

## Association testing of common variants in the insulin receptor substrate-1 gene (*IRS1*) with type 2 diabetes

J. C. Florez · M. Sjögren · C. M. Agapakis ·  
N. P. Burt · P. Almgren · U. Lindblad · G. Berglund ·  
T. Tuomi · D. Gaudet · M. J. Daly · K. G. Ardlie ·  
J. N. Hirschhorn · D. Altshuler · L. Groop

Received: 30 October 2006 / Accepted: 8 February 2007 / Published online: 19 April 2007  
© Springer-Verlag 2007

### Abstract

**Aims/hypothesis** Activation of the insulin receptor substrate-1 (*IRS1*) is a key initial step in the insulin signalling pathway. Despite several reports of association of the G972R polymorphism in its gene *IRS1* with type 2 diabetes, we and others

have not observed this association in well-powered samples. However, other nearby variants might account for the putative association signal.

**Subjects and methods** We characterised the haplotype map of *IRS1* and selected 20 markers designed to capture common variations in the region. We genotyped this comprehensive set of markers in several family-based and case-control samples of European descent totalling 12,129 subjects.

D. Altshuler and L. Groop jointly supervised this project.

**Electronic supplementary material** The online version of this article (doi:10.1007/s00125-007-0657-5) contains supplementary material, which is available to authorised users.

J. C. Florez (✉) · C. M. Agapakis · M. J. Daly · D. Altshuler  
Simches Research Building-CPZN 6820,  
Diabetes Unit/Center for Human Genetic Research,  
Massachusetts General Hospital,  
Boston, MA 02114, USA  
e-mail: jcflorez@partners.org

T. Tuomi · L. Groop  
Department of Medicine, Helsinki University Central Hospital,  
Helsinki, Finland

T. Tuomi  
Research Program for Molecular Medicine,  
University of Helsinki,  
Helsinki, Finland

J. C. Florez · M. J. Daly · D. Altshuler  
Department of Medicine, Massachusetts General Hospital,  
Boston, MA, USA

T. Tuomi  
Folkhälsan Genetic Institute, Folkhälsan Research Center,  
Helsinki, Finland

J. C. Florez · D. Altshuler  
Department of Medicine, Harvard Medical School,  
Boston, MA, USA

D. Gaudet  
University of Montreal Community Genomic Center,  
Chicoutimi Hospital,  
Chicoutimi, QC, Canada

J. C. Florez · C. M. Agapakis · N. P. Burt · M. J. Daly ·  
J. N. Hirschhorn · D. Altshuler  
Program in Medical and Population Genetics,  
Broad Institute of Harvard and MIT,  
Cambridge, MA, USA

K. G. Ardlie  
Genomics Collaborative Division, SeraCare LifeSciences,  
Cambridge, MA, USA

M. Sjögren · P. Almgren · L. Groop  
Department of Endocrinology, University Hospital MAS,  
Lund University,  
Malmö, Sweden

J. N. Hirschhorn · D. Altshuler  
Department of Genetics, Harvard Medical School,  
Boston, MA, USA

U. Lindblad · G. Berglund  
Department of Clinical Science, University Hospital MAS,  
Lund University,  
Malmö, Sweden

J. N. Hirschhorn  
Divisions of Genetics and Endocrinology, Children's Hospital,  
Boston, MA, USA

**Results** In an initial sample of 2,235 North American and Polish case-control pairs, the minor allele of the rs934167 polymorphism showed nominal evidence of association with type 2 diabetes (odds ratio [OR] 1.25, 95% CI 1.03–1.51,  $p=0.03$ ). This association showed a trend in the same direction in 7,659 Scandinavian samples (OR 1.16, 95% CI 0.96–1.39,  $p=0.059$ ). The combined OR was 1.20 ( $p=0.008$ ), but statistical correction for the number of variants examined yielded a  $p$  value of 0.086. We detected no differences across rs934167 genotypes in insulin-related quantitative traits.

**Conclusions/interpretation** Our data do not support an association of common variants in *IRS1* with type 2 diabetes in populations of European descent.

**Keywords** Common variants · Genetic association · Insulin resistance · *IRS1* · Linkage disequilibrium · Single nucleotide polymorphism · SNP · Type 2 diabetes

### Abbreviations

CEU	Caucasian samples from the HapMap project
GRR	genotypic relative risk
HOMA <sub>IR</sub>	insulin resistance by homeostasis model assessment
LD	linkage disequilibrium
MAF	minor allele frequency
OR	odds ratio
SNP	single nucleotide polymorphism
UTR	untranslated region

### Introduction

The insulin receptor substrate-1 (*IRS1*) is an early mediator in the insulin receptor signal transduction pathway [1]. Tyrosine phosphorylation of IRS proteins causes them to bind phosphatidylinositol 3-kinase, which in turn initiates a phosphorylation cascade resulting in various downstream effects of insulin [2]. Tissue-specific knockout experiments in mice have shown that *IRS1* is a necessary component of insulin action in skeletal muscle, adipose tissue and pancreatic beta cells [3]. Its gene *IRS1* has therefore been proposed as a candidate gene that might cause diabetes in humans [4].

#### Present address:

C. M. Agapakis  
Biological and Biomedical Sciences Program, Harvard University,  
Boston, MA, USA

#### Present address:

K. G. Ardlie  
Biological Samples Platform, Broad Institute of Harvard and MIT,  
Cambridge, MA, USA

Indeed, the common missense variant glycine → arginine at codon 972 (G972R) [5] was associated with type 2 diabetes in a meta-analysis of 27 studies comprising 8,827 subjects, although the statistical significance of this result was modest [6]. To test the reproducibility of this association, we recently attempted to replicate the same genetic model in large samples totalling 9,000 white individuals. Despite >95% power to obtain a  $p<0.05$  for the odds ratio (OR) of 1.25 that had been estimated in the meta-analysis, we did not observe an association of G972R with type 2 diabetes, related traits or age of onset [7]. A simultaneous independent report in over 2,000 samples also failed to confirm the association [8].

Other variants in *IRS1* may affect gene expression or protein function and thereby induce insulin resistance; in such a scenario, varying degrees of linkage disequilibrium (LD) in different populations may have given rise to discrepant association signals. In order to assess whether common genetic variation in *IRS1* increases risk of type 2 diabetes, we set out to characterise the haplotype structure of *IRS1*, select a set of markers that capture common variation in the region and genotype them in several large family-based and case-control samples of northern European ancestry.

### Subjects and methods

**Clinical samples** To maximise genotyping efficiency, we performed this association study in two stages. In stage 1, we genotyped 20 tag single nucleotide polymorphisms (SNPs) in two case-control diabetes samples of European descent obtained from Genomics Collaborative (Cambridge, MA, USA). One sample comprised 1,226 case-control pairs from the USA and one comprised 1,009 case-control pairs from Poland, both matched for age, sex and grandparental country of origin. A subset of these samples has been formally examined for the presence of population stratification [9]. For stage 2, we selected the only nominally significant result obtained in stage 1 ( $p<0.05$ ) and genotyped it in Scandinavian samples from the Botnia Study [10], which include 211 trios, 874 siblings discordant for type 2 diabetes and a case-control sample totalling 755 subjects matched for age, BMI and region of origin. Cases with severe impaired glucose tolerance (studied in previous reports from our group [7, 11–15]) were excluded. In addition, we studied: (1) an individually matched case-control sample totalling 254 subjects from the Saguenay Lac-St Jean region in Quebec (Canada); (2) a case-control sample from Sweden totalling 948 subjects matched for sex, age and BMI; and (3) an additional Scandinavian case-control sample including 1,999 cases with type 2 diabetes (from a southern Sweden Diabetes Registry [16]) and 2,260 unrelated ethnically matched

control subjects from the Malmö Diet and Cancer Study (<http://www.mdcs.mas.lu.se/>), who had fasting blood glucose <5.6 mmol/l and no known family history of type 2 diabetes. These samples were validated by replication of the three most widely reproduced associations in type 2 diabetes: (1) the P12A variant in the peroxisome proliferator-activated receptor  $\gamma$  (*PPARG*) [11], (2) the E23K variant in the islet ATP-sensitive potassium inwardly-rectifying channel (*KCNJ11*) [7, 12] and (3) common variants in the transcription factor seven-like 2 gene (*TCF7L2*) [17]. All subjects gave informed consent. Appropriate institutional review board approvals were obtained and all investigations were carried out according to the Declaration of Helsinki. The phenotypic characteristics of all patient sub-samples are presented in Table 1.

**Genotyping** Genotyping was generally performed by allele-specific primer extension of multiplex products with detection by matrix-assisted laser desorption ionisation-time of flight mass spectroscopy on the Sequenom platform [18]. The Scandinavian case-control sample was genotyped by the allelic discrimination method on an ABI7900 machine (Applied Biosystems, Foster City, CA, USA). Average genotyping success (taking into account all SNPs

and all samples) was 96.9% and our consensus rate was 99.98%, based on 8,940 duplicate genotypes.

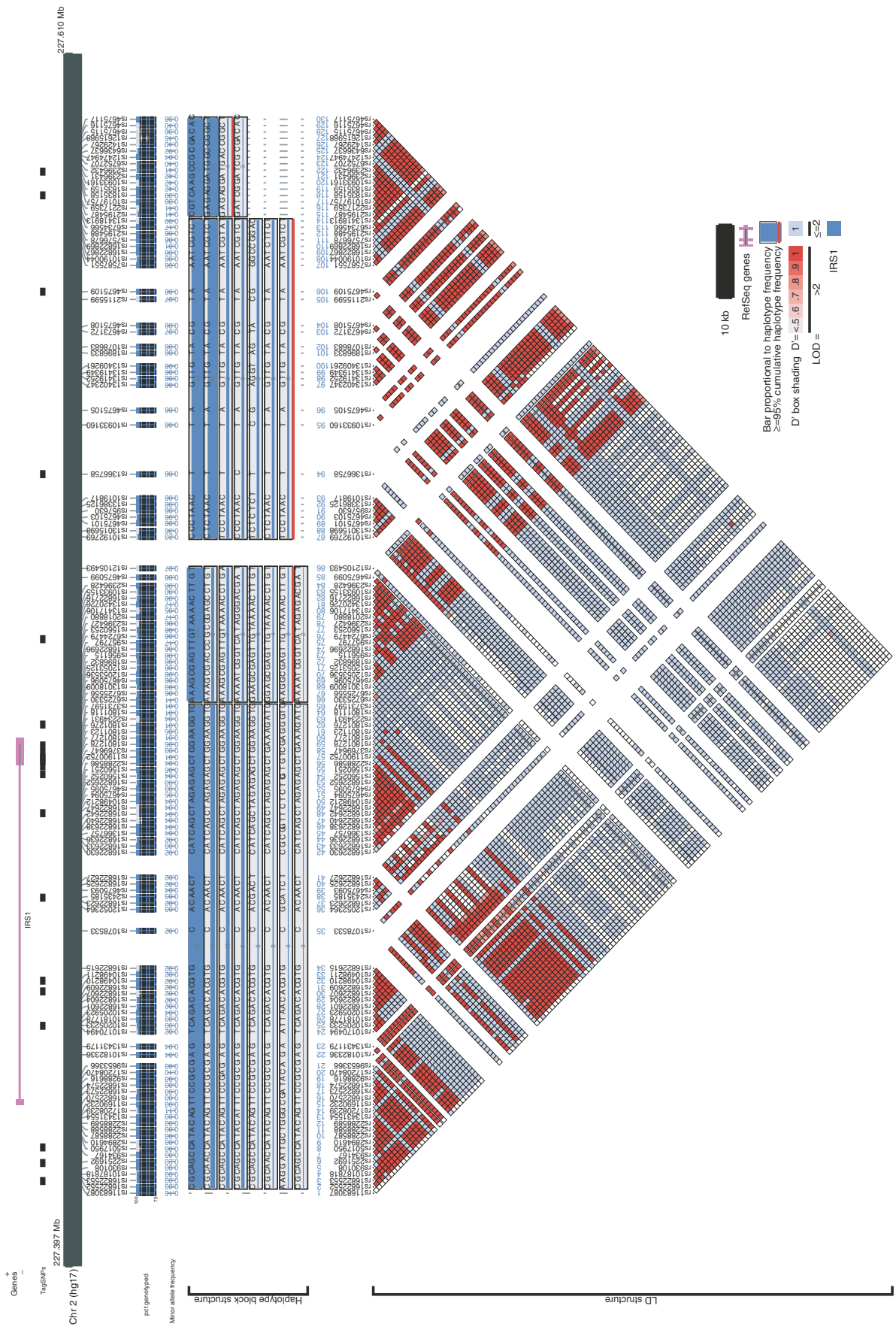
**Haplotype structure** To evaluate the haplotype structure of the *IRS1* gene, we first downloaded data for the CEU (Caucasian) samples from phase 1 of the HapMap project [19]. We targeted a segment that would begin ~20 kb upstream of the *IRS1* transcription start site and end ~10 kb downstream from the end of the 3' untranslated region (UTR), expanding this region in both directions until we noted decay of LD as defined by the end of a haplotype block [20]. Additional SNPs were genotyped in the HapMap CEU plate and integrated into the map to refine areas of low SNP density or clarify the extent of LD. Results were updated when genotypes from phase 2 of the HapMap became available.

**Tag SNP selection** Tag SNPs were selected with Tagger [21] (<http://www.broad.mit.edu/mpg/tagger/>), based on data from phase 1 of the HapMap, complemented by limited additional genotyping in the HapMap CEU plate. By setting a threshold  $r^2 \geq 0.8$ , forcing in G972R (rs1801278) as a tag and using an 'aggressive' strategy as implemented in Tagger, we obtained 20 single-marker tags only (equivalent

**Table 1** Characteristics of patient samples

Population	n (male/female)	Age (years)	BMI (kg/m <sup>2</sup> )	Fasting PG (mmol/l)	HbA <sub>1c</sub> (%)	PG at 2 h OGTT (mmol/l)
USA case-control						
Type 2 diabetes	644/582	63±11	33±7	9.8±3.0	8.0±3.1	nd
Normal glucose tolerance	644/582	61±10	27±5	5.1±0.9	nd	nd
Poland case-control						
Type 2 diabetes	422/587	62±10	30±5	8.9±4.0	7.9±1.3	nd
Normal glucose tolerance	422/587	59±7	26±4	4.8±1.2	nd	nd
Botnia trios						
Probands	99/112	40±9	28±5	7.5±3.0	nd	9.8±2.9
Parents	185/186					
Botnia sibships						
Type 2 diabetes	212/247	65±10	28±7	9.8±3.3	nd	15.5±5.8
Normal glucose tolerance	167/235	63±10	27±3	5.4±0.4	nd	6.1±1.1
Botnia case-control						
Type 2 diabetes	220/182	61±10	29±5	9.8±3.1	nd	15.8±5.2
Normal glucose tolerance	187/166	60±10	26±4	5.4±0.5	nd	5.8±1.2
Canada case-control						
Type 2 diabetes	70/57	53±8	29±5	6.4±1.8	6.5±1.9	12.8±2.1
Normal glucose tolerance	70/57	52±8	29±4	5.1±0.6	5.1±0.6	6.1±1.1
Sweden case-control						
Type 2 diabetes	252/228	67±11	28±4	10.1±3.0	6.7±1.5	nd
Normal glucose tolerance	244/224	66±12	27±4	5.3±0.5	nd	nd
Scandinavia case-control						
Type 2 diabetes	1,008/991	61±11	30±6	12.5±4.1	7.8±2.1	nd
Normal glucose tolerance	1,009/1,251	59±5	25±4	4.8±0.4	4.8±0.4	nd

Data are presented as mean±SD. Plasma glucose (PG) was measured at baseline (fasting) and 2 h after an OGTT. nd, not determined



**Fig. 1** Linkage disequilibrium (*LD*) plot across the *IRS1* gene region. The horizontal black line depicts the chromosomal segment analysed in the HapMap CEU sample and the purple line indicates the *IRS1* gene (5' to 3' is right to left). The tag SNP locations are indicated by hatch marks above the black line. The LD plot in the bottom part of the figure is based on the measure *D'*: each diamond represents the magnitude of LD for a single pair of markers, with red indicating that LD is strong (*D'* >0.8) and statistically significant (LOD >2.0). Due to the high number of consecutive rare SNPs in the first exon of *IRS1*, there were not enough observations to merit the statistical definition of haplotype blocks according to Gabriel et al. [20]; however, the 'spine

of LD' definition as implemented in Haploview revealed two large haplotype blocks, which begin 93.6 kb upstream of the *IRS1* transcription start site and end 12.4 kb downstream of the 3' UTR respectively (a third smaller haplotype block is noted further upstream). These haplotypes are shown by the blue line above the LD plot, with the thickness of the blue line indicating their frequency in the CEU reference sample. pct genotyped, per cent genotyped. Figure prepared using LocusView version 2.0 (T. Petryshen, A. Kirby, M. Ainscow, Broad Institute of Harvard and MIT, Cambridge, MA, USA; unpublished software)

to the 'pairwise' approach), which were carried forward in the disease samples.

**Statistical analysis** To examine the association of alleles with type 2 diabetes we used simple  $\chi^2$  analysis in the case-control samples, the transmission disequilibrium test [22] in the diabetic trios and the discordant allele test [23] in the sib pairs; the first two were implemented in Haploview (<http://www.broad.mit.edu/mpg/haploview/>) [24]. Results from the various samples were combined by Mantel–Haenszel meta-analysis of the odds ratios [25]. Homogeneity of ORs among study samples was tested using an asymptotic Breslow–Day statistic [26]. Adjustment for covariates was performed using the software program whap (<http://pngu.mgh.harvard.edu/purcell/whap/>) and the test

for genotype-BMI interaction was performed with PLINK (<http://pngu.mgh.harvard.edu/~purcell/plink/>).

In order to correct for the multiple variants examined, we performed 10,000 permutations in the US case-control sample, obtaining an empiric *p* value based on the number of times the best  $\chi^2$  statistic was exceeded. The ratio of this empiric *p* value to the best nominal *p* value is an indicator of the statistical correction needed to account for the number of markers tested; this correction factor (10.95 $\times$ ) is lower than a pure Bonferroni correction (20 $\times$ ) due to the correlation among variants brought about by LD in the region. The correction factor was applied to the overall meta-analysis *p* value and used in our power calculations, which were performed with the program of Purcell et al. [27], available at <http://pngu.mgh.harvard.edu/~purcell/gpc/>.

**Table 2** Association study of *IRS1* SNPs (stage 1)

SNPs	M/m	MAF	USA ( <i>n</i> =1,226 C/C pairs)			Poland ( <i>n</i> =1,009 C/C pairs)			Overall ( <i>n</i> =2,235 C/C pairs)		
			OR	95% CI	<i>p</i> value	OR	95% CI	<i>p</i> value	OR	95% CI	<i>p</i> value
rs1835159	A/G	0.41	1.09	0.98–1.23	0.12	1.01	0.89–1.14	0.90	1.05	0.97–1.15	0.22
rs2217359	G/A	0.41	1.10	0.98–1.24	0.09	1.01	0.89–1.14	0.93	1.06	0.97–1.15	0.19
rs4675109	A/G	0.09	1.19	0.97–1.45	0.09	0.95	0.78–1.17	0.63	1.06	0.92–1.23	0.39
rs1366758	T/C	0.08	0.89	0.72–1.10	0.29	0.89	0.74–1.07	0.23	0.89	0.78–1.03	0.11
rs2018880	A/G	0.17	1.09	0.94–1.27	0.26	0.96	0.82–1.14	0.66	1.03	0.92–1.15	0.59
rs6725330	A/G	0.12	1.04	0.87–1.24	0.65	1.21	0.96–1.54	0.11	1.10	0.96–1.27	0.19
rs3731597	G/A	0.01	0.86	0.50–1.47	0.57	1.20	0.61–2.33	0.60	0.98	0.64–1.48	0.91
rs1801276	G/C	0.02	1.02	0.70–1.49	0.91	0.88	0.63–1.22	0.45	0.94	0.73–1.20	0.62
rs1801123	A/G	0.10	1.00	0.83–1.21	0.98	1.13	0.89–1.43	0.32	1.05	0.91–1.22	0.52
rs1801278	G/A	0.06	0.79	0.63–1.00	0.05	1.11	0.85–1.46	0.44	0.92	0.77–1.09	0.33
rs3769647	G/C	0.06	1.08	0.85–1.35	0.53	1.19	0.89–1.58	0.23	1.12	0.94–1.34	0.22
rs6754575	T/C	0.03	0.70	0.49–0.99	0.04	1.13	0.74–1.74	0.57	0.85	0.65–1.11	0.23
rs10498212	A/T	0.07	1.05	0.85–1.31	0.63	1.26	0.96–1.66	0.09	1.13	0.95–1.34	0.16
rs2435185	A/G	0.02	1.19	0.79–1.80	0.42	1.27	0.76–2.10	0.36	1.22	0.88–1.68	0.23
rs10498211	T/A	0.02	0.74	0.51–1.06	0.10	1.10	0.72–1.67	0.67	0.87	0.66–1.15	0.34
rs10498210	G/A	0.03	0.72	0.51–1.02	0.06	1.07	0.71–1.60	0.75	0.85	0.65–1.11	0.22
rs10170494	T/A	0.05	1.16	0.90–1.49	0.26	1.17	0.85–1.63	0.34	1.16	0.95–1.42	0.14
rs934167	C/T	0.06	1.22	0.96–1.56	0.11	1.29	0.94–1.77	0.12	1.25	1.03–1.51	0.03
rs2251692	G/A	0.26	1.03	0.91–1.18	0.60	0.88	0.76–1.01	0.07	0.96	0.87–1.06	0.43
rs11683087	A/G	0.19	0.93	0.81–1.07	0.32	0.98	0.84–1.15	0.80	0.95	0.86–1.06	0.37

Twenty single nucleotide polymorphisms (*SNPs*) that capture common genetic variation in *IRS1* were genotyped in two white case-control (C/C) samples obtained from Genomics Collaborative. Overall results were combined by Mantel–Haenszel meta-analysis of the odds ratios (OR), expressed as the minor allele (*m*) vs the major allele (*M*). MAF, minor allele frequency in the US sample

We note that although we chose a two-stage strategy for the purposes of SNP selection and genotyping efficiency, the SNP promoted to the second stage was analysed in all samples jointly; we therefore applied a statistical correction for the total number of variants tested at the joint analysis stage.

**Quantitative trait comparisons** A 75-g OGTT with insulin and glucose measurements at 0, 30, 60 and 120 min was performed in a subset of the control Botnian subjects ( $n=850$ , 415 female). Plasma glucose was measured by a glucose oxidase method on a glucose analyser (Beckman Instruments, Fullerton, CA, USA) and fasting insulin was measured by radioimmunoassay. The insulinogenic index was calculated as: [(insulin at 30 min) – (insulin at 0 min)]/[(glucose at 30 min) – (glucose at 0 min)] [28]; the whole-body insulin sensitivity index was calculated as:  $10,000/\sqrt{[(\text{mean OGTT glucose}) \times (\text{mean OGTT insulin}) \times (\text{fasting glucose}) \times (\text{fasting insulin})]}$  [29]; insulin resistance by homeostasis model assessment ( $\text{HOMA}_{\text{IR}}$ ) was calculated as:  $(\text{fasting serum insulin} \times \text{fasting plasma glucose})/22.5$  [30]; and insulin AUC was calculated by the trapezoidal method as:  $(\text{value at 30 min} \times 30) + (\text{value at 60 min} \times 45) + (\text{value at 120 min} \times 30) - (\text{value at 0 min} \times 105)$ . Values were log-transformed to approximate normality and compared by *t* test between major allele homozygotes at rs934167 and minor allele heterozygotes (no minor allele homozygotes were detected in this subset).

## Results

**Characterisation of common sequence variation at *IRS1*** The 130 polymorphic SNPs compiled by us span 183 kb and contain the entire *IRS1* gene, from ~105 kb upstream of the transcription start site to ~13 kb downstream of the end of the 3' UTR. The average spacing between SNPs is 1.4 kb and the maximum gap 8.6 kb. There is significant LD in the region, which results in limited haplotype diversity and a sizeable efficiency gain when attempting to capture common genetic variation (Fig. 1).

Due to the temporal sequence of data availability, we used the haplotype plot constructed with genotypes from phase 1 of the HapMap (supplemented with additional genotyping in the CEU plate) to select our tag SNPs. After genotyping was completed in the disease samples, phase 2 of the HapMap became available; we therefore evaluated our original set of tags against the updated dataset. We observed that our 20 tag SNPs capture 85% of all common variants (minor allele frequency [MAF]  $\geq 0.05$ ) with  $r^2 \geq 0.8$  and 94% of all such variants with  $r^2 \geq 0.4$  (see Electronic supplementary material [ESM] Table 1).

**Power calculations** We established a nominal *p* value of 0.05 at stage 1 as the threshold for SNPs to be carried forward to stage 2. Assuming a type 2 diabetes prevalence of 8%, our initial (stage 1) sample of 2,235 US and Polish

**Table 3** Genotype counts for rs934167 in stage 2 (Scandinavian and Canadian samples)

					MAF OR (95% CI)		<i>p</i> value
Family-based	Transmitted	Untransmitted	Excess in affected	Excess in unaffected			
Trios	17	20			0.05	0.85 (0.45–1.62)	0.57
Discordant sibs			13	5	0.03	2.6 (0.96–7.02)	0.06
Case-control subjects	C/C		C/T	T/T			
Botnia							
Cases	380		21	1	0.03		
Controls	322		30	0	0.04	0.66 (0.38–1.15)	0.14
Sweden							
Cases	466		34	0	0.03		
Controls	476		30	0	0.03	1.13 (0.65–1.89)	0.65
Canada							
Cases	110		16	0	0.06		
Controls	114		11	1	0.05	1.25 (0.59–2.64)	0.57
Scandinavia							
Cases	1,771		178	6	0.05		
Controls	1,265		101	2	0.04	1.28 (1.00–1.63)	0.048
Overall						1.16 (0.96–1.39)	0.059

The transmission disequilibrium test and the discordant allele test were performed in the Botnia trios and the discordant sibpairs, respectively. Odds ratios (OR, with 95% confidence intervals) were calculated for the minor allele (C) versus the major allele (T) and pooled by Mantel–Haenszel analysis. MAF, minor allele frequency. *p* values are two-sided and uncorrected, except for the overall *p* value which is one-sided

**Table 4** Genotype-phenotype correlations according to genotype at rs934167

Quantitative trait	rs934167 genotype		
	C/C ( <i>n</i> =779)	C/T ( <i>n</i> =71)	<i>p</i> value ( <i>t</i> test)
Body mass index (kg/m <sup>2</sup> )	26.3±3.7	27.3±4.0	0.03
Fasting insulin (pmol/l)	56.9±34.0	61.1±38.9	0.75
Insulinogenic index ([pmol/l]/[mmol/l])	151.4±242.4	136.1±135.4	0.69
2-h insulin (pmol/l)	307.0±243.1	350.0±312.6	0.36
Insulin AUC (nmol/l)	35.4±22.9	39.6±26.4	0.11
HOMA <sub>IR</sub> (pmol/l × mmol/l)	13.9±9.0	16.0±10.4	0.45
Whole-body ISI ([pmol/l×mmol/l] <sup>-1</sup> )	14.1±8.4	14.1±15.5	0.18

Untransformed data are presented as mean±SD. We calculated various parameters related to insulin secretion and sensitivity as described, log-transformed them to improve normality and compared them by simple two-tailed *t* tests across rs934167 genotypes in 850 Scandinavian non-diabetic subjects for whom we had OGTT data (no minor allele homozygotes were detected in this subset). *n* = number of individuals. Insulin AUC was calculated during an OGTT; ISI, insulin sensitivity index calculated from the OGTT. *p* values are uncorrected for the number of hypotheses examined. To convert glucose mmol/l to mg/dl, multiply by 18.01; to convert insulin pmol/l to mU/ml, divide by 6.945.

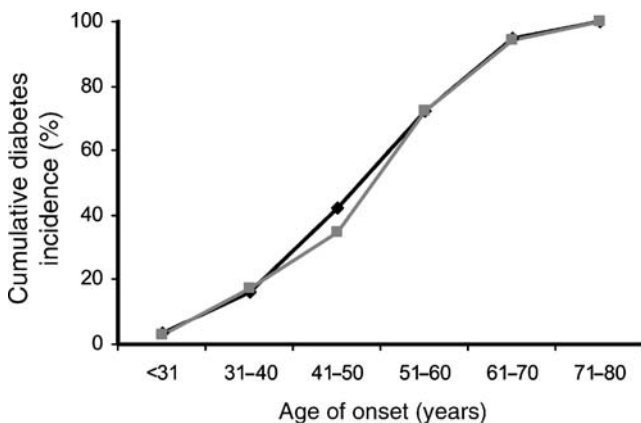
case-control pairs had >86% power to detect associations of SNPs with MAF ≥5% and a genotypic relative risk (GRR) ≥1.3 under an additive model; power drops to ~56% for a GRR of 1.2. Due to the statistical correction necessary to account for the number of markers tested (see **Methods**), we required a nominal *p* value <0.0046 at the joint analysis stage to declare a statistically significant association; our combined sample (US, Poland and Scandinavia) had ~73% power to detect associations of SNPs with MAF ≥5% and GRR ≥1.2 under an additive model, achieving >98% power for GRR ≥1.3.

For our quantitative trait comparisons, and taking fasting insulin as an example, our sample of 850 control subjects had 99.8% power to detect a 1-SD increase in fasting insulin caused by an allele of MAF 4% (e.g. rs934167),

assuming genotype affects 5% of the variance; if genotype only affects 1% of the variance, power would be reduced to 42.6%.

**Association study** The results of stage 1 of our association study are presented in Table 2; genotype counts for this stage are shown in ESM Table 2. After a combined meta-analysis of the US and Polish samples, only one SNP (rs934167) showed nominal evidence of association (OR 1.25, 95% CI 1.03–1.51, *p*=0.03). We therefore took rs934167 forward to stage 2: in a meta-analysis of the subsamples studied at this stage, rs934167 showed a trend in the same direction and approached nominal statistical significance (Table 3). Joint analysis of all samples resulted in an OR of 1.20 (95% CI 1.05–1.37) and a two-sided *p* value of 0.008, but after applying a statistical correction for multiple hypothesis testing, the resulting *p* value was 0.086. No heterogeneity was detected among our subsamples (*p*=0.29).

**Genotype-phenotype correlations** It is possible that, while we were not able to detect a statistically robust effect on risk of type 2 diabetes, this *IRS1* variant (which is in perfect LD with several SNPs in the 3' UTR) might influence glycaemic quantitative traits, either by decreasing insulin secretion in the pancreatic beta cells or by increasing peripheral insulin resistance. We detected no significant differences in fasting insulin, insulinogenic index, 2-h insulin, insulin AUC, HOMA<sub>IR</sub> or insulin sensitivity index between major allele homozygotes and heterozygotes at rs934167 (Table 4). A nominal effect of rs934167 on BMI in the subset of Botnian normoglycaemic subjects was not replicated in the control subjects from the USA (C/C vs C/T, 27.4±5.2 vs 27.3±5.4 kg/m<sup>2</sup>, *p*=0.71), Poland (C/C vs C/T, 26.1±3.6 vs 26.4±3.1 kg/m<sup>2</sup>, *p*=0.45) or Scandinavia (C/C vs C/T, 25.4±3.7 vs 25.5±3.2 kg/m<sup>2</sup>, *p*=0.65). Because of the nominal association of rs934167 with BMI



**Fig. 2** Cumulative incidence of type 2 diabetes, by age of onset and genotype at rs934167. The 1,210 Scandinavian subjects for whom we had precise age of onset of type 2 diabetes were stratified by rs934167 genotype and the proportion developing type 2 diabetes was plotted over time. Due to the small number of T/T homozygotes (*n*=3), we analysed C/C homozygotes (black line with diamonds) vs T carriers (T/X, grey line with squares). There was no statistically significant difference in the proportion of T carriers diagnosed before age 50 (*p*=0.26)

in the Botnian subsample, we further adjusted the initial association of this SNP with type 2 diabetes for log-transformed BMI; this adjustment attenuated the nominal statistical significance obtained in the unadjusted analysis ( $p=0.24$  and  $0.11$  in the US and Polish case-control samples, respectively). In addition, we did not detect any interaction between BMI and genotype at rs934167 ( $p=0.48$  and  $0.41$  in the US and Polish case-control samples, respectively). Finally, we noted no effect of genotype at rs934167 on diabetes age of onset for the 1,210 Scandinavian subjects for whom we had precise information on age at diagnosis (Fig. 2).

## Discussion

Our study was a logical next step in our exploration of the hypothesis that common genetic variants in *IRS1* influence insulin resistance. Having been unable to reproduce the specific association of G972R with type 2 diabetes suggested by Jellema et al. [6], we wondered whether other nearby variants (in LD with G972R in some samples but not in others) might account for the association signal noted by other groups. Our study design was well powered to detect associations with similar allele frequencies and putative GRR as proposed for G972R; in addition, its ability to capture common variants in the region was reasonably comprehensive. Nevertheless, we were unable to conclusively identify such an associated variant.

The non-coding SNP rs934167 showed a modest nominal association in the US and Polish case-control pairs, with a similar trend noted in the Scandinavian subsamples. There is little-to-no LD between rs934167 and rs1801278 (G972R) in white subjects ( $D'=0.08$ ,  $r^2=0$ ), such that our finding, if real, cannot be considered to support the proposed G972R association. Although rs934167 is located downstream of the *IRS1* gene, it is a perfect proxy for several SNPs located in its 3' UTR, which might conceivably affect mRNA stability and thereby influence *IRS1* levels. To determine conclusively whether rs934167 is associated with type 2 diabetes would require independent confirmation in adequately powered samples, estimated to require ~4,700 case-control pairs; even larger samples would probably be needed to detect an effect of rs934167 on glycaemic traits. If the association were confirmed, the question of whether rs934167 (or another variant in LD with it) impacts *IRS1* mRNA levels could be examined in cells isolated from individuals with different genotypes at that locus.

Although our attempt to capture common genetic variation at *IRS1* was adequate, the possibility remains that

as-yet uncaptured polymorphisms do increase risk of type 2 diabetes. Upcoming large-scale whole-genome association scans and their analytical integration should allow for a focused test of all common variants in the region, including rs934167 and those with more modest effects or those which were not perfectly captured here. On the other hand, if an aggregate of multiple rare polymorphisms at *IRS1*, rather than common variants, leads to type 2 diabetes, alternative association methodologies in large resequenced samples will be required.

**Acknowledgements** J. C. Florez is supported by NIH Research Career Award 1 K23 DK65978-03. T. Tuomi is a Research Fellow at the Academy of Finland. D. Altshuler is a Charles E. Culpeper Scholar of the Rockefeller Brothers Fund and a Burroughs Wellcome Fund Clinical Scholar in Translational Research. D. Altshuler, M. J. Daly and J. N. Hirschhorn are recipients of The Richard and Susan Smith Family Foundation/ADA Pinnacle Program Project Award. L. Groop, T. Tuomi and the Botnia Study are principally supported by the Sigrid Juselius Foundation, the Academy of Finland, the Finnish Diabetes Research Foundation, The Folkhalsan Research Foundation, the Swedish Medical Research Council and the Novo Nordisk Foundation. We thank the members of the Altshuler, Hirschhorn, Daly and Groop labs for helpful discussions. We also thank the participants of the studied cohorts for contributing their DNA samples and phenotypic measurements.

**Duality of interest** K. G. Ardlie was formerly employed by Genomics Collaborative, a division of SeraCare Life Sciences, which owns a sample repository that may be requested and used for research of diabetes and its complications; some of those samples have been used in this study. T. Tuomi is a member of the Nordic Cardiovascular Board. The other authors have no duality of interest to disclose with regard to this manuscript.

## References

1. Sun XJ, Rothenberg P, Kahn CR et al (1991) Structure of the insulin receptor substrate IRS-1 defines a unique signal transduction protein. *Nature* 352:73–77
2. Burks DJ, White MF (2001) IRS proteins and  $\beta$ -cell function. *Diabetes* 50(Suppl. 1):S140–S145
3. Nandi A, Kitamura Y, Kahn CR, Accili D (2004) Mouse models of insulin resistance. *Physiol Rev* 84:623–647
4. Sesti G, Federici M, Hribal ML, Lauro D, Sbraccia P, Lauro R (2001) Defects of the insulin receptor substrate (IRS) system in human metabolic disorders. *FASEB J* 15:2099–2111
5. Almind K, Bjorbaek C, Vestergaard H, Hansen T, Echwald S, Pedersen O (1993) Aminoacid polymorphisms of insulin receptor substrate-1 in non-insulin-dependent diabetes mellitus. *Lancet* 342:828–832
6. Jellema A, Zeegers MP, Feskens EJ, Dagnelie PC, Mensink RP (2003) Gly972Arg variant in the insulin receptor substrate-1 gene and association with Type 2 diabetes: a meta-analysis of 27 studies. *Diabetologia* 46:990–995
7. Florez JC, Sjögren M, Burt N et al (2004) Association testing in 9,000 people fails to confirm the association of the insulin receptor substrate-1 G972R polymorphism with type 2 diabetes. *Diabetes* 53:3313–3318



8. Zeggini E, Parkinson J, Halford S et al (2004) Association studies of insulin receptor substrate 1 gene (*IRS1*) variants in type 2 diabetes samples enriched for family history and early age of onset. *Diabetes* 53:3319–3322
9. Ardlie KG, Lunetta KL, Seielstad M (2002) Testing for population subdivision and association in four case-control studies. *Am J Hum Genet* 71:304–311
10. Groop L, Forsblom C, Lehtovirta M et al (1996) Metabolic consequences of a family history of NIDDM (the Botnia study): evidence for sex-specific parental effects. *Diabetes* 45:1585–1593
11. Altshuler D, Hirschhorn JN, Klannemark M et al (2000) The common PPAR $\gamma$  Pro12Ala polymorphism is associated with decreased risk of type 2 diabetes. *Nat Genet* 26:76–80
12. Florez JC, Burt N, de Bakker PIW et al (2004) Haplotype structure and genotype-phenotype correlations of the sulfonylurea receptor and the islet ATP-sensitive potassium channel gene region. *Diabetes* 53:1360–1368
13. Florez JC, Agapakis CM, Burt NP et al (2005) Association testing of the protein tyrosine phosphatase 1B gene (*PTPNI*) with type 2 diabetes in 7,883 people. *Diabetes* 54:1884–1891
14. Florez JC, Wiltshire S, Agapakis CM et al (2006) High-density haplotype structure and association testing of the insulin-degrading enzyme (*IDE*) gene with type 2 diabetes in 4,206 people. *Diabetes* 55:128–135
15. Sun MW, Lee JY, de Bakker PIW et al (2006) Haplotype structures and large-scale association testing of the 5' AMP-activated protein kinase genes *PRKAA2*, *PRKAB1*, and *PRKAB2* with type 2 diabetes. *Diabetes* 55:849–855
16. Lindholm E, Agardh E, Tuomi T, Groop L, Agardh C-D (2001) Classifying diabetes according to the new WHO clinical stages. *Eur J Epidemiol* 17:983–989
17. Saxena R, Gianniny L, Burt NP et al (2006) Common single nucleotide polymorphisms in *TCF7L2* are reproducibly associated with type 2 diabetes and reduce the insulin response to glucose in nondiabetic individuals. *Diabetes* 55:2890–2895
18. Tang K, Fu DJ, Julien D, Braun A, Cantor CR, Koster H (1999) Chip-based genotyping by mass spectrometry. *Proc Natl Acad Sci USA* 96:10016–10020
19. The International HapMap Consortium (2005) A haplotype map of the human genome. *Nature* 437:1299–1320
20. Gabriel SB, Schaffner SF, Nguyen H et al (2002) The structure of haplotype blocks in the human genome. *Science* 296:2225–2229
21. de Bakker PIW, Yelensky R, Pe'er I, Gabriel SB, Daly MJ, Altshuler D (2005) Efficiency and power in genetic association studies. *Nat Genet* 37:1217–1223
22. Spielman RS, McGinnis RE, Ewens WJ (1993) Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *Am J Hum Genet* 52:506–516
23. Boehnke M, Langefeld CD (1998) Genetic association mapping based on discordant sib pairs: the discordant-alleles test. *Am J Hum Genet* 62:950–961
24. Barrett JC, Fry B, Maller J, Daly MJ (2005) Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 21:263–265
25. Lohmueller K, Pearce CL, Pike M, Lander ES, Hirschhorn JN (2003) Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. *Nat Genet* 33:177–182
26. Reis IM, Hirji KF, Afifi AA (1999) Exact and asymptotic tests for homogeneity in several 2 $\times$ 2 tables. *Stat Med* 18:893–906
27. Purcell S, Cherny SS, Sham PC (2003) Genetic power calculator: design of linkage and association genetic mapping studies of complex traits. *Bioinformatics* 19:149–150
28. Byrne CD, Wareham NJ, Brown DC et al (1994) Hypertriglyceridemia in subjects with normal and abnormal glucose tolerance: relative contributions of insulin secretion, insulin resistance and suppression of plasma non-esterified fatty acids. *Diabetologia* 37:889–896
29. Matsuda M, DeFronzo R (1999) Insulin sensitivity indices obtained from oral glucose tolerance testing: comparison with the euglycemic insulin clamp. *Diabetes Care* 22:1462–1470
30. Matthews DR, Hosker JP, Rudenski AS, Naylor BA, Treacher DF, Turner RC (1985) Homeostasis model assessment: insulin resistance and beta-cell function from fasting plasma glucose and insulin concentrations in man. *Diabetologia* 28:412–419