

# Candidate loci for insulin sensitivity and disposition index from a genome-wide association analysis of Hispanic participants in the Insulin Resistance Atherosclerosis (IRAS) Family Study

N. D. Palmer · C. D. Langefeld · J. T. Ziegler · F. Hsu · S. M. Haffner · T. Fingerlin · J. M. Norris · Y. I. Chen · S. S. Rich · T. Haritunians · K. D. Taylor · R. N. Bergman · J. I. Rotter · D. W. Bowden

Received: 4 May 2009 / Accepted: 5 October 2009 / Published online: 10 November 2009  
© Springer-Verlag 2009

## Abstract

**Aims/hypothesis** The majority of type 2 diabetes genome-wide association studies (GWAS) to date have been performed in European-derived populations and have identified few variants that mediate their effect through insulin resistance. The aim of this study was to evaluate two quantitative, directly assessed measures of insulin resistance, namely insulin sensitivity index ( $S_I$ ) and insulin disposition index (DI), in Hispanic-American participants using an agnostic, high-density single nucleotide polymorphism (SNP) scan, and to validate these findings in additional samples.

**Methods** A two-stage GWAS was performed in Hispanic-American samples from the Insulin Resistance Atherosclerosis

Family Study. In Stage 1, 317,000 SNPs were assessed using 229 DNA samples. SNPs with evidence of association with glucose homeostasis and adiposity traits were then genotyped on the entire set of Hispanic-American samples ( $n=1,190$ ). This report focuses on the glucose homeostasis traits:  $S_I$  and DI.

**Results** Although evidence of association did not reach genome-wide significance ( $p=5 \times 10^{-7}$ ), in the combined analysis SNPs had admixture-adjusted  $p$  values of  $p_{\text{ADD}}=0.00010$ – $0.0020$  with 8 to 41% differences in genotypic means for  $S_I$  and DI.

**Conclusions/interpretation** Several candidate loci were identified that are nominally associated with  $S_I$  and/or DI in Hispanic-American participants. Replication of these

**Electronic supplementary material** The online version of this article (doi:10.1007/s00125-009-1586-2) contains supplementary material, which is available to authorised users.

N. D. Palmer · D. W. Bowden (✉)  
Department of Biochemistry,  
Centers for Human Genomics & Diabetes Research,  
Wake Forest University School of Medicine,  
1 Medical Center Blvd,  
Winston-Salem, NC 27157, USA  
e-mail: dbowden@wfubmc.edu

C. D. Langefeld · J. T. Ziegler · F. Hsu  
Department of Public Health Sciences,  
Wake Forest University School of Medicine,  
Winston-Salem, NC, USA

S. M. Haffner  
Department of Medicine,  
University of Texas Health Sciences  
Center at San Antonio,  
San Antonio, TX, USA

T. Fingerlin · J. M. Norris  
Department of Epidemiology, Colorado School of Public Health,  
University of Colorado Denver,  
Denver, CO, USA

Y. I. Chen · T. Haritunians · K. D. Taylor · J. I. Rotter  
Medical Genetics Institute,  
Burns and Allen Cedars-Sinai Research Institute,  
Los Angeles, CA, USA

S. S. Rich  
Center for Public Health Genomics, University of Virginia,  
Charlottesville, VA, USA

R. N. Bergman  
Department of Physiology and Biophysics,  
Keck School of Medicine, University of Southern California,  
Los Angeles, CA, USA

findings in independent cohorts and additional focused analysis of these loci is warranted.

**Keywords** Disposition index · Hispanic-Americans · Insulin sensitivity · Type 2 diabetes

### Abbreviations

AIM	Ancestry-informative marker
AIR	Acute insulin response
CEU	Centre d'Etude du Polymorphisme (Utah residents with northern and western European ancestry)
DI	Insulin disposition index
EBV	Epstein–Barr virus
GWAS	Genome-wide association study
HOMA <sub>IR</sub>	HOMA of insulin resistance
IRAS-FS	Insulin Resistance Atherosclerosis Family Study
LD	Linkage disequilibrium
MAF	Minor allele frequency
PC	Principal components
S <sub>I</sub>	Insulin sensitivity index
SNP	Single nucleotide polymorphism

### Introduction

Recent genome-wide association studies (GWAS) have made major contributions to the understanding of complex genetic traits. Several studies in samples of type 2 diabetes patients and controls [1–6] have had an extraordinary impact on the current understanding of genetic susceptibility to type 2 diabetes, primarily in European-derived populations. As with many major technical advances, these results have raised additional questions. To date, there have been few type 2 diabetes GWAS in US minority populations [7]. In addition, evaluations of consensus type 2 diabetes-associated single nucleotide polymorphisms (SNPs) from European-derived studies suggest that the influence of these polymorphisms in US minorities, especially African-American participants, may be limited [7–9]. A feature of these GWAS has been that most of the type 2 diabetes genes identified probably mediate their influence on type 2 diabetes susceptibility through the beta cell. This contrasts with the widely accepted belief that insulin resistance is a major component of type 2 diabetes susceptibility [10–14]. One possibility is that GWAS evaluation of type 2 diabetes patients compared with non-diabetic controls may not be an efficient way to locate genes that code for other risk factors for type 2 diabetes, e.g. reduced insulin action and/or inability of the beta cells

to compensate for insulin resistance, i.e. impaired insulin disposition.

The purpose of this study was to evaluate two quantitative, directly assessed measures of insulin resistance, namely insulin sensitivity index (S<sub>I</sub>) and insulin disposition index (DI), in a non-European population. Herein, we present results of a two-stage GWAS in Hispanic-Americans from the Insulin Resistance Atherosclerosis Family Study (IRAS-FS). Through an unbiased approach using a high-density SNP scan and follow-up genotyping, we have identified novel loci that could potentially contribute to variation in glucose homeostasis. This report complements the results published by Rich et al. for the phenotype acute insulin response (AIR) [15] and Norris et al. for the adiposity phenotypes [16], both in the same cohort. We acknowledge that these results are preliminary and that replication of these findings in independent cohorts is essential.

### Methods

*IRAS-FS participants* Study design, recruitment and phenotyping have been described previously [17]. IRAS-FS is a multi-centre study designed to identify the genetic determinants of quantitative measures of glucose homeostasis. Members of large families of Hispanic ancestry ( $n=1,334$  in 92 pedigrees from San Antonio, TX and San Luis Valley, CO) were recruited and presented in this report. The institutional review boards at each participating analysis and clinical site approved the study protocol and all participants provided written informed consent.

A clinical examination was performed, which included a frequently sampled IVGTT, anthropometric measurements and collection of samples for blood chemistry and biomarker analysis. Measures of glucose homeostasis were derived using mathematical modelling methods (MINMOD) [18] from glucose and insulin values obtained during the IVGTT [19–21]. These estimates included: S<sub>I</sub>, AIR and DI ( $DI=AIR \times S_I$ ). This is a report of the results for S<sub>I</sub> and DI.

A subset of IRAS-FS Hispanic-Americans ( $n=229$  in 34 families from San Antonio, TX) was chosen for the GWAS. This subset consisted of participants without type 2 diabetes, for whom complete data for glucose homeostasis and obesity phenotypes were available, and whose age, BMI and sex composition were consistent with the IRAS-FS collection. Samples chosen represented a genetically homogenous population as assessed from Structure analysis [22] using microsatellite markers from the genome-wide linkage scan [23, 24]. DNA used in the genotyping was obtained from Epstein–Barr virus (EBV)-transformed lymphoblastoid cell lines.

**Genome-wide association study** Genotyping was performed using 1.5 µg of genomic DNA (15 µl of 100 ng/µl stock) with Illumina Infinium II HumanHap 300 BeadChips (Illumina, San Diego, CA, USA) at Cedars-Sinai Medical Center according to a standardised protocol [25]. Genotypes were called on the basis of clustering raw intensity data for the two dyes using Illumina BeadStudio software. Repeat genotyping of DNA samples was performed once if the overall call rate was <98% and that sample was rejected if there was no improvement. The average sample call rate was 99.76%. Consistency of genotyping was checked using 18 repeat samples; the concordance rate was 100%. SNPs with Hardy–Weinberg equilibrium  $p < 0.001$ , minor allele frequency (MAF) less than 0.05 or more than 5% missing genotypes were excluded from subsequent analysis. Genotypes with GenCall scores <0.15 were set to missing (0.25%). For highly associated SNPs, clustering was repeated to exclude spurious significance. All genotypes were oriented to the forward strand.

**Validation genotyping in the entire IRAS-FS Hispanic sample** SNPs with evidence of association in the GWAS were validated in the entire Hispanic cohort (excluding participants with type 2 diabetes). A total of 1,536 SNPs was chosen for genotyping on all Hispanic samples ( $n=1,190$ ). Genotyping was performed at Cedars-Sinai Medical Center using the Illumina Golden Gate assay. SNPs with low call frequencies (<98%) were manually re-clustered (~15% of SNPs). Of the 1,536 SNPs, 3.5% were excluded due to call frequency of <0.7 and/or cluster separation of <0.3. The average SNP call frequency was 99.48%. Duplicate genotyping of 12 samples was 100% concordant. The minimum acceptable sample call rate was 95%; the average sample call rate was 99.5%. SNP selection for this second stage was based upon: (1) identification of the most strongly associated 50 to 100 SNPs for each trait of interest (Electronic supplementary material [ESM] Table 1) from the initial GWAS; (2) tag SNPs in genes with high evidence of association across more than one phenotype; these were selected using the HapMap Centre d'Etude du Polymorphisme (Utah residents with northern and western European ancestry) (CEU) reference population to capture common variation within the associated linkage disequilibrium (LD) block; and (3) ancestry-informative markers (AIMs) for Hispanic populations [26, 27]. In total, 118 and 96 SNPs for  $S_1$  and DI, respectively, were selected and successfully genotyped in the validation study.

**Follow-up locus-specific genotyping** Loci with evidence of association from the GWAS and validation genotyping were targeted for additional genotyping using tag SNPs. Genotyping was performed using iPLEX Gold SBE assays

on the Sequenom genotyping system (Sequenom MassArray; Sequenom, San Diego, CA, USA). Locus-specific primers were designed using MassArray Assay Design 3.0 software and resulting mass spectrograms were analysed by the MassArray Typer software. The minimum acceptable call frequency was 95%. We included 51 blind duplicates to evaluate genotyping accuracy; the concordance rate was 100%. SNPs were chosen to capture common variation within LD blocks as defined by the CEU population of the HapMap project [28]. Specifically, genotype data from the genomic interval containing the candidate gene  $\pm 5$  kb was exported from the HapMap database and imported into Haploview [29]. For genes with few LD blocks, i.e. *VIPRI*, *SLC1A4* and both *P2RY2* and *P2RY6*, SNPs were selected to tag the entire genic region with a mean  $r^2=0.80$  and with forced inclusion of previously genotyped SNPs. For larger genes, i.e. *MAG11*, *KLHL25*, *MYH13*, *RGS7*, *EFCAB7* and *PGMI*, SNP selection focused on the LD block containing SNPs associated in the validation genotyping, with additional SNPs being selected to tag the block with a mean  $r^2=0.80$  with forced inclusion of previously genotyped SNPs.

**Statistical methods** For quality control, each SNP was examined for Mendelian inconsistencies using PedCheck [30] and 1,657 SNPs exhibiting inconsistencies were converted to 'missing'. Maximum likelihood estimates of allele frequencies were computed using the largest set of unrelated Hispanic-American individuals ( $n=34$ ); SNP genotypes were tested for departures from Hardy–Weinberg Equilibrium. SNPs with no evidence of a difference in  $S_1$  or DI values between individuals with and without missing genotype data ( $p > 0.05$ ), and with no evidence of departure from Hardy–Weinberg equilibrium ( $p > 0.001$ ) were included in subsequent analyses.

To test for association between individual SNPs and the traits of interest, i.e.  $S_1$  and DI, differences in trait values by genotype were tested using the variance components model that explicitly models the correlation among related individuals as implemented in SOLAR [31]. X-chromosome SNPs were not used in the primary analyses when using this method. For statistical testing,  $S_1$  and DI were transformed using log and signed-square root, respectively, to best approximate the distributional assumptions of the test and to minimise heterogeneity of the variance. The primary statistical inference was the additive genetic model that was used to rank SNPs. All tests and levels of significance were computed after adjustment for age, sex and BMI.

Analysis of validation and locus-specific genotyping data followed the same analytical framework as the GWAS, except that covariate adjustment included a term for the site of recruitment (San Antonio, San Luis Valley) and one for admixture. For admixture analysis, a collection of AIMs

was used. These were selected from the literature on studies performed in Hispanics [26, 27]. The GWAS had 80 SNPs (including 14 on the X chromosome) and the validation genotyping had 149 SNPs (including 23 on the X chromosome). The 149 AIMS were available on 1,279 participants, and these data were merged with HapMap data for Centre d'Etude du Polymorphisme Humain ( $n=90$ ) and Yoruba ( $n=90$ ) populations.

A principal components (PC) analysis was performed on the 149 AIMS as well as on the 80 AIMS in common between the GWAS (317,000 SNP panel) and validation (1536 SNP panel) experiments. The total proportion of variance explained by the first three PCs with the 80 AIMS (PC1, 10.2%; PC2, 5.1%; PC3, 2.7%) differed little from the proportion of variance explained by the 149 AIMS (PC1, 10.3%; PC2, 4.8%; PC3, 1.9%). However, overall differences were seen between the Hispanic-American sites with respect to PC2 ( $p=2.35 \times 10^{-53}$ ). In addition, Hispanic-Americans from the sites differed in measures of glucose homeostasis ( $S_I$ ,  $p=0.0006$ ; DI,  $p=1.8 \times 10^{-11}$ ). For  $S_I$  and DI, the proportion of variance explained by the centre of ascertainment was 0.01% and 1.59%, respectively; thus, all results are presented with adjustment for admixture in addition to age, sex, BMI and recruitment centre.

## Results

**Study participants** Hispanic-American participants ( $n=229$ ) from the San Antonio population with complete phenotypic data and DNA obtained from EBV-transformed cell lines were used in the GWAS (Stage 1). A sample of 814 participants with DNA and baseline data was used for validation (Stage 2). The total sample of 1,043 Hispanic-

American participants included 59.4% women, average age 41.1 years, mean  $S_I$   $2.16 \times 10^{-5} \text{ min}^{-1} (\text{pmol/l})^{-1}$ , mean DI of  $1321.7 \times 10^{-5} \text{ min}^{-1}$  and mean BMI of  $28.4 \text{ kg/m}^2$ . Table 1 summarises relevant demographic measures showing the comparability of Stage 1 and 2 samples. Specifically, there was no significant difference in age ( $p=0.67$ ), sex ( $p=0.40$ ), BMI ( $p=0.084$ ) or DI ( $p=0.058$ ), and only a modest difference in  $S_I$  ( $p=0.045$ ).  $S_I$  and DI had a modest genetic correlation of  $0.38 \pm 0.12$  in the overall sample.

**GWAS for  $S_I$  and DI** A total of 309,200 SNPs met quality control criteria and were evaluated for association with  $S_I$  and DI. SNPs were ranked using  $p$  values from the additive genetic model. The quantile–quantile plot for the stage 1 GWAS indicated that the majority of SNPs exhibited a  $-\log_{10}(p \text{ value}) < 2$  and that the observed distribution matched the expectation for the majority of the data (Fig. 1). The highest-ranking SNPs associated with  $S_I$  and DI were chosen for genotyping (ESM Table 2) on all Hispanic-American participants in the IRAS-FS ( $n=1,190$ ). A total of 611 SNPs with evidence of association with  $S_I$ , DI and other glucose homeostasis phenotypes ( $S_G$  and AIR) or SNPs that tag genes associated with multiple phenotypes were included in a 1536 custom chip. For  $S_I$  and DI, 145 and 98 SNPs, respectively, were chosen for validation (ESM Table 1).

**Candidate genes/regions of association for  $S_I$  and DI** The most significantly associated SNPs for  $S_I$  and DI are presented in Table 2, ordered by chromosomal position as determined by dbSNP ([www.ncbi.nlm.nih.gov/projects/SNP/](http://www.ncbi.nlm.nih.gov/projects/SNP/)) using NCBI Build 36.1 (hg18). These hits: (1) were selected for follow-up from the GWAS based on significance; (2) had nominal evidence of association in the

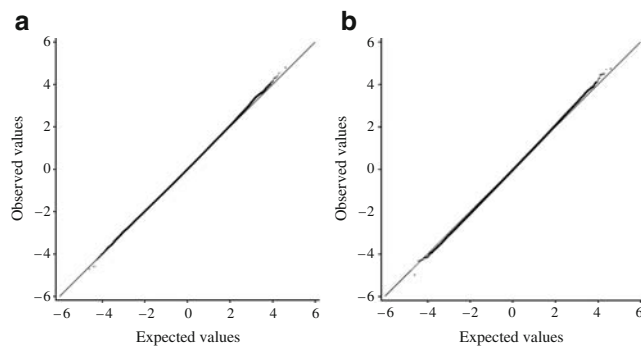
**Table 1** Demographics for IRAS-FS Hispanic-American samples

Variables	GWAS sample (stage 1)	Validation sample (stage 2)	Combined sample
Participant characteristics	229	814	1,043
Male sex, $n$ (%)	83 (36.2)	341 (41.9)	424 (40.6)
Female sex, $n$ (%)	146 (63.8)	473 (58.1)	619 (59.4)
Age (years)	41.3 $\pm$ 13.9	41.1 $\pm$ 13.9	41.1 $\pm$ 13.9
BMI ( $\text{kg/m}^2$ )	29.4 $\pm$ 5.9	28.1 $\pm$ 5.8	28.4 $\pm$ 5.8
Glucose homeostasis			
$S_I$ ( $\times 10^{-5} \text{ min}^{-1} [\text{pmol/l}]^{-1}$ )	1.87 $\pm$ 1.91	2.25 $\pm$ 1.95	2.16 $\pm$ 1.94
AIR (pmol/l)	722.9 $\pm$ 598.1	780.2 $\pm$ 678.2	766.9 $\pm$ 660.5
DI ( $\times 10^{-5} \text{ min}^{-1}$ )	1,035.4 $\pm$ 930.2	1,408.9 $\pm$ 1,318.9	1,321.7 $\pm$ 1,248.7
Fasting glucose (mmol/l)	5.22 $\pm$ 0.55	5.18 $\pm$ 0.52	5.19 $\pm$ 0.53
Fasting insulin (pmol/l)	112 $\pm$ 77	103 $\pm$ 78	104 $\pm$ 78

Values are reported as the mean  $\pm$  standard deviation unless otherwise noted

Participants with type 2 diabetes were excluded





**Fig. 1** Quantile–quantile plots of **a** insulin sensitivity and **b**  $S_1$  for the initial GWAS analysis ( $n=229$ ). Plots compare observed vs expected values of the Z test statistics under the null hypothesis of no association across the genome, and are reported with adjustment for age, sex and BMI

validation sample ( $p<0.05$ ) (ESM Table 3); and (3) showed consistency in directionality (beta coefficient) with respect to the same allele (all analyses are presented with respect to the minor allele). Of the 145 (31.7%) selected for validation, 46 were nominally associated with  $S_1$  ( $p<0.05$ ) in the combined analysis (ESM Table 4). Two of the strongest associations for  $S_1$ , rs7091573 and rs6560787, are within 2 kb of each other on chromosome 10 ( $D'=1.00$ ,  $r^2=0.72$ ) near the genomic location of a cDNA for a hypothesised gene *AK097474*. Additionally, two non-genic SNPs on chromosome 15 (rs7174900 and rs7172316;  $D'=1.00$ ,  $r^2=1.00$ ) were also significantly associated. Haplotype analysis of high scoring, closely linked SNPs did not provide more strongly associated findings than single SNP analysis (data not shown). Admixture-adjusted  $p_{\text{Additive}}$  values for the top hits in the total Hispanic-American population were in the range of  $p_{\text{Additive}}=1.0\times 10^{-4}$  to  $p_{\text{Additive}}=1.3\times 10^{-3}$ , which are comparable in magnitude to the  $p$  values observed in the GWAS.

Of the 98 SNPs selected for validation with DI, 31 (31.6%) were nominally associated ( $p<0.05$ ) in the combined 1536 SNP analysis (ESM Table 4). The most highly associated SNP overall for DI was rs217463 (admixture-adjusted  $p_{\text{Additive}}=6.89\times 10^{-4}$ ) in the *EFCAB7* gene. Similar to the results with  $S_1$ , most of the high scoring DI SNPs had broadly comparable  $p$  values in the GWAS. Overall,  $p$  values for SNPs most highly associated with DI were of a comparable magnitude to those for the SNPs most highly associated with  $S_1$ .

Several genes were chosen for additional genotyping and analysis: for  $S_1$ , *RGS7* (regulator of G protein signalling 7); for DI, *SLC1A4* and *EFCAB7/PGM1* (phosphoglucosyltransferase 1); and for both  $S_1$  and DI, *MAG11* and *VIPRI* (vasoactive intestinal peptide receptor 1). Tagging SNPs were chosen to cover these genes or, in the case of very large genes, e.g. *MAG11* (685 kb), to cover the LD block containing the

associated SNPs. Overall 47 additional SNPs in these six loci were genotyped. Although not striking, this genotyping resulted in additional evidence of association: (1) with  $S_1$ , with a trend towards association for *RGS7* ( $p_{\text{Additive}}=0.063$  for rs7531569); (2) with DI for *SLC1A4* (rs2075209,  $p_{\text{Additive}}=0.00015$ ) and *EFCAB7/PGM1* (rs855315,  $p_{\text{Additive}}=0.0016$ ); plus two additional SNPs: rs11208250 and rs855325); and (3) with  $S_1$  and DI for *VIPRI* (rs7627240,  $p_{\text{Additive}}=0.00026$ ; and two additional SNPs for  $S_1$ : rs421558 and rs417387) and *MAG11* (rs884067,  $p_{\text{Additive}}=0.00064$ ) (ESM Tables 5 and 6). These additional SNPs are not highly correlated ( $r^2<0.49$ ) with the initial high scoring SNP or each other.

## Discussion

The aim of this study was to survey the genome for evidence of association with two important quantitative measures of glucose homeostasis:  $S_1$  and DI. The initial genotyping was performed in 229 participants from the IRAS-FS San Antonio, TX, population as a rapid and cost-effective method for scanning the genome. To our knowledge this is the first report of a GWAS of  $S_1$  and DI as direct measures of glucose homeostasis. Insulin resistance is an important component of type 2 diabetes risk and an independent risk factor for complications such as atherosclerosis. Disposition index is a strong predictor of conversion to type 2 diabetes [10, 32] and thus a phenotype of potentially crucial importance in understanding the genetic underpinnings of type 2 diabetes. While DI has frequently been interpreted as a beta cell functional response to insulin resistance, its phenotype may also capture more central or other tissue effects that regulate glucose homeostasis. The signalling mechanism involved in beta cell compensation is still not clearly delineated [33], leaving the possibility that extrapancreatic factors are changed in persons at risk of type 2 diabetes. Such putative signals related to DI-associated loci without a clear link to  $S_1$  or AIR may be of greatest interest for follow-up, since they could identify central regulatory pathways of glucose homeostasis.

Genes identified by GWAS from study designs comparing allele frequencies between type 2 diabetes-affected participants and non-type 2 diabetes controls appear primarily, if not solely, to be beta cell genes [1–6]. While these observations can be interpreted as a lack of insulin-resistance risk variants, several lines of evidence suggest that GWAS designed around contrasting type 2 diabetes with non-type 2 diabetes participants may not be the best way to identify genes that influence insulin sensitivity. Insulin resistance is common in adults. For example, 45 to

**Table 2** Association results for  $S_I$  and DI

SNP per characteristic	Chr	Position	Alleles <sup>a</sup>	MAF <sup>b</sup>	Nearest gene ( $\pm 10$ kb)	GWAS <sup>c</sup>	Validation <sup>d</sup>	Combined <sup>d</sup>
$S_I$								
rs6794189	3	65618513	A/C	0.14	<i>MAG11</i>	0.030	0.0039	$3.92 \times 10^{-4}$
rs2613675	8	21095515	A/G	0.29		$7.88 \times 10^{-5}$	0.019	$3.04 \times 10^{-4}$
rs279910	9	965112	A/G	0.35	<i>DMRT3</i>	$6.08 \times 10^{-4}$	0.039	0.0013
rs7091573 <sup>e</sup>	10	2053806	T/C	0.33	<i>AK097474</i> (hypothetical gene)	$6.21 \times 10^{-4}$	0.0052	$1.03 \times 10^{-4}$
rs6560787 <sup>e</sup>	10	2051961	G/A	0.42	<i>AK097474</i> (hypothetical gene)	$4.85 \times 10^{-4}$	0.013	$2.17 \times 10^{-4}$
rs7174900 <sup>f</sup>	15	71077956	T/C	0.03		$7.78 \times 10^{-4}$	0.022	$4.16 \times 10^{-4}$
rs7172316 <sup>f</sup>	15	71074205	T/G	0.03		$7.78 \times 10^{-4}$	0.029	$5.40 \times 10^{-4}$
rs7181017	15	84145916	C/T	0.09	<i>KLHL25</i>	$4.13 \times 10^{-4}$	0.036	0.0011
DI								
rs217463	1	63814375	A/G	0.35	<i>EFCAB7</i>	$3.22 \times 10^{-5}$	0.020	$6.89 \times 10^{-4}$
rs3004318	1	63673327	T/C	0.35	<i>ALG6</i>	$3.22 \times 10^{-5}$	0.027	0.0011
rs7772334	6	41397828	G/T	0.32		$6.27 \times 10^{-5}$	0.039	0.0011
rs3809738 <sup>g</sup>	17	10217371	A/G	0.19	<i>MYH13</i>	$1.87 \times 10^{-4}$	0.048	0.0020

Results are additive  $p$  values for SNPs in the GWAS ( $n=229$ ), Validation ( $n=814$ ) and Combined ( $n=1,043$ ) samples

SNPs are ranked by significance with respect to the admixture-adjusted, combined sample additive  $p$  value

<sup>a</sup> Major/minor alleles; the minor allele is the reference allele for all analyses

<sup>b</sup> Determined from the maximum set of unrelated individuals in the combined sample ( $n=229$ )

<sup>c</sup> Covariates: age, sex, recruitment centre and BMI

<sup>d</sup> Covariates: age, sex, recruitment centre, BMI and admixture adjustment

<sup>e</sup>  $D'=1.00$ ;  $r^2=0.72$

<sup>f</sup>  $D'=1.00$ ;  $r^2=1.00$

<sup>g</sup> Hardy–Weinberg equilibrium  $p=6.55 \times 10^{-4}$  in the Validation sample  
Chr, chromosome

55% of the non-diabetic European-American, Hispanic-American and African-American participants in IRAS have  $S_I$  values  $<1.0$  (data not shown). Thus, as many as half of these non-diabetic, middle-aged adults are significantly insulin-resistant, yet few of the previous type 2 diabetes GWAS have detailed high-quality measures of insulin sensitivity to identify insulin resistance in their control participants. As an alternative approach, the use of surrogate measures of insulin resistance, e.g. fasting insulin or HOMA of insulin resistance ( $HOMA_{IR}$ ), provides some improvement, but in IRAS, Spearman's correlation of fasting insulin and  $S_I$  was  $-0.61$ , while that for  $HOMA_{IR}$  and  $S_I$  was  $-0.68$  among those with normal glucose tolerance [34]. For Hispanic-Americans in the IRAS-FS, Spearman's correlation of  $HOMA_{IR}$  and  $S_I$  was  $-0.71$ , and that for  $HOMA_{IR}$  and DI was  $-0.46$  (data not shown). Importantly, these surrogates correlate especially poorly in participants with glucose intolerance/insulin resistance, with: (1) Spearman's correlation of  $HOMA_{IR}$  and  $S_I$  of  $-0.39$ , and that for  $HOMA_{IR}$  and DI of  $-0.34$  in Hispanic-American participants; and (2)  $S_I < 1.0$  and Spearman's correlations for fasting insulin and  $S_I$  of  $-0.40$  and  $-0.34$  in persons with impaired glucose tolerance and type 2

diabetes, respectively (IRAS-FS, data not shown). In addition, minimal model-based assessment of  $S_I$  has been shown to have greater heritability and a different genetic basis than  $HOMA_{IR}$  or fasting insulin [35]. It is important to note that minimal model-based measurement of  $S_I$  is a direct measure of insulin sensitivity, rather than a surrogate, and reflects dynamic measurements of insulin sensitivity compared with  $HOMA_{IR}$ , which is basal state measurement.

To carry out this study, we chose a research design in which a 317,000 SNP GWAS was performed on 229 Hispanic-American participants from one clinical centre (San Antonio, TX). From the analysis, a set of 1,536 high scoring SNPs were chosen for validation in the IRAS-FS Hispanic-American sample, in which we have high-quality metabolic measures. Ideally we would have carried out a GWAS of the entire sample set. However, it should be noted that numerous important genes have been identified from GWAS analysis of small initial samples. Examples of such gene–disease associations are: the complement factor H gene and macular degeneration [36] with 224 patients and 134 controls; the *NOS1AP* gene and cardiac repolarisation [37] (200 participants); *TNFSF15* and Crohn's disease

[38] (94 participants); and *CDKN2A* and *CDKN2B* genes with coronary heart disease [39] (322 patients, 312 controls). While high levels of noise are to be expected in GWAS, it is encouraging to note that ten of the final top 16 SNPs for  $S_1$  and ten of the final top 13 SNPs for DI from the 1536 SNP analysis scored high for these traits in the GWAS stage analysis. Nevertheless, this approach is clearly not perfect, for the highest scoring DI SNP, rs2540970 in *SLC1A4*, was not associated with DI in the GWAS ( $p=0.47$ ). This result might thus have been excluded from further consideration, but it is noteworthy that methodological studies by Skol et al. [40] have suggested that an approach using joint analysis of data is more efficient than replication-based analysis for two-stage GWAS. In follow-up genotyping, two additional SNPs in *SLC1A4*, rs2075209 and rs6546119, also showed evidence of association with DI ( $p<0.0043$ ), suggesting *SLC1A4* may indeed be associated with DI. With these provisos, we emphasise that the evidence for association does not meet genome-wide significance and consider the loci reported here as candidates for future detailed evaluation, rather than as confirmed  $S_1$  and DI genes. Relevant to this issue, we have generated quantile–quantile plots (Fig. 1) from the GWAS analysis. These plots compare the observed vs expected values of the  $Z$  test statistics under the null hypothesis of no association across the genome. As expected, the majority of SNPs exhibit a  $-\log_{10} p$  value of  $p<2.0$ . The observed distribution of  $p$  values matches expectations for the majority of the observed data, but departs from the null distribution, albeit modestly, for  $S_1$  and DI at  $p<10^{-3}$ , suggesting that at least some of the loci detected by us are genuine  $S_1$  or DI loci.

Several loci from the GWAS received additional locus-specific follow-up: *RGS7* for  $S_1$ , *SLC1A4* and the *EFCAB7/PGMI* gene cluster for DI, and *MAG11* and *VIPR1* for both  $S_1$  and DI. This analysis did not provide additional compelling evidence that these loci are associated with  $S_1$  or DI, but nine additional SNPs had nominal evidence of association with  $p$  values ranging from  $p=0.023$  to  $p=0.00015$  (ESM Tables 5 and 6). Examination of the top association signals for overlap with previously reported linkage signals [24] did not provide any additional support for the identified loci. Although IRAS-FS, with 181 type 2 diabetes-affected Hispanic participants, has limited power to detect association with type 2 diabetes, SNPs rs6794189 in *MAG11* and rs10793057 near *P2RY2* showed evidence of association with type 2 diabetes in discrete trait analysis with  $p$  values of  $p=0.0099$  and  $p=0.0016$ , respectively (data not shown), lending additional support to their relevance.

Gene families are represented in the overall results of the GWAS analysis, with *MAG12*, in the same family as *MAG11*, nominally associated with DI (data not shown)

and several other traits in the study (J. I. Rotter, L. Wagenknecht, A. Hanley, X. Guo, C. D. Langefeld, F. Hsu, T. Haritunians, unpublished data). SNP rs321983 in *MAG12* was nominally associated with type 2 diabetes ( $p=0.0072$ ) in the Starr County, TX, Mexican-American 100,000 type 2 diabetes GWAS [7]. The only other SNP to overlap with previous type 2 diabetes GWAS was rs10504553 in *TCEB1*, the 11th-ranked SNP for  $S_1$  and 14th for DI in our study, and ranked 20th in the Amish type 2 diabetes 100,000 SNP GWAS [41]. We have reviewed in detail the available results from other type 2 diabetes GWAS, e.g. Diabetes Genetics Initiative, Finland–United States Investigation of NIDDM Genetics and Wellcome Trust Case Control Consortium, along with those from subsequent meta-analyses [42] and including supplementary material to identify type 2 diabetes-associated SNPs that overlap with our study. No overlap was observed, either with the individual studies or with meta-analysis results.

In spite of the numerous advantages of performing GWAS studies in the IRAS-FS sample, there are limitations. First, there are very few comparable studies with minimal-model assessed  $S_1$  and DI, especially in Hispanic-American samples. Minimal model assessment of glucose homeostasis measures in diabetic patients has limitations, so it should be noted that IRAS-FS participants with a type 2 diabetes diagnosis were excluded from these analyses. The admixture-adjusted, additive  $p$  values reported by us do not meet genome-wide significance.  $S_1$  and DI in our Hispanic families were moderately heritable, with  $S_1$  ranging from 0.29 to 0.38 and DI ranging from 0.20 to 0.37 [35]. The effect sizes reflected in the genotypic means for each SNP as summarised in ESM Table 7 range from 8 to 41% of a standard deviation, which is consistent with power estimates for this study design. Quantitative measures of glucose homeostasis have greater power on a per-individual basis than discrete traits. Under an additive model,  $MAF=0.15$  and  $\alpha=0.0001$ , we estimated a power of 0.90 to detect a 0.30 standard deviation change in the genotypic means.

In summary, using a GWAS approach, we identified several genic and non-genic loci that are candidates for association with  $S_1$  and DI in a sample of Hispanic-American families. To obtain more compelling evidence of association with  $S_1$  and DI, additional replication samples will be required. In addition, with the study design we used (317,000 SNPs in 229 participants), the genome was not covered in detail, so other loci influencing  $S_1$  and DI are likely to be unidentified.

**Acknowledgements** This research was supported in part by: NIH grants HL060894, HL060931, HL060944, HL061019 and HL061210; by the University of Virginia Harrison Chair in Public Health Sciences (to S. S. Rich), the Cedars-Sinai Board of Governors' Chair in Medical Genetics (to J. I. Rotter), the General Clinical Research

Centers Program, NCRR grant M01RR00069 and the Diabetes Endocrinology Research Center Grant DDK063491. M. Brown and A. Williams (both of Wake Forest University, Winston-Salem, NC, USA) assisted in the data analysis.

**Duality of interest** The authors declare that there is no duality of interest associated with this manuscript.

## References

- 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
- 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
- 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
- Sladek R, Rocheleau G, Rung J et al (2007) A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature* 445:881–885
- Steinthorsdottir V, Thorleifsson G, Reynisdottir I et al (2007) A variant in CDKAL1 influences insulin response and risk of type 2 diabetes. *Nat Genet* 39:770–775
- 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
- Hayes MG, Pluzhnikov A, Miyake K et al (2007) Identification of type 2 diabetes genes in Mexican Americans through genome-wide association studies. *Diabetes* 56:3033–3044
- Lewis JP, Palmer ND, Hicks PJ et al (2008) Association analysis of European-derived type 2 diabetes SNPs from whole genome association studies in African Americans. *Diabetes* 57:2220–2225
- Palmer ND, Goodarzi MO, Langefeld CD et al (2008) Quantitative trait analysis of type 2 diabetes susceptibility loci identified from whole genome association studies in the Insulin Resistance Atherosclerosis Family Study. *Diabetes* 57:1093–1100
- Bergman RN, Ader M, Huecking K, van Citters G (2002) Accurate assessment of beta-cell function: the hyperbolic correction. *Diabetes* 51(Suppl 1):S212–S220
- DeFronzo RA (1988) Lilly lecture 1987. The triumvirate: beta-cell, muscle, liver. A collusion responsible for NIDDM. *Diabetes* 37:667–687
- Haffner SM, D'Agostino R, Saad MF et al (1996) Increased insulin resistance and insulin secretion in nondiabetic African-Americans and Hispanics compared with non-Hispanic whites. The Insulin Resistance Atherosclerosis Study. *Diabetes* 45:742–748
- Haffner SM, Howard G, Mayer E et al (1997) Insulin sensitivity and acute insulin response in African-Americans, non-Hispanic whites, and Hispanics with NIDDM: the Insulin Resistance Atherosclerosis Study. *Diabetes* 46:63–69
- Muoio DM, Newgard CB (2008) Mechanisms of disease: molecular and metabolic mechanisms of insulin resistance and beta-cell failure in type 2 diabetes. *Nat Rev Mol Cell Biol* 9:193–205
- Rich SS, Goodarzi MO, Palmer ND et al (2009) A genome-wide association scan for acute insulin response to glucose in Hispanic-Americans: the Insulin Resistance Atherosclerosis Family Study (IRAS FS). *Diabetologia* 52:1326–1333
- Norris JM, Langefeld CD, Talbert ME et al (2009) Genome-wide association study and follow-up analysis of adiposity traits in Hispanic Americans: The IRAS Family Study. *Obesity* (Silver Spring) 17:1932–1941
- Henkin L, Bergman RN, Bowden DW et al (2003) Genetic epidemiology of insulin resistance and visceral adiposity. The IRAS Family Study design and methods. *Ann Epidemiol* 13:211–217
- Pacini G, Bergman RN (1986) MINMOD: a computer program to calculate insulin sensitivity and pancreatic responsiveness from the frequently sampled intravenous glucose tolerance test. *Comput Methods Programs Biomed* 23:113–122
- Bergman RN, Finegood DT, Ader M (1985) Assessment of insulin sensitivity in vivo. *Endocr Rev* 6:45–86
- Bergman RN, Ider YZ, Bowden CR, Cobelli C (1979) Quantitative estimation of insulin sensitivity. *Am J Physiol* 236:E667–E677
- Steil GM, Volund A, Kahn SE, Bergman RN (1993) Reduced sample number for calculation of insulin sensitivity and glucose effectiveness from the minimal model. Suitability for use in population studies. *Diabetes* 42:250–256
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155:945–959
- Rich SS, Bowden DW, Haffner SM et al (2005) A genome scan for fasting insulin and fasting glucose identifies a quantitative trait locus on chromosome 17p: the insulin resistance atherosclerosis study (IRAS) family study. *Diabetes* 54:290–295
- Rich SS, Bowden DW, Haffner SM et al (2004) Identification of quantitative trait loci for glucose homeostasis: the Insulin Resistance Atherosclerosis Study (IRAS) Family Study. *Diabetes* 53:1866–1875
- Gunderson KL, Steemers FJ, Ren H et al (2006) Whole-genome genotyping. *Methods Enzymol* 410:359–376
- Choudhry S, Coyle NE, Tang H et al (2006) Population stratification confounds genetic association studies among Latinos. *Hum Genet* 118:652–664
- Seldin MF, Tian C, Shigeta R et al (2007) Argentine population genetic structure: large variance in Amerindian contribution. *Am J Phys Anthropol* 132:455–462
- International HapMap Consortium (2003) The International HapMap Project. *Nature* 426:789–796
- Barrett JC, Fry B, Maller J, Daly MJ (2005) Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 21:263–265
- O'Connell JR, Weeks DE (1998) PedCheck: a program for identification of genotype incompatibilities in linkage analysis. *Am J Hum Genet* 63:259–266
- Almasy L, Blangero J (1998) Multipoint quantitative-trait linkage analysis in general pedigrees. *Am J Hum Genet* 62:1198–1211
- Bergman RN, Finegood DT, Kahn SE (2002) The evolution of beta-cell dysfunction and insulin resistance in type 2 diabetes. *Eur J Clin Invest* 32(Suppl 3):35–45
- Kim SP, Catalano KJ, Hsu IR, Chiu JD, Richey JM, Bergman RN (2007) Nocturnal free fatty acids are uniquely elevated in the longitudinal development of diet-induced insulin resistance and hyperinsulinemia. *Am J Physiol Endocrinol Metab* 292:E1590–E1598
- Saad MF, Anderson RL, Laws A et al (1994) A comparison between the minimal model and the glucose clamp in the assessment of insulin sensitivity across the spectrum of glucose tolerance. Insulin Resistance Atherosclerosis Study. *Diabetes* 43:1114–1121
- Bergman RN, Zaccaro DJ, Watanabe RM et al (2003) Minimal model-based insulin sensitivity has greater heritability and a different genetic basis than homeostasis model assessment or fasting insulin. *Diabetes* 52:2168–2174



36. Edwards AO, Ritter R 3rd, Abel KJ, Manning A, Panhuysen C, Farrer LA (2005) Complement factor H polymorphism and age-related macular degeneration. *Science* 308:421–424
37. Arking DE, Pfeufer A, Post W et al (2006) A common genetic variant in the NOS1 regulator NOS1AP modulates cardiac repolarization. *Nat Genet* 38:644–651
38. Yamazaki K, McGovern D, Ragoussis J et al (2005) Single nucleotide polymorphisms in TNFSF15 confer susceptibility to Crohn's disease. *Hum Mol Genet* 14:3499–3506
39. McPherson R, Pertsemlidis A, Kavaslar N et al (2007) A common allele on chromosome 9 associated with coronary heart disease. *Science* 316:1488–1491
40. Skol AD, Scott LJ, Abecasis GR, Boehnke M (2006) Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. *Nat Genet* 38:209–213
41. Rampersaud E, Damcott CM, Fu M et al (2007) Identification of novel candidate genes for type 2 diabetes from a genome-wide association scan in the Old Order Amish: evidence for replication from diabetes-related quantitative traits and from independent populations. *Diabetes* 56:3053–3062
42. 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