ARTICLE

A role for coding functional variants in *HNF4A* in type 2 diabetes susceptibility

B. Jafar-Mohammadi · C. J. Groves · A. P. Gjesing · B. M. Herrera · W. Winckler ·
H. M. Stringham · A. P. Morris · T. Lauritzen · A. S. F. Doney · A. D. Morris ·
M. N. Weedon · A. J. Swift · J. Kuusisto · M. Laakso · D. Altshuler · A. T. Hattersley ·
F. S. Collins · M. Boehnke · T. Hansen · O. Pedersen · C. N. A. Palmer ·
T. M. Frayling · DIAGRAM Consortium · A. L. Gloyn · M. I. McCarthy

Abstract

Aims/hypothesis Rare mutations in the gene *HNF4A*, encoding the transcription factor hepatocyte nuclear factor 4α (HNF-4A), account for ~5% of cases of MODY and more frequent variants in this gene may be involved in multifactorial forms of diabetes. Two low-frequency, non-synonymous variants in *HNF4A* (V255M, minor allele frequency [MAF] ~0.1%; T130I, MAF ~3.0%)—known to influence downstream HNF-4A target gene expression—are of interest, but previous type 2 diabetes association reports were inconclusive. We aimed to evaluate the contribution of these variants to type 2 diabetes susceptibility through large-scale association analysis.

B. Jafar-Mohammadi · C. J. Groves · B. M. Herrera ·
A. L. Gloyn · M. I. McCarthy (✉)
Oxford Centre for Diabetes, Endocrinology and Metabolism
(OCDEM), University of Oxford, Churchill Hospital,
Old Road, Headington,
Oxford OX3 7LJ, UK
e-mail: mark.mccarthy@drl.ox.ac.uk

B. Jafar-Mohammadi · C. J. Groves · A. L. Gloyn ·
M. I. McCarthy
Oxford NIHR Biomedical Research Centre, Churchill Hospital,
Oxford, UK

A. P. Gjesing · T. Hansen · O. Pedersen
Steno Diabetes Center and Hagedorn Research Institute,
Copenhagen, Denmark

B. M. Herrera · A. P. Morris · M. I. McCarthy
Wellcome Trust Centre for Human Genetics,
University of Oxford,
Oxford, UK

W. Winckler · D. Altshuler
Broad Institute of Harvard and Massachusetts
Institute of Technology (MIT),
Cambridge, MA, USA

H. M. Stringham · M. Boehnke
Center for Statistical Genetics, Department of Biostatistics,
University of Michigan,
Ann Arbor, MI, USA

T. Lauritzen
Department of General Practice, University of Aarhus,
Aarhus, Denmark

A. S. F. Doney · A. D. Morris
Diabetes Research Group, Ninewells Hospital and Medical
School, University of Dundee,
Dundee, UK

M. N. Weedon · A. T. Hattersley · T. M. Frayling
Genetics of Complex Traits, Institute of Biomedical and Clinical
Science, Peninsula Medical School,
Exeter, UK

M. N. Weedon · A. T. Hattersley · T. M. Frayling
Diabetes Genetics, Institute of Biomedical and Clinical Science,
Peninsula Medical School,
Exeter, UK

Methods We genotyped both variants in at least 5,745 cases and 14,756 population controls from the UK and Denmark. We also undertook an expanded association analysis that included previously reported and novel genotype data obtained in Danish, Finnish, Canadian and Swedish samples. A meta-analysis incorporating all published association studies of the T130I variant was subsequently carried out in a maximum sample size of 14,279 cases and 26,835 controls.

Results We found no association between V255M and type 2 diabetes in either the initial ($p=0.28$) or the expanded analysis ($p=0.44$). However, T130I demonstrated a modest association with type 2 diabetes in the UK and Danish samples (additive per allele OR 1.17 [95% CI 1.08–1.28]; $p=1.5\times10^{-4}$), which was strengthened in the meta-analysis (OR 1.20 [95% CI 1.10–1.30]; $p=2.1\times10^{-5}$).

Conclusions/interpretation Our data are consistent with T130I as a low-frequency variant influencing type 2 diabetes risk, but are not conclusive when judged against stringent standards for genome-wide significance. This study exemplifies the difficulties encountered in association testing of low-frequency variants.

Keywords HNF4A · Low-frequency variants · T130I · Type 2 diabetes · V255M

Abbreviations

ADDITION	Anglo–Danish–Dutch study of Intensive Treatment in People with Screen-Detected Diabetes in Primary Care
DIAGRAM	Diabetes Genetics Replication and Meta-analysis Consortium
FUSION	Finland–United States Investigation of Non Insulin Dependent Diabetes Mellitus Genetics
GWAS	Genome-wide association study
HNF-4A	Hepatocyte nuclear factor 4α
LF	Low-frequency
MAF	Minor allele frequency
METSIM	Metabolic Syndrome in Men (study)
UKT2DGC	UK Type 2 Diabetes Genetics Consortium

Introduction

The heritability seen in type 2 diabetes remains largely unexplained, despite substantial progress in identifying genetic variants conferring increased risk of this condition. To date, ~40 such variants have been identified, largely through genome-wide association studies (GWAS) [1–8]. However, the sibling relative risk (λ_S) of type 2 diabetes conferred by all these variants combined is ~1.15, well below the epidemiological estimate (~3.0) [8, 9].

Of late there has been great interest in the potential role of low-frequency (LF) variants in terms of individual susceptibility to complex diseases such as type 2 diabetes. As GWAS to date have focused on the detection of common variant associations, the contribution to type 2 diabetes risk of variants with a minor allele frequency (MAF) below 5% remains largely unexplored.

A logical place to initiate the search for LF variants influencing multifactorial type 2 diabetes lies in exploring those genes already implicated in diabetes pathogenesis

A. J. Swift
National Human Genome Research Institute,
National Institutes of Health,
Bethesda, MD, USA

J. Kuusisto · M. Laakso
Department of Medicine, University of Kuopio and Kuopio
University Hospital,
Kuopio, Finland

D. Altshuler
Centre for Human Genetic Research,
Massachusetts General Hospital,
Boston, MA, USA

D. Altshuler
Department of Medicine, Harvard Medical School,
Boston, MA, USA

D. Altshuler
Department of Molecular Biology, Harvard Medical School,
Boston, MA, USA

D. Altshuler
Department of Genetics, Harvard Medical School,
Boston, MA, USA

D. Altshuler
Diabetes Unit, Massachusetts General Hospital,
Boston, MA, USA

F. S. Collins
National Institutes of Health,
Bethesda, MD, USA

T. Hansen
Faculty of Health Sciences, University of Southern Denmark,
Odense, Denmark

O. Pedersen
Faculty of Health Sciences, University of Copenhagen,
Copenhagen, Denmark

O. Pedersen
Faculty of Health Science, University of Aarhus,
Aarhus, Denmark

C. N. A. Palmer
Biomedical Research Centre, Ninewells Hospital and Medical
School, University of Dundee,
Dundee, UK

because they contain either rare mutations causal for monogenic forms of diabetes, or common variants associated with multifactorial type 2 diabetes. In genes implicated because of their role in monogenic diabetes, there is evidence that large-effect mutations are compatible with life, and that they result in a phenotype with substantial similarities to (and clinical overlap with) to type 2 diabetes. It is likely, therefore, that variants with less dramatic effects on function and/or expression, where they exist, result in less extreme clinical phenotypes including multifactorial type 2 diabetes. As existing GWAS and linkage study data argue against the possibility of common variants (MAF >5%) of medium- to large-effect size, variants with such effect sizes are also likely to be rare (MAF <0.1%) or of low frequency (MAF 0.1–5%). LF variants have been implicated in the pathogenesis of other complex diseases, such as type 1 diabetes [10], although their contribution to type 2 diabetes predisposition is as yet uncertain.

Rare, highly penetrant mutations in the gene *HNF4A*, encoding the transcription factor hepatocyte nuclear factor 4α (HNF-4A), account for approximately 5% of cases of MODY [11]. Though *HNF4A* is expressed in multiple tissues, its expression in the pancreatic beta cells and liver is of particular interest. In pancreatic beta cells, HNF-4A is required for glucose metabolism as well as normal insulin gene expression and secretion [12]. In the liver, HNF4-A is required for hepatic gluconeogenesis [13]. Several studies have shown linkage between multifactorial type 2 diabetes and the region of chr20q where *HNF4A* is located [14–17]. Previous candidate gene analyses have demonstrated weak evidence of association ($p\sim0.01$) between common variants in the P1 and P2 promoters of *HNF4A* and multifactorial type 2 diabetes [17, 18], but these have not been substantiated in GWAS to date [1–5, 8]. As common variants in *HNF4A* do not explain the findings of linkage studies, it is possible that this region harbours more penetrant LF variants that might explain this observation [19].

HNF4A has been extensively re-sequenced, not least as part of clinical diagnostic screening for MODY. These re-sequencing efforts have, inter alia, identified two LF coding non-synonymous variants of particular interest: V255M (c.763G>A p.Val255Met) and T130I (c.389C>T p.Thr130Ile, rs1800961). V255M was first described following re-sequencing of Danish samples but no evidence of association to type 2 diabetes was seen in analysis of 1,434 cases and 4,790 controls [20]. T130I, positioned in the DNA binding domain of *HNF4A*, showed modest ($p=0.04$) association with type 2 diabetes in the same sample [20], though subsequent efforts at replication failed to confirm this [21]. One arm of a meta-analysis of the association of *HNF4A* genetic variants to type 2 diabetes [22] also included some previous association studies of T130I (by our estimation including approximately 3,500 cases and 3,700 controls for this variant), and demonstrated a modest association ($p=0.045$) [22]. Most recently, and of particular interest given the relationship between lipids and type 2 diabetes, a significant association between T130I and HDL-cholesterol levels has been demonstrated ($p=8\times10^{-10}$) in a GWAS meta-analysis incorporating 30,714 individuals [23].

Both variants have been shown to be functional based on studies of the transcriptional regulation of HNF-4A target genes in a range of cell lines and primary mouse hepatocytes [20, 24–26]. We therefore reasoned that they remain interesting candidates for assessment in larger samples to more clearly establish their likely contribution to type 2 diabetes susceptibility.

Methods

Individuals studied Three categories of samples were included. Category 1 consisted of samples specifically genotyped for this study. Category 2 comprised samples with previously reported genotyping information for these single nucleotide polymorphisms. Category 3 included samples for which only summary statistics were available from previous published reports.

Category 1 samples were derived from three sources (two UK samples and one Danish sample). UK sample 1 ('UK1', n=4,124 cases, 5,126 controls) included the UK Type 2 Diabetes Genetics Consortium (UKT2DGC) collection recruited in Tayside, Scotland: these have been previously described [1, 27]. UK sample 2 ('UK2') comprised type 2 diabetes cases (n=1,853 for V255M; 1,193 for T130I) ascertained from a subset of the Diabetes UK Warren 2 repository [28]. The controls for UK2 were taken from the population-based British 1958 Birth Cohort (n=7,133), and the UK Blood Services Collection (n=3,087) [27].

Danish sample 1 ('DK1', n=2,646 cases) was also included in category 1 for the study of T130I. DK1 represents samples collected in the Steno Diabetes Centre and Danish samples from the Anglo–Danish–Dutch study of Intensive Treatment in People with Screen-Detected Diabetes in Primary Care (ADDITION) [20, 29]. The new samples in DK1 were combined with the previously reported case and control data from DK2 (described below) to generate a combined DK analysis of 3,771 cases and 4,727 controls.

Category 2 included samples from Denmark, Sweden, Finland and Canada. Danish sample 2 ('DK2', n=1,397 cases; 4,865 controls) was previously genotyped for T130I and V255M [20]. Two samples from the Finland–United States Investigation of Non Insulin Dependent Diabetes Mellitus Genetics (FUSION) study were included for T130I

(FUSION sample 1, 'FS1', [n=1,160 cases; 1,173 controls] and FUSION sample 2, 'FS2', [n=1,211 cases; 1,264 controls]) [4]. FS1 and FS2 represent the FUSION GWAS and replication samples, respectively, and have been included in a type 2 diabetes [2] and a lipid GWAS [23] and subsequent follow-up of significant findings. The numbers of individuals quoted for FS1 and FS2 differ slightly from those in the reference article as a consequence of DNA availability, the withdrawal of some individuals and the updated type 2 diabetes status of others. The recruitment criteria for these samples have been reported by Zeggini et al. [4]. We also included samples from the Metabolic Syndrome in Men (METSIM) study ('MS1', n=801 cases; 3,043 controls) recruited in Finland [30]. The T130I genotype data for MS1 were included as part of the lipid GWAS follow-up [23], though type 2 diabetes data have not been published. Previously reported genotyping results for T130I from three samples from the Broad Institute were also included [21]. These comprised a Canadian sample (Broad sample 1, 'BR1', n=127 cases; 127 controls), a combined Swedish/Finnish sample (Broad sample 2, 'BR2', n=490 cases; 490 controls) and a Swedish sample (Broad sample 3, 'BR3', n=514 cases; 514 controls).

All studies were approved by local ethics committees and were performed in accordance with the principles of the Helsinki Declaration II. Informed consent was obtained from all individuals before participation. Detailed descriptions of category 1 and 2 samples are included in the Electronic supplementary material (ESM). All participant characteristics are summarised in Table 1.

For a more complete study of T130I, the results from these category 1 and 2 samples were included in a meta-analysis together with those from all previously published studies (category 3) for which summary statistics were available. For these category 3 samples we had no access to genotype information. These included a Pima Indian sample ('PI1', n=573 cases; 464 controls) [31] and a Japanese sample ('JP1', n=423 cases; 354 controls) [24] included in the meta-analysis by Sookoian et al. [22] in addition to a Mexican sample ('MX1', n=100 cases; 75 controls) [32].

Genotyping and quality control Genotyping of T130I in the UK samples was carried out using a TaqMan assay on the ABI 7900HT platform (Applied Biosystems, Warrington, Cheshire, UK). A KBioscience allele-specific PCR (KAS-Par) assay (KBioscience, Hoddesdon, UK) was used in the genotyping of V255M in the UK samples and T130I in DK1. The quality of the genotyping was assured by: (1) assessing the genotyping pass rate (>96% globally); (2) evaluating the estimated error rate based on completed duplicate pairs (UK samples: 0.00% for V255M, n=314 duplicate pairs and 0.18% for T130I, 268 duplicate pairs; DK1: 0.20% based on 521 duplicate samples); and (3) assessing for departure (p<0.05) from Hardy–Weinberg equilibrium (none detected). Genotyping methods and quality control measures for DK2 [20], Broad samples

Table 1 Study participant characteristics

Sample name	Short form	Sample name in previous publications[a]	Participants		Male (%)		Mean age[b] (years) ± SD		Mean BMI (kg/m^2) ± SD	
			Controls	Cases	Controls	Cases	Controls	Cases	Controls	Cases
UK sample 1	UK1	UKT2DGC sample	5,126	4,124	51	55	60±13	58±12	27±5	31±6
UK sample 2	UK2	Diabetes UK Warren 2 repository, British 1958 Birth Cohort, UK Blood Services Collection of common controls	10,220	1,853	50	61	42±7	52±7	27±6[c]	32±7
Danish sample 1	DK1	Inter99, SDC1, SDC2	–	2,646	–	59	–	57±10	–	31±6
Danish sample 2	DK2	SDC2 and ADDITION	4,865	1,397	47	60	46±9	52±10	26±4	30±5
FUSION sample 1	FS1	Stage 1 genotyping sample for DIAGRAM meta-analysis	1,173	1,160	49	56	64±7	54±9	27±4	30±5
FUSION sample 2	FS2	Stage 2 genotyping sample for DIAGRAM meta-analysis	1,264	1,211	61	60	59±8	55±9	27±4	31±5
METSIM	MS1	METSIM	3,043	801	100	100	59±6	58±7	26±4	30±5
Broad sample 1	BR1	Canada sample	127	127	55	55	52±8	52±8[d]	29±4	29±5
Broad sample 2	BR2	Scandinavia sample	490	490	52	52	60±10	60±10[d]	27±4	28±5
Broad sample 3	BR3	Sweden sample	514	514	52	52	66±12	66±12[d]	28±4	28±4

[a] Full details available in the ESM

[b] Mean age refers to age at recruitment for controls and age at diagnosis for cases

[c] Based on the British 1958 Birth Cohort (n=7,133) and Panel 2 of the UK Blood Services Collection (n=1,643)

[d] Based on age at recruitment (age of diagnosis not available)

Table 2 Association testing of the T130I variant of *HNF4A*

Study sample	n cases/controls	Cases: genotype counts (genotype frequencies [%])			Controls: genotype counts (genotype frequencies [%])			Statistical analysis		
		CC	CT	TT	CC	CT	TT	OR Het (95% CI)	OR Hom (95% CI)	p_{add}
UK1	4,040/5,012	3,771 (93.34)	266 (6.59)	3 (0.07)	4,711 (94.00)	291 (5.81)	10 (0.19)	1.09 (0.92–1.29)	1.18 (0.84–1.65)	0.333
UK2	1,158[a]/9,744	1,069 (92.31)	89 (7.69)	0 (0.00)	9,175 (94.16)	563 (5.78)	6 (0.06)	1.32 (1.04–1.66)	1.74 (1.08–2.77)	0.020
DK1	2,447/0	2,240 (91.54)	200 (8.17)	7 (0.29)	–	–	–	–	–	–
DK2	1,324/4,727	1,216 (91.84)	106 (8.01)	2 (0.15)	4,412 (93.33)	305 (6.45)	10 (0.21)	1.21 (0.97–1.51)	1.47 (0.93–2.29)	0.090
DK1 and DK2	3,771/4,727	3,456 (91.65)	306 (8.11)	9 (0.24)	4,412 (93.33)	305 (6.45)	10 (0.21)	1.26 (1.07–1.47)	1.57 (1.15–2.16)	0.004
Category 1 (UK1, UK2, DK1, DK2)	7,645/14,756							1.20 (1.08–1.33)	1.44 (1.17–1.77)	5×10^{-4}
FS1	1,160/1,172	1,057 (91.12)	98 (8.44)	5 (0.43)	1,087 (92.75)	82 (7.00)	3 (0.26)	1.24 (0.93–1.66)	1.54 (0.86–2.77)	0.134
FS2	1,181/1,237	1,069 (90.52)	109 (9.23)	3 (0.25)	1,142 (92.32)	93 (7.52)	2 (0.16)	1.25 (0.94–1.67)	1.57 (0.89–2.78)	0.122
MS1	774/2,957	720 (93.02)	52 (6.72)	2 (0.26)	2,760 (93.34)	191 (6.46)	6 (0.20)	1.05 (0.77–1.43)	1.11 (0.59–2.03)	0.758
BR1	111/110	107 (96.40)	4 (3.60)	0 (0.00)	106 (96.36)	4 (3.64)	0 (0.00)	0.99 (0.18–5.46)	–	1.000
BR2	488/483	460 (94.26)	26 (5.33)	2 (0.41)	451 (93.37)	31 (6.42)	1 (0.21)	0.90 (0.54–1.51)	0.81 (0.29–2.29)	0.711
BR3	500/500	461 (92.20)	36 (7.20)	3 (0.60)	459 (91.80)	39 (7.80)	2 (0.40)	0.98 (0.63–1.52)	0.96 (0.39–2.31)	1.000
Categories 1 and 2 (UK1, UK2, DK1, DK2, FS1, FS2, MS1, BR1, BR2, BR3)	13,183/25,942							1.17 (1.08–1.28)	1.38 (1.17–1.63)	1.5×10^{-4}
Categories 1, 2 and 3 (UK1, UK2, DK1, DK2, FS1, FS2, MS1, BR1, BR2, BR3, PI1, JP1, MX1)	14,279/26,835							1.20 (1.10–1.30)		2.1×10^{-5}

n is total number successfully genotyped only

[a] Full sample set not available in Oxford at the time of genotyping ($n=1,193$ cases available for genotyping)

CC, common homozygote; CT, heterozygote; TT, rare homozygote

OR het, OR for heterozygote carriers compared with CC; OR hom, OR for rare allele homozygotes compared with CC; p_{add}, p value for additive model

[21], FUSION samples [2, 23] and MS1 [23] have been previously reported.

Statistical analysis No heterogeneity of genotype counts was seen between category 1 cases when assessed by an exact Pearson χ^2 test using StatXact (v6.0: Cytel Software Corp., Cambridge, MA, USA). The same was true of controls. We subsequently carried out a primary association analysis of category 1 samples (UK1, UK2, DK1 and DK2 for T130I, UK1 and UK2 for V255M) (each as a separate stratum) followed by a secondary association analysis including category 1 and category 2 samples (UK1, UK2, DK1, DK2, FS1, FS2, MS1, BR1, BR2 and BR3 for T130I and UK1, UK2 and DK2 for V255M) as separate strata. We used an exact Cochran–Armitage trend test (StatXact v6.0) for all association analyses in this report. The ORs and sample sizes for each stratum from this study were subsequently used in a meta-analysis of T130I incorporating the previously defined category 3 samples [22, 32] using an additive model performed with the Genome Wide Association Meta-Analysis (GWAMA) software package (www.well.ox.ac.uk/gwama). Though these samples are geographically disparate, a low level of heterogeneity of the T130I association effect sizes was detected using GWAMA ($I^2=61\%$; Q statistic p value$=0.10$; quantified by the comparison of the samples in categories 1 and 2 with category 3 samples).

Power calculations derived from QUANTO [33], using the previously reported OR for T130I [20], indicated 99% power to detect an effect size of 1.3 (for $\alpha=0.001$) for this variant in our expanded association analysis incorporating UK, Danish, FUSION, METSIM and Broad samples. The power for V255M is lower (12% power to detect the same
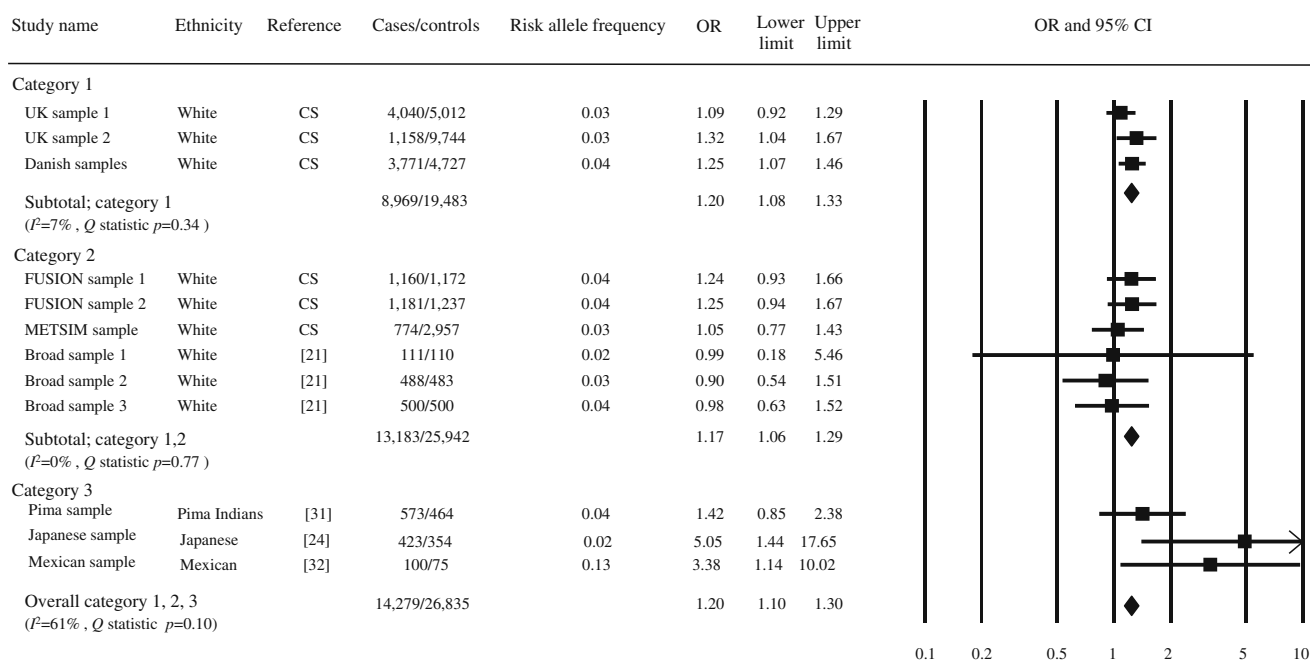
Study name	Ethnicity	Reference	Cases/controls	Risk allele frequency	OR	Lower limit	Upper limit	OR and 95% CI
Category 1								
UK sample 1	White	CS	4,040/5,012	0.03	1.09	0.92	1.29	
UK sample 2	White	CS	1,158/9,744	0.03	1.32	1.04	1.67	
Danish samples	White	CS	3,771/4,727	0.04	1.25	1.07	1.46	
Subtotal; category 1 (I^2=7% , Q statistic p=0.34)			8,969/19,483		1.20	1.08	1.33	
Category 2								
FUSION sample 1	White	CS	1,160/1,172	0.04	1.24	0.93	1.66	
FUSION sample 2	White	CS	1,181/1,237	0.04	1.25	0.94	1.67	
METSIM sample	White	CS	774/2,957	0.03	1.05	0.77	1.43	
Broad sample 1	White	[21]	111/110	0.02	0.99	0.18	5.46	
Broad sample 2	White	[21]	488/483	0.03	0.90	0.54	1.51	
Broad sample 3	White	[21]	500/500	0.04	0.98	0.63	1.52	
Subtotal; category 1,2 (I^2=0% , Q statistic p=0.77)			13,183/25,942		1.17	1.06	1.29	
Category 3								
Pima sample	Pima Indians	[31]	573/464	0.04	1.42	0.85	2.38	
Japanese sample	Japanese	[24]	423/354	0.02	5.05	1.44	17.65	
Mexican sample	Mexican	[32]	100/75	0.13	3.38	1.14	10.02	
Overall category 1, 2, 3 (I^2=61% , Q statistic p=0.10)			14,279/26,835		1.20	1.10	1.30	

0.1 0.2 0.5 1 2 5 10

Fig. 1 Meta-analysis of the association studies of the T130I variant to type 2 diabetes. The plot was generated using Comprehensive Meta Analysis software version 2.2050 (Biostat, Englewood, NJ, USA). CS, current study

effect size [for α=0.001]) because of the much lower MAF. For this variant, we had 80% power to detect an effect size of 3.0 (for α=0.001).

Results

T130I was successfully genotyped in 7,645 cases and 14,756 controls in category 1. This variant had a MAF of

3.76% in cases and 3.00% in controls. In category 1 samples, a modest association with type 2 diabetes was conferred by the T allele of this variant (additive per allele OR 1.20 [95% CI 1.08–1.33]; p=5×10^{-4}). The expanded association analysis incorporating the category 1 and 2 samples marginally increased the strength of this association (OR 1.17 [95% CI 1.08–1.28]; p=1.5×10^{-4}) (Table 2). The meta-analysis (Fig. 1) incorporating all available studies of T130I further increased the strength of this association (n=14,279 cases; 26,835 controls; OR 1.20 [95% CI 1.10–

Table 3 Association testing of the V255M variant of *HNF4A*

Study sample	n		Cases, genotype counts (genotype frequencies [%])			Controls, genotype counts (Genotype frequencies [%])			Statistical analysis	
	Cases/controls		GG	GA	AA	GG	GA	AA	OR het (95% CI)	p_{add}
UK1	3,949/4,925		3,942 (99.82)	7 (0.18)	0 (0.00)	4,911 (99.72)	14 (0.28)	0 (0.00)	0.62 (0.21–1.65)	0.38
UK2	1,796/10,119		1,794 (99.89)	2 (0.11)	0 (0.00)	10,103 (99.84)	16 (0.16)	0 (0.00)	0.76 (0.08–2.99)	0.76
Category 1 (UK1, UK2)	5,745/15,044								0.64 (0.26–1.44)	0.28
DK 2	1,360/4,791		1,357 (99.78)	3 (0.22)	0 (0.00)	4,781 (99.79)	10 (0.21)	0 (0.00)	1.05 (0.19–4.11)	1.00
Categories 1 and 2 (UK1, UK2, DK2)	7,105/19,835								0.73 (0.34–1.47)	0.44

n, total number successfully genotyped only

AA, rare homozygote, GA, heterozygote; GG, common homozygote

OR het, OR for heterozygote carriers compared with GG; p_{add}, p value for additive model

1.30]; $p=2.1\times10^{-5}$). There was no evidence for heterogeneity in our meta-analysis ($I^2=61\%$; Q statistic p value=0.10).

The V255M variant was successfully genotyped in 5,745 cases and 15,044 controls in the UK study. The MAF for V255M was far lower than for T130I (cases 0.08%; controls 0.10%), and no type 2 diabetes association was observed in either the UK sample ($p=0.28$) or the larger association analysis incorporating Danish genotyping data ($p=0.40$) (Table 3).

We did not find any evidence for linkage disequilibrium ($r^2<0.005$) between either T130I or V255M and the common variants in the promoter region of *HNF4A* that had previously shown a weak association to type 2 diabetes susceptibility.

Discussion

When genes implicated in diabetes pathogenesis undergo monogenic screening, the variants discovered tend to be put into two categories: they are either considered to be causal for monogenic diabetes or neutral 'polymorphisms'. The latter group has been assumed to have no role in disease susceptibility. With the large sample sizes now available, it is possible to go back to some of these coding variants that are clearly not causal for monogenic diabetes and re-examine whether they could, nevertheless, be influencing susceptibility to common forms of diabetes. Previous functional and association studies had highlighted two coding LF variants within *HNF4A* as interesting candidates in this respect and we have carried out the largest association analysis to date for the V255M and T130I variants of *HNF4A* to better understand their role in type 2 diabetes pathogenesis.

We found no association between the V255M variant of *HNF4A* and type 2 diabetes risk in the UK samples or in our larger analysis. It is worth emphasising that the low MAF of this variant means that our power to detect association was limited to large effect sizes only.

In contrast, evidence of association between the T130I variant of HNF4A and type 2 diabetes risk was found in our analysis of category 1 and 2 samples. The evidence for association was increased when we added category 3 samples, reaching a p of 2.1×10^{-5}. To determine whether there was any additional evidence for association available from recent large-scale genome-wide association meta-analyses for type 2 diabetes, we examined data from the recently published Diabetes Genetics Replication and Meta-analysis Consortium (DIAGRAM)+ meta-analysis [8], after excluding samples already in our meta-analysis. T130I (rs1800961) is represented on Illumina arrays but is neither present on nor can it be reliably imputed into genome-wide genotypes obtained on early Affymetrix platforms, limiting

the data available from DIAGRAM+ to 3,590 cases and 32,326 controls. In these samples, there was a directionally consistent but non-significant association with T130I ($p=0.13$) such that the combined analysis (17,869 cases and 59,197 controls) reached $p=1.0\times10^{-5}$.

However, this association fails—by some margin—to reach widely accepted thresholds for genome-wide significance which, in the context of LF variants, should be even more stringent than those required for common variants (perhaps around $\alpha=5\times10^{-9}$), given the larger number of independent tests that are possible once lower-frequency variants are considered. We estimate that to achieve this level of significance for T130I (using the effect size [an OR of 1.20] observed in our expanded meta-analysis) would require almost 100,000 samples (in fact, 48,697 cases and 48,697 controls). Recent evidence (achieving such levels of genome-wide significance) that T130I is associated with altered HDL-cholesterol levels raises the prior odds that the type 2 diabetes association we observe here is genuine, as does the strong biological candidacy of this gene given its proven causal role in monogenic forms of diabetes. The previously reported association with lipid levels also raises an interesting question as to whether or not the type 2 diabetes association is mediated by a direct influence of the variant (or a causal variant with which it is in linkage disequilibrium) on beta cell function, or a primary effect on lipid physiology. The latter question could, in principle, be answered by a suitably scaled Mendelian randomisation experiment. The broad transcriptional effects of HNF-4A would be consistent with pleiotropic effects of the variant on multiple systems.

As efforts are increasingly aimed at understanding the full allelic spectrum of variants involved in multifactorial disease pathogenesis, large-scale genotyping will be required to clarify the role played by LF variants. As our results exemplify, the numbers required for association testing of such variants are substantial when the effect size is modest. It is clear that large collaborative efforts will be needed to maximise the samples available for any such studies.

References

1. Zeggini E, Weedon MN, Lindgren CM et al (2007) Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. Science 316:1336–1341

2. Scott LJ, Mohlke KL, Bonnycastle LL et al (2007) A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. Science 316:1341–1345

3. Saxena R, Voight BF, Lyssenko V et al (2007) Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. Science 316:1331–1336

4. Zeggini E, Scott LJ, Saxena R et al (2008) Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. Nat Genet 40:638–645

5. Dupuis J, Langenberg C, Prokopenko I et al (2010) New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. Nat Genet 42:105–116

6. Saxena R, Hivert MF, Langenberg C et al (2010) Genetic variation in GIPR influences the glucose and insulin responses to an oral glucose challenge. Nat Genet 42:142–148

7. Kong A, Steinthorsdottir V, Masson G et al (2009) Parental origin of sequence variants associated with complex diseases. Nature 462:868–874

8. Voight BF, Scott LJ, Steinthorsdottir V et al (2010) Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. Nat Genet 42:579–589

9. Lango H, Palmer CN, Morris AD et al (2008) Assessing the combined impact of 18 common genetic variants of modest effect sizes on type 2 diabetes risk. Diabetes 57:3129–3135

10. Nejentsev S, Walker N, Riches D, Egholm M, Todd JA (2009) Rare variants of IFIH1, a gene implicated in antiviral responses, protect against type 1 diabetes. Science 324:387–389

11. Ellard S, Bellanne-Chantelot C, Hattersley AT (2008) Best practice guidelines for the molecular genetic diagnosis of maturity-onset diabetes of the young. Diabetologia 51:546–553

12. Bartoov-Shifman R, Hertz R, Wang H, Wollheim CB, Bar-Tana J, Walker MD (2002) Activation of the insulin gene promoter through a direct effect of hepatocyte nuclear factor 4α. J Biol Chem 277:25914–25919

13. Rhee J, Inoue Y, Yoon JC et al (2003) Regulation of hepatic fasting response by PPARγ coactivator-1α (PGC-1): requirement for hepatocyte nuclear factor 4α in gluconeogenesis. Proc Natl Acad Sci USA 100:4012–4017

14. Weedon MN, Owen KR, Shields B et al (2004) Common variants of the hepatocyte nuclear factor-4α P2 promoter are associated with type 2 diabetes in the U.K. population. Diabetes 53:3002–3006

15. Ji L, Malecki M, Warram JH, Yang Y, Rich SS, Krolewski AS (1997) New susceptibility locus for NIDDM is localized to human chromosome 20q. Diabetes 46:876–881

16. Zouali H, Hani EH, Philippi A et al (1997) A susceptibility locus for early-onset non-insulin dependent (type 2) diabetes mellitus maps to chromosome 20q, proximal to the phosphoenolpyruvate carboxykinase gene. Hum Mol Genet 6:1401–1408

17. Love-Gregory LD, Wasson J, Ma J et al (2004) A common polymorphism in the upstream promoter region of the hepatocyte nuclear factor-4α gene on chromosome 20q is associated with type 2 diabetes and appears to contribute to the evidence for linkage in an Ashkenazi Jewish population. Diabetes 53:1134–1140

18. Silander K, Mohlke KL, Scott LJ et al (2004) Genetic variation near the hepatocyte nuclear factor-4α gene predicts susceptibility to type 2 diabetes. Diabetes 53:1141–1149

19. Lillioja S, Wilton A (2009) Agreement among type 2 diabetes linkage studies but a poor correlation with results from genome-wide association studies. Diabetologia 52:1061–1074

20. Ek J, Rose CS, Jensen DP et al (2005) The functional Thr130Ile and Val255Met polymorphisms of the hepatocyte nuclear factor-4α (HNF4A): gene associations with type 2 diabetes or altered beta-cell function among Danes. J Clin Endocrinol Metab 90:3054–3059

21. Winckler W, Graham RR, de Bakker PI et al (2005) Association testing of variants in the hepatocyte nuclear factor 4α gene with risk of type 2 diabetes in 7,883 people. Diabetes 54:886–892

22. Sookoian S, Gemma C, Pirola CJ (2010) Influence of hepatocyte nuclear factor 4α (HNF4α) gene variants on the risk of type 2 diabetes: a meta-analysis in 49,577 individuals. Mol Genet Metab 99:80–89

23. Kathiresan S, Willer CJ, Peloso GM et al (2009) Common variants at 30 loci contribute to polygenic dyslipidemia. Nat Genet 41:56–65

24. Zhu Q, Yamagata K, Miura A et al (2003) T130I mutation in HNF-4α gene is a loss-of-function mutation in hepatocytes and is associated with late-onset type 2 diabetes mellitus in Japanese subjects. Diabetologia 46:567–573

25. Moller AM, Urhammer SA, Dalgaard LT et al (1997) Studies of the genetic variability of the coding region of the hepatocyte nuclear factor-4α in Caucasians with maturity onset NIDDM. Diabetologia 40:980–983

26. Navas MA, Munoz-Elias EJ, Kim J, Shih D, Stoffel M (1999) Functional characterization of the MODY1 gene mutations HNF4 (R127W), HNF4(V255M), and HNF4(E276Q). Diabetes 48:1459–1465

27. Wellcome Trust Case Control Consortium (2007) Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 447:661–678

28. Wiltshire S, Hattersley AT, Hitman GA et al (2001) A genome-wide scan for loci predisposing to type 2 diabetes in a U.K. population (the Diabetes UK Warren 2 Repository): analysis of 573 pedigrees provides independent replication of a susceptibility locus on chromosome 1q. Am J Hum Genet 69:553–569

29. Lauritzen T, Griffin S, Borch-Johnsen K, Wareham NJ, Wolfenbuttel BH, Rutten G (2000) 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–S11

30. Stancakova A, Javorsky M, Kuulasmaa T, Haffner SM, Kuusisto J, Laakso M (2009) Changes in insulin sensitivity and insulin release in relation to glycemia and glucose tolerance in 6,414 Finnish men. Diabetes 58:1212–1221

31. Muller YL, Infante AM, Hanson RL et al (2005) Variants in hepatocyte nuclear factor 4α are modestly associated with type 2 diabetes in Pima Indians. Diabetes 54:3035–3039

32. Menjivar M, Granados-Silvestre MA, Montufar-Robles I et al (2008) High frequency of T130I mutation of HNF4A gene in Mexican patients with early-onset type 2 diabetes. Clin Genet 73:185–187

33. Gauderman W, Morrison J (2006) QUANTO 1.1: A computer program for power and sample size calculations for genetic-epidemiology studies. Available from http://hydra.usc.edu/gxe, accessed 15 August 2008