GENOME-WIDE INTERACTION WITH SELECTED TYPE 2 DIABETES LOCI REVEALS NOVEL LOCI FOR TYPE 2 DIABETES IN AFRICAN AMERICANS

JACOB M. KEATON[1,2,3], JACKLYN N. HELLWEGE[2,3], MAGGIE C. Y. NG[2,3], NICHOLETTE D. PALMER[2,3,4,5], JAMES S. PANKOW[6], MYRIAM FORNAGE[7], JAMES G. WILSON[8], ADOLFO CORREA[8], LAURA J. RASMUSSEN-TORVIK[9], JEROME I. ROTTER[10], YII-DER I. CHEN[10], KENT D. TAYLOR[10], STEPHEN S. RICH[11], LYNNE E. WAGENKNECHT[5,12], BARRY I. FREEDMAN[3,5,13], DONALD W. BOWDEN[2,3,4]

[1]*Molecular Genetics and Genomics Program, Wake Forest School of Medicine, Medical Center Blvd, Winston-Salem, NC, 27157, US*

[2]*Center for Genomics and Personalized Medicine Research, Wake Forest School of Medicine, Medical Center Blvd, Winston-Salem, NC, 27157, US*

[3]*Center for Diabetes Research, Wake Forest School of Medicine, Wake Forest School of Medicine, Medical Center Blvd, Winston-Salem, NC, 27157, US*

[4]*Department of Biochemistry, Wake Forest School of Medicine, Medical Center Blvd, Winston-Salem, NC, 27157, US*

[5]*Center for Public Health Genomics, Wake Forest School of Medicine, Medical Center Blvd, Winston-Salem, NC, 27157, US*

[6]*Division of Epidemiology and Community Health, University of Minnesota, Minneapolis, MN, 55455, US*

[7]*Institute of Molecular Medicine and Human Genetics Center, University of Texas Health Science Center at Houston, 7000 Fannin St #1200, Houston, TX, 77030, US*

[8]*University of Mississippi Medical Center, 2500 N State St, Jackson, MS, 39216, US*

[9]*Department of Preventive Medicine, Northwestern University Feinberg School of Medicine, 303 E Chicago Ave, Chicago, IL, 60611, US*

[10]*Institute for Translational Genomics and Population Sciences, Los Angeles BioMedical Research Institute, Harbor-UCLA Medical Center, 1000 W Carson St, Torrance, CA, 90502, US*

[11]*Center for Public Health Genomics, University of Virginia, Charlottesville, VA, 22904, US*

[12]*Division of Public Health Sciences, Wake Forest School of Medicine, Medical Center Blvd, Winston-Salem, NC, 27157, US*

[13]*Department of Internal Medicine - Section on Nephrology, Wake Forest School of Medicine, Medical Center Blvd, Winston-Salem, NC, 27157, US*

Type 2 diabetes (T2D) is the result of metabolic defects in insulin secretion and insulin sensitivity, yet most T2D loci identified to date influence insulin secretion. We hypothesized that T2D loci, particularly those affecting insulin sensitivity, can be identified through interaction with known T2D loci implicated in insulin secretion. To test this hypothesis, single nucleotide polymorphisms (SNPs) nominally associated with acute insulin response to glucose (AIR_g), a dynamic measure of first-phase insulin secretion, and previously associated with T2D in genome-wide association studies (GWAS) were identified in African Americans from the Insulin Resistance Atherosclerosis Family Study (IRASFS; n=492 subjects). These SNPs were tested for interaction, individually and jointly as a genetic risk score (GRS), using GWAS data from five cohorts (ARIC, CARDIA, JHS, MESA, WFSM; n=2,725 cases, 4,167 controls) with T2D as the outcome. In single variant analyses, suggestively significant ($P_{interaction} < 5\times10^{-6}$) interactions were observed at several loci including *DGKB* (rs978989), *CDK18* (rs12126276), *CXCL12* (rs7921850), *HCN1* (rs6895191), *FAM98A* (rs1900780), and *MGMT* (rs568530). Notable beta-cell GRS interactions included two SNPs at the *DGKB* locus (rs6976381; rs6962498). These data support the hypothesis that additional genetic factors contributing to T2D risk can be identified by interactions with insulin secretion loci.

1. Introduction

Although common variants examined in genome-wide association studies (GWAS) have identified ~80 loci associated with T2D risk, these variants explain only about 15% of T2D heritability[1,2]. A portion of the missing heritability may be explained by epistasis, which occurs when a genetic risk factor is modified by other factors in an individual's genetic background[3]. Epistasis, or gene-gene interaction, analyses may facilitate the detection of novel loci when non-additive effects exist, but may also provide novel insights illuminating biological mechanisms underlying complex diseases such as T2D[4].

T2D is characterized by impaired insulin secretion arising from pancreatic beta-cell dysfunction and insulin resistance in skeletal muscle, hepatic, and other peripheral tissues, leading to decreased plasma glucose uptake. However, documented T2D loci primarily map to genes influencing insulin secretion or other aspects of beta-cell biology[1]. Given the underlying bimodal pathophysiology, T2D may be a particularly well-suited disease model for hypothesis-driven investigation of epistatic interactions. Genetic insults to both insulin secretion and insulin sensitivity may jointly increase an individual's T2D risk in a non-additive manner. Considering the higher prevalence rate of T2D, insulin resistance, and obesity, African Americans are optimal for the study of genetic interactions that contribute to T2D risk.

In an effort to identify interactions contributing to T2D and to discover novel insulin sensitivity loci, we hypothesized that T2D risk loci, particularly those affecting insulin sensitivity, could be identified by interaction analyses with known T2D loci implicated in insulin secretion. In cross-sectional meta-analyses of five T2D studies (ARIC, CARDIA, JHS, MESA, and WFSM), we tested whether 5 SNPs from known T2D loci implicated in insulin secretion, or a genetic risk score summarizing these SNPs, modified genome-wide SNP associations with T2D risk.

2. Research Design and Methods

2.1 *Subjects*

Two sources of data were analyzed in this study. Primary inferences of association with insulin secretion were derived from African American participants (n=492 individuals from 42 families) in the Insulin Resistance Atherosclerosis Family Study (IRASFS), a metabolically well-characterized cohort[5]. Glucose homeostasis traits were measured by the frequently sampled intravenous glucose tolerance test (FSIGT)[5]. Briefly, a 50% glucose solution (0.3g/kg) and regular human insulin (0.03units/kg) were injected intravenously at 0 and 20 minutes, respectively. Blood was collected at −5, 2, 4, 8, 19, 22, 30, 40, 50, 70, 100, and 180 minutes for measurement of plasma glucose and insulin. AIR_g was calculated as the increase in insulin at 2–8 minutes above the basal (fasting) insulin level after the bolus glucose injection at 0-1 minute. Insulin sensitivity (S_I) was calculated by mathematical modeling using the MINMOD program (version 3.0 [1994])[6]. Disposition index (DI) was calculated as the product of S_I and AIR_g.

Inferences of genome-wide epistatic interaction with insulin secretion loci for T2D susceptibility were derived from African American participants from the Atherosclerosis Risk in Communities Study (ARIC; n = 955 T2D cases, 414 controls), Coronary Artery Risk Development in Young Adults (CARDIA; n = 94 T2D cases, 654 controls), Jackson Heart Study (JHS; n = 333 T2D cases, 1,450 controls), Multi-Ethnic Study of Atherosclerosis (MESA; n = 411 T2D cases, 793 controls), and the Wake Forest School of Medicine (WFSM; n = 932 T2D cases, 856 controls) cohorts for a total of 2,725 T2D cases and 4,167 controls[7–12]. T2D was diagnosed according to the American Diabetes Association criteria with at least one of the following: fasting glucose ≥126 mg/dL, 2-h oral glucose tolerance test glucose ≥200 mg/dL, random glucose ≥200 mg/dL, use of oral hypoglycemic agents and/or insulin, or physician diagnosed diabetes. Subjects diagnosed before 25 years of age were excluded. Normal glucose tolerance was defined as fasting glucose <100 mg/dL and 2-h oral glucose tolerance test glucose <140 mg/dL (if available) without reported use of diabetes medications. Control subjects <25 years of age were excluded.

IRB approval was obtained at all sites and all participants provided written informed consent. Descriptions of the T2D study cohorts are summarized in the Supplementary Methods.

2.2 *Genotyping, imputation, and quality control*

For the IRASFS samples, genotyping and quality control were performed at the Wake Forest Center for Genomics and Personalized Medicine Research using the Illumina Infinium HumanExome BeadChip v1.0 as previously described[13]. Briefly, the exome chip contained 247,870 variants (92% protein coding). In addition, the chip included 64 SNPs associated with T2D from previous GWAS in Europeans, many of which have been implicated in insulin secretion (exome chip design: http://genome.sph.umich.edu/wiki/Exome_Chip_Design). Sample and autosomal SNP call rates were ≥99%, and SNPs with poor cluster separation (<0.35) were excluded. Mendelian errors were identified using PedCheck[14] and resolved by removing conflicting genotypes. Hardy–Weinberg Equilibrium (HWE) was assessed in unrelated samples (n = 39) using PLINK (http://pngu.mgh.harvard.edu/purcell/plink)[15] to reduce biases introduced by familial allele frequencies. All variants were in accordance with HWE ($P > 1x10^{-5}$).

The T2D study samples were genotyped using the Affymetrix Genome-Wide Human SNP Array 6.0. For the ARIC, CARDIA, JHS, and MESA cohorts, genotyping and quality control were completed by the National Heart, Lung, and Blood Institute's (NHLBI's) Candidate Gene Association Resource (CARe) at the Broad Institute[16]. Genotyping for the WFSM study was performed at the Center for Inherited Disease Research (CIDR). For all T2D studies, imputation was performed using MACH with the function –mle (version 1.0.16, http://www.sph.umich.edu/csg/abecasis/MaCH/) to obtain missing genotypes and replace genotypes inconsistent with reference haplotypes as previously described[17]. SNPs with call rate ≥ 95% and minor allele frequency (MAF) ≥ 1% that passed study-specific quality control were used for imputation[16,18]. A 1:1 HapMap II (NCBI Build 36) CEU:YRI (European:African) consensus haplotype was used as reference. A total of 2,713,329 to 2,907,086 autosomal SNPs from each GWAS with call rate ≥95%, MAF ≥ 1%, and Hardy-Weinberg P-value ≥ 0.0001 for genotyped SNPs and MAF ≥ 1% and RSQ ≥ 0.5 for imputed SNPs were included in subsequent data analyses.

2.3 *Principal component analysis*

For IRASFS, admixture was estimated using principal components (PCs) from 39 ancestry informative markers (AIMs) and including HapMap CEU and YRI samples for comparison[19]. Only PC1 correlated with HapMap populations, and was thus used as a covariate in all analyses.

For the T2D studies, PCs were computed for each study using high-quality SNPs as previously described[13,16–18,20]. The first PC was highly correlated ($r^2 > 0.87$) with global African-European ancestry, as measured by ANCESTRYMAP[21], STRUCTURE[22], or FRAPPE[23]. The African American T2D study samples had an average of 80% African ancestry. By analyzing unrelated samples from all studies using SMARTPCA[20], only the first PC appeared to account for substantial genetic variation (data not shown), whereas the subsequent PCs may reflect sampling noise and/or relatedness in samples[21]. The first PC (PC1) was used as a covariate in all analyses to adjust for population substructure.

2.4 *Analysis of association with measures of glucose homeostasis in IRASFS*

To approximate a normal distribution, trait values were transformed by square root (AIR_g, DI) or natural logarithm plus a constant (S_I). Measured genotype association analyses of exome chip variants with AIR_g, S_I, and DI were performed under an additive model using the variance components method implemented in Sequential Oligogenic Linkage Analysis Routines (SOLAR)[24] with adjustment for age, gender, body mass index (BMI), and PC1.

2.5 *Genetic risk score construction*

We further explored our interaction approach by constructing genetic risk scores (GRS), both weighted and unweighted, summarizing the effects of SNPs associated with both T2D and insulin secretion (T2D-IS SNPs). The T2D-IS GRS was created using the T2D risk alleles for T2D-IS SNPs defined from the literature (Table 1). The unweighted risk score was calculated by summation of the number of risk alleles for each individual across all selected SNPs. The weighted T2D-IS GRS was calculated as the sum of risk alleles at each locus multiplied by the

natural log of their T2D odds ratio (OR) defined from the literature[2,25–28]. Missing genotypes for a given SNP were imputed as the average number of risk alleles across all samples. The association of each GRS with both AIR_g and DI, a combinatorial measure of first-phase insulin secretion and insulin sensitivity, were evaluated in IRASFS using the variance components method implemented in SOLAR[24] adjusted for age, gender, and ancestry proportions.

Table 1. Characteristics and single-SNP AIR_g association results for T2D-IS SNPs in published GWAS and IRASFS

T2D-IS SNP	Chr	Position*	Gene	Published GWAS				IRASFS AIR_g			
				T2D Risk Allele	Other Allele	T2D OR†	PMID‡	RAF§	Beta	SE‖	P
rs7593730	2	161171454	RBMS1	T	C	1.11	20418489	0.39	-1.38	0.86	0.086
rs864745	7	28180556	JAZF1	T	C	1.10	18372903	0.72	-1.52	0.91	0.096
rs5215	11	17408630	KCNJ11	C	T	1.08	24509480	0.15	-2.60	1.18	0.033
rs1552224	11	72433098	ARAP1	A	C	1.14	20581827	0.06	-3.05	1.69	0.077
rs7119	15	77777632	HMG20A	C	T	1.24	22885922	0.52	-1.50	0.81	0.059

*NCBI build 37. †Reported odds ratio. ‡PubMed ID. §Risk allele frequency. ||Standard error.

2.6 *Analysis of interaction for T2D risk in the African American T2D case-control studies*

A logistic regression test for additive allelic interaction adjusted for age, gender, and PC1 was used for all interaction analyses with T2D as the outcome. Additional models included adjustment for BMI, and individuals with missing values were excluded (n = 110). In each study, genome-wide interaction tests were performed in PLINK between each SNP in the genome with each candidate SNP (i.e. insulin secretion SNP) and GRS (i.e. insulin secretion risk score). An example PLINK command is provided in the Supplementary Methods. Interaction results with extreme values (absolute β or SE > 10), primarily due to low cell counts, were excluded. Across interaction analyses with all SNPs and risk scores, the number of SNPs excluded as outliers ranged from 0 to 17,000. Interaction results were combined by fixed-effect inverse variance weighting for each candidate SNP or GRS in METAL (http://www.sph.umich.edu/csg/abecasis/metal/). Each meta-analysis contained results for 486,148 to 2,965,304 SNPs.

3. Results

3.1 *Candidate beta-cell function SNP selection*

The characteristics of IRASFS subjects are shown in Supplementary Table 1. Samples included 492 African Americans with mean age 41.2 years and mean BMI 29.1 kg/m^2. Average African ancestry proportion was 0.75. FSIGT was performed for all subjects without T2D (n = 492) to assess measures including insulin secretion (AIR_g), insulin sensitivity index (S_I), and disposition index (DI).

We identified 5 SNPs (Table 1) from established T2D risk loci from published GWAS[25–28] in which the T2D risk alleles were trending towards association ($P < 0.10$) with AIR_g in IRAS-FS (T2D-IS SNPs). Selected SNPs were identical to the published T2D GWAS index SNPs with the exception of rs7119 (*HMG20A*), which is in strong linkage disequilibrium with the GWAS

index SNP rs7178572 in the current study ($r^2 \geq 0.73$ in all cohorts) and is suggestively associated with T2D ($P = 5.24x10^{-7}$) in individuals from Southeast Asia[29].

3.2 *Interaction analysis*

The selected SNPs were examined for genome-wide first order multiplicative interactions with 1) individual insulin secretion SNPs and 2) risk scores summarizing these insulin secretion SNPs. To maximize power, these analyses were performed in five studies (ARIC, CARDIA, JHS, MESA, and WFSM) including 2,725 T2D cases and 4,167 non-diabetic controls and results were meta-analyzed. Representative meta-analysis q-q plots are provided in Supplementary Figures 1 and 2. A flowchart summarizing experimental workflow is provided in Supplementary Figure 3.

The characteristics of T2D case (n = 2,725) and control subjects (n = 4,167) for each study cohort are shown in Supplementary Table 2. Mean age at examination ranged from 38.2 (CARDIA) to 67.6 (MESA) years. Mean age at diagnosis for T2D cases ranged from 35.0 (CARDIA) to 54.6 (MESA) years. In all cohorts except WFSM, BMI was >3 kg/m^2 higher in cases compared to controls.

3.3 *T2D-IS SNP interactions*

Five T2D-IS SNPs were tested for genome-wide interactions for T2D risk in the ARIC, CARDIA, JHS, MESA, and WFSM cohorts. Individual T2D-IS SNP results were meta-analyzed across cohorts. While no interactions were observed at a genome-wide significance level, a total of 21 SNP-pairs demonstrated suggestive evidence of interaction ($P_{interaction} < 5x10^{-6}$; Table 2). The most significant T2D-IS SNP interaction observed was between rs7119 at the *HMG20A* locus (T2D-IS SNP) and rs6487610 (interacting SNP; $P_{interaction} = 3.83x10^{-7}$). This interacting SNP is located in an intron of *SMCO2*, which encodes single-pass membrane protein with coiled-coil domains 2. Top interactions with T2D-IS SNPs overall were robust against BMI adjustment (Table 2), with similar p-values. Other notable interacting SNPs included rs978989 (*DGKB*), rs12126276 (*CDK18*), rs7921850 (*CXCL12*), rs6895191 (*HCN1*), rs1900780 (*FAM98A*), and rs568530 (*MGMT*).

Table 2. Top meta-analyzed interactions with T2D-IS SNPs regressed on T2D risk in ARIC, CARDIA, JHS, MESA, and WFSM

T2D-IS SNP (Gene)	Intxn SNP* (Gene)	Chr	Position†	MAF‡	β_{intxn}§	P_{intxn}§	P_{het}‖	$\beta_{intxn\_adj\_bmi}$¶	$P_{intxn\_adj\_bmi}$¶
rs5215 (*KCNJ11*)	rs3024370 (*F13A1*)	6	6250967	0.48	-0.52	3.01E-06	0.71	-0.56	2.32E-06
rs5215 (*KCNJ11*)	rs7842913 (*FUT10*)	8	33089041	0.07	-2.77	4.58E-06	1.00	-2.75	4.57E-06
rs7119 (*HMG20A*)	rs12121207 (*ATG4C*)	1	63232384	0.44	-0.29	2.68E-06	0.20	-0.28	1.43E-05
rs7119 (*HMG20A*)	rs1900780 (*FAM98A*/*MYADML*)	2	33901094	0.33	0.36	3.46E-06	0.76	0.37	6.92E-06
rs7119 (*HMG20A*)	rs978989 (*DGKB*)	7	14954759	0.27	0.33	2.72E-06	0.23	0.33	4.27E-06
rs7119 (*HMG20A*)	rs6487610 (*SMCO2*)	12	27628742	0.38	0.32	3.83E-07	0.42	0.32	8.45E-07
rs7119 (*HMG20A*)	rs7965793 (*ANKS1B*)	12	100175468	0.31	0.44	1.05E-06	0.76	0.47	7.74E-07
rs7119 (*HMG20A*)	rs1496811 (Intergenic)	18	38952563	0.49	0.27	4.95E-06	0.98	0.27	1.24E-05
rs7119 (*HMG20A*)	rs4812424 (Intergenic)	20	38654372	0.35	-0.47	4.68E-07	0.14	-0.46	1.51E-06
rs7119 (*HMG20A*)	rs6105151 (*ESF1*)	20	13691752	0.34	0.30	2.08E-06	0.42	0.32	7.23E-07
rs7593730 (*RBMS1*)	rs6895191 (*HCN1*)	5	45877674	0.28	0.32	2.80E-06	0.39	0.32	6.91E-06

rs7593730 (*RBMS1*)	rs4705321 (*SH3TC2*/*ABLIM3*)	5	148508860	0.31	0.30	4.13E-06	0.58	0.28	2.91E-05
rs7593730 (*RBMS1*)	rs16872382 (*ZFPM2*)	8	106108691	0.03	-0.97	7.34E-07	0.85	-0.99	8.49E-07
rs7593730 (*RBMS1*)	rs12865410 (Intergenic)	13	104785227	0.35	-0.30	9.69E-07	0.46	-0.32	6.44E-07
rs7593730 (*RBMS1*)	rs12863474 (Intergenic)	13	104784409	0.37	0.33	1.29E-06	0.89	0.36	4.48E-07
rs864745 (*JAZF1*)	rs12126276 (*CDK18*)	1	205494508	0.18	-0.92	1.31E-06	0.68	-0.92	2.98E-06
rs864745 (*JAZF1*)	rs12343907 (*GLT6D1*)	9	138498904	0.35	-0.34	1.44E-06	0.87	-0.34	2.04E-06
rs864745 (*JAZF1*)	rs7921850 (*CXCL12*)	10	44704401	0.37	-0.33	2.52E-06	0.56	-0.31	1.37E-05
rs864745 (*JAZF1*)	rs568530 (*MGMT*)	10	131018864	0.41	0.32	3.27E-06	0.30	0.32	1.03E-05
rs864745 (*JAZF1*)	rs16973790 (*WRD72*/*UNC13C*)	15	54188148	0.15	0.55	3.13E-06	0.27	0.51	3.09E-05
rs864745 (*JAZF1*)	rs12483006 (*SLC37A1*)	21	43953851	0.07	-0.66	1.95E-06	0.58	-0.64	8.17E-06

*SNP interacting with selected T2D-IS SNP. †NCBI build 37. ‡Minor allele frequency. §Meta-analyzed effect size and p-value from interaction models adjusted for age, gender, and PC1. ||Heterogeneity p-values across studies from interaction models adjusted for age, gender, and PC1. ¶ Meta-analyzed effect size and p-value from interaction models adjusted for age, gender, PC1, and BMI.

3.4 *GRS validation and interaction analysis*

Each GRS was tested for association with AIR_g and DI under an additive model using the variance components method with adjustment for age, gender, and PC1 in IRASFS (Supplementary Table 3). The weighted T2D-IS GRS was not associated with AIR_g; it was associated with DI with or without BMI adjustment ($P = 4.43\text{x}10^{-2}$ and $4.51\text{x}10^{-2}$, respectively). Since the weighted risk score was associated with measures of glucose homeostasis, analysis of this risk score was emphasized in the tests for genome-wide interaction in the ARIC, CARDIA, JHS, MESA, and WFSM cohorts.

Meta-analyzed estimates of genome-wide interactions with the weighted T2D-IS GRS are presented in Table 3. No interactions met conventional GWAS thresholds for significance. However, eight interactions with the weighted T2D-IS GRS reached a suggestive level of significance ($P_{interaction} < 5\text{x}10^{-6}$; Table 3). The most significant T2D-IS GRS interaction was with rs12434405 (Table 3, $P_{interaction} = 9.60\text{x}10^{-7}$). This is an intronic SNP in the gene *CEP128*, which encodes centrosomal protein 128kDa. Further, the T2D-IS GRS interaction analysis identified two SNPs at the *DGKB* locus, rs6976381 and rs6962498 ($r^2 \geq 0.75$ in all cohorts). This locus was identified in single variant interaction analyses with T2D-IS SNP rs7119 (*HMG20A*), though through a different interacting SNP (rs978989). Two SNPs at the *FAM98A* locus, rs6543772 and rs11687252, were also identified in this analysis. This locus was implicated in single variant analyses with T2D-IS SNP rs7119 (*HMG20A*) through the interacting SNP rs1900780. Top interactions with the T2D-IS GRS were also robust against BMI adjustment.

Table 3. Top meta-analyzed interactions with weighted T2D-IS GRS regressed on T2D risk in ARIC, CARDIA, JHS, MESA, and WFSM

Intxn SNP* (Gene)	Chr	Position†	MAF‡	β_{intxn}§	P_{intxn}§	P_{het}‖	$\beta_{intxn\_adj\_bmi}$¶	$P_{intxn\_adj\_bmi}$¶
rs6543722 (*FAM98A*)	2	33832523	0.39	-1.20	2.82E-06	0.79	-1.22	3.52E-06
rs11687252 (*FAM98A*)	2	33834496	0.38	-1.17	3.27E-06	0.68	-1.19	3.70E-06
rs6851672 (*DKK2*)	4	107907908	0.03	3.70	4.79E-06	0.82	3.63	9.62E-06
rs6976381 (*DGKB*)	7	15048814	0.18	-1.67	1.21E-06	0.73	-1.66	2.18E-06

rs6962498 (*DGKB*)	7	15050305	0.14	-1.77	3.71E-06	0.54	-1.77	6.65E-06
rs17082105 (*PCDH9*)	13	67685156	0.18	1.45	3.46E-06	0.86	1.51	2.65E-06
rs12434405 (*CEP128*)	14	81044614	0.12	-1.90	9.60E-07	0.12	-1.87	2.49E-06
rs16951940 (Intergenic)	16	80021664	0.03	3.40	2.29E-06	0.84	3.43	4.58E-06

*SNP interacting with the weighted T2D-IS GRS. †NCBI build 37. ‡Minor allele frequency. §Meta-analyzed effect size and p-value from interaction models adjusted for age, gender, and PC1. ||Heterogeneity p-values across studies from interaction models adjusted for age, gender, and PC1. ¶ Meta-analyzed effect size and p-value from interaction models adjusted for age, gender, PC1, and BMI.

4. Discussion

Meta-analyses of five African American T2D studies did not reveal genome-wide statistically significant ($P_{\text{interaction}} < 5\times10^{-8}$) first-order interactions with insulin secretion SNPs or composite risk scores. However, the observed interactions ($P_{\text{interaction}} < 5\times10^{-6}$) suggest that a candidate insulin secretion SNP/GRS interaction approach is a valid method for identifying insulin sensitivity and T2D risk loci. For example, analyses with the T2D-IS SNP rs864745 (*JAZF1*) revealed an interaction with rs7921850, an intergenic SNP downstream of the *CXCL12* gene encoding chemokine (C-X-C motif) ligand 12 (also known as stromal cell-derived factor 1). CXCL12 is an adipocyte-derived chemotactic factor that recruits macrophages and is required for the establishment of obesity-induced adipose tissue inflammation and systemic insulin resistance in mice[30].

Several genes related to pancreatic beta-cell function were also identified; suggesting interactions are not limited to insulin resistance as in our initial hypothesis. Evaluations of the T2D-IS SNP rs7119 (*HMG20A*) and the T2D-IS GRS identified interactions with rs978989 and rs6976381, respectively, intergenic SNPs downstream of the *DGKB* gene. Variants at *DGKB* have been associated with T2D, fasting glucose, and pancreatic islet beta-cell function as measured by HOMA-B[27,31]. Variants near *DGKB* disrupt islet-specific enhancer activity[32]. Several other variants detected in our analyses show interactions with similar biological relationships to insulin secretion and T2D.

Interestingly, we observed interactions discrete for individual loci. For example, analyses with rs864745 (*JAZF1*), a locus involved in transcriptional repression, showed an interaction with rs568530, an intergenic SNP upstream of *MGMT*, which encodes O-6-Methylguanine-DNA Methyltransferase. These observations may reflect different, input-dependent physiological characteristics of interaction results, and may lead to mechanistic insights about the underlying causes of T2D and defects in glucose homeostasis in expanded analyses.

Although results varied widely between interaction analyses, interactions with two loci, *DGKB* and *FAM98*, were replicated in multiple analyses. Functional characteristics of *FAM98* related to T2D and glucose homeostasis pathophysiology are not evident in the current literature.

Previous GWAS have largely ignored epistatic contributions to T2D risk due to the heavy multiple testing burden and computational challenges of exhaustive analytical approaches, and when they have considered this contribution, results have not been striking. For example, a recent genome-wide scan for two-locus interactions in the Wellcome Trust Case Control Consortium T2D GWAS data did not reveal any significant epistatic signals at a Bonferroni-

corrected p-value threshold of $2.14x10^{-11}$ after adjusting for the main effects of the most strongly associated T2D locus, *TCF7L2*[33]. Further, Herold et al. estimated that analysis of all pairwise interactions among 550,000 SNPs in 1,200 samples on a 3 GHz computer would require a running time of 120 days[34]. The interaction analysis presented here overcomes the issue of a heavy multiple testing burden by using a candidate SNP approach. A recent study by Becker et al. demonstrated that a multiple test correction of 0.4m, where m is the number of SNP pairs tested, is sufficiently conservative for large-scale allelic interaction tests[35]. Further, Babron et al. show that a correction for the effective number of SNP pairs is equally sufficient[36]. Li et al. previously demonstrated that the effective number of SNPs for an imputed dataset is ~10^6. These findings suggest that a significance threshold of $1x10^{-8}$ is appropriate for this study.

We did not detect interactions even at the conventional GWAS threshold of $5x10^{-8}$ in the current study. In part, this likely reflects the challenge of inherently reduced power of interaction models due to the low frequency of compound genotypes[37]. Computational resources required for this study were equivalent to the requirements for running 12 GWAS (5 candidate insulin secretion SNPs plus a GRS, with and without BMI adjustment). This is a significant reduction compared to exhaustive approaches examining genome-wide interactions with all available SNP pairs.

In summary, our findings demonstrate that genome-wide interaction studies with selected insulin secretion variants is a powerful approach for the detection of T2D risk, insulin secretion, and insulin sensitivity loci. The use of a high-quality measure of first-phase insulin secretion, AIR_g, to identify candidate interaction SNPs yielded compelling associations. These results justify an expansion of the current study and further investigation of putative insulin sensitivity loci, namely *CXCL12*.

Acknowledgements

The authors would like to acknowledge the contributions of the involved research institutions, study investigators, field staff, and study participants of ARIC, CARDIA, JHS, MESA, and WFSM.

Genotyping services for the WFSM study were provided by CIDR. CIDR is fully funded through a federal contract from the National Institutes of Health (NIH) to The Johns Hopkins University (Contract HHSC268200782096C). The work at Wake Forest was supported by NIH grants K99-DK-081350 (N.D.P.), R01-DK-066358 (D.W.B.), R01-DK-053591 (D.W.B.), R01-HL-56266 (B.I.F.), and R01-DK-070941 (B.I.F.), and in part by the General Clinical Research Center of the WFSM Grant M01-RR-07122. This work was also supported by the NHLBI.

The following four parent studies have contributed parent study data, ancillary study data, and DNA samples through the Massachusetts Institute of Technology-Broad Institute (N01-HC-65226) to create this genotype/phenotype database for wide dissemination to the biomedical research community: ARIC, CARDIA, JHS, and MESA.

The Atherosclerosis Risk in Communities (ARIC) Study is carried out as a collaborative study supported by National Heart, Lung, and Blood Institute contracts (HHSN268201100005C, HHSN268201100006C, HHSN268201100007C, HHSN268201100008C, HHSN268201100009C, HHSN268201100010C, HHSN268201100011C, and HHSN268201100012C), R01HL087641, R01HL59367 and R01HL086694; National Human

Genome Research Institute contract U01HG004402; and National Institutes of Health contract HHSN268200625226C. The authors thank the staff and participants of the ARIC study for their important contributions. Infrastructure was partly supported by Grant Number UL1RR025005, a component of the National Institutes of Health and NIH Roadmap for Medical Research. The authors thank the staff and participants of the ARIC study for their important contributions.

The Coronary Artery Risk Development in Young Adults (CARDIA) Study is conducted and supported by the National Heart, Lung, and Blood Institute in collaboration with the University of Alabama at Birmingham (HHSN268201300025C & HHSN268201300026C), Northwestern University (HHSN268201300027C), University of Minnesota (HHSN268201300028C), Kaiser Foundation Research Institute (HHSN268201300029C), and Johns Hopkins University School of Medicine (HHSN268200900041C). CARDIA is also partially supported by the Intramural Research Program of the National Institute on Aging. Genotyping was funded as part of the NHLBI Candidate-gene Association Resource (N01-HC-65226) and the NHGRI Gene Environment Association Studies (GENEVA) (U01-HG004729, U01-HG04424, and U01-HG004446). This manuscript has been reviewed and approved by CARDIA for scientific content.

The Jackson Heart Study (JHS) is supported by contracts HHSN268201300046C, HHSN268201300047C, HHSN268201300048C, HHSN268201300049C, HHSN268201300050C from the National Heart, Lung, and Blood Institute and the National Institute on Minority Health and Health Disparities. The authors thank the participants and data collection staff of the Jackson Heart Study.

Multi-Ethnic Study of Atherosclerosis (MESA), and the MESA SHARe project are conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with MESA investigators. Support for MESA is provided by contracts N01-HC-95159, N01-HC-95160, N01-HC-95161, N01-HC-95162, N01-HC-95163, N01-HC-95164, N01-HC-95165, N01-HC-95166, N01-HC-95167, N01-HC-95168, N01-HC-95169, UL1-TR-001079, UL1-TR-000040, and DK063491. The MESA CARe data used for the analyses described in this manuscript were obtained through Genetics (CMP00068). Funding for CARe genotyping was provided by NHLBI Contract N01-HC-65226.

The views expressed in this manuscript are those of the authors and do not necessarily represent the views of the National Heart, Lung, and Blood Institute; the National Institutes of Health; or the U.S. Department of Health and Human Services.

Supplementary Material

Supplementary methods, tables, and figures can be found at http://csb.wfu.edu/SupplementaryData_online.docx.

References

1. Prasad, R. B. & Groop, L. Genetics of type 2 diabetes-pitfalls and possibilities. *Genes* **6,** 87–123 (2015).
2. Morris, A. P. *et al.* Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat. Genet.* **44,** 981–990 (2012).

3. Cordell, H. J. Detecting gene-gene interactions that underlie human diseases. *Nat. Rev. Genet.* **10,** 392–404 (2009).
4. Moore, J. H. The ubiquitous nature of epistasis in determining susceptibility to common human diseases. *Hum. Hered.* **56,** 73–82 (2003).
5. Henkin, L. *et al.* Genetic epidemiology of insulin resistance and visceral adiposity. The IRAS Family Study design and methods. *Ann. Epidemiol.* **13,** 211–217 (2003).
6. Pacini, G. & Bergman, R. N. MINMOD: a computer program to calculate insulin sensitivity and pancreatic responsivity from the frequently sampled intravenous glucose tolerance test. *Comput. Methods Programs Biomed.* **23,** 113–122 (1986).
7. The ARIC Investigators. The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. The ARIC investigators. *Am. J. Epidemiol.* **129,** 687–702 (1989).
8. Friedman, G. D. *et al.* CARDIA: study design, recruitment, and some characteristics of the examined subjects. *J. Clin. Epidemiol.* **41,** 1105–1116 (1988).
9. Taylor, H. A. *et al.* Toward resolution of cardiovascular health disparities in African Americans: design and methods of the Jackson Heart Study. *Ethn. Dis.* **15,** S6-4–17 (2005).
10. Bild, D. E. *et al.* Multi-ethnic study of atherosclerosis: objectives and design. *Am. J. Epidemiol.* **156,** 871–881 (2002).
11. McDonough, C. W. *et al.* A genome-wide association study for diabetic nephropathy genes in African Americans. *Kidney Int.* **79,** 563–572 (2011).
12. Palmer, N. D. *et al.* A genome-wide association search for type 2 diabetes genes in African Americans. *PloS One* **7,** e29202 (2012).
13. Hellwege, J. N. *et al.* Genome-wide family-based linkage analysis of exome chip variants and cardiometabolic risk. *Genet. Epidemiol.* **38,** 345–352 (2014).
14. O'Connell, J. R. & Weeks, D. E. PedCheck: a program for identification of genotype incompatibilities in linkage analysis. *Am. J. Hum. Genet.* **63,** 259–266 (1998).
15. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81,** 559–575 (2007).
16. Lettre, G. *et al.* Genome-wide association study of coronary heart disease and its risk factors in 8,090 African Americans: the NHLBI CARe Project. *PLoS Genet.* **7,** e1001300 (2011).
17. Ng, M. C. Y. *et al.* Transferability and fine mapping of type 2 diabetes loci in African Americans: the Candidate Gene Association Resource Plus Study. *Diabetes* **62,** 965–976 (2013).
18. Hester, J. M. *et al.* Implication of European-derived adiposity loci in African Americans. *Int. J. Obes. 2005* **36,** 465–473 (2012).
19. Palmer, N. D. *et al.* Evaluation of DLG2 as a positional candidate for disposition index in African-Americans from the IRAS Family Study. *Diabetes Res. Clin. Pract.* **87,** 69–76 (2010).
20. Patterson, N., Price, A. L. & Reich, D. Population structure and eigenanalysis. *PLoS Genet.* **2,** e190 (2006).
21. Patterson, N. *et al.* Methods for high-density admixture mapping of disease genes. *Am. J. Hum. Genet.* **74,** 979–1000 (2004).
22. Pritchard, J. K., Stephens, M. & Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics* **155,** 945–959 (2000).

23. Keene, K. L. *et al.* Exploration of the utility of ancestry informative markers for genetic association studies of African Americans with type 2 diabetes and end stage renal disease. *Hum. Genet.* **124,** 147–154 (2008).
24. Almasy, L. & Blangero, J. Multipoint quantitative-trait linkage analysis in general pedigrees. *Am. J. Hum. Genet.* **62,** 1198–1211 (1998).
25. Qi, L. *et al.* Genetic variants at 2q24 are associated with susceptibility to type 2 diabetes. *Hum. Mol. Genet.* **19,** 2706–2715 (2010).
26. Zeggini, E. *et al.* Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. *Nat. Genet.* **40,** 638–645 (2008).
27. DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium *et al.* Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. *Nat. Genet.* **46,** 234–244 (2014).
28. Voight, B. F. *et al.* Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nat. Genet.* **42,** 579–589 (2010).
29. Sim, X. *et al.* Transferability of type 2 diabetes implicated loci in multi-ethnic cohorts from Southeast Asia. *PLoS Genet.* **7,** e1001363 (2011).
30. Kim, D. *et al.* CXCL12 secreted from adipose tissue recruits macrophages and induces insulin resistance in mice. *Diabetologia* **57,** 1456–1465 (2014).
31. Dupuis, J. *et al.* New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nat. Genet.* **42,** 105–116 (2010).
32. Pasquali, L. *et al.* Pancreatic islet enhancer clusters enriched in type 2 diabetes risk–associated variants. *Nat. Genet.* **46,** 136–143 (2014).
33. Bell, J. T. *et al.* Genome-wide association scan allowing for epistasis in type 2 diabetes. *Ann. Hum. Genet.* **75,** 10–19 (2011).
34. Herold, C., Steffens, M., Brockschmidt, F. F., Baur, M. P. & Becker, T. INTERSNP: genome-wide interaction analysis guided by a priori information. *Bioinforma. Oxf. Engl.* **25,** 3275–3281 (2009).
35. Becker, T., Herold, C., Meesters, C., Mattheisen, M. & Baur, M. P. Significance levels in genome-wide interaction analysis (GWIA). *Ann. Hum. Genet.* **75,** 29–35 (2011).
36. Babron, M.-C., Etcheto, A. & Dizier, M.-H. A New Correction for Multiple Testing in Gene-Gene Interaction Studies. *Ann. Hum. Genet.* (2015). doi:10.1111/ahg.12113
37. Lucas, G. *et al.* Hypothesis-Based Analysis of Gene-Gene Interactions and Risk of Myocardial Infarction. *PLoS ONE* **7,** (2012).