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, Lijun Ma, Neeraj K Sharma, Craig L Hanis, Steven C Elbein, Sandra J Hasstedt and Swapan K Das, for the American Diabetes Association GENNID Study Group

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. 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. To date, the only gene verified as being associated with T2D in populations of African descent is TCF7L2. However, other studies have cast doubt on whether these and other markers associated with T2D represent the same level of risk in AA populations. Results from two case–control association studies evaluating these GWAS-derived T2D single-nucleotide polymorphisms (SNPs) in AAs remain conflicting and warrant further study.

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. Family-based association studies allow for better control of population stratification and heterogeneity compared with case–control association studies. In
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, 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 Elbeim et al.11 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 data27,28 under a confidence interval model of linkage disequilibrium-block based on HapMap (CEU, YRI and ASW) and an AA ESRD cohort genotype data.9 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 data27,28 under a confidence interval model of linkage disequilibrium-block structure around the index SNP (pair-wise tagging with an r² > 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) according to 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. A33B002Z, 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.30 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 rs1092393) 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 (~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 rs12235372, 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 0.0097, respectively. Thus, discriminatory ability of three SNPs combined in predicting T2D risk was slightly higher 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-locd drop support interval 77–102 Mb) and chromosome 7 (logarithm of odds = 2.62 at 24 cM, 1-locd drop support interval 14–29 Mb). SNPs rs10490072 in BCL11A and rs864745 in JAZF1 are within the 1-locd 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 GTGTTG 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 CCCCGAGC for SNPs rs11894442, rs6718203, rs17402905, rs17897712, 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

| Nearest gene | Chromosome | Physical location (Mb) | SNP | Allele | MAF<sub>CEU</sub> | P-value |
|--------------|------------|------------------------|-----|--------|-------------------|---------|
| SNPs         |            |                        |     |        | (allele)          |         |
|              |            |                        |     |        |                  |         |
| ADAM30-NOTCH2 | 1          | 120                    | rs2641348 | C       | 0.341             | 0.097    |
| ADAMTS9      | 3          | 65                     | rs4607103 | T       | 0.32               | 0.109    |
| BCL11A       | 2          | 61                     | rs10490072 | A       | 0.094             | 0.271    |
| CDC123       | 10         | 12                     | rs11127562 | T       | 0.16               | 0.261    |
| CDKAL1       | 6          | 21                     | rs10946398 | A       | 0.204             | 0.336 (C) |
| CDKN2A/2B    | 9          | 22                     | rs10811661 | C       | 0.05               | 0.199    |
| HK1          | 10         | 71                     | rs906217 | A       | 0.072             | 0.434    |
| DCD          | 12         | 53                     | rs1151388 | T       | 0.235             | 0.257    |
| EXT2         | 11         | 44                     | rs11037909 | T       | 0.149             | 0.283    |
| FLJ9370      | 4          | 113                    | rs1704417 | T       | 0.343             | 0.239    |
| FTO          | 16         | 52                     | rs8050136 | A       | 0.431             | 0.465    |
| GOK          | 7          | 44                     | rs730497 | G       | 0.204             | 0.195    |
| HHEX         | 10         | 94                     | rs1111875 | G       | 0.23               | 0.416    |
| HHEX         | 10         | 94                     | rs5015480 | C       | 0.37               | 0.42     |
| HK1          | 10         | 71                     | rs906217 | T       | 0.356             | 0.438 (T) |
| IGF2BP2      | 1          | 38                     | rs502897 | A       | 0.457             | 0.296 (T) |
| JAZF1        | 7          | 28                     | rs864745 | A       | 0.247             | 0.487 (A) |
| KCN11        | 11         | 17                     | rs5219  | T       | 0.065             | 0.46     |
| KCNQ1        | 11         | 3                      | rs2237892 | C       | 0.091             | 0.075    |
| LOC387761    | 11         | 42                     | rs7480010 | A       | 0.135             | 0.279 (G) |
| MTRNR1B      | 11         | 92                     | rs1387153 | C       | 0.361             | 0.272    |
| NOTCH2       | 1          | 120                    | rs10932931 | C       | 0.036             | 0.093    |
| PKN2         | 1          | 89                     | rs6698181 | G       | 0.109             | 0.365    |
| SLC30A8      | 8          | 118                    | rs13266634 | C       | 0.098             | 0.239    |
| TCF7L2       | 10         | 115                    | rs7903146 | C       | 0.478             | 0.279    |
| TCF7L2       | 10         | 115                    | rs12255372 | G       | 0.302             | 0.248    |
| THADA        | 2          | 4                      | rs7578597 | T       | 0.26               | 0.124    |
| TSPAN8-LGR5  | 12         | 70                     | rs7961581 | T       | 0.186             | 0.252    |
| VEGLA        | 6          | 44                     | rs9472138 | C       | 0.188             | 0.239    |
| WFS1         | 4          | 6                      | rs10010131 | G       | 0.335             | 0.323    |

Abbreviations: AA, African American; Adj, adjusted for; AOD, age of diagnosis; BMI, body mass index; GWASs, genome-wide association studies; SNPs, single-nucleotide polymorphisms; T2D, type 2 diabetes; WHR, waist-hip ratio. Statistically significant associations are marked in bold.

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

| GWAS region (nearest gene) | SNP   | chrom  | NCBI location | ADJ_T2D general P | non-ADJ general P | ADJ_T2D family P | non-ADJ family P | PROBE_ID | cdsn location | Transcript |
|---------------------------|-------|--------|---------------|-------------------|-------------------|-----------------|-----------------|----------|--------------|------------|
| 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 | 44477940    | NUDCD3     |
| CDKN2A                    | rs10811661 | chr09 | 22124094      | 2.70E-08          | 2.27E-08          | 0.0583          | 0.0571         | ILMN_1664466 | 21333931    | KLHL9      |
|                           |        |        |               | 0.0031            | 0.002             | 0.0553          | 0.0592         | ILMN_1804184 | 44620441    | TMED4      |
| EKT2                      | rs11037909 | chr11 | 44212190      | 0.0433            | 0.043             | 0.0095          | 0.0092         | ILMN_1744295 | 21982666    | CDKN2A     |
| DCD                       | rs1153188 | chr12  | 53385263      | 0.0008            | 0.001             | 0.0763          | 0.0671         | ILMN_1804184 | 44624444    | CD2        |

Abbreviations: ADJ_T2D, additional adjustment for T2D; cdsn, 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).

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. A previous case-control association study by Lewis et al. 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 22044 Caucasian subjects. 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 JAZFI. The SNP rs864745 has been characterized as a risk factor in European populations by Zeggini et al. in a large meta-analysis. Deletion of the JAZFI 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. The rs864745 was associated with JAZFI expression in muscle in our prior population-based sample of mixed ethnicity using RT-PCR, where the association was largely contributed by African Americans. Our gene expression arrays were unable to detect significant expression of the JAZFI transcripts in GENNID transformed lymphoblast cell lines. The SNPs in the JAZFI 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. 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 JAZFI 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. 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.
