, Wellcome Trust Sanger Institute, unpublished data). The results of SNP calling were displayed and low-frequency variants were manually reviewed in a specific implementation of GAP4 (Staden Sequence Analysis Package software). All regions produced usable sequences for >90% of samples.

Tagging SNP selection. Details are provided in the supplemental Results and supplemental Fig. 2. Linkage disequilibrium (LD) was calculated using Haploview version 4.0 (http://www.broad.mit.edu/mpg/haploview), and pairwise tagging SNPs were selected by Tagger using $r^2 \geq 0.8$, force including nonsynonymous variants.

Genotyping. Tagging SNPs were genotyped using the Sequenom iPlex platform, whereas rs1046320 and rs7691824 were genotyped using TaqMan MGB chemistry (Applied Biosystems, Foster City, CA) according to the manufacturers' instructions (conditions and primers available upon request).

Quality control. Variants were excluded if they departed from Hardy-Weinberg equilibrium ($P < 0.001$) or had low call rates ($n < 85\%$) and/or if there was discrepancy in the call rate between case and control subjects ($P < 0.001$) (details provided in the online appendix).

Real-time PCR. The relative expression of rs1046320 alleles was assessed using total RNA from 10 HapMap lymphoblastoid cell lines (12) heterozygous for rs1046320 (six CEU, two YRI, and two CHN) and a TaqMan RNA-to-CT 2-Step Kit according to the manufacturer's instructions (primers and probes available upon request). Genomic DNA was used as a control.

Statistical analyses. Statistical analyses were conducted in StataSE 9. Logistic regression was used to assess the contribution of individual SNPs under a log-additive model (1 df) to risk of type 2 diabetes using study as a categorical covariate. Log-likelihood ratio tests were used to assess whether associated SNPs independently contributed to risk of type 2 diabetes by comparing the log likelihood of a nested model (2 df) containing an associated SNP and study with that of the full model (3 df) also containing the test SNP. The difference in prevalence of type 2 diabetes in carriers versus noncarriers of rare variants was analyzed using Fisher exact test. Power was calculated using the Power and Sample Size Program (22) and Quanto version 1.1.1 (http://hydra.usc.edu/gxe). Fixed-effects meta-analyses were performed using the metan command, combining summary estimates (log odds ratios and lower and upper CIs for each study), weighted using the inverse-variance method. An expectation-maximization algorithm was used to estimate haplotype frequencies, and GENEBPM software was used to cluster haplotypes by allelic make-up and risk of type 2 diabetes to obtain a Bayes' factor (BF) in favor of association (23).

RESULTS

WFS1 tagging SNPs are associated with type 2 diabetes. Of 31 tagging SNPs, 24 passed quality control and captured 81% of the common (MAF >0.05) and/or nonsynonymous *WFS1* variants in the Cambridgeshire case-control samples and 98% of the common *WFS1* region variants in HapMap CEU trios.

Eight SNPs were nominally associated with type 2 diabetes risk ($P < 0.05$) in a pooled analysis of Cambridgeshire and EPIC-Norfolk studies (Table 1). The LD between these eight SNPs (supplemental Table 1) and the consistency of their effect size suggests that they are linked to similar extents with the real causal allele(s). However, we were unable to demonstrate that any of the associated SNPs were contributing to type 2 diabetes risk independently of the other seven (supplemental Table 2).

Replication of the rs1046320 association. In our data, the strongest association with type 2 diabetes risk (rs1046320, $P = 0.008$) mapped within a putative functional region (3′ UTR). We therefore genotyped it in four further case-control studies—Exeter, Ashkenazi, ADDITION/Ely, and Västerbotten studies (3,753 case and 4,198 control subjects)—to improve the accuracy of the effect-size estimate and compare it with rs10010131, the strongest SNP from the original study (10,11). In meta-analyses, rs1046320 and rs10010131 demonstrated similar magnitudes of association with type 2 diabetes risk (OR 0.856 [95% CI 0.804–0.912], $P = 1.25 \times 10^{-6}$ and 0.854 [0.800–0.912], $P = 2.58 \times 10^{-6}$, respectively) (Fig. 1). The high correlation between these SNPs in our samples ($r^2 = 0.92$) suggests >100,000 samples would be required to have 80% power to distinguish between their effects with a significance level $P < 10^{-3}$ (supplemental Fig. 4). In an assessment of the possible function of rs1046320, we found no difference in allele-specific expression in lymphoblastoid cell lines from 10 heterozygous HapMap individuals (data not shown).

Haplotype analysis. To test whether haplotypes tag the causal variant(s) better than individual SNPs, we estimated the frequency of haplotypes across 20 genotyped SNPs (excluding three variants with MAF <0.01) in Cambridgeshire and EPIC-Norfolk samples. We found nine haplotypes with MAF >0.01 (supplemental Table 3) that fell into two clusters according to allelic make-up and type 2 diabetes risk (supplemental Fig. 3). One cluster contained haplotypes that are protective against type 2 diabetes relative to the most common haplotype. There were six SNPs (including rs10010131 and rs1046320) that partition the two clusters entirely and, due to high LD between them, were each sufficient to separate the clusters. When the analysis was repeated with haplotypes made up of each SNP in turn, we found the evidence in favor of association was stronger for the single SNPs than for the haplotypes. For the overall haplotype analysis, the estimated log10 BF was 0.64, whereas for the single SNPs, the strongest log10 BF was 1.07 for rs1046320. This suggests that haplotypes made up of SNPs in this study do not tag the causal variant(s) any better than any of the individual SNPs. However, this does not preclude the possibility of independent causal variants in the region that we cannot tag with our SNPs (either pairwise or through the use of haplotypes).

Imputation of nontagging SNPs. To increase coverage in the region, we imputed missing genotypes at 66 additional loci (supplemental Methods) (Fig. 2). In this analy-

TABLE 1
Association of *WFS1* tagging SNPs with type 2 diabetes risk in a pooled analysis of Cambridgeshire and EPIC case-control studies

SNP	Nucleotide change (major minor)	Protein consequence	MAF	Odds ratio (95% CIs)	*P* odds ratio*
rs13107806	C T	Conserved upstream	0.427	0.90 (0.79–1.02)	0.11
rs10937714	T C	Intron 1	0.212	0.93 (0.79–1.09)	0.35
rs4689391	A G	Intron 2	0.423	0.90 (0.79–1.03)	0.11
rs752854	T C	Intron 2	0.344	0.87 (0.76–1.00)	*0.05*
WFS1_3	C G	Intron 3	0.051	0.89 (0.66–1.21)	0.46
rs4688989	C T	Intron 3	0.402	0.86 (0.75–0.98)	*0.03*
rs5018648	G C	Intron 4	0.412	0.85 (0.74–0.97)	*0.01*
rs10010131	**G A**	**Intron 4**	**0.398**	**0.87 (0.76–0.98)**	*0.02*
WFS1_K193Q	A C	K193Q	0.004	1.00 (0.36–2.81)	>0.99
rs13101355	C T	Intron 5	0.4	0.85 (0.75–0.97)	*0.02*
rs7672995	G C	R228R	0.316	0.84 (0.73–0.97)	*0.02*
rs6446482	G C	Intron 6	0.405	0.87 (0.77–0.99)	*0.03*
rs12511742	G T	Intron 6	0.072	0.93 (0.72–1.20)	0.58
rs3821943	T C	Intron 7	0.457	0.91 (0.81–1.03)	0.15
rs1801212	A G	I333V	0.28	0.90 (0.78–1.03)	0.14
rs35031397	C G	L432V	0.004	1.10 (0.39–3.09)	0.86
rs1801208	G A	R456H	0.046	1.25 (0.92–1.69)	0.15
WFS1_A559T	G A	A559T	0.005	0.66 (0.25–1.75)	0.40
rs2230719	C T	A575A	0.076	0.92 (0.72–1.18)	0.51
rs734312	A G	H611R	0.455	0.93 (0.82–1.05)	0.25
rs1802453	G A	3′ UTR	0.089	0.93 (0.74–1.17)	0.54
rs1046320	**A G**	**3′ UTR**	**0.419**	**0.83 (0.72–0.95)**	*0.008*
rs1046322	G A	3′ UTR	0.119	1.01 (0.82–1.23)	0.95

This list excludes tagging SNP WFS1_S855P as the minor allele was only present in one case and one control subject and therefore could not be analyzed. Bold text indicates significant *P* values. Bold italicized text highlights the strongest statistically associated SNP from the original study cohorts described in Sandhu et al. (10) and the strongest statistically associated SNP in Cambridgeshire and EPIC case-control fine-mapping studies (rs1046320). *Outcome of a logistic regression analysis.

sis, rs1046320 remained the most strongly associated SNP in Cambridgeshire and EPIC-Norfolk studies, except for the imputed intronic rs7691824. However, genotyping of rs7691824 in Cambridgeshire and EPIC studies showed that there were no carriers among our samples.

Deep sequencing summary. Sequencing of exons, splice junctions, and conserved noncoding regions in 1,235 type 2 diabetic case and 1,668 control subjects from the Cambridgeshire and ADDITION/Ely studies revealed 290 variants (supplemental Table 4). Of 250 rare (MAF <0.01)

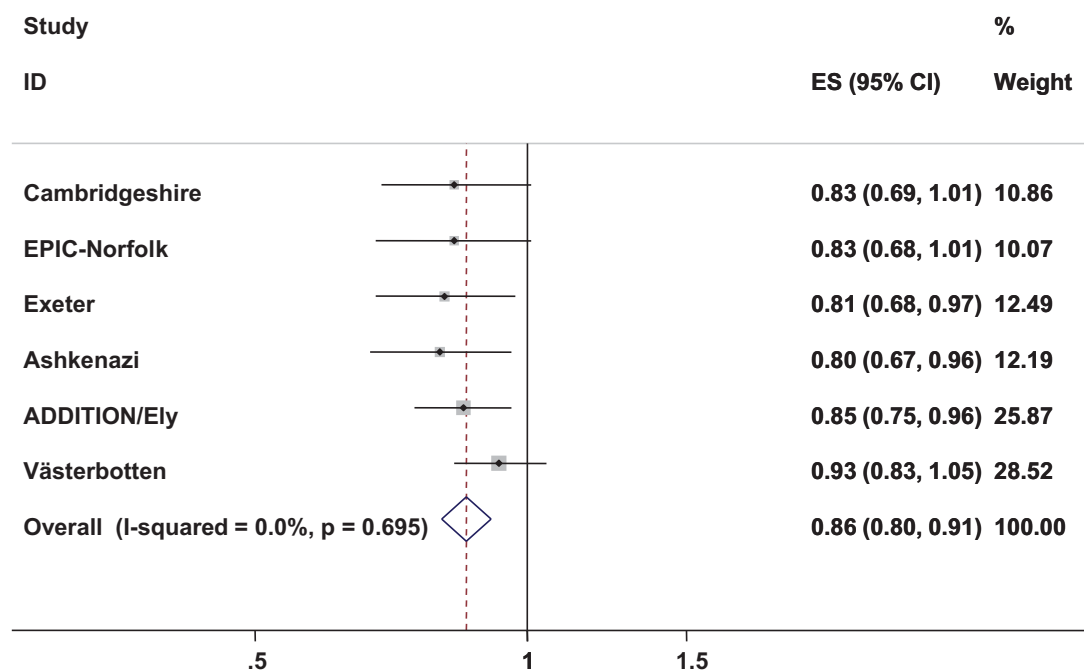| Study ID		ES (95% CI)	% Weight
Cambridgeshire		0.83 (0.69, 1.01)	10.86
EPIC-Norfolk		0.83 (0.68, 1.01)	10.07
Exeter		0.81 (0.68, 0.97)	12.49
Ashkenazi		0.80 (0.67, 0.96)	12.19
ADDITION/Ely		0.85 (0.75, 0.96)	25.87
Västerbotten		0.93 (0.83, 1.05)	28.52
Overall (I-squared = 0.0%, p = 0.695)		0.86 (0.80, 0.91)	100.00

FIG. 1. Meta-analysis of rs1046320 in U.K., Ashkenazi, ADDITION/Ely, and Västerbotten case-control studies. The overall OR is 0.856 (95% CI 0.804–0.912), $P = 1.25 \times 10^{-6}$. Summary statistics from previously published work (11) show that rs10010131 has a comparable overall OR of 0.854 (0.800–0.912), $P = 2.58 \times 10^{-6}$, in the same populations. ES, effect size.
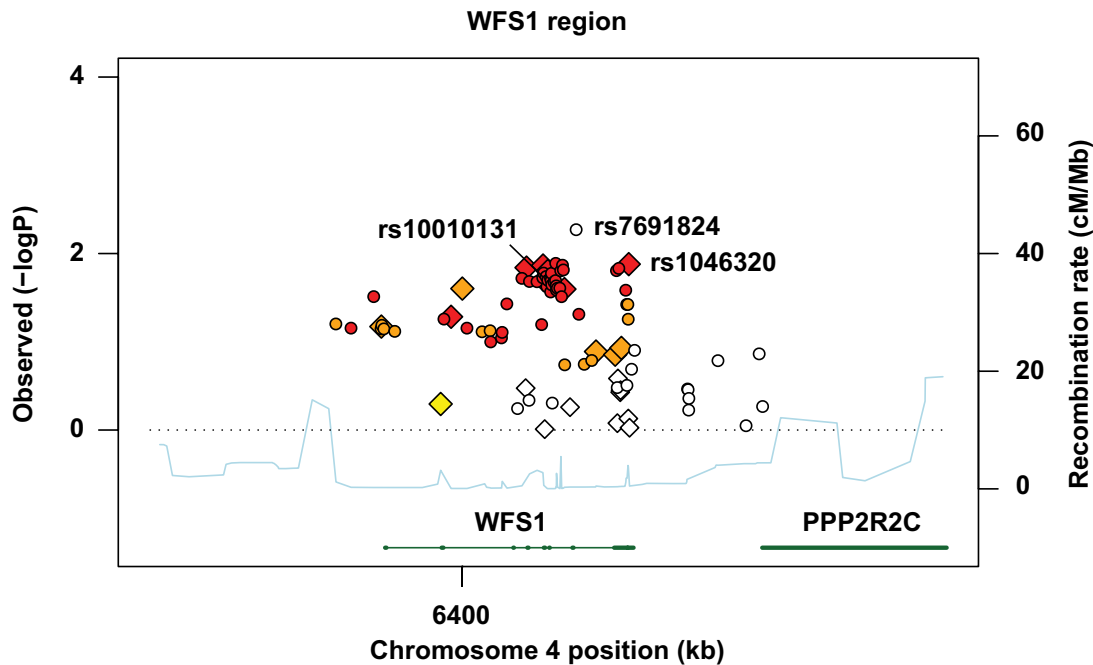
WFS1 region

FIG. 2. The statistical strength of the association of *WFS1* tagging (◇) and imputed (○) SNPs in the context of estimated recombination rates (blue line) and pairwise correlation between rs10010131 and surrounding markers. Red represents $r^2 > 0.85$, orange represents $0.5 < r^2 < 0.85$, yellow represents $0.2 < r^2 < 0.5$, and white represents $r^2 < 0.2$.

changes, 94% were novel, demonstrating the value of deep resequencing for identifying rare changes.

Analysis of variants with MAF <0.01. Given the sample size, our study is underpowered to detect effects of each rare variant tested individually. For example, 76 of the 82 rare, nonsynonymous variants have MAF <0.001, for which we have only 25% power to detect an OR of ~3. To improve power, we collapsed rare (MAF <0.01) variants together by comparing the prevalence of type 2 diabetes in carriers versus noncarriers. We collapsed only nonsynonymous variants in the first instance, as their relative paucity at higher MAFs in the population suggest they are enriched for functional changes under negative selective pressure (supplemental Fig. 5). However, there was no significant increase in risk of type 2 diabetes in carriers compared with noncarriers (OR 1.04 [95% CI 0.79–1.37], Fisher exact test $P = 0.79$) (Table 2). Adding rare variants in conserved noncoding sequences and TargetScan (http://www.targetscan.org/)-predicted miRNA seed sequences to the rare nonsynonymous changes made no material difference ($P = 0.67$) (Table 2).

A comparative study of synonymous variants (MAF <0.01), assumed to be functionally neutral, yielded similar results (Table 2). Further exploratory analyses, including examination of mutation load, PANTHER scores, and combined analysis of rare and intermediate frequency variants, also did not yield significant results (supplemental Results and supplemental Table 6).

Predicting variants with deleterious effects on protein function. To avoid diluting effects of rare (MAF <0.01) nonsynonymous variants on disease risk by pooling them with neutral nonsynonymous changes, we restricted analysis to nonsynonymous changes most likely to impact protein function. Variants were selected based on three criteria *1*) previous biochemical evidence that the variant causes loss of wolframin function, *2*) previous genetic evidence for involvement in Wolfram syndrome, and *3*) predicted deleterious functional effects by three

programs: SIFT (http://sift.jcvi.org/), PolyPhen (http://genetics.bwh.harvard.edu/pph/), and PANTHER (http://www.pantherdb.org/tools/csnpScoreForm.jsp). Using these criteria, we inferred 23 functionally important mutations (supplemental Table 5), but carriers were at type 2 diabe-

TABLE 2

Number of case and control subjects carrying nonsynonymous, synonymous, or inferred functional mutations with MAF <0.01 versus wild type

Types of variation (MAF <0.01)	Cases	Controls	Total	*P*
Nonsynonymous				
Noncarriers	1,128	1,529	2,657	
Carriers	107	139	246	
Total	1,235	1,668	2,903	0.79
Nonsynonymous and variants in predicted miRNA sites and conserved noncoding sequence*				
Noncarriers	1,125	1,527	2,652	
Carriers	110	141	251	
Total	1,235	1,668	2,903	0.67
Synonymous				
Noncarriers	1,173	1,596	2,769	
Carriers	62	72	134	
Total	1,235	1,668	2,903	0.37
Inferred mutations†				
Noncarriers	1,189	1,605	2,794	
Carriers	46	63	109	
Total	1,235	1,668	2,903	>0.99

*Rare variants in predicted miRNA sites are in red italics in supplemental Table 4 and variants in conserved noncoding sequence are in blue italics in supplemental Table 4. †Nonsynonymous variants with MAF <0.01 and predicted functional based on previous biochemical evidence that the variant causes loss of wolframin function, previous genetic evidence for involvement in Wolfram syndrome, and predicted deleterious functional effects by web-based programs SIFT, PolyPhen, and PANTHER.

tes risk comparable with that of noncarriers (OR 0.99 [95% CI 0.65–1.48], $P > 0.99$) (Table 2).

Analysis of variants with MAF 0.01–0.1. Two nonsynonymous SNPs, V871M and R456H, had MAFs of 0.013 and 0.042, respectively. In single SNP analyses of pooled Cambridgeshire and ADDITION/Ely studies, neither were associated with type 2 diabetes ($P = 0.13$ and $P = 0.25$, respectively).

DISCUSSION

We performed a comprehensive fine-mapping and low-frequency variant analysis for *WFS1*, a locus associated with type 2 diabetes risk (9–11). Using a sequencing, SNP-tagging, and genotyping approach, we identified a number of putative causal variants for type 2 diabetes association. However, due to strong LD between the SNPs within the candidate interval, we were unable to distinguish between their effects on disease risk. None of the associated SNPs have obvious functional properties, and real-time PCR revealed no difference in allele-specific expression of rs1046320 (the strongest associated in this study) in lymphoblastoid cell lines, suggesting this SNP is unlikely to affect mRNA stability or processing in this tissue. However, we cannot rule out rs1046320-associated expression changes in other tissues.

Deep resequencing of *WFS1* exons, splice junctions, and conserved noncoding sequences in 1,235 type 2 diabetic case and 1,668 control subjects revealed no statistically significant differences in the cumulative frequency of rare (MAF <0.01) nonsynonymous variants ($P = 0.79$). Given that ~8% of study participants carried at least one rare nonsynonymous change, we had >80% power to detect ORs >1.43 at $P < 0.05$. This study was therefore well powered to detect previously reported effect sizes for rare variants on complex traits (the average being OR 3.74) (24). Restricting the analysis to those variants most likely to be functional reduced the frequency of the exposure (carrier status) to ~4%, but retained >80% power to detect ORs >1.65. Still, there were no statistical differences between case and control subjects ($P > 0.99$), suggesting rare variants in *WFS1* do not have a large impact (ORs >2) on type 2 diabetes risk. It is worth noting that our analyses assumed all rare variants have the same direction of effect. Our power to detect significant effects on type 2 diabetes would have been reduced if the variants were a mixture of protective and susceptibility alleles. Finally, our study had >80% power to detect moderate effect sizes of intermediate-frequency SNPs V871M and R456H on risk of type 2 diabetes (ORs >1.93 and >1.45, respectively), though neither were statistically associated with type 2 diabetes ($P = 0.13$ and $P = 0.25$). Selecting case subjects enriched for early onset/family history of the disease might have increased our power to find rarer variants of slightly higher penetrance that might segregate in the family. However, this kind of analysis was not feasible in our study, as we have no DNA from family members.

Our attempts to refine the *WFS1* association signal demonstrate that while high LD is useful for minimizing the amount of genotyping required to discover a genetic association, it can compromise attempts to further refine the association signal. Studying populations with different and/or weaker patterns of LD may help refine signals. For example, the LD block spanning the *WFS1* gene is more fragmented in HapMap samples of African descent, and correlation between SNPs is generally weaker ($r^2 = 0.204$

between SNPs rs10010131 and rs1046320 in YRI HapMap samples compared with $r^2 > 0.92$ in CEU samples). In this setting, studies with ~10,000 samples (compared with >100,000) would be well powered to distinguish their effects (supplemental Fig. 4), assuming that this locus is associated with type 2 diabetes in this population. An alternative strategy is to test SNPs within the candidate region for association with proximal traits, which may provide greater power to distinguish between SNP effects (25).

Limitations of our fine-mapping study design are that we were underpowered to detect associations with SNPs at MAF <0.05, and we limited sequencing to regions most likely to harbor functional variation. Though we were able to impute 66 additional known SNPs in the region (most common and in high LD [$r^2 > 0.8$] with directly genotyped SNPs), 7 had MAF <0.05 and were not well correlated ($r^2 < 0.8$) with genotyped SNPs. As illustrated by follow-up genotyping of the imputed SNP (rs7691824), monomorphic in our samples, imputation of rare variants is less accurate. This could potentially lead to false negative results if rare variants of poorer imputation quality have larger effect sizes than more common SNPs.

In conclusion, we have undertaken the most comprehensive fine-mapping and rare variant analysis in a type 2 diabetes gene to date. We identified six SNPs that have comparable associations with type 2 diabetes ranging from ORs of 0.85 to −0.87. We also show that low-frequency variants in putative functional regions of *WFS1* are not associated with diabetes risk in our U.K. populations. Future whole exome/genome resequencing studies should consider that functionality of rare variants is difficult to predict and that pooling variants and candidate genes together for purposes of analysis might diminish the power to detect true risk alleles.

REFERENCES

1. Inoue H, Tanizawa Y, Wasson J, Behn P, Kalidas K, Bernal-Mizrachi E, Mueckler M, Marshall H, Donis-Keller H, Crock P, Rogers D, Mikuni M,

Kumashiro H, Higashi K, Sobue G, Oka Y, Permutt MA. A gene encoding a transmembrane protein is mutated in patients with diabetes mellitus and optic atrophy (Wolfram syndrome). Nat Genet 1998;20:143–148

2. Strom TM, Hörtnagel K, Hofmann S, Gekeler F, Scharfe C, Rabl W, Gerbitz KD, Meitinger T. Diabetes insipidus, diabetes mellitus, optic atrophy and deafness (DIDMOAD) caused by mutations in a novel gene (wolframin) coding for a predicted transmembrane protein. Hum Mol Genet 1998;7:2021–2028

3. Osman AA, Saito M, Makepeace C, Permutt MA, Schlesinger P, Mueckler M. Wolframin expression induces novel ion channel activity in endoplasmic reticulum membranes and increases intracellular calcium. J Biol Chem 2003;278:52755–52762

4. Takeda K, Inoue H, Tanizawa Y, Matsuzaki Y, Oba J, Watanabe Y, Shinoda K, Oka Y. WFS1 (Wolfram syndrome 1) gene product: predominant subcellular localization to endoplasmic reticulum in cultured cells and neuronal expression in rat brain. Hum Mol Genet 2001;10:477–484

5. Takei D, Ishihara H, Yamaguchi S, Yamada T, Tamura A, Katagiri H, Maruyama Y, Oka Y. WFS1 protein modulates the free Ca(2+) concentration in the endoplasmic reticulum. FEBS Lett 2006;580:5635–5640

6. Fonseca SG, Fukuma M, Lipson KL, Nguyen LX, Allen JR, Oka Y, Urano F. WFS1 is a novel component of the unfolded protein response and maintains homeostasis of the endoplasmic reticulum in pancreatic beta-cells. J Biol Chem 2005;280:39609–39615

7. Ueda K, Kawano J, Takeda K, Yujiri T, Tanabe K, Anno T, Akiyama M, Nozaki J, Yoshinaga T, Koizumi A, Shinoda K, Oka Y, Tanizawa Y. Endoplasmic reticulum stress induces Wfs1 gene expression in pancreatic beta-cells via transcriptional activation. Eur J Endocrinol 2005;153:167–176

8. Wolfram DJ, Wagener HP. Diabetes mellitus and simple optic atrophy among siblings: report of four cases. Mayo Clin Proc 1938;13:715–718

9. Minton JA, Hattersley AT, Owen K, McCarthy MI, Walker M, Latif F, Barrett T, Frayling TM. Association studies of genetic variation in the WFS1 gene and type 2 diabetes in U.K. populations. Diabetes 2002;51:1287–1290

10. 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 2007;39:951–953

11. 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 2008;51:458–463

12. Stranger BE, Forrest MS, Clark AG, Minichiello MJ, Deutsch S, Lyle R, Hunt S, Kahl B, Antonarakis SE, Tavare S, Deloukas P, Dermitzakis ET. Genome-wide associations of gene expression variation in humans. PLoS Genet 2005;1:e78

13. Fraser FC, Gunn T. Diabetes mellitus, diabetes insipidus, and optic atrophy. An autosomal recessive syndrome? J Med Genet 1977;14:190–193

14. 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 2001;36:1056–1061

15. 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 2004;159:73–82

16. Lauritzen T, Griffin S, Borch-Johnsen K, Wareham NJ, Wolffenbuttel BH, Rutten G; the Anglo-Danish-Dutch Study of Intensive Treatment in People with Screen Detected Diabetes in Primary Care. 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 2000;24(Suppl. 3):S6–S11

17. 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 1998;80:235–241

18. Frayling TM, Walker M, McCarthy MI, Evans JC, Allen LI, Lynn S, Ayres S, Millauer B, Turner C, Turner RC, Sampson MJ, Hitman GA, Ellard S, Hattersley AT. Parent-offspring trios: a resource to facilitate the identification of type 2 diabetes genes. Diabetes 1999;48:2475–2479

19. Knight B, Shields BM, Hattersley AT. The Exeter Family Study of Childhood Health (EFSOCH): study protocol and methodology. Paediatr Perinat Epidemiol 2006;20:172–179

20. Owen K, Hattersley AT. Maturity-onset diabetes of the young: from clinical description to molecular genetic characterization. Best Pract Res Clin Endocrinol Metab 2001;15:309–323

21. 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 2001;50:681–685

22. Dupont WD, Plummer WD Jr. Power and sample size calculations: a review and computer program. Control Clin Trials 1990;11:116–128

23. Morris AP. Direct analysis of unphased SNP genotype data in population-based association studies via Bayesian partition modelling of haplotypes. Genet Epidemiol 2005;29:91–107

24. Bodmer W, Bonilla C. Common and rare variants in multifactorial susceptibility to common diseases. Nat Genet 2008;40:695–701

25. Plagnol V, Smyth DJ, Todd JA, Clayton DG. Statistical independence of the colocalized association signals for type 1 diabetes and RPS26 gene expression on chromosome 12q13. Biostatistics 2009;10:327–334