A Common Functional Regulatory Variant at a Type 2 Diabetes Locus Upregulates *ARAP1* Expression in the Pancreatic Beta Cell

Jennifer R. Kulzer,[1] Michael L. Stitzel,[2,3] Mario A. Morken,[2] Jeroen R. Huyghe,[4] Christian Fuchsberger,[4] Johanna Kuusisto,[5] Markku Laakso,[5] Michael Boehnke,[4] Francis S. Collins,[2] and Karen L. Mohlke[1,*]

Genome-wide association studies (GWASs) have identified more than 70 loci associated with type 2 diabetes (T2D), but for most, the underlying causal variants, associated genes, and functional mechanisms remain unknown. At a T2D- and fasting-proinsulin-associated locus on 11q13.4, we have identified a functional regulatory DNA variant, a candidate target gene, and a plausible underlying molecular mechanism. Fine mapping, conditional analyses, and exome array genotyping in 8,635 individuals from the Metabolic Syndrome in Men study confirmed a single major association signal between fasting proinsulin and noncoding variants ($p = 7.4 \times 10^{-50}$). Measurement of allele-specific mRNA levels in human pancreatic islet samples heterozygous for rs11603334 showed that the T2D-risk and proinsulin-decreasing allele (C) is associated with increased *ARAP1* expression ($p < 0.02$). We evaluated four candidate functional SNPs for allelic effects on transcriptional activity by performing reporter assays in rodent pancreatic beta cell lines. The C allele of rs11603334, located near one of the *ARAP1* promoters, exhibited 2-fold higher transcriptional activity than did the T allele ($p < 0.0001$); three other candidate SNPs showed no allelic differences. Electrophoretic mobility shift assays demonstrated decreased binding of pancreatic beta cell transcriptional regulators PAX6 and PAX4 to the rs11603334 C allele. Collectively, these data suggest that the T2D-risk allele of rs11603334 could abrogate binding of a complex containing PAX6 and PAX4 and thus lead to increased promoter activity and *ARAP1* expression in human pancreatic islets. This work suggests that increased *ARAP1* expression might contribute to T2D susceptibility at this GWAS locus.

Introduction

Genome-wide association studies (GWASs) have identified more than 70 loci associated with type 2 diabetes (T2D [MIM 125853])[1–9] and ten loci associated with fasting proinsulin levels.[10,11] A 11q13.4 locus near *ARAP1* (MIM 606646), *PDE2A* (MIM 602658), *STARD10*, *ATG16L2*, and *FCHSD2* is strongly associated with T2D (rs1552224, $p = 1.4 \times 10^{-22}$),[1] fasting proinsulin (rs11603334, $p = 3.2 \times 10^{-102}$),[10] and 32,33-split proinsulin (rs11603334, $p = 1.2 \times 10^{-25}$).[10] This locus is also nominally associated with the insulinogenic index (rs1552224, $p = 2 \times 10^{-6}$) and both insulin ($p = 0.001$) and glucose ($p = 2 \times 10^{-5}$) levels at 30 min during an oral glucose tolerance test.[12] The clustering of multiple phenotypic associations related to proinsulin processing and insulin secretory response during an oral glucose challenge suggests that the affected target gene(s) might play a role in pancreatic beta cell function. Currently, the functional DNA variant(s), the affected gene(s), and the underlying molecular genetic mechanism(s) contributing to these associations are unknown.

SNPs rs1552224 and rs11603334 are in perfect linkage disequilibrium (LD) with each other ($r^2 = 1$, 1000 Genomes European ancestry [EUR]) and are located within the first exon and 5′ UTR of *ARAP1* RefSeq isoforms NM_015242.4 and NM_001135190.1 (rs1552224, +305 nt from the transcription start site [TSS]; rs11603334, +418 nt from the TSS) at 72.11 Mb (hg18, UCSC Genome Browser) on chromosome 11. A third *ARAP1* RefSeq isoform, NM_001040118.2, is expressed from an alternative promoter and TSS located at 72.14 Mb. We designate the promoter at 72.11 Mb as P1 and the promoter at 72.14 Mb as P2.

We hypothesized that functional variant(s) at this locus are in high LD ($r^2 \geq 0.8$) with rs1552224 and rs11603334. None of the variants in high LD with these SNPs are within gene coding regions where they could alter protein function, suggesting that functional common SNP(s) at this locus might influence gene regulation. Genes within the LD region containing rs1552224 and rs11603334 have reported expression in human pancreas, islets, and flow-sorted beta cells;[1] however, islet expression of *ATG16L2*, *FCHSD2*, and *PDE2A* might not be above background.[13–15] None of these genes have been demonstrated to play roles in insulin processing or secretion. ARAP1 activates Arf and Rho GTPases, which regulate membrane trafficking and actin cytoskeleton reorganization.[16] STARD10 binds and transfers membrane phospholipids.[17] PDE2A is a cyclic nucleotide phosphodiesterase that degrades second messengers cGMP and cAMP.[18] ATG16L2 shares sequence homology with ATG16, a protein required for autophagy in yeast.[19] FCHSD2 is named for its FCH and SH3 protein domains.[20]

Here, we show data supporting rs11603334 as a functional variant regulating *ARAP1* expression. Dense

[1]Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA; [2]National Human Genome Research Institute, National Institutes of Health, Bethesda, MD 20892, USA; [3]The Jackson Laboratory for Genomic Medicine, Farmington, CT 06030, USA; [4]Department of Biostatistics and Center for Statistical Genetics, School of Public Health, University of Michigan, Ann Arbor, MI 48109, USA; [5]Department of Medicine, University of Eastern Finland and Kuopio University Hospital, 70210 Kuopio, Finland
*Correspondence: mohlke@med.unc.edu
http://dx.doi.org/10.1016/j.ajhg.2013.12.011.

fine-mapping data and conditional analyses support a single association signal. We demonstrate that the T2D-risk allele (C) of rs11603334 is associated with increased *ARAP1* transcript levels in primary human pancreatic islets, disrupts binding of a protein complex containing transcriptional regulators, and increases transcriptional activity at the *ARAP1* P1 promoter. These data suggest that increased *ARAP1* P1 promoter activity and *ARAP1* expression might be molecular consequences of the T2D-associated variants in this region.

Material and Methods

Study Population and Phenotypes

In the proinsulin association analyses, we included 8,635 Finnish men without diabetes and not taking diabetes medication (mean age = 57.2 years [range = 45–74 years]; mean body mass index [BMI] = 26.8 kg/m^2 [range = 16.2–51.6 kg/m^2]; fasting plasma glucose levels < 7 mmol/l; and plasma glucose levels after 2 hr oral glucose tolerance test < 11.1 mmol/l) from the population-based Metabolic Syndrome in Men (METSIM) study.[21] Fasting plasma-specific proinsulin (Human Proinsulin RIA Kit, Linco Research; no cross-reaction with insulin or C-peptide) and fasting insulin (ADVIA Centaur Insulin IRI, 02230141, Siemens Medical Solutions Diagnostics; minimal cross-reaction with proinsulin or C-peptide) were measured by immunoassay. For the T2D association analyses, 1,389 T2D cases and 5,748 normoglycemic controls were included. The study was approved by the ethics committee of the University of Kuopio and Kuopio University Hospital, and informed consent was obtained from all study participants.

Genotyping

Samples were genotyped with the Illumina HumanOmniExpress BeadChip. In sum, 681,789 autosomal SNPs passed quality control (Hardy-Weinberg equilibrium [HWE] $p \geq 1 \times 10^{-6}$ in the total sample; call frequency ≥ 0.97). The same samples were previously genotyped with the Illumina HumanExome BeadChip,[11] which focuses on protein-altering variants selected from >12,000 exome and genome sequences.

Imputation and Reference Panel

We used a two-step imputation strategy wherein samples were prephased with ShapeIT version 2[22] before genotypes were imputed with Minimac.[23] To increase imputation quality, we used 5,474 haplotypes from the 2,737 central-northern European samples sequenced within the Genetics of Type 2 Diabetes (GoT2D) study as a reference panel (C.F., J. Flannick, K. Gaulton, H. Kang, and the GoT2D Consortium, unpublished data).

Fine-Mapping and Conditional Analysis

Assuming an additive genetic model and adjusting for age, BMI, and log-transformed fasting plasma insulin, we tested SNP associations with log-transformed fasting plasma proinsulin levels by using a linear mixed model with an empirical kinship matrix to account for relatedness as implemented in EMMAX.[24] We analyzed both raw residuals and rank-based inverse-normal-transformed residuals to assess robustness of rare-variant associations with outliers. Genotyped variants with minor allele count ≥ 5 (minor allele frequency [MAF] ~ 0.03%) and HWE $p \geq 1 \times 10^{-6}$ and imputed variants with imputation quality score $R^2 > 0.3$ and MAF ≥ 0.03% were included in the analyses. To identify additional independent signals in the region, we performed conditional analyses on previously reported lead SNP rs11603334 or on our fine-mapped lead SNP rs7109575 by using allele count (genotyped variants) or allelic dosage (imputed variants) as a covariate in the model. SNP associations with T2D were tested similarly but with adjustment for age only. Kang et al.[24] describe and motivate the use of a linear mixed model for analysis of a binary outcome.

Allelic Expression Imbalance

Human islets from nondiabetic organ donors were obtained from the National Disease Research Interchange and the Islet Cell Resource Center Integrated Islet Distribution Program. DNA and RNA were obtained from 87 primary human pancreatic islet samples with the use of the PureGene (DNA) or RNAeasy (RNA) kits (QIAGEN). Reverse transcription of the RNA was performed with the Superscript III First-Strand Synthesis System (Life Technologies). Common, transcribed SNPs (MAF ≥ 0.1) with the highest LD values with previously reported lead SNPs rs11603334 and rs1552224 were selected for testing allelic expression of *ARAP1*, *STARD10*, *PDE2A*, and *FCHSD2*. LD plots including the transcribed SNPs were created with the use of the HapMap Genome Browser (Figure S1, available online).[25] High-quality genomic DNA (gDNA) and mRNA were available for five islet samples heterozygous for rs11603334 and rs1552224; one additional high-quality heterozygous sample each was available for mRNA only or gDNA only. For each transcribed SNP, the relative proportions of each allele comprising the gDNA and cDNA were quantified with Sequenom iPLEX MALDI-TOF mass-spectrometry-based genotyping (Sequenom). Primers were designed with MassARRAY (Sequenom) (Table S1). We designed primers within a single exon to avoid amplicon size differences between gDNA and cDNA, except for rs2291289, which is located near an exon-intron boundary. To control for assay variation, we analyzed the proportions of each SNP allele measured in the cDNA relative to measurements taken in the gDNA. Data are reported as the percentage of total gDNA or cDNA containing a given transcribed SNP allele. The statistical significance of the differences in allelic representation was determined on the basis of LD scenarios as previously described.[26] In brief, two-sided t tests were used when the transcribed SNP and lead SNPs were in perfect LD ($D' = 1.0$, $r^2 = 1.0$) or when the transcribed SNP and lead SNPs were in intermediate LD with low pairwise correlation ($D' \approx 1.0$, $r^2 < 0.2$). For all t tests, F tests were first used for determining equal or unequal variance. Nonparametric Wilcoxon pairwise tests were used instead of t tests when gDNA or cDNA measurement data were not normally distributed. One-sided F tests were used when the transcribed SNP and lead SNPs were in low LD ($D' < 0.2$, $r^2 < 0.2$).

Regulatory Variant Selection

All known variants in high LD ($r^2 \geq 0.8$, EUR) with rs11603334 and rs1552224 were identified with the use of Phase 1 data from the 1000 Genomes Project.[27] Variants were prioritized on the basis of location relative to regions of potential regulatory function, as indicated by the following: open chromatin in primary human pancreatic islets detected by FAIRE[28] and DNase hypersensitivity,[29] accessible chromatin in primary human pancreatic islets detected by chromatin immunoprecipitation sequencing (ChIP-seq) of active histone H3 lysine methylation or acetylation modifications (H3K4me3, H3K4me1, and H3K9ac)

as reported in the Human Epigenome Atlas,[30] and transcription factor binding in ENCODE tissues or cell lines detected by ChIP-seq.[31]

Cell Culture

The MIN6 mouse insulinoma beta cell line[32] was grown in Dulbecco's modified Eagle's medium (Sigma) supplemented with 10% fetal bovine serum (FBS), 1 mM sodium pyruvate, and 100 μM 2-mercaptoethanol. The INS-1-derived 832/13 rat insulinoma beta cell line[33] (a gift from C.B. Newgard, Duke University) was grown in RPMI 1640 culture medium (Corning Cellgro) supplemented with 10% FBS, 2 mM L-glutamine, 1 mM sodium pyruvate, 10 mM HEPES, and 0.05 mM 2-mercaptoethanol. Both lines were maintained at 37°C and 5% CO_2. One day prior to transfection, MIN6 cells were seeded at a density of 200,000 cells per well on 24-well plates. 832/13 cells were seeded at a density of 300,000 cells per well on 24-well plates or 150,000 cells per well on 48-well plates.

Dual Luciferase Transcriptional Reporter Assay

Genomic regions including variant positions were PCR amplified from human gDNA with 5 PRIME Mastermix (5 PRIME) or the Expand High Fidelity PCR System (Roche) with the primers listed in Table S2. Amplified DNA was subcloned with restriction enzymes XhoI, KpnI, and/or NheI and T4 DNA Ligase (New England Biolabs) into the multiple cloning site upstream of the firefly luciferase gene in the pGL4.10 promoterless vector (Promega). Site-directed mutagenesis was performed with the QuikChange II XL Site-Directed Mutagenesis Kit (Agilent) and the primers listed in Table S2. For each construct, two to ten independent clones were selected. The fidelity and genotype of each gDNA construct were verified by sequencing. Equal amounts, between 600 and 800 ng, of each gDNA construct or empty pGL4.10 promoterless vector were cotransfected with 80 ng of *Renilla* luciferase vector into duplicate wells of MIN6 or 832/13 cells with Lipofectamine 2000 (Life Technologies). Transfected cells were incubated at 37°C and 5% CO_2 overnight, and the transfection media were replaced with fresh culture media after 24 hr. Forty-eight hours posttransfection, cell lysates were collected and assayed for luciferase activity with the Dual-Luciferase Reporter Assay System (Promega). We normalized firefly luciferase activity to that of *Renilla* luciferase to control for differences in transfection efficiency. Normalized data are reported as the fold change (±SD) in activity relative to that of the empty pGL4.10 promoterless vector control. Two-sided Student's t tests were used for comparing luciferase activity between genotypes or haplotypes. F tests were used for determining equal or unequal variance. Two-way ANOVA was used for simultaneous comparisons of haplotypes resulting from two independently tested SNPs.

Electrophoretic Mobility Shift Assay and Transcription Factor Prediction

Four sets of complementary 21-mer oligonucleotides centered on rs11603334 (C/T) or rs1552224 (T/G) were generated; each contained either of the two alternate alleles, both with and without biotin end labeling (Integrated DNA Technologies). We annealed each set to create double-stranded oligonucleotides by incubating 50 pmol of each single-stranded oligonucleotide in buffer containing 10 mM Tris, 1 mM EDTA, and 50 mM NaCl at 95°C for 5 min and then gradually cooled them to 4°C. Double-stranded oligonucleotide sequences are specified in Table S3. Nuclear protein lysates were extracted from MIN6 and 832/13 cell pellets with NE-PER Nuclear and Cytoplasmic Extraction Reagents (Thermo Scientific), and protein concentrations were determined with the Pierce BCA Protein Assay (Thermo Scientific). Electrophoretic mobility shift assays (EMSAs) were carried out with the LightShift Chemiluminescent EMSA Kit (Thermo Scientific) according to the manufacturer's instructions. Each protein-DNA binding reaction contained 1× binding buffer, 1 μg poly(dIdC), 4 μg nuclear protein lysates, and 100 fmol biotin-labeled double-stranded oligonucleotide in 20 μl total reaction volume. For reactions demonstrating DNA competition, 60-fold excess unlabeled double-stranded oligonucleotide was preincubated with nuclear protein lysates in the reaction mixture for 20 min before the addition of biotin-labeled oligonucleotide. For EMSA reactions with supershift, 4 μg antibody was added to the final reaction mixture and the samples were incubated for an additional 30 min. We used JASPAR,[34] TRANSFAC,[35] CONSITE,[36] PWMSCAN,[37] and Tfsitescan[38] to computationally predict transcription factors that might differentially bind at rs11603334 by contrasting the predictions and/or scores generated for each SNP allele. Transcription factor ChIP-seq data for the region overlapping rs11603334 were obtained from the ENCODE project.[31] Antibody to PAX6 (PRB-278P) was purchased from Covance. Antibodies to PAX4 (M-45X, sc-98942X) and AP-2β (H-87X, sc-8976X) were purchased from Santa Cruz Biotechnology. Antibodies to HIF1α (MA1-516) and HIF1β (MA1-515) were purchased from Thermo Scientific. Additional antibodies to the following factors were also purchased and tested for supershift: YY1 (39071, 39345), Active Motif; NFIC (ab86570), Abcam; and AP-2α (C-18, sc-184X), YY1 (C-20, sc-281X; H-414, sc-1703X), MAX (C-124, sc-765X), p53 (DO-1, sc-126X), OCT3/4 (C-10, sc-5279X), SP1 (E-3, sc-17824X), USF-1 (C-20, sc-229X), EGR-1 (C-19, sc-189X), RXRα (D-20, sc-553X), E4BP4 (V-19, sc-9549X), PDX1 (N-18, sc-14662X), CEBPβ (C-19, sc-150X), CUX1 (C-20, sc-6327X), and p300 (C-20, sc-585X), Santa Cruz Biotechnology. Protein-DNA complexes were separated by electrophoresis on 6% DNA retardation gels (Life Technologies) in 0.5 X TBE (Lonza), transferred to nylon membranes (Thermo Scientific), and UV crosslinked before chemiluminescent detection.

Results

Fine Mapping of the 11q13.4 Locus, Associated with Fasting Proinsulin and T2D

To fine map the association signal for fasting proinsulin, we performed association analyses in 8,635 Finnish subjects from the METSIM study with SNPs imputed from a reference panel of 2,737 sequenced central-northern European individuals and SNPs directly genotyped on an exome array.[11] The lead SNP was rs7109575 ($p = 7.4 \times 10^{-50}$; n = 8,635), a noncoding variant located at the *ARAP1* P2 TSS (Figure 1 and Table S4) and in high LD ($r^2 = 0.94$) with previously reported lead proinsulin-associated SNP rs11603334[10] ($p = 2.2 \times 10^{-48}$; n = 8,630). Conditional analyses using rs11603334 or rs7109575 as covariates largely attenuated the proinsulin association signal (all SNPs, $p > 1.0 \times 10^{-5}$; Figure 1, Figure S2, Table S4, and data not shown). One rare missense variant genotyped in a 1 Mb region on the exome array was associated ($p < 0.05$) with proinsulin after conditional analysis (rs202137453,

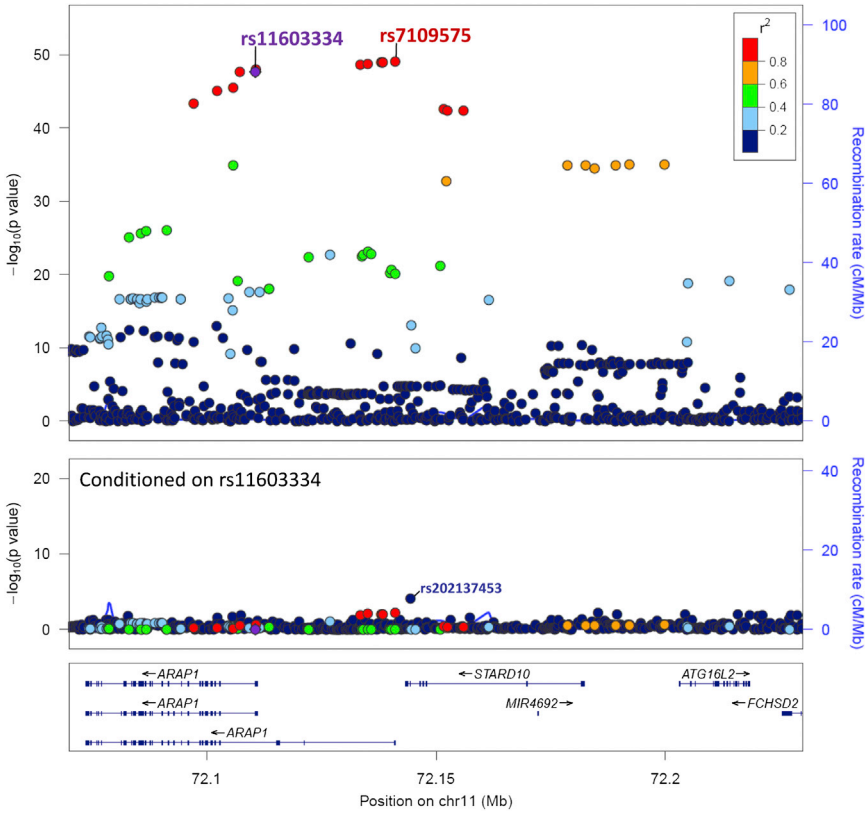

Figure 1. Variants Associated with Proinsulin in the METSIM Study

LocusZoom[39] plots of fasting proinsulin association adjusted for age, BMI, and fasting insulin. The robust association with proinsulin levels (top panel: rs7109575, $p = 7.4 \times 10^{-50}$) is attenuated upon inclusion of rs11603334 as a covariate (bottom panel, only showing y axis to 10^{-20}). Each SNP is colored according to its LD (r^2) with rs11603334. Genomic coordinates refer to hg18 (UCSC Genome Browser).

MAF = 0.003, p = 0.015, $p_{\text{conditional rs11603334}} = 8.5 \times 10^{-5}$, Table S4). T2D association using 1,389 case and 5,748 control subjects (rs7109575, $p = 2.2 \times 10^{-3}$; rs11603334, $p = 3.6 \times 10^{-3}$) was also attenuated after conditional analysis ($p > 0.05$; Table S4, Figure S3, and data not shown). Taken together, fine-mapping and conditional analyses provide evidence supporting a functional role for common variant(s) in high LD with rs7109575 and rs11603334.

The Proinsulin-Decreasing and T2D-Risk Alleles Are Associated with Increased *ARAP1* mRNA Levels in Primary Human Pancreatic Islets

To determine whether expression of *ARAP1*, *STARD10*, *PDE2A*, or *FCHSD2* is associated with the T2D- and fasting-proinsulin-associated lead SNPs, we evaluated allelic expression imbalance (AEI) in primary human pancreatic islets. AEI maximizes sensitivity and power to detect modest *cis*-acting effects on mRNA expression through quantification of allele-specific transcript levels within heterozygous samples, thereby controlling for differential *trans*-acting and environmental exposures between individuals,[40] as well as differential sample purity and viability between islet preparations.[41,42] The allelic composition of the total transcript level for each gene was quantitatively determined with Sequenom MALDI-TOF mass spectrometry, a sensitive approach that conserves limited biological sample, and statistically analyzed relative to the allelic composition of the gDNA (Material and Methods). Located in the first exon and 5′ UTR of *ARAP1* isoforms transcribed from P1, previously reported lead SNPs rs11603334 and rs1552224 demonstrated AEI (Figure 2). In six samples heterozygous for rs11603334 and rs1552224, we observed *ARAP1* expression differences of 7%–12% between alleles. For both SNPs, the T2D-risk alleles were associated with higher *ARAP1* transcript levels (rs11603334, p = 0.0002; rs1552224, p = 0.015). Because of lower LD between the lead SNPs and a third *ARAP1* transcribed SNP located in a downstream exon, rs2291289 ($r^2 = 0.12$, $D' = 1.0$, rs11603334, EUR), only three samples were heterozygous for both rs2291289 and rs11603334 (Table 1). The gDNA-to-cDNA comparison showed no AEI, whereas a comparison among cDNA samples showed the nonrisk allele associated with modestly increased transcript levels (Figure S4A). We also attempted to evaluate AEI for nearby genes containing transcribed SNPs in lower LD with rs11603334 by applying previously described statistical methods.[26] We did not observe significant evidence of AEI for *STARD10*, *PDE2A*, or *FCHSD2* (Figures S4B–S4D and data not shown); zero to four samples were heterozygous for each transcribed SNP and rs11603334 (Table 1). Overall, these results suggest that one or more functional variants in high LD with rs11603334 and rs1552224 could increase the transcriptional activity or message stability of *ARAP1* isoforms containing these SNPs.

Candidate Regulatory Variants at the *ARAP1* Promoters

The originally identified lead SNPs rs11603334 and rs1552224 at the P1 promoter overlap ENCODE transcription factor binding sites within a region of open and accessible chromatin that is marked by DNase hypersensitivity and active histone H3 lysine modifications H3K4me3 and H3K9ac in human pancreatic islets (Figure 3 and Table S5). The lead fine-mapped SNP in this study, rs7109575 ($r^2 = 0.94$ with rs11603334, EUR), is located within a region of islet DNase hypersensitivity

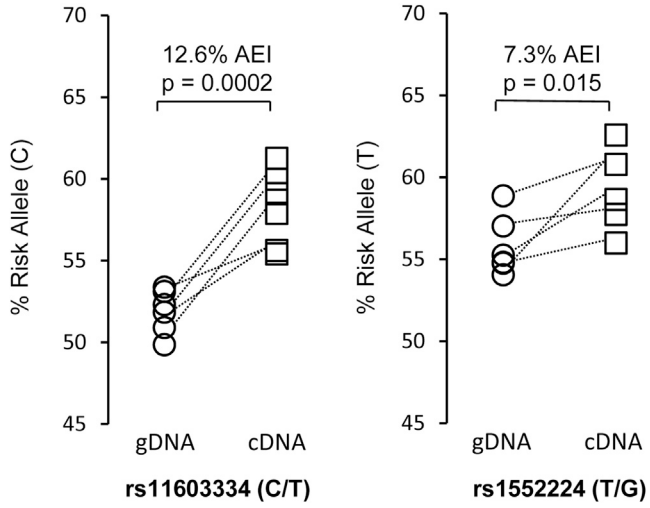

Figure 2. The T2D-Risk Alleles of rs11603334 and rs1552224 Are Associated with Increased *ARAP1* Expression
Allelic expression analysis in human pancreatic islet gDNA and cDNA from six donors each, including five matched pairs (matched pairs only: rs11603334, $p = 0.0008$; rs1552224, $p = 0.06$). The proportion of total gDNA and cDNA containing the T2D-risk alleles of rs11603334 (left) and rs1552224 (right) from heterozygous samples are quantified on the y axis. The cDNA containing the T2D-risk allele was expressed 12.6% and 7.3% more than the cDNA containing the nonrisk allele in the rs11603334 and rs1552224 assays, respectively. The p values were calculated with a two-sided t test.

and clustered transcription factor binding at the P2 promoter (Figure S5 and Table S5). Among the 17 additional variants in high LD with rs11603334 ($r^2 \geq 0.8$, EUR; Table S5), rs77464186 ($r^2 = 0.95$ with rs11603334, EUR) is the next most promising functional candidate because of its location within a region of islet DNase hypersensitivity and clustered transcription factor binding in the first intron of *ARAP1* isoforms transcribed from the P2 promoter (Figure S5 and Table S5).

The rs11603334 T2D-Risk and Proinsulin-Decreasing Allele Increases Transcriptional Reporter Activity in Rodent Pancreatic Beta Cell Lines

To interrogate allelic differences in transcriptional activity of rs11603334, rs1552224, rs7109575, and rs77464186, we cloned DNA sequences containing the T2D-risk (proinsulin-decreasing) or nonrisk (proinsulin-increasing) alleles into a promoterless luciferase vector and measured luciferase activity in the MIN6 mouse and 832/13 rat beta cell lines. Because of the perfect LD and short distance between rs1552224 and rs11603334, we first tested them together as a haplotype in the context of the *ARAP1* P1 promoter. A 1,357 bp region of DNA, defined by DNase hypersensitivity in human pancreatic islets and Pol2 binding in several other cell types (Figure 3), exhibited strong transcriptional activity in comparison to an empty vector control in both MIN6 (Figure 4A) and 832/13 (Figure 4B). The rs1552224 and rs11603334 T2D-risk haplotype (TC) demonstrated a 2-fold increase in transcriptional activity in comparison to the nonrisk haplotype (GT) in both MIN6 ($p < 0.0001$) and 832/13 ($p = 0.015$), consistent with our finding that the T2D-risk alleles are associated with increased *ARAP1* expression. The 1,357 bp region also contained two short sequences of variable CT repeat length ($CT_{6–10}$); however, the numbers of repeats at the two sites did not influence transcriptional activity (Figures S6A and S6B). A narrower, 245 bp DNA region (Figure 3) containing only variants rs1552224 and rs11603334 also exhibited strong transcriptional activity in comparison to an empty vector control (Figure 4C). As observed with the 1,357 bp region, the two-SNP T2D-risk haplotype (TC) demonstrated a 2-fold increase in transcriptional activity in comparison to the nonrisk haplotype (GT) ($p = 0.012$). To determine which of the SNPs was responsible for the haplotype effects on transcriptional activity at the P1 promoter, we used site-directed mutagenesis to synthetically create the alternate "missing" haplotypes. Compared to the nonrisk T allele, the T2D-risk allele (C)

Table 1. Transcribed SNPs Used for Evaluating Allelic Expression Imbalance for *ARAP1*, *STARD10*, *PDE2A*, and *FCHSD2*

Gene	SNP	Annotation	Major/Minor Allele	MAF	LD with rs11603334		No. of Informative Heterozygotes[a]
					D′	r^2	
ARAP1	rs11603334	5′ UTR	C/T	0.16	1	1	6
ARAP1	rs1552224	5′ UTR	T/G	0.16	1	1	6
ARAP1	rs2291289	synonymous	A/G	0.41	0.94	0.12	3
STARD10	rs2291290	3′ UTR	A/G	0.11	1	0.025	2
STARD10	rs519790	5′ UTR	C/G	0.28	0.97	0.071	2
PDE2A	rs392565	synonymous	C/T	0.30	0.011	0.0053	4
FCHSD2	rs76762469	3′ UTR	G/T	0.10	0.18[b]	0.001[b]	0

MAF, D′, and r^2 data were estimated from 1000 Genomes Phase 1 populations of European ancestry (EUR). The following abbreviations are used: LD, linkage disequilibrium; and MAF, minor allele frequency.
[a]Out of 87 total human pancreatic islet samples.
[b]These D′ and r^2 data were estimated from the 1000 Genomes Pilot CEU (Utah residents with ancestry from northern and western Europe from the CEPH collection) population.

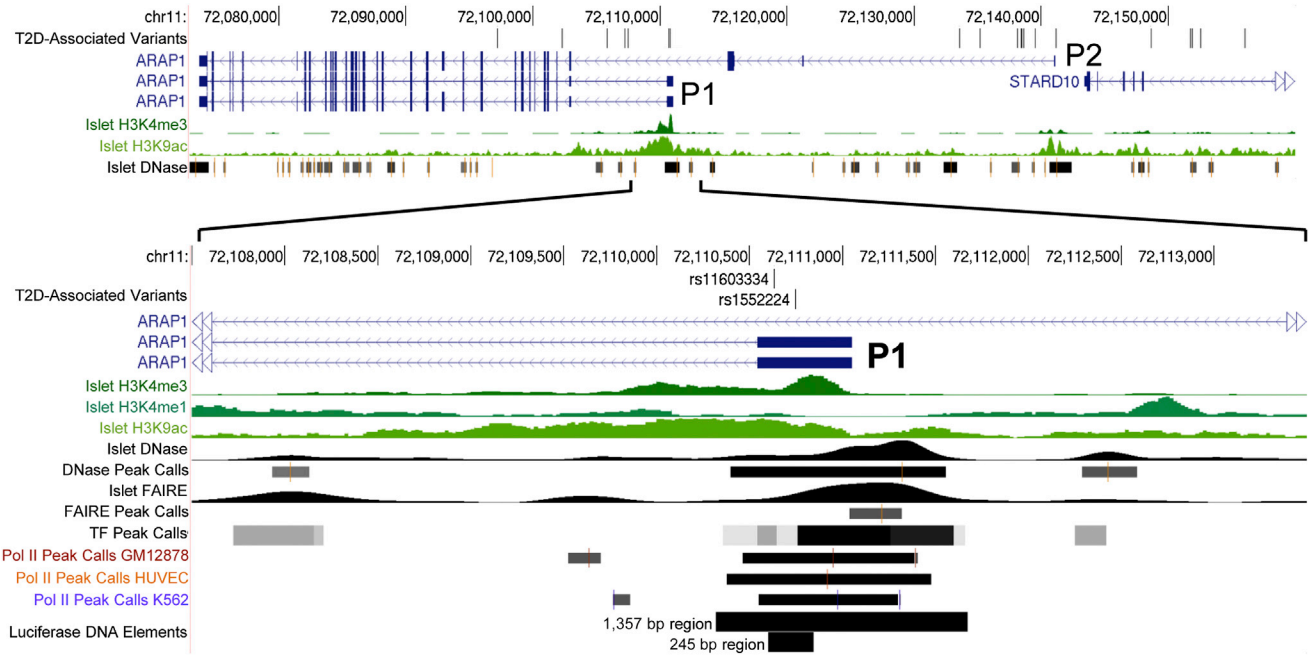

Figure 3. SNPs rs11603334 and rs1552224 Are Located in a Region with Evidence of Regulatory Potential at the *ARAP1* P1 Promoter at 72.11 Mb

UCSC Genome Browser (hg18) diagram showing rs11603334 and rs1552224 situated at the *ARAP1* P1 promoter and 18 additional variants in high LD ($r^2 \geq 0.8$). *ARAP1* is transcribed from right to left. SNPs rs11603334 and rs1552224 overlap regions of accessible chromatin in human pancreatic islets detected by H3K4me3 and H3K9ac ChIP-seq, are situated within or proximal to regions of open chromatin in human pancreatic islets detected by DNase hypersensitivity and FAIRE, and overlap binding sites for Pol2 and other transcription factors detected by ChIP-seq in several ENCODE cell lines (see also Table S5). The DNA sequences cloned for evaluating differential allelic transcriptional activity in the dual-luciferase reporter assays are indicated.

of rs11603334 replicated the previously observed 2-fold increase in transcriptional activity ($p < 0.0001$), whereas the T2D-risk and nonrisk alleles of rs1552224 (T/G) showed no significant difference (Figure 4D). We did not observe evidence of allelic differences in transcriptional activity for either rs7109575 (Figure 4E) or rs77464186 (Figure S6C) near the P2 promoter. However, compared to the empty control vector, the 287 bp region of DNA surrounding rs7109575 increased transcriptional activity 9-fold ($p < 0.0001$), consistent with promoter activity.

Decreased In Vitro Binding of PAX6 and PAX4 to the rs11603334 T2D-Risk and Proinsulin-Decreasing Allele

We performed EMSAs by using nuclear protein lysates extracted from MIN6 and 832/13 cell lines to evaluate differences in transcription factor binding to alleles of rs11603334 and rs1552224. Multiple protein-DNA complexes were observed in the EMSAs, consistent with expectations for DNA within a promoter region (Figure 5 and Figure S7). At least one protein complex was observed to bind specifically to the nonrisk (T) allele of rs11603334 in MIN6 (Figure 5A, arrow). DNA competition assays using excess unlabeled oligonucleotide containing the T allele reproducibly disrupted the protein-DNA complex more efficiently than did excess unlabeled oligonucleotide containing the C allele, further supporting allelic differences in protein binding (Figure 5A, lane 8 versus lane 9; Figure S7A, lane 7 versus lane 8). Preferential protein binding to the T allele was also detected in 832/13 for a complex with similar mobility (Figure S7A, lane 10, arrow). To identify the differentially bound protein(s), we performed EMSAs with supershift by using antibodies targeting transcription factors selected from those with computationally predicted or empirically determined (ChIP-seq) binding at rs11603334 (Material and Methods). Incubation with antibody targeting PAX6 partially disrupted the T-allele-specific protein-DNA complex (Figure 5A, lane 10; Figure 5B, lane 8), and incubation with antibodies targeting both PAX6 and PAX4 fully disrupted the complex (Figure 5B, lane 10). Among the other transcription factors tested with supershift assay, only AP-2β, HIF1α, and HIF1β were observed to bind to the 21 bp DNA sequence containing rs11603334; these proteins bound to both the C and T alleles with approximately equal affinity (Figures S7B and S7C, arrows). We did not observe clear evidence of allele-specific protein binding to rs1552224 (Figure S7D).

Discussion

For most T2D-associated loci identified through GWASs, the functional variants and affected genes are not known. In this study, we have identified a plausible functional regulatory variant, rs11603334, and demonstrated its allelic effects on expression of *ARAP1*, a gene at the 11q13.4

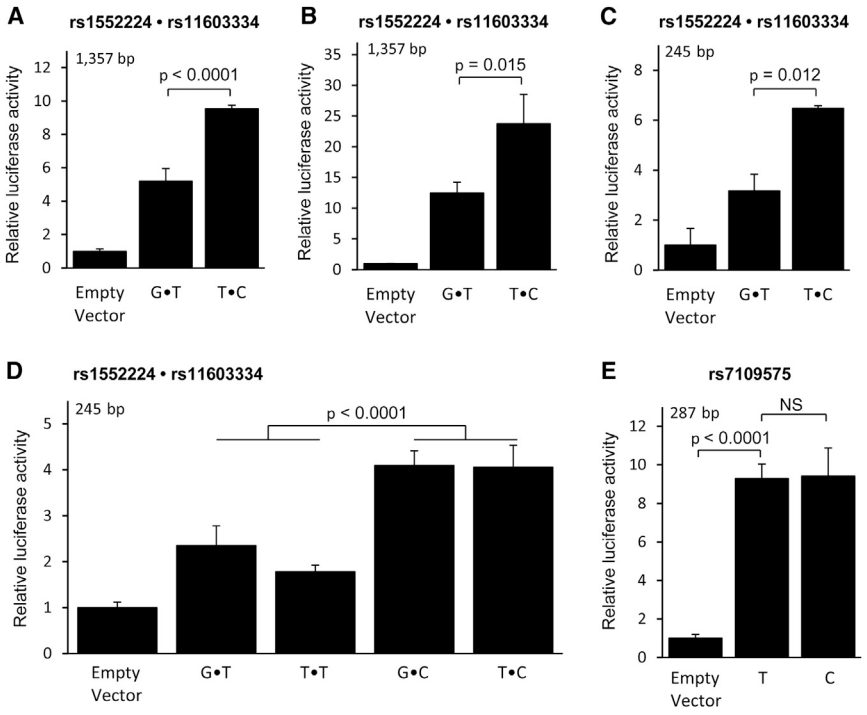

Figure 4. The T2D-Risk Allele, C, of rs11603334 Increases Transcriptional Activity of the *ARAP1* P1 Promoter

Transcriptional activity was evaluated with dual-luciferase reporter assays in the MIN6 mouse and 832/13 rat insulinoma β-cell lines 48 hr after transfection with recombinant vector containing selected regions of the human *ARAP1* promoters cloned upstream of the luciferase gene.

(A) rs1552224 and rs11603334 evaluated as a haplotype in a 1,357 bp region at the *ARAP1* P1 promoter in MIN6.

(B) rs1552224 and rs11603334 evaluated as a haplotype in the 1,357 bp region in 832/13.

(C) rs1552224 and rs11603334 evaluated as a haplotype in a narrower 245 bp region at the *ARAP1* P1 promoter in 832/13.

(D) Site-directed mutagenesis separated the effects of rs1552224 and rs11603334 in the 245 bp region in 832/13.

(E) rs7109575 evaluated in a 287 bp region at the *ARAP1* P2 promoter in 832/13.

The p values were calculated with a two-sided t test (A–C and E) or two-way ANOVA (D). Data represent the mean of two to ten clones per allele ± SD. A significance threshold of $p < 0.05$ was used. NS, not significant.

locus. We provide evidence suggesting that the C allele of rs11603334 disrupts binding of transcriptional regulators and increases transcriptional activity at the *ARAP1* P1 promoter and thus leads to higher *ARAP1* expression in pancreatic islets. Increased *ARAP1* expression might be one molecular genetic contributor to T2D susceptibility.

We conducted association analyses by using the densest set of variants to date and report rs7109575 as the lead fine-mapped SNP, confirming the primary robust association between fasting proinsulin and previously reported lead SNPs rs11603334 and rs1552224. Although not the most strongly associated SNP in this study, our proposed functional SNP (rs11603334) had an association p value ($p = 2.2 \times 10^{-48}$) similar to that of rs7109575 ($p = 7.4 \times 10^{-50}$); sampling variability, missing genotypes, or imputation uncertainty might influence the relative p values. Conditioning on either rs11603334 or rs7109575 attenuated the proinsulin association, consistent with a single signal at this locus and supporting the use of the r^2 measure of LD for identifying candidate functional SNPs. Rare variants might also contribute to variation in proinsulin levels among carriers. The variant that became more significant after adjustment for rs11603334 (rs202137453, *STARD10* c.586C>T [p.Pro196Ser]) might be a true signal or a false positive among 3,818 variants tested in this region. Further fine-mapping and conditional analyses using a larger T2D case-control population will be needed for determining whether other variants in the region also contribute to T2D risk.

We have demonstrated that the T2D-risk alleles of rs11603334 and rs1552224 are associated with increased *ARAP1* mRNA expression in primary human pancreatic islets. This finding was reproducibly evident with a small number of samples because AEI compares quantification of allele-specific mRNA levels within heterozygous individuals and thereby minimizes confounding effects, such as interindividual environmental exposures, that can mask genotypic effects in studies of expression quantitative trait loci (eQTLs). We had the best power to detect association between *ARAP1* mRNA expression and rs11603334 and rs1552224 because the SNPs are located within an exon; this exon is specific to *ARAP1* isoforms transcribed from the P1 promoter. Previously, rs519790 was reported to be a strong eQTL for *STARD10* in blood ($p = 2.7 \times 10^{-24}$);[1] however, this SNP is in low LD with rs1552224 ($r^2 = 0.071$, EUR), which shows only a modest, residual eQTL association with *STARD10* (blood, $p = 8.6 \times 10^{-7}$; liver, $p = 1.79 \times 10^{-5}$; omental adipose tissue, $p = 1.21 \times 10^{-6}$).[43] A nominal association between rs1552224 and *ARAP1* (*CENTD2*) expression has been previously reported in T cells ($p = 0.019$) in a direction consistent with our results.[44,45]

In agreement with our finding that the T2D-risk alleles of rs11603334 and rs1552224 are associated with increased *ARAP1* expression, our experiments using the transcriptional reporter assay showed that the T2D-risk allele of rs11603334 upregulates transcriptional activity at the *ARAP1* P1 promoter. Although transcriptional activity at a gene promoter might be expected, allelic differences in promoter activity cannot be predicted easily; functional assays are necessary for demonstrating allelic effects. Among three tested variants situated at the *ARAP1* P1 and P2 promoters in this study, only rs11603334 demonstrated allelic effects on promoter activity. Taken together, the results from the AEI and transcriptional reporter assays implicate increased expression of *ARAP1* isoforms containing

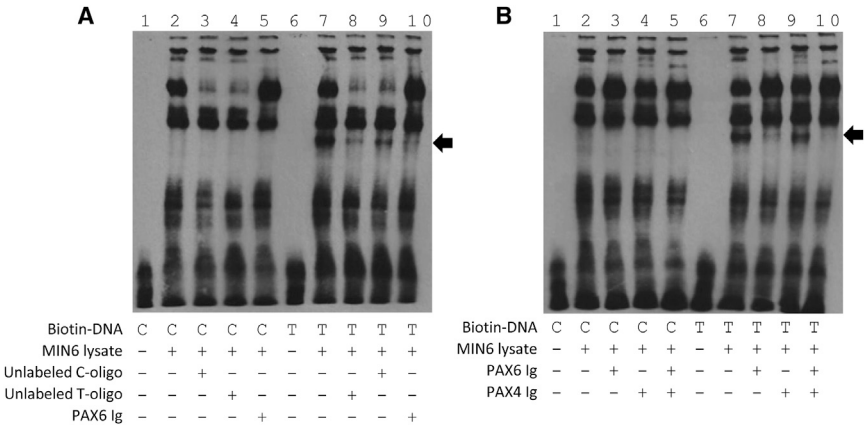

Figure 5. The T2D-Risk Allele, C, of rs11603334 Disrupts Binding of Transcriptional Regulators PAX6 and PAX4
Differential allelic protein-DNA binding was evaluated in vitro with EMSAs. Biotin-labeled 21 bp double-stranded DNA oligonucleotide centered on the reference (C) and alternate (T) alleles of rs11603334 was incubated with MIN6 nuclear lysate, and resulting protein-DNA complexes were separated by electrophoresis (A and B: lanes 2 and 7, respectively). For the competition assays, EMSA reactions were incubated with 60-fold excess unlabeled DNA oligonucleotide containing either the T or C allele (A: lanes 8 and 9, respectively). For the supershift assays, EMSA reactions were incubated with antibody targeting PAX6 (A: lane 10; B: lane 8), PAX4 (B: lane 9), or both PAX6 and PAX4 (B: lane 10).

rs11603334 and transcribed from the P1 promoter (RefSeq accession numbers NM_015242.4 and NM_001135190.1) in T2D susceptibility.

We have shown that the beta cell transcriptional regulator PAX6 preferentially binds to the mRNA-lowering allele (T) of rs11603334 in a protein-DNA complex that could also contain PAX4, suggesting a repressive role for one or both factors in transcription at the *ARAP1* P1 promoter (Figure 6). PAX6 and PAX4 have similar binding sequences, although they bind DNA with different relative affinities.[46–48] In the endocrine pancreas, PAX6 is often reported to function as a transcriptional activator, whereas PAX4 often functions as a repressor. PAX4 has been shown to inhibit the transcriptional activity of coexpressed PAX6 in the HIT-T15 hamster beta cell line and in 293T cells, possibly through competition for DNA binding sites.[47] PAX4 also represses basal transcriptional activity of the glucagon promoter in αTC1.9 and InR1G9 cells and strongly represses PAX6-activated transcription of both the glucagon[48] and insulin promoters.[49] However, repressive activity for human and mouse PAX6 has also been demonstrated for specific regulatory elements in other cell types.[50–52] The observed inhibition of transcriptional activity for *ARAP1* transcripts containing the rs11603334 T allele might be mediated by PAX4, PAX6, or both. Human coding variants in PAX4 or PAX6 have been previously implicated in impaired glucose tolerance,[53,54] early-onset diabetes,[53,55] ketosis-prone diabetes,[56] and proinsulin processing.[57] Recently, genome-wide association signals for T2D have also been identified near *PAX4* in East Asian[2] and Chinese populations.[58]

Our results do not preclude the existence of additional functional genes, variants, and/or mechanisms at this locus. Power to detect association between the lead SNPs and expression levels of *STARD10*, *PDE2A*, and *FCHSD2* was limited as a result of small numbers of SNPs and samples informative for AEI. Although 87 total islet samples were available, only six were heterozygous for rs11603334 and zero to four samples were informative for the *STARD10*, *PDE2A*, and *FCHSD2* assays. We cannot exclude possible associations between rs11603334 and expression levels of these genes. An assay to test AEI for all isoforms of *ARAP1* (with the use of rs2291289) similarly was of limited utility because of the small number of heterozygous samples (n = 3) for this SNP and wide assay variance. Moreover, *ATG16L2* could not be tested with this approach because it does not contain any common transcribed SNPs. Future analysis of a larger pancreatic islet expression data set will be required for validating our findings for *ARAP1* and assessing allelic differences in gene expression for all other genes.

We chose to investigate the mechanisms underlying the set of variants most strongly associated with proinsulin and T2D after dense fine-mapping analyses. We prioritized among highly significant noncoding variants on the basis of their locations within regions of open chromatin in pancreatic islets and successfully identified a variant with allelic differences in transcriptional activity. Variants located outside of these open chromatin regions might also be found to have regulatory function. In addition, functional noncoding variants at this locus might regulate gene expression through allelic effects on splicing, message stability, or translational control. On the basis of sequence, rs1552224 creates a possible splice site, and the rs11603334 T allele creates a methionine codon upstream of the RefSeq start codon. If used, this codon could potentially lead to attempted translation of a 90 nt upstream open reading frame, which could interfere with efficient translation of *ARAP1*.[59]

Specific roles for ARAP1 in the process of proinsulin conversion and/or insulin secretion in the beta cell have not yet been reported. ARAP1 is located near both the Golgi and the plasma membrane in other cell types[16,60–62] and has been demonstrated to regulate several members of the Arf and Rho small GTPase protein families in vitro and in cell culture.[16,63] Of the reported substrates, both ARF6 and CDC42 have been shown to play a role in glucose-stimulated insulin secretion.[64–66] Future work

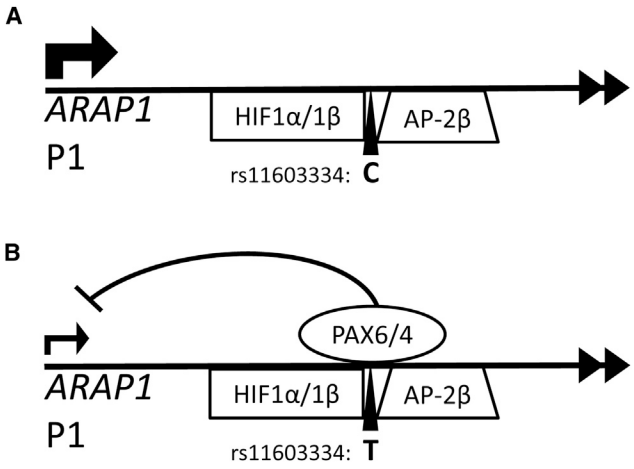

Figure 6. Proposed Model for PAX6 and PAX4 Repression of Transcriptional Activity at *ARAP1*
PAX6 and PAX4 at rs11603334 in a complex with stronger binding to the *ARAP1* mRNA-decreasing T allele (B, T2D nonrisk) than to the mRNA-increasing C allele (A, T2D risk) suggests a role for one or both of these factors in repressing transcription at the *ARAP1* P1 promoter. Transcription factors HIF1α, HIF1β, and AP-2β bind both alleles.

will address the effects of upregulated *ARAP1* expression on proinsulin processing and insulin secretion in the pancreatic beta cell. Given that our data suggest that increased *ARAP1* expression is associated with increased risk of T2D, ARAP1 antagonists might hold important therapeutic potential.

Identifying functional variants at T2D GWAS loci and elucidating mechanisms to explain how variants influence expression or function of target genes will help facilitate the discovery of biological gene candidates and novel pathways involved in disease pathology. This work contributes to a growing understanding of the molecular genetic architecture of T2D, provides a model of fine-mapping and experimental analyses that can be used for uncovering functional variants contributing to associations at GWAS loci, and paves the way for future research on the role of *ARAP1* as a diabetes candidate gene in pancreatic islets.

Supplemental Data

Supplemental Data include seven figures and five tables and can be found with this article online at http://www.cell.com/AJHG.

Acknowledgments

We thank first and foremost the human organ donors whose pancreatic islets were used in this study. Islets were obtained from the Islet Cell Resource Center Basic Science Islet Distribution Program and Integrated Islet Distribution Program and the National Disease Research Interchange. This study was supported by National Institutes of Health (NIH) grants R01DK072193 (K.L.M.), R21DA027040 (K.L.M.), R01DK093757 (K.L.M.), K99DK092251 (M.L.S.), and R01DK062370 (M.B.), NIH National Human Genome Research Institute (NHGRI) Division of Intramural Research project number Z01HG000024 (F.S.C.), and Academy of Finland grants 77299 and 124243 (M.L.). The authors gratefully acknowledge the Genetics of Type 2 Diabetes Consortium for access to the reference panel used for imputation, Michael R. Erdos (NHGRI) for assistance coordinating and preparing human islet samples, Yun Li (University of North Carolina at Chapel Hill) for providing linkage-disequilibrium data from the 1000 Genomes Project, Praveen Sethupathy (University of North Carolina at Chapel Hill) for helpful discussion and assistance generating motif prediction scores via PWM-Scan, Leslie Lange and Ethan Lange (University of North Carolina at Chapel Hill) for statistical advice, and Marie Fogarty (University of North Carolina at Chapel Hill) for helpful discussions.

Received: August 2, 2013
Accepted: December 11, 2013
Published: January 16, 2014

Web Resources

The URL for data presented herein is as follows:

Online Mendelian Inheritance in Man (OMIM), http://www.omim.org

References

1. Voight, B.F., Scott, L.J., Steinthorsdottir, V., Morris, A.P., Dina, C., Welch, R.P., Zeggini, E., Huth, C., Aulchenko, Y.S., Thorleifsson, G., et al.; MAGIC investigators; GIANT Consortium (2010). Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. Nat. Genet. *42*, 579–589.
2. Cho, Y.S., Chen, C.-H., Hu, C., Long, J., Ong, R.T.H., Sim, X., Takeuchi, F., Wu, Y., Go, M.J., Yamauchi, T., et al.; DIAGRAM Consortium; MuTHER Consortium (2012). Meta-analysis of genome-wide association studies identifies eight new loci for type 2 diabetes in east Asians. Nat. Genet. *44*, 67–72.
3. Cho, Y.S., Lee, J.-Y., Park, K.S., and Nho, C.W. (2012). Genetics of type 2 diabetes in East Asian populations. Curr. Diab. Rep. *12*, 686–696.
4. Imamura, M., Maeda, S., Yamauchi, T., Hara, K., Yasuda, K., Morizono, T., Takahashi, A., Horikoshi, M., Nakamura, M., Fujita, H., et al.; Diabetes Genetics Replication and Meta-analysis (DIAGRAM) Consortium (2012). A single-nucleotide polymorphism in ANK1 is associated with susceptibility to type 2 diabetes in Japanese populations. Hum. Mol. Genet. *21*, 3042–3049.
5. Perry, J.R.B., Voight, B.F., Yengo, L., Amin, N., Dupuis, J., Ganser, M., Grallert, H., Navarro, P., Li, M., Qi, L., et al.; MAGIC; DIAGRAM Consortium; GIANT Consortium (2012). Stratifying type 2 diabetes cases by BMI identifies genetic risk variants in LAMA1 and enrichment for risk variants in lean compared to obese cases. PLoS Genet. *8*, e1002741.
6. Morris, A.P., Voight, B.F., Teslovich, T.M., Ferreira, T., Segrè, A.V., Steinthorsdottir, V., Strawbridge, R.J., Khan, H., Grallert, H., Mahajan, A., 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 (2012). Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. Nat. Genet. *44*, 981–990.

7. Li, H., Gan, W., Lu, L., Dong, X., Han, X., Hu, C., Yang, Z., Sun, L., Bao, W., Li, P., et al.; DIAGRAM Consortium; AGEN-T2D Consortium (2013). A genome-wide association study identifies GRK5 and RASGRP1 as type 2 diabetes loci in Chinese Hans. Diabetes *62*, 291–298.
8. Tabassum, R., Chauhan, G., Dwivedi, O.P., Mahajan, A., Jaiswal, A., Kaur, I., Bandesh, K., Singh, T., Mathai, B.J., Pandey, Y., et al.; DIAGRAM; INDICO (2013). Genome-wide association study for type 2 diabetes in Indians identifies a new susceptibility locus at 2q21. Diabetes *62*, 977–986.
9. Saxena, R., Saleheen, D., Been, L.F., Garavito, M.L., Braun, T., Bjonnes, A., Young, R., Ho, W.K., Rasheed, A., Frossard, P., et al.; DIAGRAM; MuTHER; AGEN (2013). Genome-wide association study identifies a novel locus contributing to type 2 diabetes susceptibility in Sikhs of Punjabi origin from India. Diabetes *62*, 1746–1755.
10. Strawbridge, R.J., Dupuis, J., Prokopenko, I., Barker, A., Ahlqvist, E., Rybin, D., Petrie, J.R., Travers, M.E., Bouatia-Naji, N., Dimas, A.S., et al.; DIAGRAM Consortium; GIANT Consortium; MuTHER Consortium; CARDIoGRAM Consortium; C4D Consortium (2011). Genome-wide association identifies nine common variants associated with fasting proinsulin levels and provides new insights into the pathophysiology of type 2 diabetes. Diabetes *60*, 2624–2634.
11. Huyghe, J.R., Jackson, A.U., Fogarty, M.P., Buchkovich, M.L., Stančáková, A., Stringham, H.M., Sim, X., Yang, L., Fuchsberger, C., Cederberg, H., et al. (2013). Exome array analysis identifies new loci and low-frequency variants influencing insulin processing and secretion. Nat. Genet. *45*, 197–201.
12. Nielsen, T., Sparsø, T., Grarup, N., Jørgensen, T., Pisinger, C., Witte, D.R., Hansen, T., and Pedersen, O.; Diabetes Genetics Replication and Meta-analysis (DIAGRAM) Consortium (2011). Type 2 diabetes risk allele near CENTD2 is associated with decreased glucose-stimulated insulin release. Diabetologia *54*, 1052–1056.
13. Wu, C., Orozco, C., Boyer, J., Leglise, M., Goodale, J., Batalov, S., Hodge, C.L., Haase, J., Janes, J., Huss, J.W., 3rd, and Su, A.I. (2009). BioGPS: an extensible and customizable portal for querying and organizing gene annotation resources. Genome Biol. *10*, R130.
14. Su, A.I., Wiltshire, T., Batalov, S., Lapp, H., Ching, K.A., Block, D., Zhang, J., Soden, R., Hayakawa, M., Kreiman, G., et al. (2004). A gene atlas of the mouse and human protein-encoding transcriptomes. Proc. Natl. Acad. Sci. USA *101*, 6062–6067.
15. Parker, S.C.J., Stitzel, M.L., Taylor, D.L., Orozco, J.M., Erdos, M.R., Akiyama, J.A., van Bueren, K.L., Chines, P.S., Narisu, N., Black, B.L., et al.; NISC Comparative Sequencing Program; National Institutes of Health Intramural Sequencing Center Comparative Sequencing Program Authors; NISC Comparative Sequencing Program Authors (2013). Chromatin stretch enhancer states drive cell-specific gene regulation and harbor human disease risk variants. Proc. Natl. Acad. Sci. USA *110*, 17921–17926.
16. Miura, K., Jacques, K.M., Stauffer, S., Kubosaki, A., Zhu, K., Hirsch, D.S., Resau, J., Zheng, Y., and Randazzo, P.A. (2002). ARAP1: a point of convergence for Arf and Rho signaling. Mol. Cell *9*, 109–119.
17. Olayioye, M.A., Vehring, S., Müller, P., Herrmann, A., Schiller, J., Thiele, C., Lindeman, G.J., Visvader, J.E., and Pomorski, T. (2005). StarD10, a START domain protein overexpressed in breast cancer, functions as a phospholipid transfer protein. J. Biol. Chem. *280*, 27436–27442.
18. Rosman, G.J., Martins, T.J., Sonnenburg, W.K., Beavo, J.A., Ferguson, K., and Loughney, K. (1997). Isolation and characterization of human cDNAs encoding a cGMP-stimulated 3′,5′-cyclic nucleotide phosphodiesterase. Gene *191*, 89–95.
19. Matsushita, M., Suzuki, N.N., Obara, K., Fujioka, Y., Ohsumi, Y., and Inagaki, F. (2007). Structure of Atg5.Atg16, a complex essential for autophagy. J. Biol. Chem. *282*, 6763–6772.
20. Katoh, M., and Katoh, M. (2004). Identification and characterization of human FCHSD1 and FCHSD2 genes in silico. Int. J. Mol. Med. *13*, 749–754.
21. Stancáková, A., Javorský, M., Kuulasmaa, T., Haffner, S.M., Kuusisto, J., and Laakso, M. (2009). Changes in insulin sensitivity and insulin release in relation to glycemia and glucose tolerance in 6,414 Finnish men. Diabetes *58*, 1212–1221.
22. Delaneau, O., Zagury, J.-F., and Marchini, J. (2013). Improved whole-chromosome phasing for disease and population genetic studies. Nat. Methods *10*, 5–6.
23. Howie, B., Fuchsberger, C., Stephens, M., Marchini, J., and Abecasis, G.R. (2012). Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. Nat. Genet. *44*, 955–959.
24. Kang, H.M., Sul, J.H., Service, S.K., Zaitlen, N.A., Kong, S.-Y., Freimer, N.B., Sabatti, C., and Eskin, E. (2010). Variance component model to account for sample structure in genome-wide association studies. Nat. Genet. *42*, 348–354.
25. Thorisson, G.A., Smith, A.V., Krishnan, L., and Stein, L.D. (2005). The International HapMap Project Web site. Genome Res. *15*, 1592–1593.
26. Fogarty, M.P., Xiao, R., Prokunina-Olsson, L., Scott, L.J., and Mohlke, K.L. (2010). Allelic expression imbalance at high-density lipoprotein cholesterol locus MMAB-MVK. Hum. Mol. Genet. *19*, 1921–1929.
27. Abecasis, G.R., Auton, A., Brooks, L.D., DePristo, M.A., Durbin, R.M., Handsaker, R.E., Kang, H.M., Marth, G.T., and McVean, G.A.; 1000 Genomes Project Consortium (2012). An integrated map of genetic variation from 1,092 human genomes. Nature *491*, 56–65.
28. Gaulton, K.J., Nammo, T., Pasquali, L., Simon, J.M., Giresi, P.G., Fogarty, M.P., Panhuis, T.M., Mieczkowski, P., Secchi, A., Bosco, D., et al. (2010). A map of open chromatin in human pancreatic islets. Nat. Genet. *42*, 255–259.
29. Stitzel, M.L., Sethupathy, P., Pearson, D.S., Chines, P.S., Song, L., Erdos, M.R., Welch, R., Parker, S.C.J., Boyle, A.P., Scott, L.J., et al.; NISC Comparative Sequencing Program (2010). Global epigenomic analysis of primary human pancreatic islets provides insights into type 2 diabetes susceptibility loci. Cell Metab. *12*, 443–455.
30. Bernstein, B.E., Stamatoyannopoulos, J.A., Costello, J.F., Ren, B., Milosavljevic, A., Meissner, A., Kellis, M., Marra, M.A., Beaudet, A.L., Ecker, J.R., et al. (2010). The NIH Roadmap Epigenomics Mapping Consortium. Nat. Biotechnol. *28*, 1045–1048.
31. ENCODE Project Consortium (2011). A user's guide to the encyclopedia of DNA elements (ENCODE). PLoS Biol. *9*, e1001046.
32. Miyazaki, J., Araki, K., Yamato, E., Ikegami, H., Asano, T., Shibasaki, Y., Oka, Y., and Yamamura, K. (1990). Establishment of

a pancreatic beta cell line that retains glucose-inducible insulin secretion: special reference to expression of glucose transporter isoforms. Endocrinology *127*, 126–132.
33. Hohmeier, H.E., Mulder, H., Chen, G., Henkel-Rieger, R., Prentki, M., and Newgard, C.B. (2000). Isolation of INS-1-derived cell lines with robust ATP-sensitive K+ channel-dependent and -independent glucose-stimulated insulin secretion. Diabetes *49*, 424–430.
34. Portales-Casamar, E., Thongjuea, S., Kwon, A.T., Arenillas, D., Zhao, X., Valen, E., Yusuf, D., Lenhard, B., Wasserman, W.W., and Sandelin, A. (2010). JASPAR 2010: the greatly expanded open-access database of transcription factor binding profiles. Nucleic Acids Res. *38* (Database issue), D105–D110.
35. Matys, V., Fricke, E., Geffers, R., Gössling, E., Haubrock, M., Hehl, R., Hornischer, K., Karas, D., Kel, A.E., Kel-Margoulis, O.V., et al. (2003). TRANSFAC: transcriptional regulation, from patterns to profiles. Nucleic Acids Res. *31*, 374–378.
36. Sandelin, A., Wasserman, W.W., and Lenhard, B. (2004). ConSite: web-based prediction of regulatory elements using cross-species comparison. Nucleic Acids Res. *32* (Web Server issue), W249–W252.
37. Levy, S., and Hannenhalli, S. (2002). Identification of transcription factor binding sites in the human genome sequence. Mamm. Genome *13*, 510–514.
38. Ghosh, D. (2000). Object-oriented transcription factors database (ooTFD). Nucleic Acids Res. *28*, 308–310.
39. Pruim, R.J., Welch, R.P., Sanna, S., Teslovich, T.M., Chines, P.S., Gliedt, T.P., Boehnke, M., Abecasis, G.R., and Willer, C.J. (2010). LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics *26*, 2336–2337.
40. Pastinen, T. (2010). Genome-wide allele-specific analysis: insights into regulatory variation. Nat. Rev. Genet. *11*, 533–538.
41. Pisania, A., Weir, G.C., O'Neil, J.J., Omer, A., Tchipashvili, V., Lei, J., Colton, C.K., and Bonner-Weir, S. (2010). Quantitative analysis of cell composition and purity of human pancreatic islet preparations. Lab. Invest. *90*, 1661–1675.
42. Ichii, H., Inverardi, L., Pileggi, A., Molano, R.D., Cabrera, O., Caicedo, A., Messinger, S., Kuroda, Y., Berggren, P.-O., and Ricordi, C. (2005). A novel method for the assessment of cellular composition and beta-cell viability in human islet preparations. Am. J. Transplant. *5*, 1635–1645.
43. Greenawalt, D.M., Dobrin, R., Chudin, E., Hatoum, I.J., Suver, C., Beaulaurier, J., Zhang, B., Castro, V., Zhu, J., Sieberts, S.K., et al. (2011). A survey of the genetics of stomach, liver, and adipose gene expression from a morbidly obese cohort. Genome Res. *21*, 1008–1016.
44. Dimas, A.S., Deutsch, S., Stranger, B.E., Montgomery, S.B., Borel, C., Attar-Cohen, H., Ingle, C., Beazley, C., Gutierrez Arcelus, M., Sekowska, M., et al. (2009). Common regulatory variation impacts gene expression in a cell type-dependent manner. Science *325*, 1246–1250.
45. Yang, T.-P., Beazley, C., Montgomery, S.B., Dimas, A.S., Gutierrez-Arcelus, M., Stranger, B.E., Deloukas, P., and Dermitzakis, E.T. (2010). Genevar: a database and Java application for the analysis and visualization of SNP-gene associations in eQTL studies. Bioinformatics *26*, 2474–2476.
46. Smith, S.B., Ee, H.C., Conners, J.R., and German, M.S. (1999). Paired-homeodomain transcription factor PAX4 acts as a transcriptional repressor in early pancreatic development. Mol. Cell. Biol. *19*, 8272–8280.
47. Fujitani, Y., Kajimoto, Y., Yasuda, T., Matsuoka, T.A., Kaneto, H., Umayahara, Y., Fujita, N., Watada, H., Miyazaki, J.I., Yamasaki, Y., and Hori, M. (1999). Identification of a portable repression domain and an E1A-responsive activation domain in Pax4: a possible role of Pax4 as a transcriptional repressor in the pancreas. Mol. Cell. Biol. *19*, 8281–8291.
48. Ritz-Laser, B., Estreicher, A., Gauthier, B.R., Mamin, A., Edlund, H., and Philippe, J. (2002). The pancreatic beta-cell-specific transcription factor Pax-4 inhibits glucagon gene expression through Pax-6. Diabetologia *45*, 97–107.
49. Petersen, H.V., Jørgensen, M.C., Andersen, F.G., Jensen, J., F-Nielsen, T., Jørgensen, R., Madsen, O.D., and Serup, P. (2000). Pax4 represses pancreatic glucagon gene expression. Mol. Cell Biol. Res. Commun. *3*, 249–254.
50. Duncan, M.K., Haynes, J.I., 2nd, Cvekl, A., and Piatigorsky, J. (1998). Dual roles for Pax-6: a transcriptional repressor of lens fiber cell-specific β-crystallin genes. Mol. Cell. Biol. *18*, 5579–5586.
51. Shyr, C.-R., Tsai, M.-Y., Yeh, S., Kang, H.-Y., Chang, Y.-C., Wong, P.-L., Huang, C.-C., Huang, K.-E., and Chang, C. (2010). Tumor suppressor PAX6 functions as androgen receptor co-repressor to inhibit prostate cancer growth. Prostate *70*, 190–199.
52. Liu, T., Zhao, Y., Tang, N., Feng, R., Yang, X., Lu, N., Wen, J., and Li, L. (2012). Pax6 directly down-regulates Pcsk1n expression thereby regulating PC1/3 dependent proinsulin processing. PLoS ONE *7*, e46934.
53. Shimajiri, Y., Sanke, T., Furuta, H., Hanabusa, T., Nakagawa, T., Fujitani, Y., Kajimoto, Y., Takasu, N., and Nanjo, K. (2001). A missense mutation of Pax4 gene (R121W) is associated with type 2 diabetes in Japanese. Diabetes *50*, 2864–2869.
54. Yasuda, T., Kajimoto, Y., Fujitani, Y., Watada, H., Yamamoto, S., Watarai, T., Umayahara, Y., Matsuhisa, M., Gorogawa, S., Kuwayama, Y., et al. (2002). PAX6 mutation as a genetic factor common to aniridia and glucose intolerance. Diabetes *51*, 224–230.
55. Plengvidhya, N., Kooptiwut, S., Songtawee, N., Doi, A., Furuta, H., Nishi, M., Nanjo, K., Tantibhedhyangkul, W., Boonyasrisawat, W., Yenchitsomanus, P.T., et al. (2007). PAX4 mutations in Thais with maturity onset diabetes of the young. J. Clin. Endocrinol. Metab. *92*, 2821–2826.
56. Mauvais-Jarvis, F., Smith, S.B., Le May, C., Leal, S.M., Gautier, J.-F., Molokhia, M., Riveline, J.-P., Rajan, A.S., Kevorkian, J.-P., Zhang, S., et al. (2004). PAX4 gene variations predispose to ketosis-prone diabetes. Hum. Mol. Genet. *13*, 3151–3159.
57. Wen, J.H., Chen, Y.Y., Song, S.J., Ding, J., Gao, Y., Hu, Q.K., Feng, R.P., Liu, Y.Z., Ren, G.C., Zhang, C.Y., et al. (2009). Paired box 6 (PAX6) regulates glucose metabolism via proinsulin processing mediated by prohormone convertase 1/3 (PC1/3). Diabetologia *52*, 504–513.
58. Ma, R.C.W., Hu, C., Tam, C.H., Zhang, R., Kwan, P., Leung, T.F., Thomas, G.N., Go, M.J., Hara, K., Sim, X., et al.; DIAGRAM Consortium; MuTHER Consortium (2013). Genome-wide association study in a Chinese population identifies a susceptibility locus for type 2 diabetes at 7q32 near PAX4. Diabetologia *56*, 1291–1305.
59. Wethmar, K., Smink, J.J., and Leutz, A. (2010). Upstream open reading frames: molecular switches in (patho)physiology. Bioessays *32*, 885–893.
60. Símová, S., Klíma, M., Cermak, L., Sourková, V., and Andera, L. (2008). Arf and Rho GAP adapter protein ARAP1 participates in the mobilization of TRAIL-R1/DR4 to the plasma membrane. Apoptosis *13*, 423–436.

61. Daniele, T., Di Tullio, G., Santoro, M., Turacchio, G., and De Matteis, M.A. (2008). ARAP1 regulates EGF receptor trafficking and signalling. Traffic *9*, 2221–2235.
62. Yoon, H.-Y., Lee, J.-S., and Randazzo, P.A. (2008). ARAP1 regulates endocytosis of EGFR. Traffic *9*, 2236–2252.
63. Cuthbert, E.J., Davis, K.K., and Casanova, J.E. (2008). Substrate specificities and activities of AZAP family Arf GAPs in vivo. Am. J. Physiol. Cell Physiol. *294*, C263–C270.
64. Lawrence, J.T.R., and Birnbaum, M.J. (2003). ADP-ribosylation factor 6 regulates insulin secretion through plasma membrane phosphatidylinositol 4,5-bisphosphate. Proc. Natl. Acad. Sci. USA *100*, 13320–13325.
65. Wang, Z., Oh, E., and Thurmond, D.C. (2007). Glucose-stimulated Cdc42 signaling is essential for the second phase of insulin secretion. J. Biol. Chem. *282*, 9536–9546.
66. Jayaram, B., Syed, I., Kyathanahalli, C.N., Rhodes, C.J., and Kowluru, A. (2011). Arf nucleotide binding site opener [ARNO] promotes sequential activation of Arf6, Cdc42 and Rac1 and insulin secretion in INS 832/13 β-cells and rat islets. Biochem. Pharmacol. *81*, 1016–1027.