

A genome-wide association scan for acute insulin response to glucose in Hispanic-Americans: the Insulin Resistance Atherosclerosis Family Study (IRAS FS)

S. S. Rich · M. O. Goodarzi · N. D. Palmer · C. D. Langefeld · J. Ziegler ·
S. M. Haffner · M. Bryer-Ash · J. M. Norris · K. D. Taylor · T. Haritunians ·
J. I. Rotter · Y.-D. I. Chen · L. E. Wagenknecht · D. W. Bowden · R. N. Bergman

Received: 10 August 2008 / Accepted: 7 April 2009 / Published online: 9 May 2009
© Springer-Verlag 2009

Abstract

Aims/hypothesis This study sought to identify genes and regions in the human genome that are associated with the acute insulin response to glucose (AIRg), an important predictor of type 2 diabetes, in Hispanic-American participants from the Insulin Resistance Atherosclerosis Family Study (IRAS FS).

Methods A two-stage genome-wide association scan (GWAS) was performed in IRAS FS Hispanic-American samples. In the first stage, 317K single nucleotide polymorphisms (SNPs) were assessed in 229 Hispanic-American DNA samples from 34 families from San

Antonio, TX, USA. SNPs with the most significant associations with AIRg were genotyped in the entire set of IRAS FS Hispanic-American samples ($n=1,190$). In chromosomal regions with evidence of association, additional SNPs were genotyped to capture variation in genes. **Results** No individual SNP achieved genome-wide levels of significance ($p < 5 \times 10^{-7}$); however, two regions (chromosomes 6p21 and 20p11) had multiple highly ranked SNPs that were associated with AIRg. Additional genotyping in these regions supported the initial evidence of variants contributing to variation in AIRg. One region resides in a gene desert between *PXT1* and *KCTD20* on 6p21, while the

S. S. Rich and M. O. Goodarzi contributed equally to this study.

Electronic supplementary material The online version of this article (doi:10.1007/s00125-009-1373-0) contains supplementary material, which is available to authorised users.

S. S. Rich (✉)
Center for Public Health Genomics, University of Virginia,
6111 West Complex, P.O. Box 800717, Charlottesville, VA
22908, USA
e-mail: ssr4n@virginia.edu

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

N. D. Palmer · D. W. Bowden
Department of Biochemistry, Centers for Human Genomics &
Diabetes Research, Wake Forest University School of Medicine,
Winston-Salem, NC, USA

C. D. Langefeld · J. Ziegler · L. E. Wagenknecht
Division 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

M. Bryer-Ash
Department of Medicine, University of Oklahoma,
Oklahoma City, OK, USA

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

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

region on 20p11 has several viable candidate genes (*ENTPD6*, *PYGB*, *GINS1* and *RP4-691N24.1*).

Conclusions/interpretation A GWAS in Hispanic-American samples identified several candidate genes and loci that may be associated with AIRg. These associations explain a small component of variation in AIRg. The genes identified are involved in phosphorylation and ion transport, and provide preliminary evidence that these processes are important in beta cell response.

Keywords Beta cell response · Genetic association · Hispanic-American · Insulin secretion

Abbreviations

AIM	Ancestry-informative marker
AIRg	Acute insulin response to glucose
CEPH	Centre d'Etude du Polymorphisme Humain
DI	Disposition index
EBV	Epstein–Barr virus
FSIGT	Frequently sampled intravenous glucose tolerance test
GWAS	Genome-wide association scan
IRAS FS	Insulin Resistance Atherosclerosis Family Study
LD	Linkage disequilibrium
MAF	Minor allele frequency
PC	Principal components
S_G	Glucose effectiveness
S_I	Insulin sensitivity
SNP	Single nucleotide polymorphism

Introduction

An individual's risk of developing insulin resistance and type 2 diabetes is determined in part by genetic factors [1]. The transition from normal glucose homeostasis to type 2 diabetes is thought to be due primarily to increasingly dysfunctional beta cells [2–6]. While estimates of gene contribution to beta cell response vary, there is little argument that discovery of genes accounting for variation in the way beta cells respond to glucose could identify important pathways for type 2 diabetes risk prediction, intervention and treatment.

Candidate gene evaluations and genome-wide linkage scans have been used to search for genes with modest to large effect that contribute to variation in beta cell response; however, these methods have not been highly successful. The genome-wide association scan (GWAS) has become a popular approach for detecting genes with small to modest effect. GWAS studies in type 2 diabetes have been especially productive and had an extraordinary impact on

current understanding of genetic susceptibility to type 2 diabetes [7–13].

The vast majority of GWAS studies for type 2 diabetes have used cases and controls from European or European-derived populations. Although over a dozen genes/regions have been identified with robust statistical significance, it is unclear whether these genes/regions will also be found in populations of non-European origin [14, 15]. Currently, the genes identified for type 2 diabetes from populations of European origin have suggested that most, if not all, of the type 2 diabetes genes mediate their influence through the beta cell and not through insulin resistance pathways.

In the current report, we evaluate a quantitative, directly assessed measure of beta cell response 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 a high-density single nucleotide polymorphism (SNP) scan and follow-up genotyping, a series of genes and regions may have been identified that contribute to variation in the acute insulin response of the beta cell to a glucose challenge.

Methods

IRAS FS participants The description of the study design, recruitment and phenotyping for IRAS FS have been presented previously [16]. Briefly, the IRAS FS is a multi-centre study designed to identify the genetic determinants of quantitative measures of glucose homeostasis in African-American and Hispanic-American populations in the USA. Members of large families of self-reported Hispanic ancestry ($n=1268$ individuals in 92 pedigrees; San Antonio, TX; San Luis Valley, CO, USA) were recruited and used in this report. The Institutional Review Board of each participating clinical and analysis site approved the study protocol and all participants provided their written informed consent.

A clinical examination was performed that included an in-depth medical history interview, a frequently sampled intravenous glucose tolerance test (FSIGT), anthropometric measurements and collection of samples for blood chemistry and biomarker analysis. Measures of glucose homeostasis were derived using mathematical models [17] from glucose and insulin values obtained during the FSIGT [18–20]. These estimates of glucose homeostasis include insulin sensitivity (S_I), glucose effectiveness (S_G), disposition index (DI; with $DI = AIR \times S_I$) and the acute insulin response to glucose (AIRg). This report focuses on the AIRg phenotype of glucose homeostasis.

A collection of IRAS FS DNA samples from Hispanic-American participants (229 individuals from 34 families) was chosen from the San Antonio study group for the first

stage of the GWAS. These samples were from participants without type 2 diabetes who had complete data for glucose homeostasis and obesity phenotypes, but with an age, BMI and sex distribution consistent with that of the entire IRAS FS collection. The participants appear to be representative of a relatively homogeneous population based upon Structure (<http://pritch.bsd.uchicago.edu/structure.html>, accessed October 2008) analysis [21] using microsatellite polymorphisms from an earlier genome-wide linkage scan [22, 23]. DNA used in the high-throughput genotyping (317K SNPs) was obtained from Epstein–Barr virus (EBV)-transformed lymphoblastoid cell lines.

Genome-wide association scan Genotyping was performed at Cedars–Sinai Medical Center using 1.5 µg of genomic DNA (15 µl of 100 ng/µl stock) and Illumina technology (Illumina Infinium II HumanHap 300 BeadChips; Illumina, San Diego, CA, USA) and assay protocol [24]. Genotypes were determined based on clustering of the raw intensity data for the two dyes using Illumina Bead Studio software. Consistency of genotyping was checked using 18 repeat samples; the concordance rate was 100%. Repeat genotyping of DNA samples was performed once if the overall call rate was less than 98%; the sample was rejected if there was no improvement in call rate. The average sample call rate was 99.76%. SNPs with Hardy–Weinberg disequilibrium ($p < 0.001$), minor allele frequency (MAF) less than 0.05 or more than 5% missing genotypes were excluded from subsequent analysis. Genotypes with GenCall (Illumina) scores less than 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. Risk of strand ambiguities is minimal, as no C/G or A/T polymorphisms are included on the Illumina 300K HumanHap panel.

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 were chosen for genotyping on all Hispanic samples for which glucose homeostasis data were available ($n = 1190$). Genotyping was performed at Cedars–Sinai Medical Center using the Illumina Golden Gate Assay. SNPs with low call frequencies ($< 98\%$) were manually re-clustered ($\sim 15\%$ of all SNPs). Of the 1,536 SNPs, 3.5% were excluded due to call frequency less than 0.7 and/or cluster separation less than 0.3. The average SNP call frequency was 99.48%. Duplicate genotyping of 12 samples yielded a 100% concordance rate. 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 glucose homeostasis and related phenotype (S_1 , S_G , AIRg, DI) from the initial GWAS; (2) tag SNPs in genes with high evidence of association across more than one phenotype (tags for SNPs with MAF > 0.1 using Haploview version 4 [25]); and (3) ancestry-informative markers (AIMs) for Hispanic populations [26, 27].

Follow-up locus-specific genotyping Individual genes/regions with confirmed evidence of association from the GWAS and the validation genotyping were targeted for additional genotyping using tag SNPs. This genotyping was performed using iPLEX Gold SBE assays on a MassARRAY genotyping system (Sequenom, San Diego, CA, USA). Locus-specific primers were designed using the system's software (MassARRAY Assay Design 3.0; Sequenom). Mass spectrograms were analysed using MassARRAY Typer software (Sequenom). The minimum acceptable call frequency was 98%; no SNP failed this criterion as average call frequencies were $> 99.14\%$. We included 51 blind duplicate samples to evaluate genotyping accuracy; the concordance rate was 99.94%. SNPs were selected to capture common variation within linkage disequilibrium (LD) haplotype blocks as defined by the Centre d'Etude du Polymorphisme Humain (CEPH) (Utah residents with northern and western European ancestry) (CEU) population of the International HapMap project [28]. Specifically, genotype data from the genomic interval containing the candidate gene ± 5 kb were exported from the HapMap database and imported into Haploview. For genes with few LD blocks (e.g. *ETV7*), SNPs were selected to tag the entire genic region with a mean $r^2 = 0.80$ with forced inclusion of previously genotyped SNPs. For larger genes (e.g. *STK38*), SNP selection focused on the LD block containing the SNP associated in the validation genotyping; additional SNPs were 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 [29]. There were 1657 SNPs exhibiting inconsistencies and these genotypes 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 proportions. SNPs with no evidence of a difference in AIRg 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 AIRg, variance component measured genotype analyses were performed as implemented in SOLAR (Sequential

Oligogenic Linkage Analysis Routines) software [30]. X-Chromosome SNPs (used for ancestry adjustment) were not used in the primary analyses. For statistical testing, AIRg was transformed using the signed-square root to best approximate the distributional assumptions of the test and minimise heterogeneity of the variance. SNPs were ranked using p values from the additive genetic model. The primary statistical inference was the additive genetic model. All tests and levels of significance were computed after adjustment for age, sex and BMI.

Analyses for validation and locus-specific genotyping data used the same modelling framework as employed for analysis of the GWAS data, except that covariate adjustment included a term for the site of recruitment (San Antonio, San Luis Valley) and admixture. For incorporation of admixture into the combined analyses of GWAS and validation data, a collection of AIMs was used. These AIMs 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; these data were merged with HapMap data for CEPH ($n=90$) and Yoruba ($n=90$) populations.

A principal components (PC) analysis was performed on the 149 AIMs as well as the 80 AIMs common to the GWAS (317K SNP panel) and the validation (1,536 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, there were differences overall between the Hispanic-American sites with respect to PC2 ($p = 2.35 \times 10^{-53}$). Hispanic-Americans from the two sites differed in two measures of glucose homeostasis (S_G , $p = 2.65 \times 10^{-17}$; DI, $p = 6.22 \times 10^{-10}$) and two measures of adiposity (BMI, $p = 1.46 \times 10^{-12}$; visceral fat, $p = 7.37 \times 10^{-5}$), but not in AIRg (Spearman correlation 0.09, $p=0.09$). For AIRg, the proportion of variance explained by the centre of ascertainment was 2.56%; thus, all results are presented with adjustment for admixture in addition to age, sex, BMI and centre of ascertainment.

Results

Characteristics of the study participants This pilot GWAS studied 229 Hispanic-American participants who did not have type 2 diabetes, but had complete data for glucose homeostasis, critical covariates and DNA obtained from EBV-transformed cell lines from the San Antonio population. A sample of 961 participants with DNA and baseline data was used for replication. The total study sample of 1,190 Hispanic-American participants included the following characteristics: 58.6% women, average age 42.8 years, mean AIRg (excluding type 2 diabetic participants) 767 pmol/l and BMI 29.0 kg/m² (Table 1). There were no significant differences between the groups in these phenotypes.

Genome-wide association scan for AIRg A total of 309,200 SNPs met all quality control criteria and were evaluated for association with AIRg. 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 $-\log_{10}(p \text{ value})$ of less than 2; the observed distribution of p values matched the expectation for the majority of the observed data (Electronic supplementary material [ESM] Fig. 1). There was some departure from the null distribution at $p < 10^{-3}$, so this value was used as a rough cut-point for selection of SNPs for follow-up. The highest-ranking SNPs associated with AIRg were chosen for genotyping on all non-diabetic Hispanic-American participants in the IRAS FS ($n=1190$). We included 672 SNPs with evidence of association with AIRg and other glucose homeostasis phenotypes (S_I , S_G , DI) or SNPs that tag genes associated with multiple phenotypes in a 1,536 custom chip (another 461 SNPs were tested on the basis of association with adiposity and related phenotypes). For AIRg, 157 SNPs were chosen based on AIRg alone (125 SNPs) or AIRg plus other traits (32 SNPs). Of these 157 SNPs, 149 passed clustering-related quality control.

Candidate genes/regions for AIRg Results from the AIRg analyses (full admixture-adjusted analyses, initial GWAS and the independent replication analysis), representing the

Table 1 Characteristics of the non-diabetic Hispanic-American sample in the IRAS FS GWAS

Variable	GWAS ($n=229$)	Replication ($n=961$)	Total sample ($n=1,190$)
Male sex, n (%)	83 (36.2)	410 (42.7)	493 (41.4)
Age (years)	41.3±13.9 (40.3)	43.2±14.7* (41.6)	42.8±14.5 (41.3)
AIR _g (pmol/l)	722.9±598.1 (548.7)	780.2±678.2 (596.2)	766.9±660.5 (587.3)
Fasting glucose (mmol/l)	5.22±0.55 (5.13)	5.18±0.52 (5.11)	5.19±0.53 (5.11)
Fasting insulin (pmol/l)	112±77 (90)	103±78 (83)	104±78 (83)
BMI (kg/m ²)	29.4±5.9 (28.5)	28.9±6.3 (28.1)	29.0±6.2 (28.1)

Unless otherwise indicated, values are mean±SD (median)

* $p < 0.05$ compared with GWAS

most highly associated SNPs that reside in genes, are shown in Table 2. SNP locations within and in proximity to genes were determined by dbSNP (www.ncbi.nlm.nih.gov/projects/SNP/, dbSNP build 129, genome build 36.3; accessed 9 February 2009). The SNP most highly associated with AIRg was rs1566458 ($p = 5.5 \times 10^{-4}$) in *ZDHHC7* on chromosome 16q24.1. A second gene, *RGS6* (rs2238256, $p = 8.6 \times 10^{-4}$) on 14q24.3, encodes proteins that negatively regulate heterotrimeric G protein signalling and modulate neuronal, cardiovascular and lymphocytic activity. Three of the most associated genic SNPs for AIRg (rs1061632, rs2300064 and rs12190911) are located in the *KCTD20/STK38* region on 6p21.3, within the human MHC and near the SNP in the *NCR2* locus, which was also associated with AIRg. Only two associated SNPs were in exons (rs1061632 in *KCTD20* and rs3748400 in *ZCCHC14*).

The region on 6p21 (ESM Fig. 2) contains five blocks that span approximately 200 kb, based upon the MEX HapMap (genotyping of 71 Mexican Americans from Los Angeles, CA, USA). This region is covered by a five-SNP haplotype (rs12190911/rs1061632/rs2300064/rs612399/rs7772334). A smaller LD block is tagged by rs12190911 (admixture adjusted $p = 1.4 \times 10^{-3}$) and contains two genes, *ETV7* (also known as *TEL2*), a member of the E26 transformation-specific family of transcription factors, and *PXT1*. A larger LD block is tagged by rs1061632 ($p = 9.0 \times 10^{-4}$), rs2300064 ($p = 9.1 \times 10^{-4}$), rs612399

($p = 6.2 \times 10^{-3}$) and rs7772334 ($p = 7.8 \times 10^{-4}$), and contains the genes *KCTD20*, *STK38* and *SFRS3*. For the entire region, 22 SNPs were chosen to capture common variation ($r^2=0.98$) for follow-up analyses of these candidate genes in the entire IRAS FS Hispanic sample. The AIRg effect in the 6p21 region appears to be composed of three sets of two adjacent SNPs spanning a 105 kb region bounded by SNPs rs12190911 ($p=0.001$ between *PXT1* and *KCTD20*) and rs614028 ($p=0.04$ within *STK38*). Locations and genotypic means for these SNPs can be found in ESM Table 1.

A second region of associated SNPs that appear clustered is on chromosome 20p11.2-p11.1, including the genes *ENTPD6*, *PYGB*, *ABHD12*, *GINS1* and *RP4-691N24.1* (also known as *KIAA0980*) (ESM Fig. 3). In this case, however, only *ABHD12* did not have at least one SNP that was significantly associated with AIRg. *ENTPD6* (rs2179638, $p=0.001$), *PYGB* (rs6138553, $p=0.026$), *GINS1* (rs6076347, $p=0.0037$; rs2500406, $p=0.003$) and *RP4-691N24.1* (rs16987806, $p=0.001$; rs6083877, $p=0.035$) all had associations with variation in AIRg. Similar results are shown for SNPs that do not have recognised genes within 10 kb of the associated SNP (Table 3). The second most highly associated SNP in the entire data (rs7772334, $p = 7.8 \times 10^{-4}$) was 13.6 kb from *NCR2* on 6p21.1, which is involved in natural killer cell activity. Locations and genotypic means for these SNPs can be found in ESM Table 2.

Table 2 SNPs significantly associated with AIRg in the IRAS FS in genic regions ordered by p values obtained from combined first-stage (GWAS) and replication samples (admixture-adjusted)

SNP and gene variables				p values (additive model)		
SNP	Chr	Position	Gene	Admixture adjusted ($n=1190$)	GWAS ($n=233$) ^a	Replication ($n=957$) ^b
rs1566458	16	83576432	<i>ZDHHC7</i>	5.5×10^{-4}	1.2×10^{-4}	6.7×10^{-2}
rs2238256	14	71669976	<i>RGS6</i>	8.6×10^{-4}	5.2×10^{-4}	3.2×10^{-2}
rs1061632	6	36565462	<i>KCTD20</i>	9.0×10^{-4}	9.4×10^{-4}	3.7×10^{-2}
rs2300064	6	36582119	<i>STK38</i>	9.1×10^{-4}	7.8×10^{-4}	4.0×10^{-2}
rs214038	2	56377838	<i>CCDC85A</i>	1.4×10^{-3}	3.2×10^{-2}	1.5×10^{-2}
rs3748400	16	86003340	<i>ZCCHC14</i>	1.6×10^{-3}	4.0×10^{-5}	1.3×10^{-1}
rs10460887	3	180538829	<i>ZNF639/MFN1</i> ^c	2.3×10^{-3}	2.0×10^{-5}	ND
rs1340503	9	77761324	<i>PCSK5</i>	2.8×10^{-3}	1.0×10^{-5}	1.0×10^{-1}
rs6076347	20	25346790	<i>GINS1</i>	3.7×10^{-3}	3.8×10^{-4}	5.5×10^{-2}
rs10793057	11	72646308	<i>P2RY2/P2RY6</i>	4.5×10^{-3}	1.0×10^{-5}	2.2×10^{-1}

Genic region was defined by ± 10 kb from associated SNP

^a GWAS was adjusted for age, sex and BMI (not admixture)

^b Replication was adjusted for age, sex, centre, BMI and admixture

^c *ZNF639* (also known as *ZASCI*)

Chr, chromosome; ND, not done

Table 3 SNPs associated with AIRg not in genic regions ordered by *p* values from combined first-stage (GWAS) and replication samples (admixture-adjusted)

SNP and gene variables				<i>p</i> values		
SNP	Chr	Position	Flanking genes (distance)	Admixture adjusted (<i>n</i> =1190)	GWAS (<i>n</i> =233) ^a	Replication (<i>n</i> =957) ^b
rs7772334	6	41397828	<i>TREMI</i> (35 kb), <i>NCR2</i> (14 kb)	7.8×10^{-4}	8.4×10^{-4}	1.7×10^{-2}
rs1555286	20	24066582	<i>GGTLA4</i> (150 kb), <i>AK090900</i> (62 kb)	9.8×10^{-4}	8.2×10^{-3}	2.0×10^{-2}
rs12190911	6	36486465	<i>PXT1</i> (10 kb), <i>KCTD20</i> (32 kb)	1.4×10^{-3}	1.3×10^{-3}	9.8×10^{-2}
rs2797634	9	104680472	None within 500 kb	1.5×10^{-3}	2.0×10^{-5}	ND
rs517871	21	39831576	<i>SH3BGR</i> (22 kb), <i>B3GALT5</i> (19 kb)	3.2×10^{-3}	9.9×10^{-2}	ND
rs2828099	21	23675029	None within 500 kb	3.3×10^{-3}	5.0×10^{-5}	4.2×10^{-2}
rs16913885	10	61058374	<i>FAM13C1</i> (266 kb), <i>SLC16A9</i> (22 kb)	5.2×10^{-3}	0.4×10^{-5}	2.7×10^{-1}
rs716540	3	55179174	<i>CACNA2D3</i> (96 kb), <i>WNT5A</i> (296 kb)	5.8×10^{-3}	4.2×10^{-2}	3.4×10^{-2}
rs612399	6	41395781	None within 500 kb	6.2×10^{-3}	3.6×10^{-4}	9.9×10^{-2}
rs2893102	1	89544547	<i>GBP5</i> (33 kb), <i>GBP6</i> (58 kb)	7.8×10^{-3}	8.8×10^{-1}	3.1×10^{-2}

Non-genic region was defined by ± 10 kb from associated SNP

^a GWAS was adjusted for age, sex and BMI (not admixture)

^b Replication was adjusted for age, sex, centre, BMI and admixture

Chr, chromosome; ND, not done

Discussion

Minimal model analysis measurement of beta cell response and S_I derived from the FSIGT is highly correlated with corresponding measures obtained from euglycaemic- and hyperglycaemic-hyperinsulinaemic clamp studies [31, 32]. The hyperglycaemic clamp, unlike the euglycaemic clamp, provides direct estimates of beta cell function. Variation in these measures of glucose homeostasis, including AIRg, appears to be controlled more extensively by genetic factors than by other surrogate measures of S_I , such as fasting insulin or HOMA [33]. These variables of glucose homeostasis may represent important phenotypes for prediction of the overall and genetic risk of type 2 diabetes through the mechanism of beta cell action.

To our knowledge, this is the first report of GWAS analysis of a direct measure of beta cell response, namely AIRg. In this analysis, multiple regions of the human genome were identified as likely to harbour genes that contribute to variation in AIRg. Two regions, 6p21 and 20p11, were notable in having associations at more than one SNP. In each case, the region is complex, with multiple genes and extensive LD structure. The region on 6p21 contains *KCTD20* (proposed to participate in potassium ion transport) and *STK38* (involved in protein serine/threonine kinase activity, including ATP binding, magnesium ion binding and protein binding) [34, 35]. The *KCTD20* gene is of particular interest given the key role of potassium ion transport in glucose-stimulated insulin secretion [36]. However, the strongest region of significance on 6p21 is

centred on a 1.8 kb region without any known genes. It remains to be determined whether the observed SNP associations are due to a single or multiple causal variants. Preliminary analyses, in which each associated SNP was used as a covariate, suggest that at least two areas may contain independent variants, one associated with the *PXT1/KCTD20/STK38* region and one associated with the *TREMI/NCR2* region (ESM Table 3).

In contrast, the region associated with AIRg on chromosome 20p11 contains several interesting candidate genes. Similarly to the preliminary analyses in the candidate region on chromosome 6, conditioning on SNP rs1555286 led to little change in the significance of the *GINS1* SNP rs6076347, suggesting that at least two independent variants may exist (ESM Table 4). The *ENTPD6* gene encodes E-type NTPases (such as CD39) that participate in purine and pyrimidine metabolism, calcium ion binding, hydrolase activity, magnesium ion binding and nucleoside-diphosphatase activity [37]. The protein encoded by *PYGB* is a glycogen phosphorylase that catalyses the rate-determining step in glycogen degradation [38]. In addition, the rs6076347 SNP in *GINS1* is a missense mutation in exon 5 (Ile \rightarrow Val). Exposure to genistein, which is known to modulate hepatic glucose- and lipid-regulating enzyme activities in C57BL/KsJ-db/db mice [39], results in increased *GINS1* mRNA expression. However, it is not immediately apparent how the functions of these genes may relate to beta cell response.

Recent meta-analysis of GWAS in type 2 diabetes has uncovered at least 18 genes/regions that appear to be

common in the general population (MAFs of the variant associated with type 2 diabetes greater than 1%), but with relatively small effect [40]. A number of the associated SNPs have identified genes (*HHEX*, *CDKN2A/2B*, *CDKAL1*) that may be implicated in beta cell development and function. In this study, the most highly associated SNPs in the initial GWAS for several glucose homeostasis and obesity traits underwent an additional 1,536 SNP analysis in an expanded population sample of Hispanic-American ethnicity from San Antonio and San Luis Valley. Of these SNPs, four SNPs in two genes (at the $p < 0.01$ level for an additive model) were among the 18 type 2 diabetes-associated genes evaluated by GWAS in the IRAS FS sample. These two genes are *THADA* (rs7595299) and *CAMK1D* (rs2768367, rs2399866, rs1004247), and appear to have a common pathway involved in regulating the number of insulin-producing cells in the pancreas. The remaining 16 type 2 diabetes-associated genes evaluated by GWAS were not associated with AIRg in the current study.

While the GWAS results in the IRAS FS Hispanic-American sample have provided interesting candidates, there are limitations to the interpretation of study results. First, as there are very few comparable studies with similarly defined phenotypes of beta cell response, the ability to replicate these findings is limited. Second, while the associations are ranked on the basis of rigorous statistical criteria, the admixture-adjusted p values obtained do not meet genome-wide levels of significance (e.g. $p < 5 \times 10^{-7}$). Third, the effect sizes reflected in the genotypic means for each associated SNP average 25% of a standard deviation. In the full Hispanic-American sample of 1,269 individuals for SNPs with MAF=0.15, the power for this study to detect that effect size with a type 1 error (P) of 10^{-3} and 10^{-4} is 89% and 71%, respectively. The power to detect a similar effect at a genome-wide level of significance is only 32%. Further examination of populations and correlated phenotypes for replication are needed.

In summary, a multistage GWAS for AIRg as a measure of beta cell response was performed in a Hispanic-American sample from the IRAS FS. Based upon these results, we identified numerous SNPs and regions of the genome that may contain variants capable of accounting for common variation in AIRg. Two regions were examined with increased genotyping, with support being observed for a gene desert in chromosome 6p21 (containing *PXT1*, *KCTD20* and *STK38*) and for several genes in chromosome 20p11 (*ENTPD6*, *PYGB*, *GINS1* and *RP4-691N24.1*). The latter series of genes suggests a mechanism of action that could be involved in phosphorylation and ion transport. However, the power of the study is limited, the proportion of the observed heritability of AIRg explained by these genes is not substantial and there is need for replication. In conclusion, several candidate genes have been identified as

possibly contributing to variation in AIRg, a predictor of type 2 diabetes.

Acknowledgements This research was supported in part by: NIH grants HL060894, HL060931, HL060944, HL061019 and HL061210; the University of Virginia Harrison Chair in Public Health Sciences (S. S. Rich); the Cedars-Sinai Board of Governors' Chair in Medical Genetics (J. I. Rotter); the Cedars-Sinai Winnick Clinical Scholars Award (M. O. Goodarzi); the General Clinical Research Centers Program, NCCR grants M01RR00069 and M01RR00425; and the Diabetes Endocrinology Research Center Grant DK063491.

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

References

- Rich SS (1990) Mapping genes in diabetes. Genetic epidemiological perspective. *Diabetes* 39:1315–1319
- Warram JH, Martin BC, Krolewski AS, Soeldner JS, Kahn CR (1990) Slow glucose removal rate and hyperinsulinemia precede the development of type II diabetes in the offspring of diabetic parents. *Ann Int Med* 113:909–915
- Lillioja S, Mott DM, Spraul M et al (1993) Insulin resistance and insulin secretory dysfunction as precursors of non-insulin-dependent diabetes mellitus: prospective studies of Pima Indians. *New Engl J Med* 329:1988–1992
- Haffner SM, Miettinen H, Gaskill SP, Stern MP (1996) Decreased insulin action and insulin secretion predict the development of impaired glucose tolerance. *Diabetologia* 39:1201–1207
- Weyer C, Bogardus C, Mott DM, Pratley RE (1999) The natural history of insulin secretory dysfunction and insulin resistance in the pathogenesis of type 2 diabetes mellitus. *J Clin Invest* 104:787–794
- Hanley AJG, D'Agostino RB Jr, Wagenknecht LE et al (2002) Increased proinsulin levels and decreased acute insulin response independently predict the incidence of type 2 diabetes in the Insulin Resistance Atherosclerosis Study. *Diabetes* 51:1263–1270
- 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
- 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
- Diabetes Genetics Initiative of Broad Institute of Harvard and MIT, Lund University and Novartis Institutes for BioMedical Research (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
- 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
- Zeggini E, Scott L, 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
- 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

15. 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
16. 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
17. 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
18. Bergman RN, Ider YZ, Bowden CR, Cobelli C (1979) Quantitative estimation of insulin sensitivity. *Am J Physiol* 236:E667–E677
19. Bergman RN, Finegood DT, Ader M (1985) Assessment of insulin sensitivity in vivo. *Endocr Rev* 6:45–86
20. 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
21. Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155:945–959
22. 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
23. 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
24. Gunderson KL, Steemers FJ, Ren H et al (2006) Whole-genome genotyping. *Methods Enzymol* 410:359–376
25. Barrett JC, Fry B, Maller J, Daly MJ (2005) Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 21:263–265
26. Choudhry S, Coyle NE, Tang H et al (2006) Genetics of asthma in Latino Americans GALA Study. Population stratification confounds genetic association studies among Latinos. *Hum Genet* 118:652–664
27. 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
28. International HapMap Consortium (2005) A haplotype map of the human genome. *Nature* 437:1299–1320
29. O'Connell JR, Weeks DE (1998) PedCheck: a program for identification of genotype incompatibilities in linkage analysis. *Am J Hum Genet* 63:259–266
30. Almasy L, Blangero J (1998) Multipoint quantitative-trait linkage analysis in general pedigrees. *Am J Hum Genet* 62:1198–1211
31. 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. *Diabetes* 43:1114–1121
32. Korytkowski MT, Berga SL, Horwitz MJ (1995) Comparison of the minimal model and the hyperglycemic clamp for measuring insulin sensitivity and acute insulin response to glucose. *Metabolism* 44:1121–1125
33. 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
34. Tamaskovic R, Bichsel SJ, Rogniaux H et al (2003) Mechanism of Ca²⁺-mediated regulation of NDR protein kinase through autophosphorylation and phosphorylation by an upstream kinase. *J Biol Chem* 278:6710–6718
35. Bichel SJ, Tamaskovic R, Stegert MR et al (2004) Mechanism of activation of NDR (nuclear Dbf2-related) protein kinase by the hMOB1 protein. *J Biol Chem* 279:35228–35235
36. Koster JC, Permutt MA, Nichols CG (2005) Diabetes and insulin secretion: the ATP-sensitive K⁺ channel (K ATP) connection. *Diabetes* 54:3065–3072
37. Ivanenkov VV, Murphy-Piedmonte DM, Kirley TL (2003) Bacterial expression, characterization, and disulfide bond determination of soluble human NTPDase6 (CD39L2) nucleotidase: implications for structure and function. *Biochemistry* 42:11726–11735
38. Newgard CB, Littman DR, van Genderen C et al (1988) Human brain glycogen phosphorylase. Cloning, sequence analysis, chromosomal mapping, tissue expression, and comparison with the human liver and muscle isozymes. *J Biol Chem* 263:3850–3857
39. Ae Park S, Choi MS, Cho SY et al (2006) Genistein and daidzein modulate hepatic glucose and lipid regulating enzyme activities in C57BL/KsJ-db/db mice. *Life Sci* 79:1207–1213
40. Lango H, Palmer CN, Morris AD, The UK Type 2 Diabetes Genetics Consortium 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