ORIGINAL ARTICLE

Allele-Specific Transcriptional Activity at Type 2 Diabetes–Associated Single Nucleotide Polymorphisms in Regions of Pancreatic Islet Open Chromatin at the *JAZF1* Locus

Marie P. Fogarty,[1] Tami M. Panhuis,[2] Swarooparani Vadlamudi,[1] Martin L. Buchkovich,[1] and Karen L. Mohlke[1]

Translation of noncoding common variant association signals into meaningful molecular and biological mechanisms explaining disease susceptibility remains challenging. For the type 2 diabetes association signal in *JAZF1* intron 1, we hypothesized that the underlying risk variants have *cis*-regulatory effects in islets or other type 2 diabetes–relevant cell types. We used maps of experimentally predicted open chromatin regions to prioritize variants for functional follow-up studies of transcriptional activity. Twelve regions containing type 2 diabetes–associated variants were tested for enhancer activity in 832/13 and MIN6 insulinoma cells. Three regions exhibited enhancer activity and only rs1635852 displayed allelic differences in enhancer activity; the type 2 diabetes risk allele T showed lower transcriptional activity than the nonrisk allele C. This risk allele showed increased binding to protein complexes, suggesting that it functions as part of a transcriptional repressor complex. We applied DNA affinity capture to identify factors in the complex and determined that the risk allele preferentially binds the pancreatic master regulator PDX1. These data suggest that the rs1635852 region in *JAZF1* intron 1 is part of a *cis*-regulatory complex and that maps of open chromatin are useful to guide identification of variants with allelic differences in regulatory activity at type 2 diabetes loci. ***Diabetes* 62:1756–1762, 2013**

Genome-wide association studies (GWAS) have identified >50 genome-wide significant loci associated with type 2 diabetes to date (1). For many of these loci, association signals are localized to nonprotein-coding intronic and intergenic regions, which may contain variants that regulate gene transcription. A primary challenge remains in the transition from GWAS locus discovery to identification of functional variants underlying disease susceptibility. Tools are needed to detect functional variants from the set of disease-associated variants at a locus. FAIRE-seq (formaldehyde-assisted isolation of regulatory elements) and DNase-seq are two methods that identify nucleosome-depleted (open chromatin) regions comprising active DNA regulatory elements that include promoters, enhancers, silencers, and insulators (2,3). Integration of associated variants identified through GWAS with tissue-relevant genome-wide maps of open chromatin from the ENCODE Consortium (4) and epigenomic maps from the Roadmap Epigenomics Consortium (5) has great potential to facilitate identification of regulatory variants (6–8).

JAZF1 (juxtaposed with another zinc finger protein) is one such locus containing variants strongly associated with type 2 diabetes ($P = 5 \times 10^{-14}$) that are located within intron 1 (9). *JAZF1* encodes a putative transcription factor. JAZF1 protein interacts with NR2C2 (nuclear receptor subfamily 2, group C, member 2) protein and represses NR2C2-mediated transactivation (10). NR2C2, also known as TR4, is an orphan nuclear receptor targeting many genes important in metabolism (11). Mice lacking *Nr2c2* have perinatal and postnatal hypoglycemia (12) and are protected from glucose intolerance and insulin resistance induced with a high-fat diet (13). *JAZF1* locus variants have been associated with impaired β-cell function (14). The function of *JAZF1* as a transcriptional repressor of a gene negatively influencing glucose metabolism suggests that susceptibility alleles at this locus may result in decreased *JAZF1* transcription. Of note, islets from type 2 diabetes donors display decreased *JAZF1* expression, and higher levels of islet *JAZF1* are correlated with higher insulin secretion and higher glycemic control (15). Single nucleotide polymorphisms (SNPs) at the *JAZF1* association signal in intron 1 also are associated with height (16), and an independent signal ($r^2 = 0.024$ with GWAS index SNP rs849134) located >200 kb away is associated with prostate cancer (17). Additionally, expression quantitative trait locus (eQTL) analysis supports association of the type 2 diabetes risk allele with altered *JAZF1* transcript level in adipose tissue, liver, and muscle (18–20).

To gain insight into the molecular mechanisms underlying the type 2 diabetes association at *JAZF1*, we used maps of open chromatin to guide identification of variants in potential regulatory elements. Based on evidence of an effect in β-cell function and insulin secretion (14,15), experiments were performed in two available mammalian β-cell lines. We measured transcriptional activity of prioritized variants using luciferase reporter assays and report a candidate *cis*-acting SNP that displays allele-specific enhancer activity. We also evaluated allelic differences in protein binding and provide evidence of a potential molecular mechanism for *JAZF1* SNP effects.

RESEARCH DESIGN AND METHODS

Selection of SNPs for functional study. Variants were prioritized for functional study based on linkage disequilibrium (LD) and evidence of islet

From the [1]Department of Genetics, University of North Carolina, Chapel Hill, North Carolina; and the [2]Department of Zoology, Ohio Wesleyan University, Delaware, Ohio.
Corresponding author: Karen L. Mohlke, mohlke@med.unc.edu.
Received 19 July 2012 and accepted 4 December 2012.
DOI: 10.2337/db12-0972
This article contains Supplementary Data online at http://diabetes.diabetesjournals.org/lookup/suppl/doi:10.2337/db12-0972/-/DC1.
M.P.F. and T.M.P. contributed equally to this work.

open chromatin. All five SNPs in high LD ($r^2 \geq 0.8$; CEPH [Utah residents with ancestry from northern and western Europe] [CEU] 1,000G, March 2012 release) with the GWAS index SNP rs849134 and present in an islet FAIRE peak (8) or DNase peak (6) were tested for evidence of differential transcriptional activity. We also tested three SNPs based on only high LD and four other SNPs in the region (low LD SNPs; $r^2 < 0.5$; Supplementary Table 1).

Cell culture. Two insulinoma cell lines, rat-derived 832/13 (21) (C.B. Newgard, Duke University, Durham, NC) and mouse-derived MIN6 (22) were maintained at 37°C with 5% CO_2. The 832/13 cells were cultured in RPMI 1640 (Invitrogen, Carlsbad, CA) supplemented with 10% FBS, 1 mmol/L sodium pyruvate, 2 mmol/L L-glutamine, 10 mmol/L HEPES, and 0.05 mmol/L β-mercaptoethanol. MIN6 cells were cultured in DMEM (Invitrogen) supplemented with 10% FBS, 1 mmol/L sodium pyruvate, and 0.1 mmol/L β-mercaptoethanol. HepG2 hepatocellular carcinoma cells were cultured in MEM-α (Invitrogen) supplemented with 10% FBS, 1 mmol/L sodium pyruvate, and 2 mmol/L L-glutamine. Differentiated 3T3-L1 adipocyte cells were maintained in DMEM supplemented with 10% FBS, 1 μmol/L dexamethasone, 0.5 mmol/L IBMX, and 1 ug/mL insulin.

Generation of luciferase reporter constructs, transient DNA transfection, and luciferase reporter assays. The 150- to 200-bp fragments surrounding each of 12 SNPs were PCR-amplified (Supplementary Table 2) from DNA of individuals homozygous for risk and nonrisk alleles. Restriction sites for *Kpn*I and *Xho*I were added to primers during amplification, and the resulting PCR products were digested with *Kpn*I and *Xho*I and cloned in both orientations into the multiple cloning site of the minimal promoter-containing firefly luciferase reporter vector pGL4.23 (Promega, Madison, WI). Fragments are designated as forward or reverse based on their orientation in the genome with respect to the *JAZF1* coding sequence. Two to four independent clones for each allele for each orientation were isolated, verified by sequencing, and transfected in duplicate into 832/13 and MIN6 β-cell lines.

Approximately 1×10^{-5} cells per well were seeded in 24-well plates. At 80% confluency, cells were cotransfected with luciferase constructs and *Renilla* control reporter vector (phRL-TK; Promega) at a ratio of 30:1 for MIN6 using Lipofectamine 2000 (Invitrogen) and at a ratio of 10:1 for 832/13 cells using FUGENE-6 (Roche Diagnostics, Indianapolis, IN). At 48 h after transfection, cells were lysed with passive lysis buffer (Promega), and luciferase activity was measured using the Dual-Luciferase Assay System (Promega). To control for transfection efficiency, raw values for firefly luciferase activity were divided by raw *Renilla* luciferase activity values, and fold change was calculated as normalized luciferase values divided by pGL4.23 minimal promoter empty vector control values. Data are reported as the fold change in mean (±SE) relative luciferase activity per allele. A two-sided *t* test was used to compare luciferase activity between alleles. Experiments in MIN6 and 832/13 cells were performed on a second independent day and yielded comparable allele-specific results.

Electrophoretic mobility shift assay. Nuclear cell extracts were prepared from 832/13, MIN6, HepG2, and 3T3-L1 cells using the NE-PER nuclear and cytoplasmic extraction kit (Thermo Scientific) according to the manufacturer's instructions. Protein concentration was measured with a BCA protein assay (Thermo Scientific), and lysates were stored at −80°C until use. The 17-bp oligonucleotides were designed to the sequence surrounding rs1635852 risk or nonrisk alleles as follows: sense 5′ biotin-CTGATTAA[**T/C**]TCACTTAG 3′ and antisense 5′ biotin-CTAAGTGA[**G/C**]TTAATCAG3′ (SNP allele in bold). Double-stranded oligonucleotides for the risk and nonrisk alleles were generated by incubating 50 pmol complementary oligonucleotides at 95°C for 5 min, followed by gradual cooling to room temperature. Electrophoretic mobility shift assays (EMSAs) were performed using the LightShift Chemiluminescent EMSA Kit (Thermo Scientific). Binding reactions were set-up as 1× binding buffer, 50 ng/μL poly (dI·dC), 3 μg nuclear extract, and 20 fmol labeled probe in a final volume of 20 μL. For competition reactions, 67-fold excess of unlabeled double-stranded oligonucleotides for either the risk or the nonrisk allele were included. Reactions were incubated at room temperature for 25 min. For supershift assays, 4 μg polyclonal antibodies against PDX1 (SC-14662×; Santa Cruz Biotechnology) or CUX1 (SC6327×; Santa Cruz Biotechnology) were added to the binding reaction and incubation proceeded for a further 25 min. Binding reactions were subjected to nondenaturing PAGE on DNA retardation gels in 0.5× TBE (Invitrogen), transferred to nylon membranes (Invitrogen), and cross-linked on an ultraviolet light cross-linker (Stratagene). Biotin labeled DNA–protein complexes were detected by chemiluminescence. EMSAs were performed on a second independent day and yielded comparable results.

DNA affinity capture assay. Nuclear extracts (prepared as for EMSA) were dialyzed against dialysis buffer (20 mmol/L Tris/HCl [pH 7.9], 70 mmol/L KCl, 1 mmol/L EDTA) in a Slide-A-Lyzer MINI Dialysis Device (Thermo Scientific). Dialyzed nuclear extracts (300 μg) were precleared with 100 μL streptavidin-agarose dynabeads (Invitrogen) coupled to biotin-labeled scrambled control oligonucleotides. This preclearing step was performed to reduce nonspecific binding of nuclear protein. For DNA-protein binding reactions, 40 pmol of biotin-labeled probe either for rs1635852 allele (same probe as for EMSA) or for a scrambled control were incubated with 300 μg nuclear extract, binding buffer (10 mmol/L Tris, 50 mmol/L KCL, 1 mmol/L DTT), 0.5 μg/μL poly (dI·dC), and H_2O to total 450 μL at room temperature for 30 min with rotation; 100 μL (1 mg) of streptavidin-agarose dynabeads were added and the reaction was incubated for a further 20 min. Beads were washed four times in binding buffer with 0.05% NP-40 and once in binding buffer without NP-40. DNA-bound proteins were eluted in 1× reducing sample buffer (Invitrogen) by heating for 10 min at 70°C. Proteins were separated on NuPAGE denaturing gels and protein bands were stained with SYPRO-Ruby. Protein bands displaying differential binding between rs1635852 alleles (but not scrambled control) were excised from the gel and subjected to matrix-assisted laser desorption time-of-flight/time-of-flight tandem mass spectrometry (MS) and analysis at the University of North Carolina proteomics core facility. For peptide identification, all MS/MS spectra were searched against all entries in the National Center for Biotechnology Information nonredundant database using GPS Explorer software version 3.6 (ABI) and Mascot (MatrixScience) search algorithm. Mass tolerance was 80 ppm for precursor ions and 0.6 Da for fragment ions were used. In addition, two missed cleavages were allowed and oxidation of methionine was a variable modification.

RESULTS

Characterization of type 2 diabetes–associated SNPs in *JAZF1* intron 1 with regulatory potential. To distinguish potentially functional regulatory variants from proxy variants in high LD at the type 2 diabetes–associated *JAZF1* locus, we selected variants in high LD ($r^2 > 0.8$) with GWAS index SNP rs849134. To further prioritize variants for functional analysis, we used genome-wide maps of open chromatin (Fig. 1*A*, *C*) in available type 2 diabetes–relevant cell types, including pancreatic islets, liver hepatocytes, and skeletal muscle myotubes. DNase-seq and FAIRE-seq are well-established methods identifying both overlapping and unique nucleosome-depleted regions that include active regulatory elements (23). We also evaluated variant position with respect to the histone modifications H3K4me1 and H3K9ac, which are post-translational marks associated with enhancer regions (Fig. 1*D*) (24,25). Of 15 variants meeting the LD threshold, five SNPs were found to overlap an islet FAIRE peak or DNase peak and displayed some evidence of H3K4me1 or H3K9ac signal (Fig. 1). Whereas enhancer histone modification profiles served to support evidence of function, they did not help for SNP prioritization because of an overall high background signal (Fig. 1*D*). No SNPs overlapped with DNase or FAIRE peaks in the other cell types examined (data not shown).

rs1635852 displays allele-specific enhancer activity in islet cells. To evaluate transcriptional activity of the five SNPs predicted to be in islet-regulatory regions, we cloned ~200 bp surrounding each SNP allele into a minimal promoter reporter vector and measured luciferase activity in two β-cell lines, 832/13 rat insulinoma and MIN6 mouse insulinoma cells. Two to four independent clones for each allele were generated and enhancer activity was measured in duplicate for each clone. Of the five SNPs in open chromatin region, three displayed evidence of enhancer activity compared with an empty vector control in both cell lines in the forward orientations –rs1635852, rs849133, and rs849142 (Fig. 2*A*, *B*). Of these, rs1635852 showed differential allelic enhancer activity in both orientations in both cell lines. The risk allele T showed significantly decreased luciferase activity compared with the nonrisk allele C (forward: 832/13 [$P = 7.8 \times 10^{-5}$] and MIN6 [$P = 1.0 \times 10^{-5}$]; reverse: 832/13 [$P = 1.8 \times 10^{-2}$] and MIN6 [$P = 8.3 \times 10^{-4}$]; Fig. 2*A*, *B*). Enhancer activity at rs1635852 represents >2-fold and 1.7-fold increases in transcriptional activity

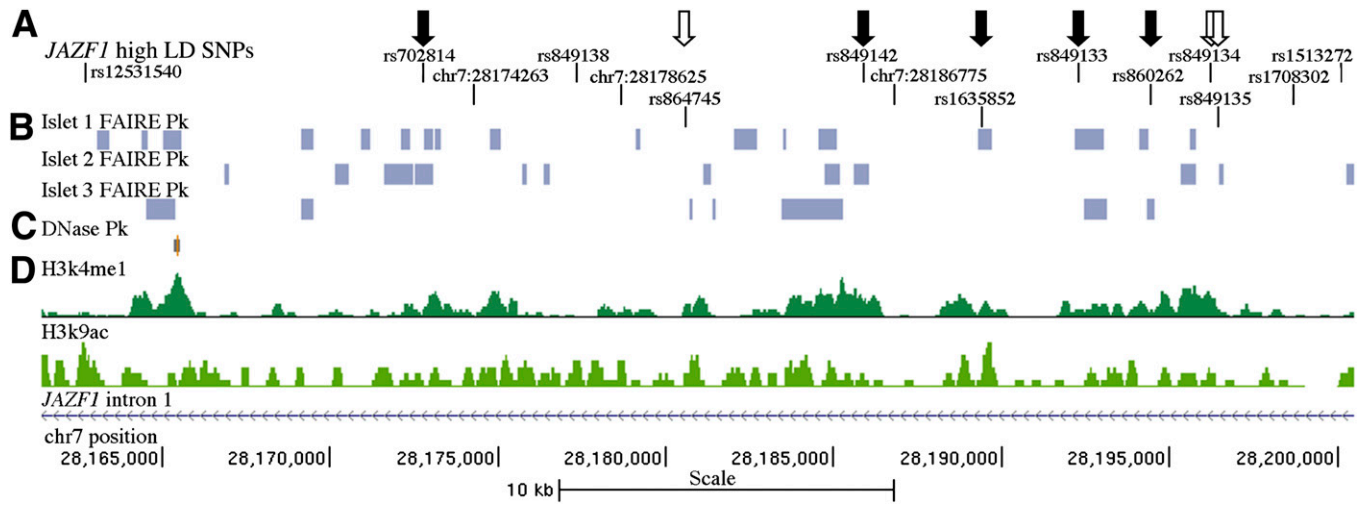

FIG. 1. Regulatory potential at type 2 diabetes–associated SNPs at the *JAZF1* locus. *A*: Twelve high LD SNPs ($r^2 \geq 0.80$ with GWAS index SNP rs849134). Closed arrows indicate five SNPs overlapping open chromatin marks tested for allele-specific transcriptional activity. Open arrows indicate three SNPs without evidence of open chromatin and tested for allele-specific transcriptional activity for comparison. Four high LD SNPs without evidence of open chromatin were not tested for allele-specific transcriptional activity. *B*: FAIRE peaks identified in three islet samples. *C*: DNase hypersensitivity peaks identified in two pooled islet samples from the ENCODE Consortium. *D*: Islet H3K4me1 and H3K9ac histone modifications from the Roadmap Epigenomics Consortium. Three high LD indels that do not overlap with open chromatin and lack reference SNP ID numbers (rs#) are shown. Four additional low LD SNPs located 12–20 kb proximal to the region shown were tested for allele-specific transcriptional activity. Image is taken from University of California, Santa Cruz, genome browser, February 2009 (GRCh37/hg19) assembly (http://genome.ucsc.edu) (38).

relative to the risk allele in forward and reverse orientation, respectively, in both cell lines. No allele-specific enhancer activity was observed for rs849133 and rs849142. Transcriptional activity was evaluated in seven other type 2 diabetes–associated SNPs in *JAZF1* intron 1 not overlapping islet FAIRE or DNase peaks, three SNPs in high LD and four SNPs in lower LD, with GWAS index SNP rs849134 (Supplementary Table 1). None of these seven SNPs showed evidence of enhancer activity in 832/13 or MIN6 cells compared with an empty vector control (Supplementary Fig. 1). Taken together, these data demonstrate that rs1635852 exhibits allelic differences in transcriptional enhancer activity and suggest it functions within a candidate *cis*-regulatory element at the *JAZF1* intronic type 2 diabetes–associated locus.

Differential protein binding to rs1635852 alleles. We next asked whether alleles of rs1635852 differentially affect DNA binding to nuclear proteins. Biotin-labeled probes surrounding the T (risk) or C (nonrisk) allele of rs1635852 were incubated with 832/13 or MIN6 nuclear lysate and subjected to EMSA. Band shifts indicative of multiple DNA–protein complexes were observed for both alleles of rs1635852 (Fig. 3*A*, *C*, *D*). In the EMSA from 832/13 nuclear extract, three protein complexes were observed for the probe containing the T allele that were either less intense (arrow a) or not present (arrow b, c) for the probe containing the C allele. In the EMSA from MIN6 nuclear extract, two protein complexes were observed for the probe containing the T allele that were either less intense (arrow a) or not present (arrow b) for the probe containing the C allele. Taken together, these data suggest differential protein binding dependent on the rs1635852 allele. Competition of labeled T allele with excess unlabeled T allele more efficiently competed away allele-specific bands than excess unlabeled C allele, demonstrating allele-specificity of the protein–DNA complexes (Fig. 3*A*). Based on these results, we hypothesized that rs1635852 is located in a binding site for a transcriptional regulator complex that may be disrupted by the rs1635852 C allele.

Identification of proteins binding rs1635852. We sought to identify factors in the protein complex binding rs1635852 using a DNA-affinity capture assay. The same biotin-labeled probes used for EMSA including the rs1635852 T or C alleles were incubated with 832/13 nuclear lysates, and the resulting DNA–protein complexes were isolated and subjected to SDS-PAGE. We observed two protein bands showing differential intensity consistent with increased binding to the T allele (Fig. 3*B*). These bands were identified as transcription factors PDX1 (pancreatic duodenal homeobox 1) and CUX1 (cut-like homeobox 1) using MALDI TOF/TOF MS.

To validate binding of rs1635852 to these transcription factors, we performed supershift experiments incubating DNA–protein complexes with antibodies for either PDX1 or CUX1. Incubation of the T allele–protein complex with PDX1 antibody completely disrupted protein–DNA binding at complex c (832/13; Fig. 3*C*, arrow c) and reduced binding to protein complex b (Fig. 3*C*, 832/13; Fig. 3*D*, MIN6, arrow b). A PDX1-mediated supershift was observed in MIN6 cells for both rs1635852 alleles (Fig. 3*D*). A consistent supershift band was not observed in 832/13 cells and may have been masked by presence of a larger DNA–protein complex (Fig. 3*C*, arrow d). No evidence of a PDX-mediated complex disruption or band supershift was observed in 3T3-L1 and HepG2 nuclear extracts (Supplementary Fig. 2*A*, *B*). In contrast, incubation with CUX1 only slightly reduced binding to complex b and c (Fig. 3*C*, *D*). These data provide evidence that the rs1635852 T allele binds PDX1 in an allele-specific manner. Consistent with this result, a search of the JASPAR CORE database (26) shows that only the sequence containing the rs1635852 T allele is predicted as a PDX1 consensus core-binding motif.

DISCUSSION

At many of the loci identified through GWAS, association signals are localized to intronic and intergenic regions and

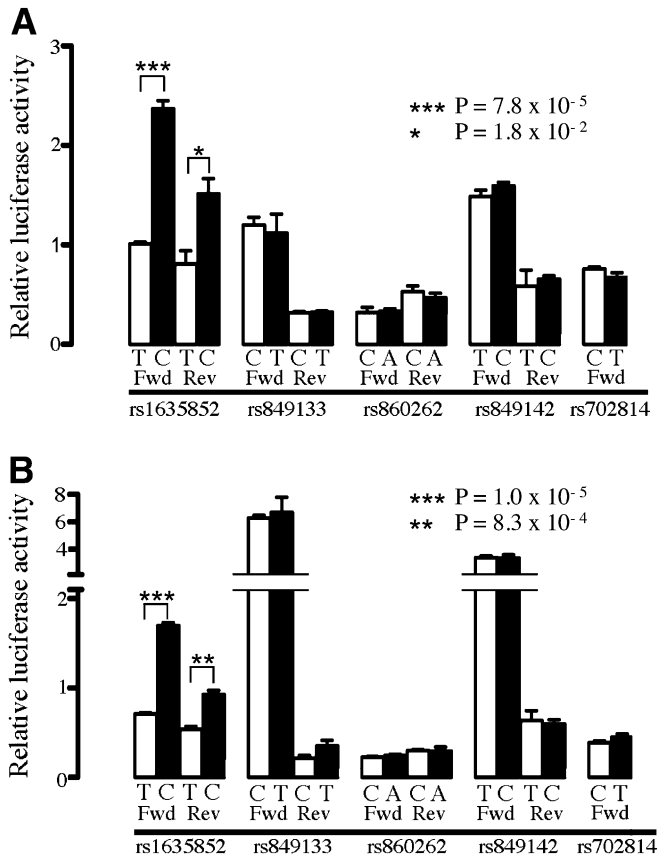

FIG. 2. rs1635852 alleles display differential transcriptional activity. *A*: Enhancer activity was tested in 832/13 cells for the type 2 diabetes risk (white bars) and nonrisk (black bars) alleles of five SNPs in candidate regulatory regions in forward (Fwd) and reverse (Rev) orientations with respect to *JAZF1*. Significant allele-specific enhancer activity was observed for rs1635852. The rs1635852 risk allele T shows less transcriptional activity than the nonrisk allele C in both orientations with respect to a minimal promoter vector. *B*: rs1635852 risk allele displays similar decreased transcriptional activity in MIN6 cells. Error bars represent SE of 2–4 independent clones for each allele. Results are expressed as fold change compared with empty vector control. *P* values were calculated by a two-sided *t* test.

are hypothesized to harbor nonpromoter regulatory elements altering gene transcription. In the current study, we focused on the type 2 diabetes–associated signal in *JAZF1* intron 1 and asked whether variants at this locus displayed allele-specific transcriptional enhancer activity consistent with a *cis*-regulatory effect. We used maps of open chromatin to guide identification of variants with allelic differences in transcriptional activity. We provide evidence that rs1635852 is a strong candidate for a differential effect on transcriptional enhancer activity, likely through altered binding in a regulatory complex containing PDX1. SNPs evaluated include the *JAZF1* lead SNPs from two recent descriptions of high-density genotyping at this locus (27,28).

A challenge in mechanistic studies of GWAS signals is differentiating among the numerous SNPs to identify those underlying the disease association. On the basis of the assumption that a common variant with modest effect likely underlies the association at *JAZF1*, we aligned high LD variants ($r^2 \geq 0.8$; $n = 14$) in *JAZF1* intron 1 with active DNA regulatory elements identified by DNase-seq and FAIRE-seq and found that five variants overlapped with islet-regulatory regions. It is important to recognize that current available data sets for FAIRE and DNase may not be complete and that SNPs in lower LD ($r^2 < 0.8$) with the index SNP also may affect gene expression or activity. Therefore, additional functional SNPs may exist at the *JAZF1* type 2 diabetes locus that were not examined in this study.

Notably, all five SNPs found in islet open chromatin regions overlapped with FAIRE peaks only, perhaps reflecting a recent observation of twice as many FAIRE sites as DNase sites when looking across multiple cell types (23). Cell-selective open chromatin tends to be located away from the transcription start site (23), suggesting that regulatory elements in intron 1 of *JAZF1* may be unique to islets. There was no evidence of overlap with open chromatin for associated SNPs in other available relevant tissue datasets examined. However, we cannot rule out that there may be shared regulatory regions in tissues for which FAIRE or DNase data are not currently available.

Choosing the correct tissue type for testing activity of functional variants is critical because many regulatory elements act in a tissue-specific manner (25,29). Based on our observation of type 2 diabetes–associated SNPs in regions of islet open chromatin, we measured transcriptional activity in two available mammalian islet cell models, 832/13 and MIN6 cells from rat and mouse, respectively. Importantly, we found similar fold changes in allelic transcriptional activity across the two cell types, suggesting that measurement of enhancer activity may be consistent across species.

Of three SNPs predicted to be located in active regulatory regions that displayed enhancer activity, only rs1635852 demonstrated allele-specific effects, making it a lead functional candidate among the SNPs tested. The T allele (risk) of rs1635852 displayed reduced enhancer activity relative to the C allele (nonrisk), suggesting that reduced expression of islet *JAZF1* may be associated with type 2 diabetes. We hypothesized that alleles of rs1635852 might differentially bind a transcription factor, resulting in altered gene transcription, and our analysis of protein binding revealed complexes that favored the rs1635852 T allele in 832/13 and MIN6 insulinoma cells. rs1635852 also showed a differential protein binding pattern in EMSA using 3T3-L1 mouse adipocyte and HepG2 human hepatocyte nuclear extracts but the pattern of binding differed from that of islet extracts with alternate DNA–protein complexes present (Supplementary Fig. 2*A*, *B*). Using DNA affinity capture, we identified PDX1 and CUX1 as two proteins that showed increased binding to the T allele and validated that in 832/13 and MIN6 cells, PDX1 antibody disrupted at least one of the protein complexes formed preferentially in the presence of the T allele. Our results suggest that the DNA sequence surrounding rs1635852 may bind protein differentially in multiple cell types but that the protein complexes involved likely differ because PDX1 binding was exclusive to insulinoma cell types. The fourth base of the PDX1 consensus core DNA binding motif (TAAT) is altered by the rs1635852 C allele. In contrast, the CUX1 consensus sequence, ATA (30), is not found in the 17 bp DNA sequence that includes rs1635852, suggesting that CUX1 may not directly bind DNA in this region, but instead may be bound to a DNA-binding protein in the complex. Additional transcription factors not identified here also may play a role in enhancer activity at rs1635852.

PDX1 plays a central role in embryonic development of pancreatic islets and in maintenance of normal glucose

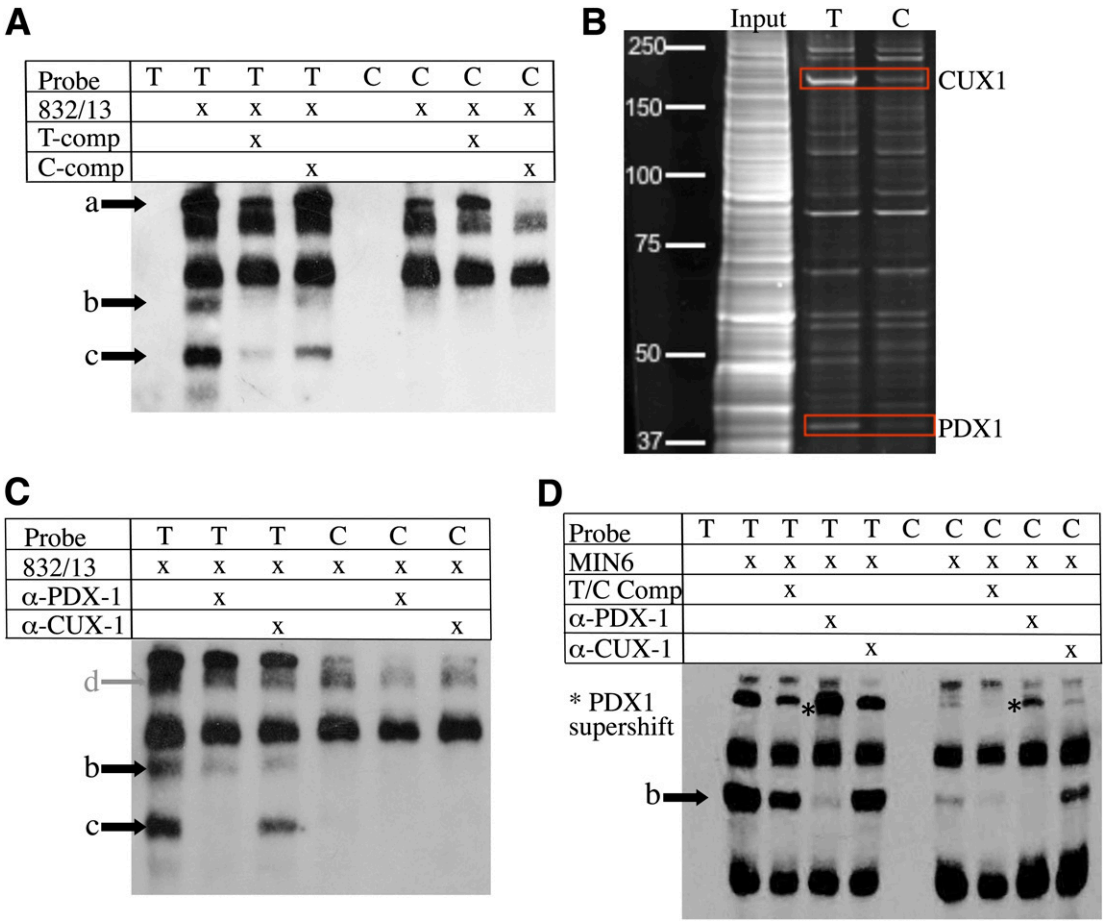

FIG. 3. Alleles of rs1635852 differentially bind PDX1 in rat 832/13 and mouse MIN6 insulinoma cells. *A*: EMSA using 832/13 nuclear extract shows differential protein–DNA binding of rs1635852 alleles. The probe containing risk allele T shows increased protein binding (arrows a, b, c) compared with the probe containing nonrisk allele C. Excess unlabeled specific probe containing the T allele (T-comp) more efficiently competed away allele-specific bands than unlabeled probe for the C allele (C-comp). To enhance visualization of protein complexes, free biotin-labeled probe is not shown. *B*: DNA affinity-capture identified differential binding of CUX1 and PDX1 at rs1635852 alleles in 832/13 cells. *C*: Incubation of 832/13 nuclear extract with PDX1 antibody disrupts the DNA–protein complex formed with T allele–containing DNA probe (arrows b and c). The presence of a nonallele-specific complex (arrow d) may mask a PDX1-mediated supershift. *D*: Incubation of MIN6 nuclear extract with PDX1 antibody disrupts the DNA–protein complex formed with T allele–containing DNA probe (arrow b) and results in a band supershift. (A high-quality color representation of this figure is available in the online issue.)

homeostasis in adult islets. Heterozygous mutations in PDX1 result in early and late onset of diabetes in humans (31–33). PDX1 binds to DNA sequences containing the consensus TAAT motif and subsequently recruits protein complexes that regulate transcription of several β-cell genes that include *INS* and *GLUT2* (34,35). Whereas the role of PDX1 as a gene activator is well-established, a recent analysis of genome-wide PDX1 occupancy using chromatin immunoprecipitation highlights a strong role for PDX1 as a transcriptional repressor (36). The same study demonstrates that PDX1 binds across 4,470 human genes, including to five regions of *JAZF1*, with 48% of binding sites being located in introns (36). Our data demonstrate reduced transcriptional activity with the rs1635852 T allele, suggesting that in this instance PDX1 may be functioning as part of a transcriptional repressor complex. Further experiments are necessary to validate binding to rs1635852 in human islets with known genotypes and to elucidate other key factors in the repressor complex.

Although *JAZF1* is not confirmed as the affected gene at this locus, eQTL analysis supports association of the type 2 diabetes risk allele with decreased *JAZF1* transcript levels in adipose tissue that is consistent with the direction of transcriptional activity we observed in islet cells (18). rs1635852 or proxy SNPs ($r^2 > 0.95$) are reported associated with *JAZF1* transcript levels in at least three tissues—liver, adipose, and muscle (18–20). In addition, *JAZF1* expression is decreased in islets from type 2 diabetes donors (15). The strongest evidence of a biological basis for *JAZF1* in type 2 diabetes susceptibility comes from observations that JAZF1 protein binds and represses action of the metabolic regulator NR2C2 (10). *Nr2c2*-deficient mice show reduced hepatic triglyceride levels, reduced lipid accumulation in adipose tissue, and are resistant to glucose intolerance and insulin resistance (13). *NR2C2* is expressed in pancreatic islets (37), but to our knowledge it has not been well-characterized. Our data suggesting differential allelic transcription in *JAZF1* may allude to a role for NR2C2 in β-cell gene regulation. Recent chromatin immunoprecipitation analysis of NR2C2 binding sites across the genome revealed that NR2C2 controls genes involved in fundamental biologic processes across diverse cell types and in targeting genes in a cell-specific manner (11). It is also possible that JAZF1 may serve as a cofactor of additional nuclear receptors, although it does not regulate transcriptional activation by two tested receptors, PPARA and RORG (10).

In summary, we demonstrate that maps of open chromatin are a useful resource to guide identification of variants in *cis*-acting regulatory elements at type 2 diabetes susceptibility loci, and we provide evidence that rs1635852, a SNP at the *JAZF1* type 2 diabetes–associated locus, differentially affects transcriptional activity through binding of a regulatory protein complex that includes PDX1.

ACKNOWLEDGMENTS

This research was funded by the National Institutes of Health (DK072193, DA027040).

No potential conflicts of interest relevant to this article were reported.

M.P.F. and T.M.P. designed research, performed research, and wrote the manuscript. S.V. and M.L.B. performed research. K.L.M. designed research and wrote the manuscript. K.L.M. is the guarantor of this work and, as such, had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis

Parts of this study were presented in abstract form at the American Society of Human Genetics meeting, San Francisco, California, 6–10 November 2012.

The authors recognize open chromatin data from the ENCODE consortium and Kyle Gaulton (Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom), Terry Furey (Department of Genetics, University of North Carolina, Chapel Hill, North Carolina), Greg Crawford (Institute for Genome Sciences and Policy, Duke University, Durham, North Carolina), and Jason Lieb (Department of Biology, University of North Carolina, Chapel Hill, North Carolina) for helpful interpretation of open chromatin data. The authors thank Doris Stoffers (Department of Medicine/Endocrinology, Diabetes and Metabolism, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania) for sharing unpublished research. The authors thank Gray Camp (Department of Cell and Molecular Physiology, University of North Carolina, Chapel Hill, North Carolina) for helpful advice on the DNA affinity experiments and the University of North Carolina Michael Hooker Proteomics Center (Chapel Hill, North Carolina) for assistance with protein identification.

REFERENCES

1. Visscher PM, Brown MA, McCarthy MI, Yang J. Five years of GWAS discovery. Am J Hum Genet 2012;90:7–24
2. Wu C, Wong YC, Elgin SC. The chromatin structure of specific genes: II. Disruption of chromatin structure during gene activity. Cell 1979;16:807–814
3. Giresi PG, Kim J, McDaniell RM, Iyer VR, Lieb JD. FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. Genome Res 2007;17:877–885
4. Consortium EP; ENCODE Project Consortium. The ENCODE (ENCyclopedia Of DNA Elements) Project. Science 2004;306:636–640
5. Zhou X, Maricque B, Xie M, et al. The Human Epigenome Browser at Washington University. Nat Methods 2011;8:989–990
6. Stitzel ML, Sethupathy P, Pearson DS, et al.; NISC Comparative Sequencing Program. Global epigenomic analysis of primary human pancreatic islets provides insights into type 2 diabetes susceptibility loci. Cell Metab 2010;12:443–455
7. Paul DS, Nisbet JP, Yang TP, et al.; Cardiogenics Consortium; MuTHER Consortium. Maps of open chromatin guide the functional follow-up of genome-wide association signals: application to hematological traits. PLoS Genet 2011;7:e1002139
8. Gaulton KJ, Nammo T, Pasquali L, et al. A map of open chromatin in human pancreatic islets. Nat Genet 2010;42:255–259
9. Zeggini E, Scott LJ, Saxena R, et al.; Wellcome Trust Case Control Consortium. Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. Nat Genet 2008;40:638–645
10. Nakajima T, Fujino S, Nakanishi G, Kim YS, Jetten AM. TIP27: a novel repressor of the nuclear orphan receptor TAK1/TR4. Nucleic Acids Res 2004;32:4194–4204
11. O'Geen H, Lin YH, Xu X, et al. Genome-wide binding of the orphan nuclear receptor TR4 suggests its general role in fundamental biological processes. BMC Genomics 2010;11:689
12. Collins LL, Lee YF, Heinlein CA, et al. Growth retardation and abnormal maternal behavior in mice lacking testicular orphan nuclear receptor 4. Proc Natl Acad Sci USA 2004;101:15058–15063
13. Kang HS, Okamoto K, Kim YS, et al. Nuclear orphan receptor TAK1/TR4-deficient mice are protected against obesity-linked inflammation, hepatic steatosis, and insulin resistance. Diabetes 2011;60:177–188
14. Grarup N, Andersen G, Krarup NT, et al. Association testing of novel type 2 diabetes risk alleles in the JAZF1, CDC123/CAMK1D, TSPAN8, THADA, ADAMTS9, and NOTCH2 loci with insulin release, insulin sensitivity, and obesity in a population-based sample of 4,516 glucose-tolerant middle-aged Danes. Diabetes 2008;57:2534–2540
15. Taneera J, Lang S, Sharma A, et al. A systems genetics approach identifies genes and pathways for type 2 diabetes in human islets. Cell Metab 2012;16:122–134
16. Johansson A, Marroni F, Hayward C, et al.; EUROSPAN Consortium. Common variants in the JAZF1 gene associated with height identified by linkage and genome-wide association analysis. Hum Mol Genet 2009;18:373–380
17. Thomas G, Jacobs KB, Yeager M, et al. Multiple loci identified in a genome-wide association study of prostate cancer. Nat Genet 2008;40:310–315
18. Voight BF, Scott LJ, Steinthorsdottir V, et al.; MAGIC investigators; GIANT Consortium. Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. Nat Genet 2010;42:579–589
19. Zhong H, Yang X, Kaplan LM, Molony C, Schadt EE. Integrating pathway analysis and genetics of gene expression for genome-wide association studies. Am J Hum Genet 2010;86:581–591
20. Sharma NK, Langberg KA, Mondal AK, Elbein SC, Das SK. 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 2011;96:E394–E403
21. Hohmeier HE, Mulder H, Chen G, Henkel-Rieger R, Prentki M, Newgard CB. Isolation of INS-1-derived cell lines with robust ATP-sensitive K+ channel-dependent and -independent glucose-stimulated insulin secretion. Diabetes 2000;49:424–430
22. Miyazaki J, Araki K, Yamato E, et al. Establishment of a pancreatic beta cell line that retains glucose-inducible insulin secretion: special reference to expression of glucose transporter isoforms. Endocrinology 1990;127:126–132
23. Song L, Zhang Z, Grasfeder LL, et al. Open chromatin defined by DNaseI and FAIRE identifies regulatory elements that shape cell-type identity. Genome Res 2011;21:1757–1767
24. Creyghton MP, Cheng AW, Welstead GG, et al. Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc Natl Acad Sci USA 2010;107:21931–21936
25. Heintzman ND, Hon GC, Hawkins RD, et al. Histone modifications at human enhancers reflect global cell-type-specific gene expression. Nature 2009;459:108–112
26. Bryne JC, Valen E, Tang MH, et al. JASPAR, the open access database of transcription factor-binding profiles: new content and tools in the 2008 update. Nucleic Acids Res 2008;36(Database issue):D102–D106
27. Saxena R, Elbers CC, Guo Y, et al.; Look AHEAD Research Group; DIAGRAM consortium. Large-scale gene-centric meta-analysis across 39 studies identifies type 2 diabetes loci. Am J Hum Genet 2012;90:410–425
28. Morris AP, Voight BF, Teslovich TM, et al.; Wellcome Trust Case Control Consortium; Meta-Analyses of Glucose and Insulin-related traits Consortium (MAGIC) Investigators; Genetic Investigation of ANthropometric Traits (GIANT) Consortium; Asian Genetic Epidemiology Network–Type 2 Diabetes (AGEN-T2D) Consortium; South Asian Type 2 Diabetes (SAT2D) Consortium; DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium. Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. Nat Genet 2012;44:981–990
29. Visel A, Blow MJ, Li Z, et al. ChIP-seq accurately predicts tissue-specific activity of enhancers. Nature 2009;457:854–858
30. Stratigopoulos G, LeDuc CA, Cremona ML, Chung WK, Leibel RL. Cut-like homeobox 1 (CUX1) regulates expression of the fat mass and obesity-associated and retinitis pigmentosa GTPase regulator-interacting protein-1-like (RPGRIP1L) genes and coordinates leptin receptor signaling. J Biol Chem 2011;286:2155–2170

31. Macfarlane WM, Frayling TM, Ellard S, et al. Missense mutations in the insulin promoter factor-1 gene predispose to type 2 diabetes. J Clin Invest 1999;104:R33–R39
32. Hani EH, Stoffers DA, Chèvre JC, et al. Defective mutations in the insulin promoter factor-1 (IPF-1) gene in late-onset type 2 diabetes mellitus. J Clin Invest 1999;104:R41–R48
33. Stoffers DA, Ferrer J, Clarke WL, Habener JF. Early-onset type-II diabetes mellitus (MODY4) linked to IPF1. Nat Genet 1997;17:138–139
34. Waeber G, Thompson N, Nicod P, Bonny C. Transcriptional activation of the GLUT2 gene by the IPF-1/STF-1/IDX-1 homeobox factor. Mol Endocrinol 1996;10:1327–1334
35. Petersen HV, Serup P, Leonard J, Michelsen BK, Madsen OD. Transcriptional regulation of the human insulin gene is dependent on the homeodomain protein STF1/IPF1 acting through the CT boxes. Proc Natl Acad Sci USA 1994;91:10465–10469
36. Khoo C, Yang J, Weinrott SA, et al. Research resource: the pdx1 cistrome of pancreatic islets. Mol Endocrinol 2012;26:521–533
37. Eizirik DL, Sammeth M, Bouckenooghe T, et al. The human pancreatic islet transcriptome: expression of candidate genes for type 1 diabetes and the impact of pro-inflammatory cytokines. PLoS Genet 2012;8:e1002552
38. Fujita PA, Rhead B, Zweig AS, et al. The UCSC Genome Browser database: update 2011. Nucleic Acids Res 2011;39(Database issue):D876–D882