Chapter 6

Large candidate gene association studies of pancreatic β-cell genes and risk of type 2 diabetes

6.1 Summary

Pancreatic β-cell dysfunction is a necessary component of type 2 diabetes and genes with important roles in pancreatic β-cell have been shown to influence disease susceptibility. I contributed to analysis of a large-scale candidate gene T2D association study of genes predicted to be involved in β-cell development, function and survival. One thousand five hundred and thirty-six non-redundant SNPs across 84 candidate genes were genotyped in three UK based case-control studies (1,484 cases and 1,856 controls) and in an Ashkenazi case-control study (930 cases and 461 controls). SNPs mapping within the Wolfram Syndrome 1 (*WFS1*) gene were associated with type 2 diabetes in both populations and in a pooled analysis of all four studies. Analysis of *WFS1* variation in HapMap samples revealed that the entire gene maps within a strong LD block extending 15 kb 5' and 3' of the gene. Further support for the association was sought by genotyping *WFS1* SNPs in three further UK studies (7119 cases and 9072 controls) and a Swedish case-control study (1296 cases and 1412 controls). An updated meta-analysis of all studies with data available from WTCCC, FUSION, and DGI genome-wide association studies demonstrated genome-wide significance of the association between rs10010131 and type 2 diabetes risk (OR = 0.89 (0.86-0.92), $P = 5.4 \times 10^{-11}$). Rare mutations in *WFS1* cause an autosomal recessive condition characterised in part by juvenile-onset non-autoimmune diabetes. These studies demonstrate that common variation in *WFS1* also predisposes to common T2D with complex inheritance patterns. The initial association studies (Sandhu et al. 2007) and Swedish replication study (Franks et al. 2008) have been published.

6.2 Introduction

Type 2 diabetes is a metabolic disease characterised by defects in insulin action and insulin secretion. Candidate gene linkage and association studies have therefore focused on genes involved in biological pathways affecting one or both of these processes. Recently, a large candidate gene association study from my group tested 152 SNPs in 71 genes for association with type 2 diabetes risk and underlying quantitative traits (Barroso et al. 2003). These genes were assigned to one of three categories: those involved in pancreatic ß-cell function, those primarily influencing insulin action and glucose metabolism in muscle, liver and adipose tissue, and those influencing other processes relevant to diabetes such as energy intake, energy expenditure and lipid metabolism. A number of SNPs showed nominal association with type 2 diabetes, most of which fell within genes involved in pancreatic ß-cell function. These results led us to hypothesise that insulin sensitivity may be more heterogeneous and influenced by environmental and other factors, while ß-cell function might have a stronger genetic component. Certainly, genes with important roles in pancreatic β-cells have been shown to influence susceptibility to type 2 diabetes, including *KCNJ11* (Gloyn et al. 2003; Nielsen et al. 2003), *HNF4α* (Love-Gregory et al. 2004; Silander et al. 2004), *TCF7L2* (Grant et al. 2006), and *TCF2* (*HNF1β*) (Winckler et al. 2007).

Hypothesis-free genome-wide association studies were not yet feasible when planning this project so we embarked on a large-scale candidate gene association study in three UK-based case-control studies and an independent replication case-control study of Ashkenazi Jewish ethnicity, prioritising genes involved in pancreatic ß-cell development, function, and survival as more likely to bestow detectable risk of type 2 diabetes. The aims of this study were:

1) To test for association between variation across candidate genes and risk of type 2 diabetes

2) To replicate any promising findings in further case-control studies

6.3 Results and Discussion

6.3.1 Type 2 diabetes association study of genes involved in pancreatic ß-cell function

This was a collaborative project involving groups lead by Dr Inês Barroso (Wellcome Trust Sanger Institute, Hinxton, UK), Prof. Nick Wareham (MRC Epidemiology Unit, Cambridge, UK), Prof. Andrew Hattersley (Peninsula Medical School, Exeter, UK), Prof. Alan Permutt (Washington University, St. Louis, USA), Prof. Benjamin Glaser (Hadassah Hebrew Medical Center, Israel) and Prof. Mark McCarthy (Wellcome Trust Centre for Human Genetics, Oxford). I contributed to the analysis of the 1536 SNPs, the genotyping of the replication studies and the writing of the paper (Sandhu et al. 2007).

6.3.1.1 Candidate gene and SNP selection

Eighty-four genes predicted to be involved in pancreatic ß-cell development, growth, function and survival were prioritised for testing in a type 2 diabetes association study by Inês Barroso (Wellcome Trust Sanger Institute, UK), Alan Permutt (Washington University, St. Louis, USA), and Andrew Hattersley (Peninsula Medical School, Exeter). All SNPs present in HapMap in November 2004 that mapped within these 84 transcripts and had been genotyped in CEU trios were selected for genotyping, except those that were in complete linkage disequilibrium ($r^2 = 1$) with a SNP already selected. In total 1536 SNPs were selected for genotyping (Table 6.1).

Table 6.1 Candidate genes and number of SNPs genotyped according to functional group

Group	Hugo Symbol	Common Name	Number of SNPs
Pancreas development/ transcription	*ACVR1*	activin A receptor, type I	12
	ACVR1b	activin A receptor, type IB	4
	ACVR2	activin A receptor, type II ; ACTRII	9
	BTC	betacellulin	5
	CDH1	cadherin 1, type 1, E-cadherin (epithelial)	10
	CDH2	cadherin 2, type 1, N-cadherin (neuronal)	80
	CDK2	cyclin-dependent kinase 2	4
	EGF	epidermal growth factor (beta-urogastrone)	7
	EGFR	epidermal growth factor receptor	37
	FGF10	fibroblast growth factor	5
	FGF7	fibroblast growth factor 7 (keratinocyte growth factor)	9
	FGFR1	fibroblast growth factor receptor 1 (fms-related tyrosine kinase 2, Pfeiffer syndrome)	4
	FGFR2	fibroblast growth factor receptor 2 (bacteria-expressed kinase, keratinocyte growth factor receptor, craniofacial dysostosis 1, Crouzon syndrome, Pfeiffer syndrome, Jackson-Weiss syndrome)	113
	FOXA3	HNF3γ	4
	FOXO1A	FKHR, FOXO1	10
	FOXO3A	FKHRL1	14
	HNF4γ	HNF4γ	9
	ISL1	ISL1 transcription factor, LIM/homeodomain, (islet-1)	1
	JAG1	jagged 1 (Alagille syndrome)	10
	JAG2	jagged 2	5
	LHX4	LIM homeobox 4	10
	LHX6	LHX6.1	8
	LMX1A	LIM homeobox transcription factor 1, alpha	29
	NEUROD1	neurod/Beta2	1
	Notch1	Notch homolog 1, translocation-associated (Drosophila)	9
	NOTCH2	Notch homolog 2 (Drosophila)	8
	NOTCH3	Notch homolog 3 (Drosophila)	8
	Notch4	Notch homolog 4 (Drosophila)	5
	NR5A2	Lrh1	30
	ONECUT1	HNF6	6
	PAX6	paired box gene 6 (aniridia, keratitis)	4
	PBX1	pre-B-cell leukaemia transcription factor 1	56
	PSEN2	presenilin 2 (Alzheimer disease 4)	5
	RARB	retinoic acid receptor, beta	95
	RORB	nuclear receptor RZR-beta	19
	TCF1	HNF1α	5
	TCF2	HNF1β	12
	TCF3	E47	3
	TGFBR1	transforming growth factor, beta receptor I (activin A receptor type II-like kinase, 53kDa)	5
	VIPR1	vasoactive intestinal peptide receptor 1	9
	VIPR2	vasoactive intestinal peptide receptor 2	9
β-cell death/ apoptosis	*BAX*	BCL2-associated X protein	3
	BCL2	B-cell leukaemia/lymphoma 2	53
	BID	BH3 interacting domain death agonist	16

Table 6.1 Candidate genes and number of SNPs genotyped according to functional group continued

Group	Hugo Symbol	Common Name	Number of SNPs
	BIK	BCL2-interacting killer (apoptosis-inducing)	4
	CASP9	caspase 9	4
	EIF2AK3	PERK	7
	GSK3B	glycogen synthase kinase 3 beta	5
	WFS1	Wolfram syndrome 1 (wolframin)	6
Insulin secretion	*ATP2B1*	ATPase, Ca++ transporting	6
	CACNA1C	calcium channel, voltage-dependent, L type, alpha 1C subunit; CaV1.2	133
	CACNA1D	calcium channel, voltage-dependent, L type, alpha 1D subunit	69
	CHGA	chromogranin A (secretogranin 2)	5
	CPE	carboxypeptidase E	19
	EIF2S1	eukaryotic translation initiation factor 2, subunit 1 alpha, 35kDa	4
	GLP1R	glucagon-like peptide 1 receptor	10
	KCNJ6	potassium inwardly-rectifying channel, subfamily J, member 6	103
	MAPK8IP1	islet-brain 1	3
	PCSK1	prohormone convertase 1	8
	PCSK2	prohormone convertase 2	45
	SGNE1	secretory granule, neuroendocrine prot 1	9
	SNAP25	SNAP25 synaptosomal-associated protein	16
	VTI1B	VTI1B vesicle transport through interaction with t-SNAREs homolog 1B (yeast) or vSNARE	4
Insulin signalling	*FGF2*	fibroblast growth factor 2 (basic); bFGF	9
	INSR	insulin receptor	27
	IRS1	insulin receptor substrate 1	5
	IRS2	insulin receptor substrate 2	7
	PIK3R1	phosphatidylinositol 3-kinase, regulatory subunit, polypeptide 1 (p85 alpha)	15
	PRKCE	PKC epsilon	109
	PRKCZ	PKC zeta	5
	PSEN1	presenilin 1 (Alzheimer disease 3)	8
	RPS6KB1	ribosomal protein S6 kinase, 70kDa, polypeptide 1	7
Enzymes and metabolism	*CAMK2A*	CAMK2A	16
	CAMK2B	CAMK2B	8
	CAMK2D	CAMK2D	28
	CAMK2G	calcium/calmodulin-dependent protein kinase (CaM kinase) II gamma	7
	CAPN3	calpain 3	11
	GCK	glucokinase	6
	GCKR	glucokinase regulatory protein	5
Inflammation	*IKBKB*	IKKb	4
	IL6	interleukin 6	2
	NFATC1	nuclear factor of activated T-cells, cytoplasmic, calcineurin-dependent 1	25
	NFKB1	NFKbeta; NFkB	10
	SOCS3	suppressor of cytokine signaling 3	2

6.3.1.2 Association testing of candidate gene SNPs in type 2 diabetes case-control studies

SNPs were genotyped by P. Deloukas's laboratory at the Wellcome Trust Sanger Institute in four case-control studies of type 2 diabetes. These included three UK-based studies: Cambridgeshire, EPIC, and Exeter (1484 cases and 1856 controls), and an Ashkenazi study (930 cases and 461 controls). Of 1536 genotyped SNPs, 1367 (89%) met the genotyping QC criteria (section 6.4.2) and were analysed in two stages. In the first stage, association with diabetes risk was tested in UK populations only (1484 cases and 1856 controls) (Figure 6.1). Eighteen (1.3%) were associated with diabetes risk at $P < 0.01$ (Table 6.2).

In the second phase of the analysis, these 18 SNPs were tested for association with diabetes risk in the Ashkenazim study set, an ethnically distinct founder population comprising 930 cases and 461 controls. Two of the originally associated SNPs (rs10010131 and rs6446482) were associated in this independent study at $P < 0.05$ (Table 6.2). These SNPs are both located in *WFS1* and are in high linkage disequilibrium ($r^2 = 0.98$) in our study populations.

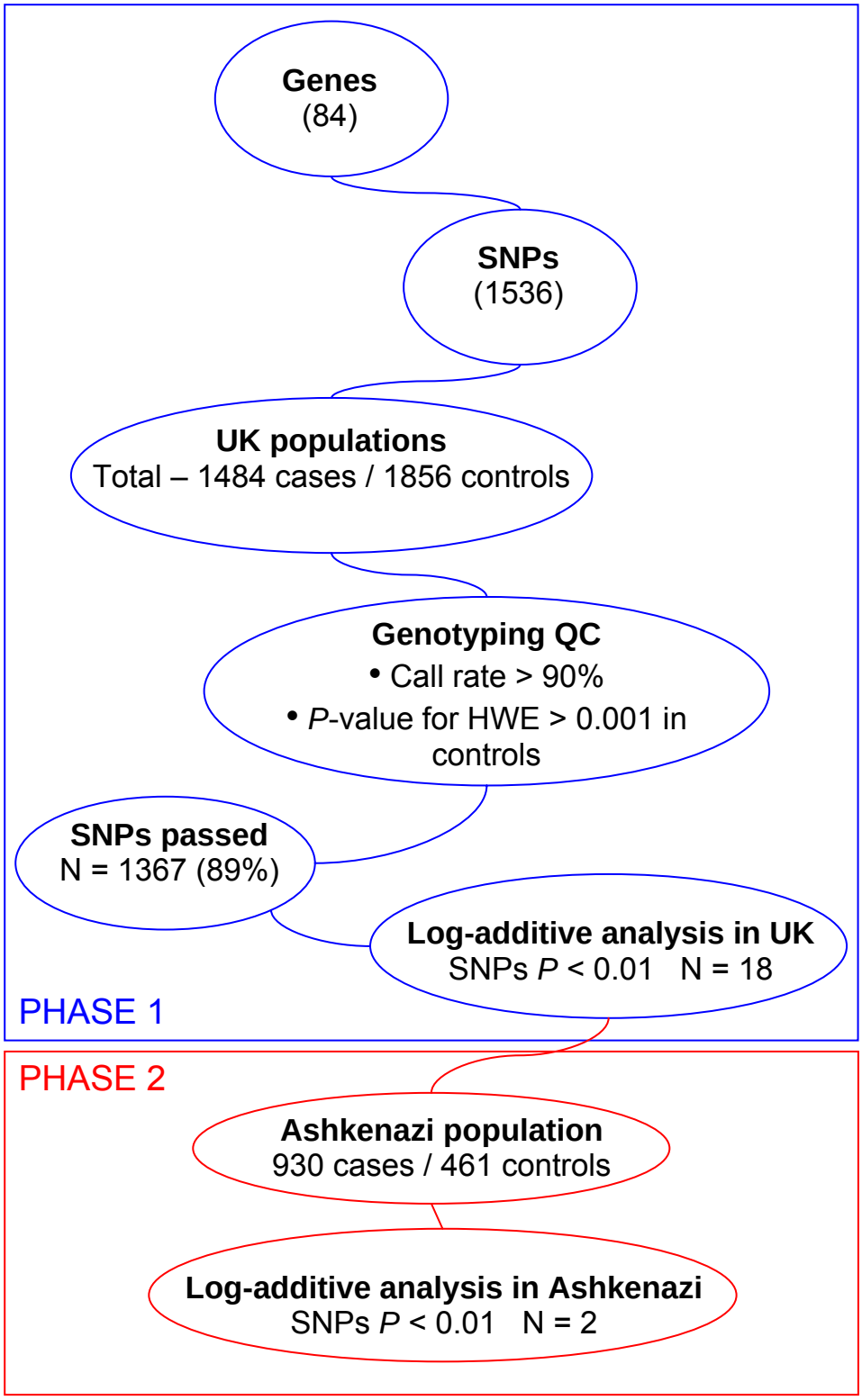


Figure 6.1 Flow chart of the procedure followed during the first and second phases of the candidate SNP analysis

Table 6.2 Statistical associations (P < 0.01) between SNPs in genes involved in beta-cell development, growth, function and survival and risk of type 2 diabetes in UK populations and for a study in an Ashkenazi population

Gene	SNP	Odds ratio (95 % CI) for stage 1: UK populations	P value for phase 1 (UK populations)*	Odds ratio (95 % CI) for Ashkenazi population	P value for r Ashkenazi population
CHGA	rs941584	0.80 (0.72-0.90)	2.8 x 10-4	1.03 (0.87-1.22)	0.75
NFATC1	rs643705	1.30 (1.12-1.52)	5.4 x 10-4	0.85 (0.71-1.02)	0.08
PAX6	rs628224	0.82 (0.71-0.93)	2.8 x 10-3	0.95 (0.74-1.22)	0.71
NFKB1	rs1609798	0.85 (0.77-0.95)	2.9 x 10-3	1.04 (0.87-1.24)	0.68
NFKB1	rs11722146	0.85 (0.76-0.95)	3.0 x 10-3	1.03 (0.86-1.22)	0.75
NFKB1	rs230498	0.86 (0.77-0.95)	3.0 x 10-3	1.03 (0.87-1.21)	0.74
WFS1	rs10010131	0.86 (0.78-0.95)	3.1 x 10-3	0.79 (0.66-0.94)	0.01
CACNA1D	rs4687736	1.18 (1.06-1.32)	3.3 x 10-3	1.00 (0.82-1.22)	0.97
EGFR	rs2075112	1.15 (1.04-1.27)	5.1 x 10-3	1.03 (0.87-1.22)	0.74
PBX1	rs7535186	0.87 (0.79-0.96)	5.3 x 10-3	0.97 (0.83-1.14)	0.76
WFS1	rs6446482	0.87 (0.79-0.96)	5.7 x 10-3	0.79 (0.66-0.95)	0.01
NFKB1	rs230539	0.87 (0.78-0.96)	5.9 x 10-3	1.03 (0.87-1.21)	0.77
TCF2	rs7501939	1.15 (1.04-1.27)	6.6 x 10-3	1.04 (0.89-1.22)	0.64
CACNA1D	rs3796347	1.15 (1.04-1.26)	7.0 x 10-3	0.96 (0.81-1.13)	0.60
CAMK2A	rs3822607	0.87 (0.78-0.96)	7.1 x 10-3	1.06 (0.89-1.26)	0.50
NFATC1	rs3826567	1.30 (1.07-1.58)	7.3 x 10-3	0.88 (0.69-1.11)	0.27
FOXA3	rs11669442	1.15 (1.04-1.28)	7.6 x 10-3	0.92 (0.77-1.08)	0.31
NUDT6	rs1048201	1.19 (1.05-1.36)	8.2 x 10-3	1.19 (0.98-1.45)	0.08

The UK population comprised up to 1,484 cases and 1,856 controls, and the Ashkenazi population comprised up to 930 cases and 461 controls. All SNPs were in Hardy-Weinberg equilibrium ($P > 0.01$ in controls). CI, confidence interval

* Based on a log additive model adjusting for study.

6.3.1.3 Association of *WFS1* SNPs with type 2 diabetes risk in three UK studies and one Ashkenazi study

Although only two of the original six variants typed across *WFS1* were associated with diabetes in the two-staged approach, in a pooled analysis of all four study sets, five SNPs showed statistical association with diabetes (Table 6.3). These SNPs were in high linkage disequilbirum (Table 6.4). As rs10010131 showed a stronger association than the other *WFS1* SNPs, likelihood ratio tests were used to assess whether rs10010131 explained all the observed associations in this region. We consecutively added the other SNPs in this region in a log additive form to a model containing rs10010131 (1df assuming no dominance at the test locus). This analysis demonstrated that none of the other SNPs in this region of high LD improved the model containing just rs10010131, and indicated that no additional genotyped SNP is independently contributing to T2D risk (Table 6.5). The reciprocal analysis, adding rs1001031 to models containing the other associated SNPs, showed that rs10010131 significantly improved the fit of all models (Table 6.5), with the exception of SNP rs6446482. The correlation between SNPs rs10010131 and rs6446482 is very high (r^2=0.98) making it difficult to distinguish between their effects on T2D risk. These analyses suggest that either rs10010131 or rs6446482 might be causal alleles, or that they are in LD with a causal allele, or both, and that the other SNPs do not independently contribute to disease risk.

There is also evidence that rs752854 might contribute to disease risk independently of rs10010131, as adding rs752854 (which was not statistically associated with disease) to the model containing rs10010131 substantially improved the fit of the model (P = 0.002, 1df) (Table 6.5).

Table 6.3 Association between SNPs located in the *WFS1* gene and risk of type 2 diabetes: all study populations comprising up to 2,414 cases and 2,317 controls

SNP	Odds ratio	95 % CI	P-value*
rs10010131	0.84	0.77-0.92	1.3×10^{-4}
rs6446482	0.85	0.78-0.93	2.7×10^{-4}
rs4689391	0.86	0.79-0.94	9.6×10^{-4}
rs3821943	0.89	0.81-0.96	5.0×10^{-3}
rs1801212	0.89	0.81-0.98	0.015
rs752854	0.94	0.85-1.03	0.164

* Based on a single locus log additive model adjusted for study. CI, confidence interval.

Table 6.4 Correlations among *WFS1* SNPs genotyped in all study populations: 2,317 controls

	rs10010131	rs6446482	rs4689391	rs3821943	rs1801212	rs752854
rs10010131	-					
rs6446482	0.98	-				
rs4689391	0.94	0.93	-			
rs3821943	0.88	0.87	0.92	-		
rs1801212	0.76	0.74	0.72	0.68	-	
rs752854	0.86	0.84	0.82	0.76	0.84	-

Correlations were not materially different in UK and Ashkenazi populations (data not shown).

Table 6.5 Log-likelihood ratio tests assessing the association among SNPs in the *WFS1* gene with risk of type 2 diabetes: all study populations comprising up to 2,414 cases and 2,317 controls

SNP	P value (1 df)*	P-value (1 df)**
rs10010131	-	-
rs6446482	0.254	0.053
rs4689391	0.524	0.047
rs3821943	0.290	5.1×10^{-3}
rs1801212	0.457	2.0×10^{-3}
rs752854	1.7×10^{-3}	4.3×10^{-6}

CI, confidence interval, all models include a study variable

* P value (log likelihood ratio test) for addition of SNP (log additive) to model containing SNP rs10010131 (2df)

** P value (log likelihood ratio test) for addition of SNP rs10010131 (log additive) to model containing SNP (2 df).

6.3.1.4 Selection of SNPs for replication in further UK cohorts

As the associations between T2D risk and SNPs rs10010131 and rs6446482 were indistinguishable given the high LD between these SNPs, and as rs752854 improves the fit of the model containing rs10010131, these three SNPs were prioritised for replication studies. We also endeavoured to discover other possible causative variants by examining variation in the region of the association signal. Using data from HapMap, and based on sequence spanning 63.4 kb (chromosome 4, 6374656-6438055), including 15 kb extending both 5' and 3' from *WFS1*, strong LD was detected across the region. The entire gene was defined by a single haplotype block of ~40 kb (Figure 6.2). Within this block, 53 SNPs had a minor allele frequency (MAF) of ≥1%. The six SNPs typed in this part of the study were all located in this region and together tagged 88% of the common variation in this block (47 of the 53 SNPs, MAF ≥1%, r^2 ≥0.8) with a mean r^2 of 0.97. One nonsynonymous SNP (R611H, rs734312) is highly correlated with SNPs rs10010131 (r^2 = 0.92) and rs6446482 (r^2 = 0.88) and thus may be a causal variant. It also showed suggestive association with type 2 diabetes in a previous study based on 479 cases and 509 controls (a subset of our samples) (Minton et al. 2002).

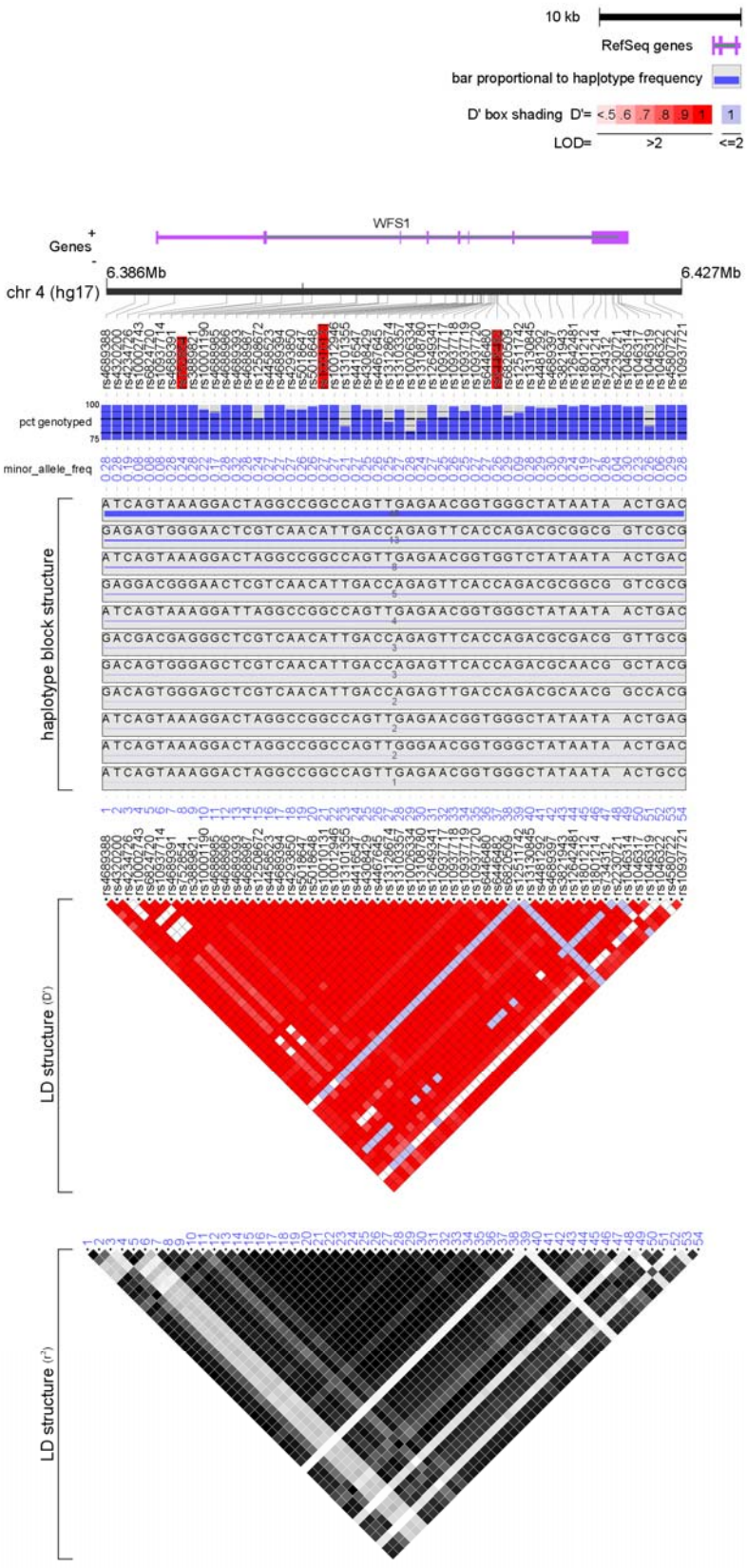


Figure 6.2 Feature map of the *WFS1* gene. Legend continued on page 181.

Figure 6.2 legend continued. The positions of the 54 SNPs genotyped in HapMap with a MAF≥ 0.01 are shown relative to the locus (purple) and chromosome 4 (black bar) (see text for details). The three SNPs typed in all studies are highlighted in red. Underlying each SNP is QC information and MAF in the CEPH samples. The middle part of the figure shows the 11 common haplotypes for the *WFS1* gene with frequency ≥ 0.01 in the CEPH samples. The thickness of the blue lines are proportional to the haplotype frequencies. 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 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(Barrett et al. 2005), available from the HapMap website (http://www.broad.mit.edu/mpg/haploview/index.php).

6.3.1.5 Replication of *WFS1* SNPs rs10010131, rs6446482, rs752854 and rs734312 in three further UK-based case-control studies

To extend support for an association between variation at *WFS1* and diabetes risk, further replication of rs10010131, rs6446482, rs752854 and the highly correlated nonsynonymous SNP (rs734312) in three further case-control studies, ADDITION (926 cases and 1,497 controls), Warren 2 (2,465 cases and 3,843 controls) and Tayside (3,728 cases and 3,732 controls) was attempted. Independent evidence for association of rs10010131, rs6446482 and rs752854, was found in each study (Table 6.6).

Table 6.6 Association between SNPs located in the *WFS1* gene and risk of type 2 diabetes: replication studies and pooled analysis

SNP	Odds ratio (95% CI)	P-value
ADDITION study (926 cases and 1,497 controls)		
rs10010131	0.87 (0.77-0.98)	0.020
rs6446482	0.87 (0.77-0.98)	0.021
rs752854	0.86 (0.76-0.97)	0.013
rs734312	0.92 (0.82-1.03)	0.163
Warren 2 study (2,465 cases and 3,843 controls)		
rs10010131	0.91 (0.84-0.98)	0.011
rs6446482	0.92 (0.86-0.99)	0.027
rs752854	0.93 (0.86-1.00)	0.060
rs734312	0.93 (0.87-1.00)	0.061
Tayside study (3,728 cases and 3,732 controls)		
rs10010131	0.93 (0.87-0.99)	0.029
rs6446482	0.92 (0.87-0.99)	0.019
rs752854	0.93 (0.86-0.99)	0.032
rs734312	0.93 (0.87-0.99)	0.019
All seven pooled studies comprising up to 9,533 cases and 11,389 controls*		
rs10010131	0.90 (0.86-0.93)	1.4×10^{-7}
rs6446482	0.90 (0.87-0.94)	3.4×10^{-7}
rs734312	0.92 (0.88-0.95)	2.0×10^{-5}
rs752854	0.92 (0.88-0.96)	1.3×10^{-4}

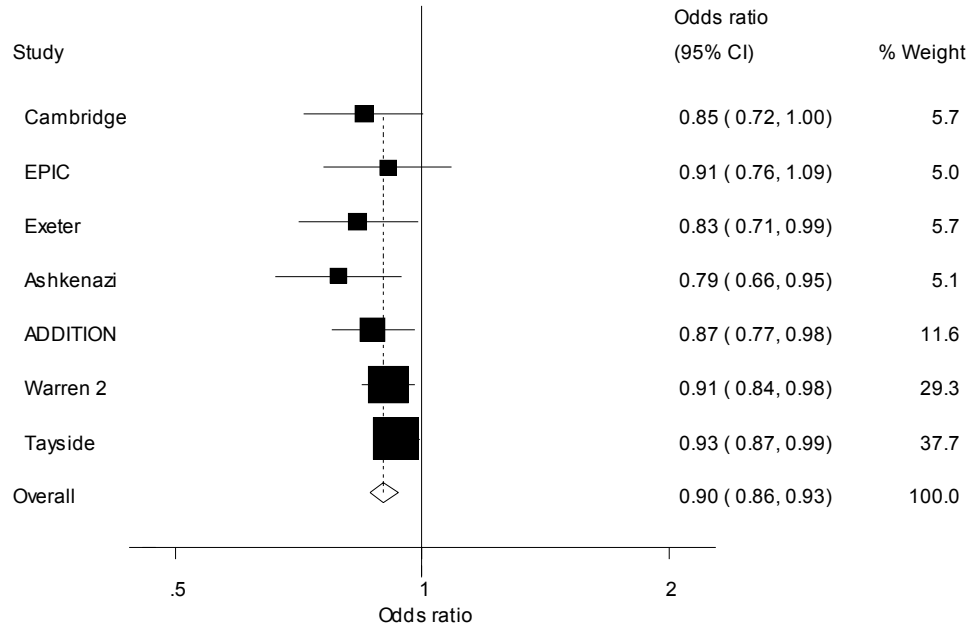
* Based on a single locus log additive model adjusted for study. CI, confidence interval.

The possible interdependency of rs10010131 and rs752854 was tested in a combined analysis of ADDITION, Warren 2 and Tayside studies. However, adding rs752854 did not improve the fit of the logistic regression model containing rs10010131 indicating that rs752854 does not interact with rs10010131 to effect risk of T2D.

6.3.1.6 Pooled analysis of *WFS1* SNPs rs10010131, rs6446482, rs752854 and rs734312 in all UK studies and the Ashkenazi study

As there was some evidence for association of the nonsynonymous SNP rs734312 in the ADDITION, Warren 2 and Tayside studies (Table 6.6) this variant was genotyped in the original four studies and a pooled analysis of all seven studies was conducted, comprising up to 9,533 cases and 11,389 controls (Table 6.6). In this analysis, rs734312 was associated with diabetes risk ($P = 2.0 \times 10^{-5}$). However, likelihood ratio tests showed that rs734312 did not contribute to a model including rs10010131 ($P = 0.88$), whereas rs10010131 substantially improved the fit of a model including rs734312 ($P = 4.9 \times 10^{-3}$), suggesting that rs734312 is unlikely to be the functional variant explaining these associations. There was also no consistent evidence for interdependency between rs10010131 or rs6446482 and rs752854 and diabetes risk in the combined study sets (data not shown). In the pooled analysis, rs10010131 (MAF = 40%) and rs6446482 (MAF = 41%) were strongly associated with diabetes risk at $P = 1.4 \times 10^{-7}$ and $P = 3.4 \times 10^{-7}$, respectively (Table 6.6). Furthermore, the magnitude of this association was highly consistent across studies (Figure 6.3), with no heterogeneity among studies (P (six degrees of freedom) = 0.59 and 0.68 for rs10010131 and rs6446482, respectively).

(a)



(b)

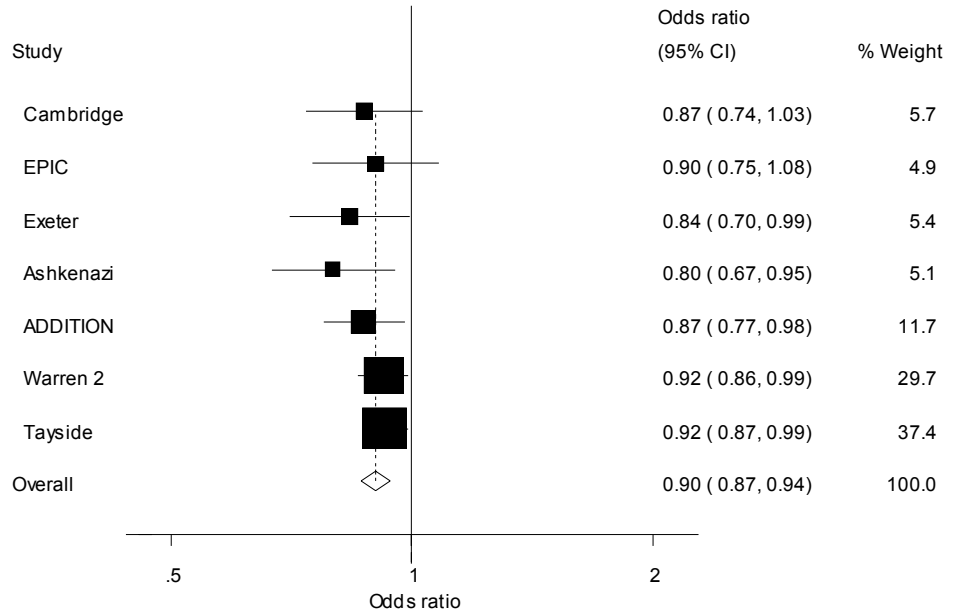


Figure 6.3 Meta-analysis of *WFS1* SNPs in all seven case-control studies, (a) Association between SNP rs10010131 and risk of type 2 diabetes in individual studies. (b) Association between SNP rs6446482 and risk of type 2 diabetes in individual studies.

6.3.1.7 Discussion

These analyses provide strong evidence for association of variation across *WFS1* and type 2 diabetes risk. Of 1536 SNPs across 84 candidate genes involved in β-cell function, two SNPs in *WFS1* were associated with type 2 diabetes in three UK-based case-control studies and in an Ashkenazi case-control study (P<0.01). In a pooled analysis of UK and Ashkenazi studies, five of the six *WFS1* SNPs tested in the candidate gene association study were significantly associated with type 2 diabetes risk (P<0.05). A meta-analysis of these studies with three further UK-based case-control studies (9533 cases and 11389 controls in total) demonstrated robust replication of the association signal in *WFS1*, the minor allele of the strongest SNP conferring a protective effect (OR = 0.90) from type 2 diabetes ($P = 1.4 \times 10^{-7}$).

Our study provides more robust statistical support for a previously reported association between rs734312 and type 2 diabetes risk. Previously, Minton et al. sequenced *WFS1* in 29 type 2 diabetic subjects and identified 5 non-synonymous SNPs, including rs734312 (R611H) and rs1801208 (R456H) (Minton et al. 2002). In 152 parent-offspring trios, R456 and H611 alleles, and the R456-H611 haplotype demonstrated borderline significant overtransmission to affected offspring from heterozygous parents. Furthermore, the H611 allele and the R456-H611 haplotype occurred significantly ($P = 0.06$ and $P = 0.023$ respectively) more frequently in 323 type 2 diabetic patients than in 357 normoglycaemic controls.

WFS1 encodes wolframin, an endoplasmic reticulum (ER) membrane protein with a role in the maintenance of ER calcium homeostasis (Takei et al. 2006). Homozygous and compound heterozygous inactivating mutations in *WFS1* cause Wolfram Syndrome (MIM222300), characterised by diabetes insipidus, young onset non-autoimmune insulin-dependent diabetes mellitus, optic atrophy and deafness (Inoue

et al. 1998; Strom et al. 1998). WFS1 deficiency also causes glucose intolerance and decreased β-cell mass in mice, which is thought to occur as a result of ER stress-induced apoptosis (Ishihara et al. 2004; Riggs et al. 2005; Yamada et al. 2006). This indicates that variation in *WFS1* might influence risk of type 2 diabetes through its effects on β-cell survival.

6.3.2 Replication of *WFS1* SNPs rs10010131, rs6446482, rs752854 and rs734312 in the Västerbotten type 2 diabetes case-control study

Although there was strong evidence for association between SNPs in *WFS1* and T2D risk in the β-cell candidate gene study (Sandhu et al. 2007) this did not reach current thresholds of genome-wide significance ($5x10^{-8}$ or 0.05 corrected for 10^6 independent tests) therefore we sought further support for the association from an independent type 2 diabetes case-control study from the county of Västerbotten in northern Sweden. In collaboration with Paul Franks' group at the Department of Public Health and Clinical Medicine, Umeå University Hospital, Umeå, Sweden, I genotyped *WFS1* SNPs rs10010131, rs6446482, rs752854 and rs734312 in 1296 Swedish cases and 1412 Swedish controls. All SNPs had genotype call rates >98%, were in Hardy-Weinberg equilibrium ($P > 0.1$), and genotyping concordance rates were 100%. The LD between SNPs in control participants was generally lower than the values reported in other populations (Figure 6.4).

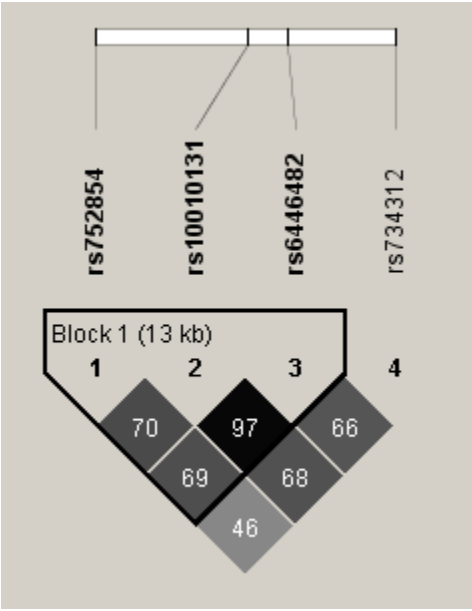


Figure 6.4 Linkage disequilibrium (r^2) between *WFS1* genotypes in the control samples of the Västerbottens type 2 diabetes case-control study (N=1,412)

SNPs rs10010131, rs6446482 and rs734312 showed borderline statistical association with type 2 diabetes (Table 6.7). The magnitude and direction of these associations were consistent with those found in UK and Ashkenazi study sets. SNP rs752854 was statistically associated with type 2 diabetes, with the minor allele conveying a protective effect (OR 0.85, 95% CI 0.75-0.96, *P* = 0.010; Table 6.7). This SNP was not the most strongly associated in UK and Ashkenazi samples, though this may reflect random statistical fluctuation rather than allelic heterogeneity between populations.

Table 6.7 Effect estimates for each of the *WFS1* SNPs in relation to glucose levels or type 2 diabetes in the Västerbottens type 2 diabetes case-control study

	Adjusted geometric means or Odds ratios (95% CI)		*P*-value
	Heterozygotes	Minor allele homozygotes	
rs10010131 (MAF = 0.43)			
Type 2 diabetes (yes vs. no)	0.87 (0.69, 1.01)	0.81 (0.63, 1.03)	0.083
rs6446482 (MAF = 0.44)			
Type 2 diabetes (yes vs. no)	0.93 (0.75, 1.16)	0.83 (0.65, 1.05)	0.098
rs752854 (MAF = 0.35)			
Type 2 diabetes (yes vs. no)	0.84 (0.64, 1.11)	0.72 (0.55, 0.96)	0.010
rs734312 (MAF = 0.48)			
Type 2 diabetes (yes vs. no)	0.81 (0.66, 1.00)	0.80 (0.64, 1.01)	0.066

Results are odds ratios (OR: type 2 diabetes) and 95% confidence intervals from additive genetic models. The frequency of T2D cases in participants homozygous for the major allele was used as the baseline risk and therefore the OR will be 1.00. Data are adjusted for age, sex and BMI. *P*-values are from tests for linear trend.

6.3.2.1 Meta-analysis of UK, Ashkenazi and Swedish type 2 diabetes case-control studies with data from genome-wide association studies

As well as data from the original UK and Ashkenazi study sets, data from three of the first five type 2 diabetes genome wide association scans were available for meta-analysis with the Swedish case-control study (Saxena et al. 2007; Scott et al. 2007; Sladek et al. 2007). We contacted relevant investigators of these GWA scans and requested summary statistics (odds ratio and 95% confidence intervals). SNP rs10010131 was available for analysis from the Finland-United States Investigation of Non-Insulin-Dependent Diabetes Mellitus Genetics (FUSION), but for Sladek et al. and the Diabetes Genetics Initiative (DGI) SNPs in high linkage disequilibrium with rs10010131 were used as proxies (Table 6.8). T2D data from the WTCCC was not included in the meta-analysis as this comprised the same samples already present in the Sandhu *et al.* paper.

Table 6.8 Details of GWA studies used in meta-analysis of *WFS1* SNP rs10010131

Study	Reference	Cases (N)	Controls (N)	SNP	Correlation with rs10010131 (r^2)*
FUSION	(Scott, Mohlke et al. 2007)	1,160	1,172	rs10010131	1.0
Sladek	(Sladek, Rocheleau et al. 2007)	686	669	rs4416547	1.0
DGI	(Saxena, Voight et al. 2007)	1,464	1,467	rs10012946	1.0

* = in the HapMap CEU panel.

The meta-analysis of the Västerbotten study, the original seven study sets, and the three additional genome-wide studies showed robust evidence for statistical association (OR = 0.89; $P = 5.4 \times 10^{-11}$; P for heterogeneity = 0.42 with 10 *df*, Figure 6.4).

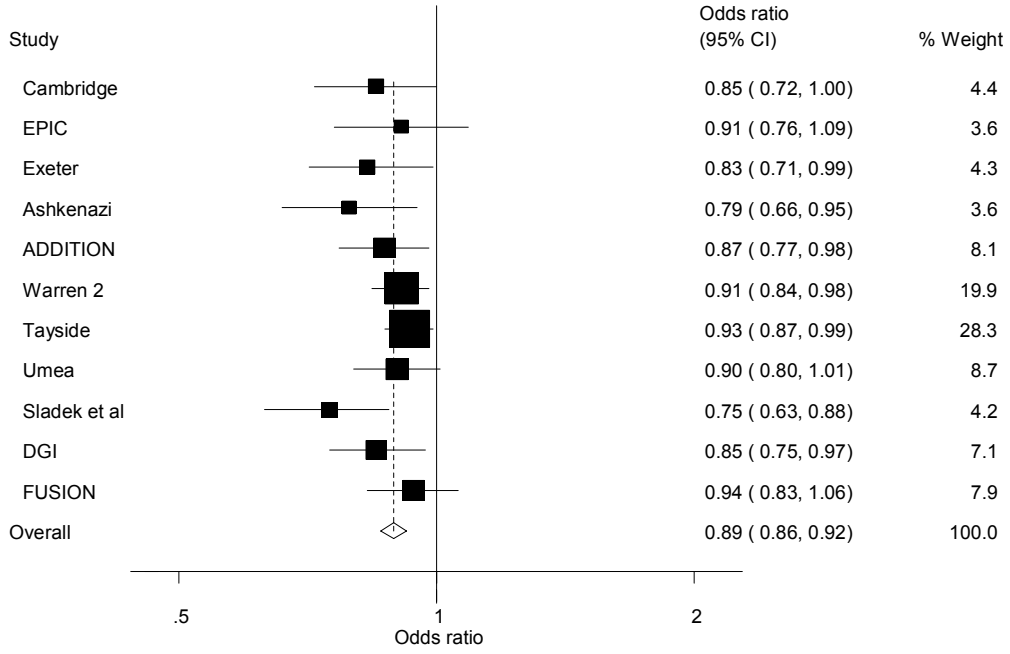


Figure 6.4 Association between *WFS1* SNPs and risk of type 2 diabetes in all seven original study sets, the Västerbotten study, and in data from three genome-wide association studies

Summary data for SNP rs10010131 was used in the meta-analysis with the exception of Sladek et al (Sladek et al. 2007) and DGI (Saxena et al. 2007) where rs4416547 and rs10012946 were substituted as proxy markers. These SNPs are perfectly correlated (r^2 = 1.0) with rs10010131 in HapMap samples.

6.3.2.2 Discussion

Four SNPs in *WFS1* were tested for association with T2D in the Västerbotten study (1296 Swedish cases and 1412 Swedish controls), and rs10010131 (or proxies in perfect LD) was tested for association with T2D in an updated meta-analysis comprising 11 different study cohorts of Northern and Western European and Ashkenazi Jewish ancestry. In the Västerbotten study, *WFS1* SNP rs752854 was statistically associated with T2D and effects of rs10010131, rs6446482 and rs734312 were similar in magnitude and direction to those previously reported (Sandhu et al. 2007). In the updated meta-analysis of data from the β-cell candidate gene study, the Västerbotten study, and three genome-wide association studies, the association of rs10010131 with T2D risk was replicated. Furthermore, the protective effect of the minor allele (OR = 0.89) reached genome-wide significance ($P = 5.4 \times 10^{-11}$). This work was published by Franks et al. (Franks et al. 2008).

The size and direction of the effects of *WFS1* SNPs rs752854, rs10010131, rs6446482 and rs734312 on T2D risk in the Västerbotten study were similar to those reported in the original β-cell candidate gene study (Sandhu et al. 2007). However, only rs752854 reached statistical significance (P<0.05). Power calculations indicate that this study had 48% probability of detecting the previously reported effect size of rs10010131 and rs6446482 (OR = 0.90) given MAF = 0.43 and 0.44 respectively, and 33% probability of detecting the effect of rs734312 (OR = 0.92) given MAF = 0.48. Therefore, the Västerbotten study on its own was underpowered to detect the previously reported effects of the *WFS1* variants.

There are several possible explanations for why rs734312 was the most strongly associated SNP in the Västerbotten study, while rs10010131 was the strongest SNP in the 6 UK and 1 Askenazi-based case-control studies. This discrepancy could simply be due to stochastic variation between these populations. Also, the difference

in statistical associations between studies could be attributable in part to different linkage disequilibrium patterns in the Västerbotten compared to the UK-based and Ashkenazi cohorts. This hypothesis is supported by the lower pair-wise LD between SNPs in the Västerbottens case-control study compared with the studies included in the original report. LD in the original study populations ranged between r^2=0.75-0.98 for pair-wise comparisons of the four SNPs, and the LD between the two SNPs showing the strongest statistical associations with type 2 diabetes were correlated at r^2=0.98. In the Västerbottens study, the LD between these two SNPs was similar (r^2=0.97), but the remaining pair-wise comparisons ranged between r^2=0.46-0.97.

By meta-analysing the Västerbottens study with the seven case-control cohorts from the Sandhu *et al.* study, and adding data from genome-wide association analyses in three additional populations, we increased power to detect effects of *WFS1* SNPs on diabetes risk. Identification of the true functional variant(s) behind the association between *WFS1* and T2D risk will require sequencing of *WFS1* to discover novel putative functional variants and further analyses of association with T2D.

6.3 Conclusion

The β-cell candidate gene study and replication study produced strong evidence for an association between *WFS1* variation and T2D risk (for rs10010131, OR = 0.89; P = 5.4 × 10−11). *WFS1* SNPs associated with T2D are highly correlated within a linkage disequilibrium block that covers the entire *WFS1* gene. It is therefore unknown whether these SNPs are individually contributing to disease risk or are simply associated with disease risk due to their correlation with tested or untested true causal variant(s). Fine-mapping of the association signal by resequencing of *WFS1* and further analysis in case-control studies will be required as a first step towards identifying functional variant(s).

6.4 Materials and Methods

6.4.1 Description of cohorts

More details of Cambridgeshire, EPIC, Exeter, ADDITION and Ely studies are provided in Chapter 2.

Discovery cohorts

6.4.1.1 Cambridgeshire case-control

This cohort comprised 552 patients aged 45–76 years with T2D and 552 controls matched for age and sex and general practice, recruited from general practitioner registers in Cambridgeshire, UK (Rathmann et al. 2001).

6.4.1.2 EPIC- Norfolk case-control study

This study is nested within the EPIC - Norfolk Study, a population based cohort study of white UK men and women aged 40-78 years and 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.

6.4.1.3 Exeter case-control study

Six hundred and one individuals with type 2 diabetes were recruited from two sources: a consecutive-case series of patients with T2D diagnosed before 45 years from North and East Devon (Owen et al. 2003), and a collection of type 2 diabetes families that had either both parents available, or one parent and at least two siblings (Frayling et al. 1999). Six hundred and ten controls from the same geographic region were also used in this study.

6.4.1.4 Ashkenazi case-control study

Of the cases, 303 are from the multiplex-affected sibships that were ascertained for the genome scan described by Permutt et al, 2001 (Permutt et al. 2001). The cases were 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.

Replication cohorts

6.4.1.5 Warren 2 case-control study

Informed consent was obtained from all participants. The subjects have been described in detail previously (Weedon et al. 2004). Briefly, all type 2 diabetes subjects were unrelated and of white UK origin who had diabetes defined either by WHO criteria (WHO Study Group 1999) or by being treated with medication for

diabetes, and were recruited from 3 sources: i) probands from type 2 diabetic sibships from the Warren 2 sibling pairs described previously (Minton et al. 2002; Wiltshire et al. 2001); ii) a collection of type 2 diabetes cases (Warren 2 cases) diagnosed between 35-65 years, but not selected on family history; iii) and probands from a collection of families that had either both parents available, or one parent and at least two siblings (Frayling et al. 1999).

Population control subjects were all white UK participants. These were recruited from 3 sources: i) the remaining parents from a consecutive birth cohort (Exeter Family Study) with normal (<6.0mmol/l) fasting glucose and/or normal HbA1c levels (< 6%; Diabetes Control and Complications trial corrected) (Minton et al. 2002); and ii) a nationally recruited population control sample of UK whites obtained from the European cell culture collection (ECACC), and iii) a follow-up study, that is ongoing, of all people born in Great Britain during one week in 1958 (National Child Development Study; http://www.cls.ioe.ac.uk/Cohort/Ncds/mainncds.htm). Cases and families where the proband had high GAD autoantibody levels (>99th percentile of the normal population) were excluded from the study. Known subtypes of diabetes (e.g. MODY) were excluded by clinical criteria and/or genetic testing.

6.4.1.6 ADDITION/Ely

The ADDITION case-control study comprised 926 cases of type 2 diabetes, aged 40-69, recruited from the UK Cambridge arm of the ADDITION trial (Lauritzen et al. 2000), and 1497 controls aged 35 to 79 years from the MRC Ely study (Wareham et al. 1998).

6.4.1.7 Tayside case-control study

The Tayside case-control study comprises 3,745 individuals with type 2 diabetes and 3,786 controls from the Wellcome Trust UK type 2 diabetes case-control collection (Go-DARTS2) which is a sub-study of Diabetes Audit and Research Tayside

(DARTS) (Doney et al. 2002; Doney et al. 2004a; Doney et al. 2003; Doney et al. 2005a; Doney et al. 2004b; Doney et al. 2005b; Morris et al. 1997). All cases were European with physician-diagnosed type 2 diabetes. They were recruited at primary or secondary care diabetes clinics or invited to participate from primary care registers from throughout the Tayside region of Scotland. The cases have not been characterised for GAD anti-bodies or MODY gene mutations. The controls were invited to participate through the primary care physicians or through their workplace occupational health departments. Controls did not have a previous physician-based diagnosis of diabetes. Control individuals with an HbA1c at recruitment of >6.2% or a fasting glucose of >=7 were removed from the analysis. All individuals in this ongoing study were recruited in Tayside between 1st October 2004 and 1st July 2006. For the purposes of this analysis, we excluded all participants below the age of 35 years, leaving 3,728 cases and 3,732 controls for analysis. The Tayside Medical Ethics Committee approved the study. Informed consent was obtained from all participants.

6.4.1.8 Västerbottens type 2 diabetes case-control study

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 (World Health Organization: 1999). 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 (World Health Organization: 1999). 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 (Franks et al. 2008).

6.4.2 Genotyping and Quality Control strategy

6.4.2.1 In Cambridgeshire, EPIC, Exeter and Ashkenazi case-control studies

1536 SNPs across 84 genes involved in pancreatic ß-cell function and survival (Table 6.1) were genotyped by P. Deloukas's laboratory at the Wellcome Trust Sanger Institute using Illumina's Golden Gate assay (Fan et al. 2003; Ke et al. 2004). DNA samples from all UK populations were whole-genome amplified by OmniPlexTM at Rubicon Genomics, Inc (Ann Arbor, MI, USA) and genomic DNA was used for the Ashkenazi samples. Samples with multiple SNP failures were repeated once and then excluded from clustering if their 50% locus Gene Call (GC) score was below 70%; these were thought to be samples of poor quality DNA. Clustering was performed on a per panel basis analysing no more than 500 samples at a time and used duplicate samples (two per 96-well plate) to improve clustering. We applied a locus cut off of 0.3 and cut off value in the genotype confidence score of 0.25. On a per panel basis we applied a call rate threshold ≥80% and removed markers that displayed more than 1 discrepancy per plate. Markers departing from Hardy Weinberg equilibrium ($\chi 2 > 10$) were flagged at this point. One thousand three hundred and sixty-seven (90%) of SNPs were polymorphic (minor allele frequency (MAF) of ≥ 0.1%), and met the following QC criteria: genotype call-rates greater than 90%; no significant difference in genotype call-rate between cases and controls; a P value ≥ 0.001 for tests of deviation from Hardy-Weinberg Equilibrium in controls. For those SNPs that survived QC where discordancy between replicate samples were found these samples were blanked out prior to analysis.

6.4.2.2 In ADDITION and Tayside studies

The ADDITION and Tayside studies were genotyped using a custom TaqMan® SNP assay (Applied Biosystems, UK) at Strangeways Research Laboratory, University of Cambridge, and at the Biomedical Research Centre, University of Dundee, respectively. Allele calling was done on the ABI PRISM® 7900HT Sequence Detection System (Applied Biosystems, UK). No SNPs deviated from Hardy-Weinberg equilibrium, tested in controls (P>0.05). Genotype call rates were >99%.

6.4.2.3 In the Warren 2 study

The Warren 2 study was genotyped by Kbiosciences (Herts, U.K.) using a KASPar assay system (details of the methods used can be found at http://www.kbioscience.co.uk). No SNPs deviated from Hardy-Weinberg equilibrium, tested in controls (P>0.05). Genotype call rates were >95%.

6.4.2.4 In the Västerbottens study

Västerbottens study genomic DNA samples were diluted to 4ng/µl. SNPs were assayed using the Taqman® MGB chemistry (Applied Biosystems, Foster City, CA) as described in Chapter 2. Genotyping success and concordance rates were >98% and 100% for all SNPs, respectively.

6.4.3 Statistical analysis

Statistical analyses for the β-cell candidate gene association study were conducted using Stata v8.2. Hardy-Weinberg was assessed using the $\chi 2$ statistic (1 df). Standard log likelihood ratio tests were used to assess the contribution of individual SNPs under a log additive model (1 df) to risk of type 2 diabetes in logistic regression analysis. An admixture maximum likelihood permutation method was used to correct for multiple testing (Tyrer et al. 2006). We also used log likelihood ratio tests to assess whether statistically associated SNPs independently contributed to risk of type 2 diabetes, comparing the log likelihood of a nested model (2 df) with that of the

full model (3 df). For the pooled analysis, we used logistic regression with study as categorical covariate (equivalent to a Mantel-Haenszel model weighted by study). Heterogeneity among studies was assessed using the Q statistic. Statistical analyses for the Västerbottens study were conducted using the SAS software v9.1 (SAS Inst., Carey, NC). Hardy Weinberg Equilibrium (HWE) was assessed using the likelihood ratio test with 1 df. Conditional logistic regression models were fitted to assess the associations between each of the *WFS1* genotypes and type 2 diabetes. Models were adjusted for age, sex, and BMI. Meta-analysis of studies was performed using STATA v8.2 using a fixed effects model and inverse-variance-weighted averages of log odds ratios to obtain a combined estimate of the overall odds ratio. Between-study heterogeneity was assessed using the $\chi 2$ statistic. In all studies, linkage disequilibrium (LD), expressed as r2, was calculated using Haploview v4.0 (http://www.broad.mit.edu/mpg/haploview) and power calculations were performed using Quanto v1.1.1 (http://hydra.usc.edu/gxe).