Journal of Human Genetics (2012) 57, 57–61

www.nature.com/jhg

ORIGINAL ARTICLE

Single nucleotide polymorphisms in *JAZF1* and *BCL11A* gene are nominally associated with type 2 diabetes in African-American families from the GENNID study

Kurt A Langberg[1,5], Lijun Ma[1,5], Neeraj K Sharma[1], Craig L Hanis[2], Steven C Elbein[1,*], Sandra J Hasstedt[3] and Swapan K Das[1], for the American Diabetes Association GENNID Study Group[4]

Prior type 2 diabetes (T2D) genome-wide association studies (GWASs) have generated a list of well-replicated susceptibility loci in populations of European and Asian ancestry. To validate the trans-ethnic contribution of the single-nucleotide polymorphisms (SNPs) involved in these GWASs, we performed a family-based association analysis of 32 selected GWAS SNPs in a cohort of 1496 African-American (AA) subjects from the Genetics of NIDDM (GENNID) study. Functional roles of these SNPs were evaluated by screening *cis*-eQTLs in transformed lymphoblast cell lines available for a sub-group of Genetics of NIDDM (GENNID) families from Arkansas. Only three of the 32 GWAS-derived SNPs showed nominally significant association with T2D in our AA cohort. Among the replicated SNPs rs864745 in *JAZF1* and rs10490072 in *BCL11A* gene (P=0.006 and 0.03, respectively, after adjustment for body mass index) were within the 1-lod drop support interval of T2D linkage peaks reported in these families. Genotyping of 19 tag SNPs in these two loci revealed no further common SNPs or haplotypes that may be a stronger predictor of T2D susceptibility than the index SNPs. Six T2D GWAS SNPs (rs6698181, rs9472138, rs730497, rs10811661, rs11037909 and rs1153188) were associated with nearby transcript expression in transformed lymphoblast cell lines of GENNID AA subjects. Thus, our study indicates a nominal role for *JAZF1* and *BCL11A* variants in T2D susceptibility in AAs and suggested little overlap in known susceptibility
to T2D between European- and African-derived populations when considering GWAS SNPs alone.
Journal of Human Genetics (2012) **57,** 57–61; doi:10.1038/jhg.2011.133; published online 24 November 2011

Keywords: association; diabetes; eQTL; gene expression; linkage disequilibrium; SNP

INTRODUCTION

Type 2 diabetes (T2D) is a widespread epidemic, which disproportionately affects minority populations in the United States, such as African-American (AA) populations, compared with populations of European descent.[1] Genetic, environmental and cultural factors may contribute to this disproportionate risk. Genome-wide association studies (GWASs) have discovered many common variants influencing predisposition to T2D. However, the vast majority of these studies have been performed in populations of European and Asian ancestry and little data are available for AAs.[2] To date, the only gene verified as being associated with T2D in populations of African descent is *TCF7L2*.[3] However, other studies have cast doubt on whether these and other markers associated with T2D represent the same level of risk in AA populations.[4,5] Results from two case–control association studies evaluating these GWAS-derived T2D single-nucleotide polymorphisms (SNPs) in AAs remain conflicting and warrant further study.[6,7]

From 1993 to 2003, investigators of the American Diabetes Association, through the Genetics of NIDDM (GENNID) project, ascertained 1496 individuals from 580 AA families through T2D-diagnosed siblings at multiple sites as a resource for the discovery of genes related to T2D and its complications.[8,9] Family-based association studies allow for better control of population stratification and heterogeneity compared with case–control association studies.[10] In

[1]Department of Internal Medicine, Wake Forest School of Medicine, Winston-Salem, NC, USA; [2]Human Genetics Center, School of Public Health, University of Texas Health Science Center at Houston, Houston, TX, USA and [3]Department of Human Genetics, University of Utah, Salt Lake City, UT, USA
[4]See Acknowledgement.
[5]These authors contributed equally to this work.
*Deceased.
Correspondence: Dr SJ Hasstedt, Department of Human Genetics, University of Utah, 15 N 2030 E RM 2100, Salt Lake City, UT 84112-5330, USA.
E-mail: sandy@genetics.utah.edu
or Dr SK Das, Section on Endocrinology and Metabolism, Department of Internal Medicine, Wake Forest School of Medicine, Medical Center Boulevard, Winston-Salem, NC 27157, USA.
E-mail: sdas@wakehealth.edu
Received 15 August 2011; revised 7 October 2011; accepted 1 November 2011; published online 24 November 2011

this study, we evaluated 32 GWAS-derived T2D-associated SNPs in the GENNID AA pedigrees. Our study verified the association of SNP rs10490072 (in *BCL11A*) and rs864745 (in *JAZF1*) with T2D in AAs. These two SNPs fall within the support interval of suggestive linkage peaks (at chromosomes 2 and 7, respectively) for T2D in this cohort;[8] thus we performed linkage disequilibrium-based fine mapping of these loci by genotyping 21 tag SNPs within the haplotype block that includes T2D-associated index SNPs. Finally, to develop causal models of diabetes, we sought to define the role of these polymorphisms as *cis*-regulatory elements in modulating the expression of transcripts in transformed lymphoblastoid cells available for a subset of 160 GENNID family subjects from Arkansas.

MATERIALS AND METHODS

Study cohort

The GENNID study ascertained 1496 subjects of 580 AA families through a sibling pair, each with a T2D diagnosis from 10 sites. T2D was diagnosed using the National Diabetes Data Group criteria. This study was approved by the Institutional Review Board at each participating institution. The GENNID cohort includes multigenerational families, affected sib pairs and nuclear families with affected siblings, available parents and unaffected sibs. Physical examination data and DNA were available on 1496 subjects, which after removing apparent sample discrepancies were reduced to 1450 individuals. Characteristics of this study cohort are summarized in Supplementary Table 1; see Elbein *et al*.[8] for more details.

SNP selection

We selected 32 SNPs of 27 loci for our analysis. All SNPs chosen were from prior GWASs for T2D in Caucasian and East-Asian populations, and most of them have been replicated in independent Caucasian ancestry cohorts.[11–26] Supplementary Table 2 lists the studies from which SNPs had been selected and associated with T2D and/or related traits in Caucasian and East Asian populations. Nine tag SNPs across *JAZF1* and 10 tag SNPs across *BCL11A* were further selected for genotyping in the GENNID AA sample, in addition to the GWAS index SNPs (rs864745 and rs10490072). Tag SNPs were selected based on HapMap (CEU, YRI and ASW) and an AA ESRD cohort genotype data[27,28] under a confidence interval model of linkage disequilibrium-block structure around the index SNP (pair-wise tagging with an $r^2 \geqslant 0.90$).

Genotyping

Salted out DNA samples from lymphoblastoid cell lines of GENNID AA subjects were provided by the Coriell Cell Repository (Camden, NJ, USA), quantified by picogreen, and concentrations adjusted for genotyping purposes. Supplementary Table 2 lists 32 T2D GWAS SNPs that were genotyped on different platforms. Sixteen SNPs were genotyped using Single Base Primer Extension reactions in a 12-plex format using the GenomeLab-SNPstream Genotyping System (Beckman Coulter, Inc., Fullerton, CA, USA) and another 16 SNPs were genotyped by pre-designed Taqman SNP genotyping assays (Applied Biosystems Inc., Foster City, CA, USA) using an ABI-7500 Fast real-time PCR system. SNP genotyping success rates for the SNPstream and Taqman were 99.3 and 98.9%, respectively. An additional 19 *JAZF1* and *BCL11A* haplotype block tag SNPs were genotyped on a Sequenom MassARRAY system (Sequenom Inc., San Diego, CA, USA) according to the manufacturer's iPLEX application guidelines. Details of Sequenom multiplex genotyping assays are shown in Supplementary Table 3. The genotyping calling rate was above 99%, and the genotyping reproducibility was 100% assured by 70 evenly distributed duplicate samples across the genotyping plates, as well as by two standard samples on each genotyping plate.

Transformed lymphocyte cell line culture

We used total RNA extracted from Epstein–Barr virus-transformed lymphocytes (TLs) for evaluating the role of GWAS-associated SNPs in regulating transcript level expression of nearby genes. TLs used in our study were derived from blood samples of 160 GENNID AA subjects (80 sib pairs) from Arkansas. Cells were grown under normoglycemic (5.6 mM glucose) standard culture conditions in RPMI-1640 culture media (Cat. 11875, Gibco-Invitrogen, Carlsbad, CA, USA) supplemented with 10% Benchmark fetal bovine serum (Cat. 100-106, lot no. A33B00Z, Gemini Bio-Products, West Sacramento, CA, USA).

RNA isolation and gene expression

Total RNA was isolated from TLs by using a Qiagen RNeasy Mini Kit (Qiagen, Valencia, CA, USA). RNA was quantified using a NanoDrop ND-2000 (NanoDrop Technologies Inc., Wilmington, DE, USA), and quality was assessed by an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). Genome-wide expression analysis using TL total RNA was performed as described elsewhere.[29] In brief, labeling and hybridization to Illumina HT-12 beadchip arrays (version 4; San Diego, CA, USA) was performed according to the manufacturer's instructions. Resulting data were processed and normalized using the average normalization algorithm as implemented in GenomeStudio Gene Expression Module v1.0 application software (Illumina). Background was subtracted before the scaling. Probes with detection *P*-values >0.01 were additionally excluded because of the lack of evidence for reliable quantitative expression.

Statistical analyses

Likelihood analysis as implemented in jPAP was used to test each SNP for association with T2D, age of diagnosis (AOD), body mass index (BMI) and waist–hip ratio (WHR).[30] BMI and WHR were transformed separately in males and females, using the inverse normal distribution, for which a quantile was assigned to each trait value and the corresponding inverse normal deviate assigned as the trait. Transformed BMI and WHR and untransformed AOD were each modeled as a normal density. T2D risk was modeled to account for AOD in the affected pedigree members, whereas allowing for censored observations.[9] SNP genotypes were coded as 0, 1 or 2, thereby assuming an additive effect. Analysis of all traits accounted for heritability and included SNP genotype and gender as covariates; analysis of BMI and WHR also included age as a covariate; analysis of T2D was performed separately without obesity adjustment and with adjustment for either BMI or WHR. Associations were tested through comparison of the maximum likelihood obtained when estimating the SNP effect to the maximum likelihood when fixing the SNP effect to zero. *P*-values were obtained as twice the natural logarithm of the likelihood ratio for a 1-df chi-square statistic. In each likelihood maximization, all other model parameters were estimated in analyses of BMI, WHR and AOD, whereas only the SNP effect and heritability were estimated in analysis of T2D, with all other parameters fixed at estimates obtained when correcting for ascertainment through an affected sib pair. In this paper we present association results for T2D and AOD. Merlin was used to infer the most likely haplotype for each family member for the *JAZF1* and *BCL11A* region SNPs.[31] Then association analysis was performed as before, except for testing each 2-SNP to 10-SNP (*JAZF1*) or 11-SNP (*BCL11A*) haplotype rather than single SNPs.

We assessed associations between the selected T2D GWAS SNPs and normalized quantitative expression values of local transcripts (within 1 Mb up- and downstream of tested SNP). We tested association of variable number of transcripts for each SNP, ranging from 3 transcripts (for SNP rs10923931 and rs864745) to 43 transcripts (for SNP rs1800247). The association of probe-expression level with genotype was assessed with an additive model implemented in SAS software (ver. 9.1, SAS Institute, Cary, NC, USA). The generalized estimating equations procedure was used to account for family membership. To control for potential population stratification, the association was also analyzed using a modification of the within-family association test.[32] In brief, this method partitions the association into between- and within-family components represented, respectively, by the sibship mean of the continuous numeric variable for genotype and each individual's deviation from this mean. The test of the significance of the within-family components is a test of co-transmission among siblings, which is robust to population stratification. An additional adjustment of diabetic status was also applied. The *P*-values were not adjusted for multiple comparisons. Statistical power of our gene expression cohort was modest ($\sim$92%) to detect 15% of the variation in gene expression levels (assuming a type 1 error rate=0.005, MAF$>$0.15, additive model) in general association.

RESULTS

The SNP rs864745 in *JAZF1* gene showed a nominally significant association with T2D (P=0.018) in the GENNID African-American

sample. This association was stronger after adjustment for BMI (P=0.006). A SNP (rs10490072) in *BCL11A* gene was also associated with T2D after adjustment for BMI (P=0.03) and was more strongly associated with AOD (P=0.007). *WFS1* SNP rs10010131 showed a marginal association with T2D. In all 29 of 32 other GWAS-derived SNPs, including *TCF7L2* SNPs rs7907346 and rs12255372, showed no association with T2D in this cohort (Table 1 and Supplementary Table 4). Discriminatory ability of the combined SNP information was assessed by grouping individuals based on number of risk alleles carried for three variants (rs10490072, rs864745 and rs10010131) that are associated with T2D in our cohort. The *P*-values testing for increased risk of T2D for 4+, 5+ and 6 risk alleles were 0.0011, 0.0011 and 0.00979, respectively. Thus, discriminatory ability of three SNPs combined in predicting T2D risk was slightly higher than for a single SNP.

In this sample, we earlier reported linkage for T2D on chromosome 2 (logarithm of odds=3.58 at 84 cM, 1-lod drop support interval 77–102 Mb) and chromosome 7 (logarithm of odds=2.62 at 24 cM, 1-lod drop support interval 14–29 Mb).[9] SNPs rs10490072 in *BCL11A and* rs864745 in *JAZF1* are within the 1-lod drop support interval of these two linkage peaks at chromosomes 2 and 7, respectively. Thus, 11 *BCL11A* tag SNPs (including the GWAS tag SNP rs10490072) and 10 *JAZF1* tag SNPs (including the GWAS index SNP rs864745) were genotyped in these pedigrees to identify causal variant(s) with larger effect sizes and tested haplotypic associations in these regions. The LD relationships of genotyped SNPs in these two loci in our cohort are shown in Supplementary Figure 1. For *JAZF1*, no T2D risk haplotype produced higher significance than did the rs864745-A allele alone, but the protective haplotype GGTGG for SNPs rs864745, rs849140, rs849141, rs10276381 and rs12154248 produced a nominal *P*-value of 0.000697 in analysis of T2D adjusted for BMI. Likewise for *BCL11A*, no early AOD risk haplotype produced higher significance than did the rs10490072-A allele alone, but the protective haplotype CCCCAGC for SNPs rs11894442, rs6718203, rs17402905, rs8179712, rs1011407, rs10490072 and rs12468946 produced identical significance in analysis of AOD.

Most of the Caucasian GWAS-derived T2D-associated SNPs are noncoding, residing in either intronic or intergenic regions, are not in LD with known non-synonymous SNPs, and are expected to increase diabetes susceptibility by modulating transcription as a *cis*-regulatory

Table 1 Association of SNPs identified in Caucasian GWAS studies with type 2 diabetes in GENNID AA families

				Allele				P-value			
Nearest gene	*Chromosome*	*Physical location (Mb)*	*SNP*	*Major*	*Minor*	MAF_{AA}	MAF_{CEU} (allele)	*T2D*	*T2D adj BMI*	*T2D adj WHR*	*AOD*
ADAM30-NOTCH2	1	120	rs2641348	T	C	0.341	0.097	0.503	0.502	0.834	0.719
ADAMTS9	3	65	rs4607103	C	T	0.32	0.19	0.48	0.578	0.829	0.501
BCL11A	2	61	rs10490072	A	G	0.094	0.271	0.074	**0.033**	0.262	**0.007**
CDC123	10	12	rs11257622	T	C	0.16	0.261	0.243	0.237	0.256	0.749
CDKAL1	6	21	rs10946398	C	A	0.204	0.336 (C)	0.823	0.511	0.628	0.849
CDKN2A/2B	9	22	rs10811661	T	C	0.05	0.199	0.792	0.472	0.849	0.327
CDKN2B	9	22	rs564398	A	G	0.072	0.434	0.288	0.257	0.957	0.975
DCD	12	53	rs1153188	A	T	0.235	0.257	0.751	0.95	1	0.809
EXT2	11	44	rs11037909	T	C	0.149	0.283	1	0.522	0.538	0.725
FLJ39370	4	113	rs17044137	T	A	0.343	0.239	0.356	0.464	0.663	0.157
FTO	16	52	rs8050136	C	A	0.431	0.46	0.24	0.527	0.975	0.244
GCK	7	44	rs730497	G	A	0.204	0.195	0.917	0.95	0.93	0.45
HHEX	10	94	rs1111875	G	A	0.23	0.416	0.59	0.886	0.648	0.372
HHEX	10	94	rs5015480	C	T	0.37	0.42	0.281	0.228	0.265	0.508
HK1	10	71	rs906216	T	G	0.356	0.438 (T)	0.532	0.366	0.577	0.456
IGF2BP2	3	186	rs4402960	T	G	0.457	0.296 (T)	0.864	0.699	0.759	0.503
IGF2BP2	3	187	rs1470579	C	A	0.404	0.296 (C)	0.37	0.256	0.505	0.127
JAZF1	7	28	rs864745	A	G	0.247	0.487 (A)	**0.018**	**0.006**	**0.041**	0.273
KCNJ11	11	17	rs5219	C	T	0.065	0.46	0.663	0.82	0.653	0.554
KCNQ1	11	3	rs2237892	C	T	0.091	0.075	0.287	0.448	0.374	0.399
LOC387761	11	42	rs7480010	G	A	0.135	0.279 (G)	0.862	0.965	0.965	0.48
MTNR1B	11	92	rs1387153	C	T	0.361	0.272	0.764	0.399	0.768	0.547
MTNR1B	11	92	rs10830963	C	G	0.068	0.3	0.343	0.268	0.65	0.975
NOTCH2	1	120	rs10923931	G	T	0.036	0.093	0.512	0.281	0.753	0.944
PKN2	1	89	rs6698181	C	T	0.109	0.365	0.171	0.224	0.135	0.279
SLC30A8	8	118	rs13266634	C	T	0.098	0.239	0.493	0.082	0.125	0.518
TCF7L2	10	115	rs7903146	C	T	0.478	0.279	0.34	0.663	0.95	0.467
TCF7L2	10	115	rs12255372	G	T	0.302	0.248	0.48	0.895	0.699	0.066
THADA	2	44	rs7578597	T	C	0.26	0.124	0.532	0.509	0.123	0.135
TSPAN8-LGR5	12	70	rs7961581	T	C	0.186	0.252	0.065	0.068	0.325	0.019
VEGFA	6	44	rs9472138	C	T	0.188	0.239	0.128	0.165	0.219	0.025
WFS1	4	6	rs10010131	G	A	0.335	0.323	**0.046**	**0.016**	**0.031**	**0.048**

Abbreviations: AA, African American; Adj, adjusted for; AOD, age of diagnosis; BMI, body mass index; GWASs, genome-wide association studies GWASs; MAF_{AA}, minor allele frequency in GENNID African Americans; MAF_{CEU}, minor allele frequency in HapMap Caucasian (CEU) subjects; SNP, single-nucleotide polymorphisms; T2D, type 2 diabetes; WHR, waist–hip ratio.
Statistically significant associations are marked in bold.

Table 2 Association of T2D GWAS SNPs with mRNA expression for adjacent genes

GWAS region (nearest gene)	*SNP*	*chrom*	*NCBI location Build 36*	*ADJ_T2D general P*	*non-ADJ general P*	*ADJ_T2D family P*	*non-ADJ family P*	*PROBE_ID*	*cdsm location Build 36*	*Transcript Symbol*
PKN2	rs6698181	chr01	88915893	0.0207	0.027	0.0152	0.0151	ILMN_1701114	89523977	GBP1
VEGFA	rs9472138	chr06	43919740	0.0419	0.0414	0.0019	0.0017	ILMN_1727073	42980879	MEA1
				0.0654	0.0671	0.0157	0.0183	ILMN_1789001	44223795	SLC35B2
				0.1146	0.1164	0.0002	0.0004	ILMN_1717313	44230228	NFKBIE
GCK	rs730497	chr07	44190246	0.0552	0.0614	0.008	0.0075	ILMN_1796900	44477904	NUDCD3
				0.0031	0.002	0.0553	0.0592	ILMN_1804148	44620441	TMED4
CDKN2A	rs10811661	chr09	22124094	2.70E-08	2.27E-08	0.0583	0.0571	ILMN_1664466	21333931	KLHL9
				0.0433	0.043	0.0095	0.0092	ILMN_1744295	21982666	CDKN2A
EXT2	rs11037909	chr11	44212190	0.0326	0.0346	0.0497	0.0452	ILMN_1815051	43348841	API5
				0.0589	0.0587	0.003	0.0029	ILMN_2392274	44628444	CD82
DCD	rs1153188	chr12	53385263	0.0008	0.001	0.0763	0.0671	ILMN_1804642	54576802	SMUG1

Abbreviations: ADJ_T2D, additional adjustment for T2D; cdsm, mid-point of cDNA; chrom, chromosome; family: within-family model; general, GEE general model; GWASs, genome-wide association studies GWASs; NCBI, National Center for Biotechnology Information; non-ADJ: no additional adjustment; SNP, single-nucleotide polymorphisms; T2D, type 2 diabetes.
P values were adjusted for age, gender and BMI (not corrected for multiple testing errors).

elements. Thus, we analyzed the genotypic association of 32 SNPs with the expression of 215 expressed local transcripts (represented by 274 probes within ± 1 Mb). Six SNPs (rs6698181, rs9472138, rs730497, rs10811661, rs11037909 and rs1153188) were associated with nearby transcript expression in transformed lymphoblast cell lines of GENNID subjects in both general and within-family analyses using generalized estimating equations (Table 2). The strongest association was observed for SNP rs10811661 in regulating the transcription of KLHL9 ($P=2.7\times10^{-8}$) under the general model of inheritance.

DISCUSSION

To our knowledge, this is the first study to evaluate European and Asian T2D GWAS derived polymorphisms in an AA family cohort. We studied 32 established T2D and related trait GWAS-derived SNPs and, consistent with earlier reports, most GWAS-derived SNPs showed no significant associations in these AA families.

TCF7L2 is one of the most significant diabetes susceptibility genes identified to date in various populations.[11] A previous case–control association study by Lewis *et al*.[6] reported a significant association of *TCF7L2* rs7903146 with T2D in AA populations. This association was not replicated in our family-based GENNID AA sample. The lack of significance may be the result of the relatively low power of our sample, especially when accounting for family structures.

A recent DIAGRAM+ meta-analysis showed associations of 12 new autosomal and X chromosomal loci in a large discovery cohort of 22 044 Caucasian subjects.[33] Most of the loci discovered by this meta-analysis showed odds ratio (OR) <1.1. Considering that the statistical power of DIAGRAM+ meta-analysis is much enhanced, we did not expect enough power to detect the effect of those SNPs in our limited-size sample, and have not selected those SNPs for validation in our cohort.

Among the other loci examined in this study, the one that showed the most significant association with T2D is a SNP (rs864745) in a zinc finger protein-coding gene *JAZF1*. The SNP rs864745 has been characterized as a risk factor in European populations by Zeggini *et al*.[22] in a large meta-analysis. Deletion of the *JAZF1* gene in mice leads to early growth retardation, which was associated with reduced plasma IGF-1 levels, and in adulthood to decreased muscle mass, increased fat mass and insulin resistance.[34] The rs864745 was associated with *JAZF1* expression in muscle in our prior population-based sample of mixed ethnicity using RT-PCR, where the association was largely contributed by African Americans.[35] Our gene expression arrays were unable to detect significant expression of the *JAZF1* transcripts in GENNID transformed lymphoblast cell lines. The SNPs in the *JAZF1 and BCL11A* genes were associated with T2D especially after adjustment for BMI and were within the support interval of suggestive linkage peaks for T2D in our GENNID African-American cohort.[9] The tag SNP based analysis in these regions revealed no further common SNP or haplotype that may be a stronger predictor of T2D susceptibility than the index SNPs. None of these associated SNPs explained linkage in this region, and associations were not significant after correcting for multiple testing errors. However, a role for rare variants not tagged by haplotypes generated by the common SNPs cannot be excluded by our study.

In summary, our study indicates a nominal role of *JAZF1* and *BCL11A* variants in T2D susceptibility in African-Americans. However, this work suggests little overlap in known susceptibility to T2D between European and African-derived populations if focusing on GWAS SNPs alone. Differences in linkage disequilibrium patterns may result in poor proxies for the tested Caucasian-attributed SNPs in AA populations. In addition, the small effect of the variants may require much larger populations to observe notable associations. Results from GWAS studies in African-Americans are awaited with interest, but further fine mapping studies based on deep sequencing of candidate regions of a representative AA cohort within areas of interest identified in Caucasian GWAS studies may be helpful to target ethnicity-specific genetic risk factors for T2D. Alternatively, as suggested by our study, T2D-associated SNPs may function as *cis*-regulatory elements and alter the expression of nearby genes, which may fall into certain unknown pathways that contribute to the development of T2D, where the effect size might be different across different populations because of different genetic and/or environmental backgrounds.

Lymphoblast cell lines may not be the most relevant cell types to evaluate T2D- and metabolism- related eQTLs. Thus, a limitation of the current screening of eQTLs in this study was that we only had transformed lymphoblast cell line gene expression data available for the reported GWAS SNPs. However, several studies revealed that eQTLs in tissues relevant to T2D and associated metabolic disorders (for example, adipose) significantly overlap with eQTLs in lymphoblast cell lines.[36,37] Functional studies of regulatory variants, as well as regulated genes *per se*, would be essential to uncover the T2D-susceptibility genes from multiple GWAS hits.

ACKNOWLEDGEMENTS

This work was supported by the National Institutes of Health (NIH)/National Institute of Diabetes and Digestive and Kidney Diseases Grant DK071100. We thank the American Diabetes Association for access to the samples used in this study. We thank the GENNID investigators who recruited and characterized the families and collected genetic material. The American Diabetes Association GENNID Study Group, which includes Eric Boerwinkle, PhD, University of Texas Health Science Center; John Buse, MD, PhD, University of North Carolina; Ralph DeFronzo, MD, University of Texas Health Science Center; David Ehrmann, MD, University of Chicago; Steven C Elbein, MD, University of Utah/University of Arkansas for Medical Sciences; Wilfred Fujimoto, MD, and Steven E Kahn, MB, ChB, University of Washington; Craig L Hanis, PhD, University of Texas Health Science Center; Richard A Mulivor, PhD, and Jeanne C Beck, PhD, Coriell Cell Repositories; Jill Norris, PhD, University of Colorado School of Medicine; M. Alan Permutt, MD, and Philip Behn, MD, Washington University School of Medicine; Leslie Raffel, MD, Cedars-Sinai Medical Center; and David C Robbins, MD, Medlantic Research Institute. We thank Winston Chu for his valuable assistance with DNA, SNPstream and taqman genotyping and Karen Klein (Research Support Core, Office of Research, Wake Forest School of Medicine) for critically reading and editing our manuscript. We also thank Prof. Siqun Zheng, Director, Genotyping Laboratory, and the technical staff of the Center for Human Genomics, Wake Forest School of Medicine (especially Ms Shelly Smith and Dr Ge Li) for their extensive support in Sequenom genotyping and Illumina gene expression analysis. Finally, we thank Drs Donald W Bowden and Nicholette D (Palmer) Allred for sharing tag-SNP information.

1 Brancati, F. L., Kao, W. H., Folsom, A. R., Watson, R. L. & Szklo, M. Incident type 2 diabetes mellitus in African American and white adults: the atherosclerosis risk in communities study. *JAMA.* **283,** 2253–2259 (2000).
2 McCarthy, M. I. The importance of global studies of the genetics of type 2 diabetes. *Diabetes. Metab. J.* **35,** 91–100 (2011).
3 Sale, M. M., Smith, S. G., Mychaleckyj, J. C., Keene, K. L., Langefeld, C. D., Leak, T. S. *et al.* Variants of the transcription factor 7-like 2 (TCF7L2) gene are associated with type 2 diabetes in an African-American population enriched for nephropathy. *Diabetes.* **56,** 2638–2642 (2007).
4 Elbein, S. C., Chu, W. S., Das, S. K., Yao-Borengasser, A., Hasstedt, S. J., Wang, H. *et al.* Transcription factor 7-like 2 polymorphisms and type 2 diabetes, glucose homeostasis traits and gene expression in US participants of European and African descent. *Diabetologia.* **50,** 1621–1630 (2007).
5 Elbein, S. C. Genetics factors contributing to type 2 diabetes across ethnicities. *J. Diabetes. Sci. Technol.* **3,** 685–689 (2009).
6 Lewis, J. P., Palmer, N. D., Hicks, P. J., Sale, M. M., Langefeld, C. D., Freedman, B. I. *et al.* Association analysis in African Americans of European-derived type 2 diabetes single nucleotide polymorphisms from whole-genome association studies. *Diabetes.* **57,** 2220–2225 (2008).
7 Waters, K. M., Stram, D. O., Hassanein, M. T., Le Marchand, L., Wilkens, L. R., Maskarinec, G. *et al.* Consistent association of type 2 diabetes risk variants found in Europeans in diverse racial and ethnic groups. *PLoS. Genet.* **6,** e1001078 (2010).
8 Elbein, S. C., Das, S. K., Hallman, D. M., Hanis, C. L. & Hasstedt, S. J. Genome-wide linkage and admixture mapping of type 2 diabetes in African American families from the American Diabetes Association GENNID (Genetics of NIDDM) Study Cohort. *Diabetes.* **58,** 268–274 (2009).
9 Hasstedt, S. J., Hanis, C. L., Das, S. K. & Elbein, S. C. Pleiotropy of type 2 diabetes with obesity. *J. Hum. Genet.* **56,** 491–495 (2011).
10 Ott, J., Kamatani, Y. & Lathrop, M. Family-based designs for genome-wide association studies. *Nat. Rev. Genet.* **12,** 465–474 (2011).
11 Grant, S. F., Thorleifsson, G., Reynisdottir, I., Benediktsson, R., Manolescu, A., Sainz, J. *et al.* Variant of transcription factor 7-like 2 (TCF7L2) gene confers risk of type 2 diabetes. *Nat. Genet.* **38,** 320–323 (2006).
12 Frayling, T. M., Timpson, N. J., Weedon, M. N., Zeggini, E., Freathy, R. M., Lindgren, C. M. *et al.* A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. *Science.* **316,** 889–894 (2007).
13 Sandhu, M. S., Weedon, M. N., Fawcett, K. A., Wasson, J., Debenham, S. L., Daly, A. *et al.* Common variants in WFS1 confer risk of type 2 diabetes. *Nat. Genet.* **39,** 951–953 (2007).
14 Saxena, R., Voight, B. F., Lyssenko, V., Burtt, N. P., de Bakker, P. I., Chen, H. *et al.* Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science.* **316,** 1331–1336 (2007).
15 Scott, L. J., Mohlke, K. L., Bonnycastle, L. L., Willer, C. J., Li, Y., Duren, W. L. *et al.* A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science.* **316,** 1341–1345 (2007).
16 Sladek, R., Rocheleau, G., Rung, J., Dina, C., Shen, L., Serre, D. *et al.* A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature.* **445,** 881–885 (2007).
17 Zeggini, E., Weedon, M. N., Lindgren, C. M., Frayling, T. M., Elliott, K. S., Lango, H. *et al.* Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. *Science.* **316,** 1336–1341 (2007).
18 Lango, H., Palmer, C. N., Morris, A. D., Zeggini, E., Hattersley, A. T., McCarthy, M. I. *et al.* Assessing the combined impact of 18 common genetic variants of modest effect sizes on type 2 diabetes risk. *Diabetes.* **57,** 3129–3135 (2008).
19 van Hoek, M., Dehghan, A., Witteman, J. C., van Duijn, C. M., Uitterlinden, A. G., Oostra, B. A. *et al.* Predicting type 2 diabetes based on polymorphisms from genome-wide association studies: a population-based study. *Diabetes.* **57,** 3122–3128 (2008).
20 Yasuda, K., Miyake, K., Horikawa, Y., Hara, K., Osawa, H., Furuta, H. *et al.* Variants in KCNQ1 are associated with susceptibility to type 2 diabetes mellitus. *Nat. Genet.* **40,** 1092–1097 (2008).
21 Unoki, H., Takahashi, A., Kawaguchi, T., Hara, K., Horikoshi, M., Andersen, G. *et al.* SNPs in KCNQ1 are associated with susceptibility to type 2 diabetes in East Asian and European populations. *Nat. Genet.* **40,** 1098–1102 (2008).
22 Zeggini, E., Scott, L. J., Saxena, R., Voight, B. F., Marchini, J. L., Hu, T. *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).
23 Lyssenko, V., Nagorny, C. L., Erdos, M. R., Wierup, N., Jonsson, A., Spégel, P. *et al.* Common variant in MTNR1B associated with increased risk of type 2 diabetes and impaired early insulin secretion. *Nat. Genet.* **41,** 82–88 (2009).
24 Prokopenko, I., Langenberg, C., Florez, J. C., Saxena, R., Soranzo, N., Thorleifsson, G. *et al.* Variants in MTNR1B influence fasting glucose levels. *Nat. Genet.* **41,** 77–81 (2009).
25 Bouatia-Naji, N., Bonnefond, A., Cavalcanti-Proença, C., Sparsø, T., Holmkvist, J., Marchand, M. *et al.* A variant near MTNR1B is associated with increased fasting plasma glucose levels and type 2 diabetes risk. *Nat. Genet.* **41,** 89–94 (2009).
26 Pare, G., Chasman, D. I., Parker, A. N., Nathan, D. M., Miletich, J. P., Zee, R. Y. *et al.* Novel association of HK1 with glycated hemoglobin in a non-diabetic population: a genome-wide evaluation of 14,618 participants in the Women's Genome Health Study. *PLoS. Genet.* **4,** e1000312 (2008).
27 Frazer, K. A., Ballinger, D. G., Cox, D. R., Hinds, D. A., Stuve, L. L., Gibbs, R. A. International HapMap Consortium *et al.* A second generation human haplotype map of over 3.1 million SNPs. *Nature.* **449,** 851–861 (2007).
28 McDonough, C. W., Palmer, N. D., Hicks, P. J., Roh, B. H., An, S. S., Cooke, J. N., Hester, J. M. *et al.* A genome-wide association study for diabetic nephropathy genes in African Americans. *Kidney Int.* **79,** 563–572 (2011).
29 Das, S. K., Sharma, N. K., Hasstedt, S. J., Mondal, A. K., Ma, L., Langberg, K. A. *et al.* An integrative genomics approach identifies activation of Thioredoxin/Thioredoxin Reductase-1 mediated oxidative stress defense pathway and inhibition of angiogenesis in obese nondiabetic human subjects. *J. Clin. Endocrinol. Metab.* (2011).
30 Hasstedt, S. J. jPAP: Document-driven software for genetic analysis. *Genet. Epidemiol.* **29,** 255 (2005).
31 Abecasis, G. R., Cherny, S. S., Cookson, W. O. & Cardon, L. R. Merlin–rapid analysis of dense genetic maps using sparse gene flow trees. *Nat. Genet.* **30,** 97–101 (2002).
32 Abecasis, G. R., Cardon, L. R. & Cookson, W. O. A general test of association for quantitative traits in nuclear families. *Am. J. Hum. Genet.* **66,** 279–292 (2000).
33 Voight, B. F., Scott, L. J., Steinthorsdottir, V., Morris, A. P., Dina, C., Welch, R. P. *et al.* Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nat. Genet.* **42,** 579–589 (2010).
34 Lee, H.- Y., Li, H., Guigni, B., Wang, J., Birkenfeld, A. L., Jornayvaz, F. R. *et al.* Early Growth Retardation and Insulin Resistance in JAZF1 KO Mice. *Diabetes* **60,** A103 (2011).
35 Sharma, N. K., Langberg, K. A., Mondal, A. K., Elbein, S. C. & Das, S. K. Type 2 diabetes (T2D) associated polymorphisms regulate expression of adjacent transcripts in transformed lymphocytes, adipose, and muscle from Caucasian and African-American subjects. *J. Clin. Endocrinol. Metab.* **96,** E394–E403 (2011).
36 Nica, A. C., Parts, L., Glass, D., Nisbet, J., Barrett, A., Sekowska, M. *et al.* The architecture of gene regulatory variation across multiple human tissues: the MuTHER study. *PLoS. Genet.* **7,** e1002003 (2011).
37 Bullaughey, K., Chavarria, C. I., Coop, G. & Gilad, Y. Expression quantitative trait loci detected in cell lines are often present in primary tissues. *Hum. Mol. Genet.* **18,** 4296–4303 (2009).

Supplementary Information accompanies the paper on Journal of Human Genetics website (http://www.nature.com/jhg)