Type 2 Diabetes associated genetic variants regulate chromatin accessibility in human islets

Shubham Khetan\textsuperscript{1,3}, Romy Kursawe\textsuperscript{1}, Ahrim Youn\textsuperscript{1}, Nathan Lawlor\textsuperscript{1}, Alexandria Jillette\textsuperscript{1}, Eladio J. Marquez\textsuperscript{1}, Duygu Ucar\textsuperscript{1,2,3*}, Michael L. Stitzel\textsuperscript{1,2,3*}

\textsuperscript{1}The Jackson Laboratory for Genomic Medicine, Farmington, CT, 06032, USA
\textsuperscript{2}Institute of Systems Genomics, \textsuperscript{3}Department of Genetics and Genome Sciences, University of Connecticut, Farmington, CT, 06032, USA

Corresponding authors: michael.stitzel@jax.org, duygu.ucar@jax.org
Phone numbers: 860-837-2431 (M.L.S.); 860-837-2436 (D.U.)

Running Title: Genetic regulation of islet epigenomes

Keywords: Human pancreatic islets, epigenome, chromatin accessibility, ATAC-seq, open chromatin region (OCR), type 2 diabetes (T2D), genome-wide association study (GWAS), single nucleotide polymorphism (SNP), chromatin accessibility quantitative trait locus (caQTL), beta cell, alpha cell
Abstract

Type 2 Diabetes (T2D) is a complex disorder in which both genetic and environmental risk factors contribute to islet dysfunction and failure. Genome-wide association studies (GWAS) have linked single nucleotide polymorphisms (SNPs), most of which are noncoding, in >200 loci to islet dysfunction and T2D. Identification of the putative causal variants, their target genes, and whether they lead to gain- or loss-of-function remains challenging. Here, we profiled chromatin accessibility in pancreatic islet samples from 19 genotyped individuals and identified 2949 SNPs associated with in vivo cis-regulatory element (RE) use (i.e., chromatin accessibility quantitative trait loci (caQTL)). Among the caQTLs tested (n=13) using luciferase reporter assays in MIN6 beta cells, more than half exhibited effects on enhancer activity that were consistent with in vivo chromatin accessibility changes. Importantly, islet caQTL analysis nominated putative causal SNPs in 13 T2D-associated GWAS loci, linking 7 and 6 T2D risk alleles, respectively, to gain or loss of in vivo chromatin accessibility. By investigating the effect of genetic variants on chromatin accessibility in islets, this study is an important step forward in translating T2D-associated GWAS SNP into functional molecular consequences.
Introduction

Type 2 diabetes (T2D) is a complex disease resulting from the combined effects of an individual’s genetic predisposition and environmental exposures (1,2). It ultimately manifests when islets cannot secrete sufficient insulin to compensate for insulin resistance in peripheral tissues (3). Genome-wide association studies (GWAS) have identified single nucleotide polymorphisms (SNPs) in >200 loci that confer genetic susceptibility to T2D and/or alter quantitative measures of islet (dys)function (4,5). These SNPs are predominantly non-coding (~90%) and enriched within islet-specific cis-regulatory elements (cis-REs) (6–9), implicating perturbed islet transcription in T2D etiology (2). However, identifying the causal variant(s) in each T2D-associated GWAS locus, their molecular effects, and the genes and pathways they affect, remains critical to translate genetic associations into mechanistic understanding and treatments.

Quantitative trait locus (QTL) analyses have linked common genetic variants to in vivo gene expression changes (eQTL) for multiple cell types (10), including islets (8,11,12). However, eQTLs cannot pinpoint the causal variants among the multiple SNPs in linkage disequilibrium with each other. QTL approaches have recently been applied in cell lines to link genetic variation to epigenomic changes, such as DNaseI sensitivity (dsQTLs (13)) chromatin accessibility (caQTLs (14–16)), and histone modification levels (17). However, little is known about how genetic variation affects epigenomes of clinically relevant primary tissues such as islets.

In this study, we used the Assay for Transposase-Accessible Chromatin-sequencing (ATAC-seq) (18) to profile genome-wide chromatin accessibility in islets from 19 individuals (14 non-diabetic + 5 T2D). Using caQTL analysis, we identified genetic variants altering in vivo chromatin accessibility in islets and exhibiting concordant effects on in vitro luciferase reporter activity. Finally, we identified putative causal variants altering islet chromatin accessibility in 13 distinct T2D-associated GWAS loci.
Together, this study provides a roadmap for translating T2D-associated GWAS SNPs into functional molecular effects.

**Material and Methods**

**Study subjects and islet culture:** Fresh human cadaveric pancreatic islets were procured from Prodo Labs or the Integrated Islet Distribution Program (IIDP) (Table S1) and processed according to Institutional Review Board-approved protocols. Upon receipt, cells were transferred into PIM(S) media supplemented with PIM(ABS) and PIM(G) (Prodo Labs) and incubated overnight in a T-150 non-tissue culture treated flask (VWR) at 37 C and 5% CO\(_2\) overnight. The following day, nuclei and total RNA were isolated for ATAC-seq and RNA-seq library preparation and analysis (8). Genomic DNA was isolated from islet explant cultures using Qiagen DNeasy Blood and Tissue Kit as previously described (8).

**ATAC-seq profiling:** Islet ATAC-seq libraries were prepared as previously described (8) from 22 donors. Per donor, three replicates, each consisting of 50-100 islet equivalents (50,000-100,000 cells), were transposed. Libraries were barcoded, pooled into 3-donor batches (corresponding to 9 barcoded transposition reactions), and sequenced using 2x75 bp Illumina NextSeq500 to an average depth of 62.6 (±18.6) million paired-end reads per donor (Table S2). Low quality portions of reads were trimmed using Trimmomatic (19) and aligned to the hg19 human genome assembly using Burrows Wheeler Aligner-MEM (20). For each replicate, reads were shifted as previously described and duplicate reads were removed (21,22). Technical replicates were merged using SAMtools after confirming high correlation between them. Open chromatin regions (OCRs) were called for each islet sample using MACS2 (23) (with parameters -callpeak -nomodel -f BAMPE) at FDR 1%. Islet ATAC-seq samples with less than 30,000 OCRs...
were excluded from further analyses, yielding data for 19 individuals. OCRs on sex chromosomes and those overlapping low mapability regions (Blacklisted Regions: http://hgdownload.cse.ucsc.edu/goldenpath/hg19/encodeDCC/wgEncodeMapability/) were excluded, resulting in 154,437 autosomal OCRs detected in at least one individual using R package DiffBind (24). Deeptools was used to generate bedgraph files for UCSC genome browser sessions (with parameters --normalizeUsingRPKM --centerReads --scaleFactor=1 -bs=25).

**OCR chromatin state annotations:** Previously described chromatin states for islets (8), ENCODE, and Roadmap Epigenomics (25) cells/tissues were used to annotate islet OCRs and visualized using ggplot2 (26). OCRs overlapping ≥2 different chromatin states were assigned a single state using the following hierarchy: Active TSS > Bivalent TSS > Weak TSS > Flanking TSS > Active Enhancer-1 > Active Enhancer-2 > Weak Enhancer > Genic Enhancer > Strong Transcription > Weak Transcription > Repressed Polycomb > Weak Repressed Polycomb > Quiescent. Previously-described stretch enhancer regions (6,8) were overlapped with islet OCRs and tested for enrichment using Fisher’s exact test. For each tissue-specific test, the background set comprised stretch enhancers from all other tissues (n=30).

**Genotyping, imputation, and caQTL analysis:** Each islet donor was genotyped using Illumina Infinium Omni2.5Exome (n=11) or Omni5Exome (n=8) BeadChip arrays (Table S1). We mapped Illumina array probe sequences to the hg19 genome assembly using BWA. SNPs with ambiguous probe alignments, 1000 Genomes (1000G) phase 1 variants with minor allele frequency of ≥1% within 7 bp of the 3’ end of probes, or call rates <95% were excluded. All alleles were oriented relative to the reference. Genotype calls were merged using vcftools/0.1.12a suite (vcf-merge command). After removing
SNPs with missing data (--max-missing 1), ~2.4 million SNPs were used for imputation (1000G Phase 3 v5 (27)) and phasing (Eagle v2.3 (28)) using the Michigan Imputation Server (https://imputationserver.sph.umich.edu/index.html) (29).

VerifyBamID (30) was used to match ATAC-seq bam files to individuals’ genotypes to eliminate the possibility of a sample swap. Islet OCRs overlapping only monomorphic SNPs were removed from caQTL analyses, yielding 84,499 OCRs. Allele-specific counts were obtained for 195,207 SNPs within these OCRs, and caQTLs were detected using RASQUAL (15). To minimize confounding factors such as batch effects, we adopted the strategy described in (15) and used the first 5 principal components (PCs) as covariates in the RASQUAL model. Significant caQTLs were identified using a two-stage multiple-hypothesis-testing correction (15): 1) correcting for the multiple SNPs tested within each OCR using Bonferroni correction; then 2) correcting for the number of OCRs tested genome-wide by controlling FDR at 10% using RASQUAL’s permutation test (‘--random-permutation’) 50 times.

To visualize chromatin accessibility patterns at caQTLs, first we calculated the number of ATAC-seq reads (normalized with respect to library size) spanning each base pair for all 19 samples using BEDTools (‘genomecov’). Next, islet donors were grouped based on their genotypes for each displayed caQTL; average read counts were calculated for each genotype group and plotted using the ‘polygon’ function in R.

**Differential OCR analyses:** Differential chromatin accessibility analyses were conducted between islet ATAC-seq profiles of 5 T2D and 5 non-diabetic individuals with the most comparable demographics (Table S3). To identify statistically significant T2D disease state-associated chromatin accessibility changes, only OCRs meeting the following criteria were used for differential analyses (n=52,387): 1) present in ≥3 islet donors; 2) present in ≥ 1 T2D and ≥2 ND (or ≥1ND and ≥2 T2D) individuals. Race, sex,
and significant surrogate variables (n=2) from Surrogate Variable Analysis (SVA (31)) were used as covariates to minimize confounding factors. EdgeR (32) R package was used to identify differentially accessible OCRs.

**GWAS SNP enrichment in islet caQTLs:** The NHGRI-EBI Catalog of GWAS index SNPs for 184 diseases/traits was retrieved on January 19th, 2017 (https://www.ebi.ac.uk/gwas/) and LD-pruned using PLINK (33) version 1.9 (parameters --maf 0.05 --clump --clump-p1 0.0001 --clump-p2 0.01 --clump-r2 0.2 --clump-kb 1000) to avoid testing enrichment for multiple SNPs representing the same genetic association signal/locus per trait. For index SNPs exhibiting pairwise correlation \( r^2 > 0.2 \), only the SNP with the more significant p-value was retained. We used GREGOR (34) on this LD-pruned list to determine whether GWAS index or linked SNPs \( (r^2>0.8, \text{LD window size } = 1\text{Mb}, \text{and minimum neighbor number } = 500) \) were enriched in islet caQTLs or differentially accessible OCRs.

**Transcription factor (TF) motif enrichments:** Homer (35) findMotifsGenome.pl script (parameters: hg19, –size given) was used to identify TF motifs enriched in islet OCRs. We compared motifs in OCRs that are accessible only in islets \( (n=40,271 \text{ islet-specific OCRs}) \) to motifs in OCRs that are also accessible in adipose, CD4\(^+\) T, GM12878, or PBMCs \( (n=41,639 \text{ shared/background OCRs}) \) (Figure 1C). Motifs enriched in caQTL-containing OCRs (Figure 2D) were identified by comparing caQTL OCRs \( (n=2949) \) to all islet OCRs \( (n=154,437) \). TFs were clustered based on the similarity of their Position Weight Matrices (PWMs) using Kullback-Leibler divergence method implemented in TFBSTools (36). Motif enrichments for differential OCRs \( (n=1515) \) were calculated against all OCRs used in differential analysis \( (n=52,387) \).
**RNA-seq profiling:** Total RNA was isolated from each islet sample using Trizol (8). Stranded RNA-seq libraries were prepared from total RNA using the TruSeq stranded mRNA kit (Illumina) for the 19 individuals with high-quality ATAC-seq data; ERCC Mix 1 or Mix 2 spike-ins were randomly added to each sample (Thermo Fisher, catalog #4456740; see Table S4). RNA-seq from 10 islet samples used in differential analyses were sequenced together on an Illumina NextSeq 500 to minimize batch effects, while the remaining nine samples were sequenced on Illumina HiSeq 2500, each to an average sequencing depth of 87.2 (± 27.8) million paired-end reads (Table S4). Paired-end RNA-seq reads were trimmed to remove low quality base calls using Trimmomatic (19). Bowtie2 (37) and RSEM (38) (rsem-calculate-expression) were used to determine fragments per kilobase of transcript per million mapped reads (FPKM) and expected read counts for all Ensembl hg19 Release 70 transcripts.

**Differential gene expression analyses:** RNA-seq data from 10 islet samples (Table S3) were used for differential expression analysis. Expected read counts for autosomal genes with FPKM >5 in ≥3 RNA-seq samples (n=10,116) were used in differential analyses based on edgeR (32) models (FDR=10%). Race, sex, ERCC-spike-in and significant SVs (n=1) from SVA were used as covariates to minimize the impact of confounding factors on T2D disease state-associated gene expression changes.

**Expression QTL (eQTL) analysis:** RSEM expected read counts (38) for 9,656 expressed genes (median FPKM >5) were used to identify islet eQTLs from 19 donors using RASQUAL (15). Only SNPs within the gene body or within 50 kilobases (kb) flanking the gene body were tested. To minimize potential batch effects, we adopted the strategy described in (15) and used the first 4 principle components (PCs), in addition to age, sex, race, T2D status, and sequencing date as covariates in the RASQUAL model.
A two-stage multiple hypothesis-testing correction (15) was used to determine significant eQTLs similar to caQTLs, where only 10 permutation tests were used in step two.

**Islet caQTL-eQTL overlaps**: Quantile-quantile (QQ) plots for caQTL p-values were generated against the expectation of a uniform p-value distribution between 0 and 1. The QQ plot was generated for islet eQTL SNPs from 112 individuals (8) and caQTL SNPs from 19 individuals by conditioning on lead caQTL SNPs that were either statistically significant at FDR 10% or background sets of randomly selected non-significant ones. Random sets of non-significant SNPs (n=2545) were generated 10 times to eliminate sampling bias; a representative result from one random set is shown in Figures 2F and S2G.

**Gateway cloning of selected islet caQTL sequences and alleles**: Islet genomic DNA from individuals homozygous for the reference or alternate allele was used as templates to PCR amplify sequences containing each allele for 13 islet caQTLs (Table S5). The corresponding 26 PCR amplicons were cloned into the pDONR201 vector using BP Clonase (Invitrogen). Sequences were validated by Sanger sequencing. Each islet caQTL sequence was transferred from pDONR201 into the Gateway-modified pGL4.23F plasmid (39) with LR Clonase.

**Luciferase reporter assays**: MIN6 cells were seeded in 96-well plates at a density of 60,000 cells per well 24 hours prior to transfection as previously described (39). 0.072 pmol Gateway-modified firefly (pGL4.23, Promega) plasmid containing each islet caQTL sequence (Table S5) and 2 ng renilla (pRL-TK, Promega) plasmid were co-transfected in triplicate using Lipofectamine 2000 Transfection reagent (Life Technologies). The Dual Luciferase Reporter Assay system (Promega) was used to determine firefly and renilla
luciferase activity in each sample. Cells were lysed with 1x Passive Lysis Buffer 38-40 hours after transfection. Luminescence was measured on a Synergy2 Microplate Reader (BioTek). Firefly values were normalized to renilla to control for differences in cell number or transfection efficiency.

Results

Human pancreatic islet chromatin accessibility maps

To determine the genome-wide location of cis-regulatory elements (cis-REs) in human islets, we generated high-quality ATAC-seq profiles from 19 islet donors (Figure S1A; Tables S1 and S2). Investigating chromatin accessibility near the \( Nkx6.1 \) locus, a well-characterized beta cell-specific transcription factor (TF), revealed both open chromatin regions (OCRs) unique to islet samples (Figure 1A, orange and green rectangles) and OCRs shared with other cell types (22,40) (Figure 1A, gray rectangle). Overall, chromatin accessibility profiles from 19 islets were highly correlated to each other and to those from sorted islet alpha and beta cells (Figure 1B and S1B) (41). Notably, ATAC-seq profiles from T2D donors (n=5; Figure 1B, asterisks), did not cluster separately from non-diabetic (ND) donors, suggesting that the T2D disease state does not lead to global remodeling of human islet chromatin accessibility.

Collectively, we identified 154,437 islet OCRs accessible in at least one of the 19 individuals (Table S6). Comparison with reported chromatin state annotations in human islets (6,8) assigned 12.9% and 23.14% of these OCRs as putative promoters and enhancers, respectively (Figure S1C). Putative promoter OCRs were shared with several of 30 tissues profiled by the NIH Roadmap Epigenomics project (25). Putative enhancer OCRs were more specific to islets, consistent with previous observations of cell type-specificity of enhancers (42). To further assess the islet-specificity of detected OCRs, we compared them to stretch enhancers (SEs) - long (>3 kb) contiguous
enhancer chromatin states that govern cell-specific functions and often harbor disease-associated SNPs relevant to the cognate cell type (6). The majority (90%) of islet SEs overlapped islet OCRs (Figure S1D), significantly greater than overlaps observed between islet OCRs and SEs in other tissues (Fisher’s Exact p<2.2e-16). As anticipated, DNA sequence binding motifs of islet-specific TFs, such as PDX1 and NKX6.1, were significantly enriched in OCRs that are accessible in islet samples and not in GM12878, PBMCs, CD4+ T cells, skeletal muscle or adipose tissues (Figure 1C). Together, these observations indicate that high-quality chromatin accessibility maps of islets from multiple individuals reveal cis-REs (OCRs) important for islet-specific gene regulation.

Only 10% (n=15,917) of islet OCRs were detected in all 19 individuals (Figure 1D), which were overwhelmingly annotated as promoters (Figure 1E, red bars). In contrast, 39.3% (n=60,713) of OCRs were detected in only 1 out of 19 individuals (Figure 1D) and were found predominantly (45%) in quiescent/low signal chromatin states (Figure 1E, white bars). Though we cannot eliminate the possibility of false positives in OCR detection, these might also represent individual-specific enhancers missed in reference islet chromatin states, since references were based on data from 2-3 individuals. OCRs detected in two to eighteen individuals (Figure 1D) were mostly enhancers (Figure 1E, orange/yellow bars), suggesting that genetic differences (i.e., SNPs) between individuals may alter the chromatin accessibility, and therefore the activity, of human islet enhancers.

Genetics of chromatin accessibility in human islets

To identify genetic variants (SNPs and small insertions/deletions) that alter chromatin accessibility of islet OCRs in which they reside (Figure 2A), we genotyped islet samples and conducted chromatin accessibility quantitative trait locus (caQTL) analysis using RASQUAL (15), a method that can discover QTLs from small sample
sizes. Using RASQUAL, we identified 2949 SNPs associated with increased or decreased chromatin accessibility at FDR 10% (Figure S2A; Table S7) from 19 islet samples. For example, the rs488797 ‘C’ allele was associated with reduced OCR accessibility in an islet SE in the intron of CELF4 (Figure 2B), a gene selectively expressed in islets (8,40). ‘CC’ homozygous islet donors exhibited dramatically lower accessibility than ‘CT’ or ‘TT’ genotypes (Figure 2B, compare blue (‘CC’), pink (‘CT’), and green ‘TT’) profiles). Moreover, ATAC-seq sequences overlapping rs488797 in ‘CT’ heterozygous samples almost exclusively contained the ‘T’ allele (Figure 2B, inset), reinforcing genetics as a strong determinant of chromatin accessibility at this OCR.

The rs488797 ‘C’ allele is predicted to disrupt FOXA2 binding (Figure 2B, sequence in square brackets vs. FOXA2 binding motif). To test this, we analyzed FOXA2 ChIP-seq data from two islet donors (HI101 and HI32) (7) (Figure 2C). We leveraged FOXA2 ChIP-seq reads and genetic linkage disequilibrium (LD) to infer genotypes of these individuals in this region. Since the caQTL SNP rs488797 alters in vivo islet chromatin accessibility, we imputed its genotype using a linked SNP rs648005 (T/C) ($r^2=0.99$ with rs488797). rs648005 overlaps a distinct OCR and a FOXA2 binding site 8,178 nucleotides away but is neither an islet caQTL nor predicted to disrupt a FOXA2 motif. In HI101, FOXA2 ChIP-seq reads overlapping rs648005 contained both ‘C’ and ‘T’ alleles (Figure 2C, top) indicating that HI101 is heterozygous at rs648005 and, by extension, at rs488797 with high probability. However, FOXA2 ChIP-seq reads overlapping the caQTL SNP rs488797 exclusively contained the ‘T’ allele, consistent with the islet caQTL analysis and supporting FOXA2 motif disruption predictions. In HI32, FOXA2 ChIP-seq reads at rs648005 contained only the ‘T’ allele, suggesting that this individual is a ‘TT’ homozygote at rs648005, and therefore a ‘CC’ homozygote at rs488797 with high probability. Notably, no FOXA2 binding is observed at rs488797 for HI32, providing further support that the ‘C’ allele disrupts FOXA2 binding. Table S8
provides predicted motif disruptions from HaploReg (43) for all islet caQTL including rs488797.

Islet caQTLs were uniformly distributed across the autosomal chromosomes (Figure S2B), and the majority (>98%) were located within 200 kilobases flanking the transcription start site (TSS) of the nearest islet-expressed gene (Figure S2C). 12% of islet caQTLs were in promoters, whereas 30% overlapped enhancers (Figure S2D). Islet caQTLs were exclusively enriched in islet SEs compared to SEs in other tissues (Figure S2E). Finally, sequence motifs for islet-specific TFs, such as FOXA2, NKX6.1, and PDX1, were enriched in caQTL OCRs (Figure 2D). To validate this, we overlapped caQTL OCRs with ChIP-seq data from human islets for islet-specific TFs and ubiquitous CTCF (7). We found that FOXA2, NKX6.1, and PDX1 binding (i.e., ChIP-seq peaks) were enriched at caQTLs (Figure S2F), in contrast to CTCF, whose binding sites were not enriched at islet caQTLs. Together, these results suggest that motif enrichment analyses likely reflect actual binding of these TFs at caQTL OCRs. Surprisingly, sequence motifs of oxidative stress-responsive TFs, such as BACH1, BACH2, and NRF2, were also enriched in caQTL OCRs, suggesting that some caQTLs may modulate stress/stimulus-responsive cis-RE activity.

To determine if caQTL alleles altering in vivo chromatin accessibility elicit concordant effects on in vitro enhancer activity, we selected a subset of caQTLs (n=13) that were nearby genes exhibiting islet-specific expression (8) (e.g., Figure 2B). We cloned DNA sequences containing each islet caQTL allele (Table S5) and measured their enhancer activity using luciferase reporter assays in MIN6 mouse beta cells. We observed allelic effects on luciferase activity for 8/13 caQTLs tested (Figure 2E). Importantly, for all 8/13 caQTLs, the allele that increased in vivo chromatin accessibility also increased in vitro enhancer activity (Figure 2E). Finally, we studied whether caQTL variants were also associated with variability in islet gene expression levels using islet
expression QTL (eQTL) data from this cohort (n=19). As shown in Figure 2F, caQTL variants exhibited more significant allelic effects on islet gene expression than randomly selected variants in OCRs (non-caQTLs). We observed the same trend comparing these caQTLs to eQTLs detected in a larger (n=112; Figure S2G), independent cohort (8). Importantly, for 84% of caQTL-eQTL pairs in our cohort (37/44; Figure S2H), we observed a concordant direction-of-effect (Pearson r=0.691), i.e., higher chromatin accessibility is associated with increased gene expression and vice versa (Figure S2H; Q1 and Q3), linking chromatin accessibility effects of these variants to downstream changes in islet gene expression.

Chromatin accessibility changes in T2D versus non-diabetic islet samples

To assess potential environmental effects of T2D disease state on the islet epigenome, we compared chromatin accessibility between 5 T2D donors and 5 non-diabetic (ND) donors with comparable demographics (Figure S3A; Table S3; e.g., age, race, sex). After completing surrogate variable analysis (SVA; (31)) to remove unwanted variation in the data (Figures S3B; e.g. batch effect, sex, post-mortem interval, drug treatment), we identified 1515/52,387 (2.8%) OCRs that were differentially accessible between T2D and ND islet samples at FDR=10% (Methods; Figures 3A and S3C; 609 at FDR 5%, and 79 at FDR 1%), where 714 have increased (opening OCRs) and 801 have decreased (closing OCRs) accessibility in T2D compared to ND samples (Figure 3A; Table S9). There was a remarkable difference in the chromatin state annotation of opening and closing OCRs. Closing OCRs, e.g., the one highlighted near BHLHE41 (Figure 3B, grey rectangle), mostly overlapped enhancers (48%), while opening OCRs extensively overlapped promoters (70%) (Figure 3C). This difference was also reflected in TF motif enrichments, where opening and closing OCRs were enriched for distinct motifs (Figure S3D). Interestingly, motifs for PDX1 and TFs that regulate stress
responses, such as ATF3/JUNB, AP-1, and BACH1, were enriched in closing OCRs, which may represent epigenomic signatures of previously described molecular perturbations in dysfunctional and T2D islets, including PDX1 export from the nucleus (44), perturbation of oxidative stress responses (45,46), and inactivation of beta cell survival pathways (47).

The overwhelming majority (>99%) of T2D disease state-associated changes in chromatin accessibility were quantitative, not qualitative, i.e., OCRs did not completely appear/disappear with T2D disease state. (Figure 3D). 654 genes were associated with opening OCRs (predominantly enhancers) and 622 genes were associated with closing OCRs (predominantly promoters). Differentially accessible OCRs at gene promoters exhibited modest positive correlation with the corresponding gene’s expression (Figure S3E). T2D disease state-associated OCRs were not enriched for any GO terms or KEGG/Wiki pathways. Differential gene expression analyses from the same ND and T2D samples revealed few significant changes (Table S9), where only 120 and 54 genes were up- or down-regulated, respectively, with T2D disease state at FDR 10% (Table S10).

Finally, given the significant impact of genetics on islet chromatin accessibility, we asked which T2D disease state-associated chromatin accessibility changes may be driven by genetic differences. Interestingly, 6% of the differentially accessible OCRs overlapped islet caQTLs (39 opening OCRs, 51 closing OCRs) (Figures 3D, S3F), including the opening OCR that contains the caQTL variant rs1463768. 4/5 T2D islet samples were heterozygous or homozygous for opening ‘G’ allele for this variant, while all 5 ND donors were homozygous for the closing ‘T’ allele (Figure 3E). rs1463768-containing sequences didn’t show allelic differences in luciferase reporter activity in MIN6 cells (Figure 2E). Therefore, it remains uncertain whether genotype, environment
(i.e., T2D disease state), or genotype-environment interactions are responsible for islet chromatin accessibility changes at this and other overlapping loci.

**T2D-associated GWAS SNPs altering islet chromatin accessibility**

The vast majority (>90%) of GWAS SNPs associated with T2D (4,48) and metabolic measures of islet (dys)function (49,50) are non-coding and overlap islet SEs (6,7). To test whether T2D- and islet (dys)function-associated GWAS SNPs alter chromatin accessibility in islets, we assessed overlaps of GWAS index and linked SNPs (Material and Methods) with islet caQTLs. Among 184 diverse trait- and disease-associated SNP sets tested, only those associated with T2D (2.97 fold), fasting plasma glucose (13.46 fold), and BMI-adjusted fasting glucose-related (7.43-fold) traits were significantly enriched in islet caQTLs (Figure 4A; p<5.43e-04, FDR=5%). In contrast, DNaseI sensitivity QTLs (13) in lymphoblastoid cell lines (LCLs) were enriched for mostly autoimmune disease-associated GWAS SNPs (Figure S4A), emphasizing the specificity of T2D-associated GWAS SNP enrichments in islet caQTLs.

We identified SNPs in 13 T2D-associated loci that alter islet chromatin accessibility, thereby nominating these as putative causal/functional SNPs (Figures 4B and S4B). caQTL SNP alleles for 4/13 T2D-associated loci (ADCY5, ZMIZ1, MTNR1B, RNF6) were previously linked to altered in vitro enhancer activity (51), in vivo chromatin accessibility (52), or in vivo steady state gene expression in islets (8,11,12,53). Importantly, T2D-associated risk alleles for these four loci exhibit concordant effects on chromatin accessibility and gene expression in islets, i.e. same direction-of-effect (Figure 4B). For 6/13 T2D-associated caQTLs, the risk allele decreased chromatin accessibility, designated as loss-of-function (Figure 4B). This included the T2D-associated caQTL SNP rs11708067 in the third intron of ADCY5, which overlaps an islet SE. The risk allele
for this variant is associated with reduced chromatin accessibility (Figure 4C), consistent with recent reports linking the rs11708067 risk allele to decreased transcriptional reporter activity in rodent beta cells (MIN6 and 832/13) and to decreased ADCY5 expression in non-diabetic human islets in vivo (12,51). The T2D risk allele was associated with increased chromatin accessibility for the remaining 7/13 islet caQTLs (Figure 4B), designated as gain-of-function. For example, the T2D risk allele ('A') at rs6937795 increased in vivo islet chromatin accessibility in the IL20RA locus (Figure 4D) and conferred 2.5-fold higher transcriptional reporter activity than the non-risk ‘C’ allele (Figure 2E). While targeted approaches are required to establish causality, our analyses nominate rs6937795 as a strong candidate for causal SNP in the T2D-associated IL20RA locus.

**Discussion**

In this study, we integrated ATAC-seq data and genotypes from 19 islet donors to link 2949 SNPs with altered in vivo chromatin accessibility. Allelic effects on in vivo chromatin accessibility correlated well with effects on in vitro enhancer activity; 8/13 caQTLs tested showed concordant allelic effects in luciferase reporter assays. Though we cannot eliminate the possibility of false positive associations for the remaining five caQTLs, these loci may also represent 1) alpha cell-specific, 2) species-specific, or 3) poised/primed cis-REs (16), which need to be tested in future studies in human cells.

The data suggest that islet caQTL variants modulate regulatory programs important for islet cell identity and function. They were enriched in islet-specific TF motifs, TF ChIP-seq peaks (7), and islet stretch enhancers (6). They were specific to islets, as only 2.3% (68/2949) of the islet caQTL variants altered chromatin accessibility in iPSCs or macrophages (data not shown) (16,54). Islet caQTL SNPs were linked to more significant effects on islet gene expression levels than variants that do not
significantly impact chromatin accessibility (i.e., non-caQTL SNPs). Increasing the cohort size and separating islet cell types in future studies should lead to increased convergence between islet caQTLs and islet eQTLs. Furthermore, studying islets under stress conditions could identify and link primed enhancers and response eQTLs, which have been reported in other cell types (16,55).

T2D disease state-associated changes in chromatin accessibility were limited and quantitative, i.e., few OCRs completely lost or gained accessibility with T2D, suggesting that the T2D disease state does not lead to extensive remodeling of steady state chromatin accessibility in islets. However, we acknowledge that T2D disease state-associated epigenetic changes may be masked by multiple factors, including: 1) relatively low HbA1c values for T2D donors (5.3%-7.4%), 2) cell type-specific changes hidden by whole islet measurements; 3) steady-state, normoglycemic culture conditions of the islets that may mask changes elicited by the diabetic state; and 4) limited power due to cohort size (n=10) and genetic diversity. We found that 6% of differentially accessible OCRs associated with T2D disease state overlapped caQTLs. Future studies integrating genotype, environment and their interaction in larger, genetically stratified cohorts should contribute to more precise understanding of epigenomic changes associated with T2D disease state.

This study demonstrates the utility of using islet caQTL analyses to identify and prioritize putative functional variants among hundreds of linked, ‘credible set’ T2D-associated SNPs (4,9,48). Even with a relatively small cohort (n=19), we identified putative causal variants at 13 T2D GWAS loci, based on their chromatin accessibility effects. These include four loci (ADCY5, MTNR1B, RNF6, and ZMIZ1) in which the same or linked (r^2>0.8) SNP functions as an islet eQTL (8,11,12,53). Importantly, the risk allele exhibited concordant effects on islet chromatin accessibility and gene expression for each locus. Finally, we identified allelic effects on both in vivo and in vitro islet
enhancer activity for multiple new loci, such as rs6937795 in the $IL20RA$ locus, and link the risk alleles at each locus to increased or decreased activity. This study provides new understanding of genetic variant effects on islet chromatin accessibility and enumerates targets for site-specific and hypothesis-driven investigation.

**Acknowledgements.** We are indebted to the anonymous pancreatic islet organ donors and their families, without whom this entire study would not be possible. A subset of human pancreatic islets was provided by the NIDDK-funded Integrated Islet Distribution Program (IIDP) at City of Hope, NIH Grant # 2UC4DK098085. We thank Jane Cha, Jackson Laboratory for Genomic Medicine, for help generating artwork for the figures, members of the Stitzel and Ucar labs for helpful discussion and critiques during study design and execution, and Taneli Helenius, Jackson Laboratory for Genomic Medicine, and anonymous reviewers, whose comments, questions, and suggested edits greatly improved the quality and clarity of this manuscript.

**Funding.** This study was made possible by generous financial support from Startup, Director’s Innovation (to M.L.S. and D.U.), and the Doug Coleman Research Fund at The Jackson Laboratory, by the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) under award number DK092251 (to M.L.S.), by the Assistant Secretary of Defense for Health Affairs, through the Peer Reviewed Medical Research Program under Award No. W81XWH-16-1-0130 (to M.L.S.), and by the American Diabetes Association Pathway to Stop Diabetes Accelerator Award 1-18-ACE-15 (to M.L.S.). Opinions, interpretations, conclusions, and recommendations are solely the responsibility of the authors and do not necessarily represent the official views of the National Institutes of Health, Department of Defense, or American Diabetes Association.
**Duality of Interest.** The authors declare they do not have financial or other conflicts of interest.

**Author Contributions.** S.K., D.U, and M.L.S conceived the study and designed experiments. R.K and M.L.S collected and prepared each islet sample for genotyping and sequencing. S.K. analyzed the data. A.Y., N.L, and E.M.C. contributed to bioinformatics and statistical analyses of the data. S.K. and A.J cloned and tested caQTL allelic effects using luciferase reporters. S.K, D.U., and M.L.S wrote the manuscript. All authors reviewed and edited the final manuscript. M.L.S. 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.

**Accession Numbers and code.** The accession number for human islet ATAC-seq and RNA-seq data reported in this paper is NCBI Sequence Read Archive: SRP117935. All supplementary tables and code used to generate the figures can be found on our GitHub page: https://github.com/UcarLab/DiabetesSubmission

**Prior Presentation.** Parts of this study were presented at the 2017 Pancreatic Diseases Gordon Research Conference, Waterville Valley, NH, June 18-23, 2017 and the Boston-Ithaca Islet Club Meeting, Worcester, MA, April 28-29, 2018.
References:

1. Halban PA, Polonsky KS, Bowden DW, Hawkins MA, Ling C, Mather KJ, et al. β-cell failure in type 2 diabetes: postulated mechanisms and prospects for prevention and treatment. Diabetes Care. 2014 Jun;37(6):1751–8.

2. Lawlor N, Khetan S, Ucar D, Stitzel ML. Genomics of Islet (Dys)function and Type 2 Diabetes. Trends Genet TIG. 2017 Apr;33(4):244–55.

3. Ashcroft FM, Rorsman P. Diabetes mellitus and the β cell: the last ten years. Cell. 2012 Mar 16;148(6):1160–71.

4. Fuchsberger C, Flannick J, Teslovich TM, Mahajan A, Agarwala V, Gaulton KJ, et al. The genetic architecture of type 2 diabetes. Nature. 2016 Aug 4;536(7614):41–7.

5. Mohlke KL, Boehnke M. Recent advances in understanding the genetic architecture of type 2 diabetes. Hum Mol Genet. 2015 Oct 15;24(R1):R85-92.

6. Parker SCJ, Stitzel ML, Taylor DL, Orozco JM, Erdos MR, Akiyama JA, et al. Chromatin stretch enhancer states drive cell-specific gene regulation and harbor human disease risk variants. Proc Natl Acad Sci U S A. 2013 Oct 29;110(44):17921–6.

7. Pasquali L, Gaulton KJ, Rodríguez-Seguí SA, Mularoni L, Miguel-Escalada I, Akerman I, et al. Pancreatic islet enhancer clusters enriched in type 2 diabetes risk-associated variants. Nat Genet. 2014 Feb;46(2):136–43.

8. Varshney A, Scott LJ, Welch RP, Erdos MR, Chines PS, Narisu N, et al. Genetic regulatory signatures underlying islet gene expression and type 2 diabetes. Proc Natl Acad Sci U S A. 2017 Feb 13;

9. Gaulton KJ, Ferreira T, Lee Y, Raimondo A, Mägi R, Reschen ME, et al. Genetic fine mapping and genomic annotation defines causal mechanisms at type 2 diabetes susceptibility loci. Nat Genet. 2015 Dec;47(12):1415–25.

10. GTEx Consortium. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science. 2015 May 8;348(6235):648–60.

11. Fadista J, Vikman P, Laakso EO, Mollet IG, Esguerra JL, Taneera J, et al. Global genomic and transcriptomic analysis of human pancreatic islets reveals novel genes influencing glucose metabolism. Proc Natl Acad Sci U S A. 2014 Sep 23;111(38):13924–9.

12. van de Bunt M, Manning Fox JE, Dai X, Barrett A, Grey C, Li L, et al. Transcript Expression Data from Human IsletsLinks Regulatory Signals from Genome-
Wide Association Studies for Type 2 Diabetes and Glycemic Traits to Their Downstream Effectors. PLoS Genet. 2015 Dec;11(12):e1005694.

13. Degner JF, Pai AA, Pique-Regi R, Veyrieras J-B, Gaffney DJ, Pickrell JK, et al. DNase I sensitivity QTLs are a major determinant of human expression variation. Nature. 2012 Feb 16;482(7385):390–4.

14. Gaffney DJ, McVicker G, Pai AA, Fondufe-Mittendorf YN, Lewellen N, Michelini K, et al. Controls of nucleosome positioning in the human genome. PLoS Genet. 2012;8(11):e1003036.

15. Kumasaka N, Knights AJ, Gaffney DJ. Fine-mapping cellular QTLs with RASQUAL and ATAC-seq. Nat Genet. 2016 Feb;48(2):206–13.

16. Alasoo K, Rodrigues J, Mukhopadhyay S, Knights AJ, Mann AL, Kundu K, et al. Shared genetic effects on chromatin and gene expression indicate a role for enhancer priming in immune response. Nat Genet. 2018 Jan 29;

17. McVicker G, van de Geijn B, Degner JF, Cain CE, Banovich NE, Raj A, et al. Identification of genetic variants that affect histone modifications in human cells. Science. 2013 Nov 8;342(6159):747–9.

18. Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat Meth. 2013 Dec;10(12):1213–8.

19. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinforma Oxf Engl. 2014 Aug 1;30(15):2114–20.

20. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinforma Oxf Engl. 2009 Jul 15;25(14):1754–60.

21. Lawlor N, Youn A, Kursawe R, Ucar D, Stitzel ML. Alpha TC1 and Beta-TC-6 genomic profiling uncovers both shared and distinct transcriptional regulatory features with their primary islet counterparts. Sci Rep. 2017 Sep 20;7(1):11959.

22. Ucar D, Márquez El, Chung C-H, Marches R, Rossi RJ, Uyar A, et al. The chromatin accessibility signature of human immune aging stems from CD8(+) T cells. J Exp Med. 2017 Oct 2;214(10):3123–44.

23. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, et al. Model-based analysis of ChIP-Seq (MACS). Genome Biol. 2008;9(9):R137.
24. Ross-Innes CS, Stark R, Teschendorff AE, Holmes KA, Ali HR, Dunning MJ, et al. Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. Nature. 2012 Jan 4;481(7381):389–93.

25. Roadmap Epigenomics Consortium, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, et al. Integrative analysis of 111 reference human epigenomes. Nature. 2015 Feb 19;518(7539):317–30.

26. Wickham H. Ggplot2: elegant graphics for data analysis. New York: Springer; 2009. 212 p. (Use R!).

27. 1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, et al. A global reference for human genetic variation. Nature. 2015 Oct 1;526(7571):68–74.

28. Loh P-R, Palamara PF, Price AL. Fast and accurate long-range phasing in a UK Biobank cohort. Nat Genet. 2016;48(7):811–6.

29. Das S, Forer L, Schönherr S, Sidore C, Locke AE, Kwong A, et al. Next-generation genotype imputation service and methods. Nat Genet. 2016;48(10):1284–7.

30. Jun G, Flickinger M, Hetrick KN, Romm JM, Doheny KF, Abecasis GR, et al. Detecting and estimating contamination of human DNA samples in sequencing and array-based genotype data. Am J Hum Genet. 2012 Nov 2;91(5):839–48.

31. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. Bioinforma Oxf Engl. 2012 Mar 15;28(6):882–3.

32. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinforma Oxf Engl. 2010 Jan 1;26(1):139–40.

33. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007 Sep;81(3):559–75.

34. Schmidt EM, Zhang J, Zhou W, Chen J, Mohlke KL, Chen YE, et al. GREGOR: evaluating global enrichment of trait-associated variants in epigenomic features using a systematic, data-driven approach. Bioinforma Oxf Engl. 2015 Aug 15;31(16):2601–6.

35. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell. 2010 May 28;38(4):576–89.
36. Tan G, Lenhard B. TFBSTools: an R/bioconductor package for transcription factor binding site analysis. Bioinforma Oxf Engl. 2016 15;32(10):1555–6.

37. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012 Mar 4;9(4):357–9.

38. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics. 2011 Aug 4;12:323.

39. Stitzel ML, Sethupathy P, Pearson DS, Chines PS, Song L, Erdos MR, et al. Global epigenomic analysis of primary human pancreatic islets provides insights into type 2 diabetes susceptibility loci. Cell Metab. 2010 Nov 3;12(5):443–55.

40. Scott LJ, Erdos MR, Huyghe JR, Welch RP, Beck AT, Wolford BN, et al. The genetic regulatory signature of type 2 diabetes in human skeletal muscle. Nat Commun. 2016;7:11764.

41. Ackermann AM, Wang Z, Schug J, Naji A, Kaestner KH. Integration of ATAC-seq and RNA-seq identifies human alpha cell and beta cell signature genes. Mol Metab. 2016 Mar;5(3):233–44.

42. Heintzman ND, Hon GC, Hawkins RD, Kheradpour P, Stark A, Harp LF, et al. Histone modifications at human enhancers reflect global cell-type-specific gene expression. Nature. 2009 May 7;459(7243):108–12.

43. Ward LD, Kellis M. HaploReg v4: systematic mining of putative causal variants, cell types, regulators and target genes for human complex traits and disease. Nucleic Acids Res. 2016 Jan 4;44(D1):D877-881.

44. Guo S, Dai C, Guo M, Taylor B, Harmon JS, Sander M, et al. Inactivation of specific β cell transcription factors in type 2 diabetes. J Clin Invest. 2013 Aug;123(8):3305–16.

45. Abebe T, Mahadevan J, Bogachus L, Hahn S, Black M, Oseid E, et al. Nrf2/antioxidant pathway mediates β cell self-repair after damage by high-fat diet-induced oxidative stress. JCI Insight. 2017 Dec 21;2(24).

46. Kondo K, Ishigaki Y, Gao J, Yamada T, Imai J, Sawada S, et al. Bach1 deficiency protects pancreatic β-cells from oxidative stress injury. Am J Physiol Endocrinol Metab. 2013 Sep 1;305(5):E641-648.

47. Gurzov EN, Barthson J, Marhfour I, Ortiz F, Naamane N, Igoillo-Esteve M, et al. Pancreatic β-cells activate a JunB/ATF3-dependent survival pathway during inflammation. Oncogene. 2012 Mar 29;31(13):1723–32.
48. Scott RA, Scott LJ, Mägi R, Marullo L, Gaulton KJ, Kaakinen M, et al. An Expanded Genome-Wide Association Study of Type 2 Diabetes in Europeans. Diabetes. 2017 May 31;

49. Wood AR, Jonsson A, Jackson AU, Wang N, van Leewen N, Palmer ND, et al. A Genome-Wide Association Study of IVGTT-Based Measures of First Phase Insulin Secretion Refines the Underlying Physiology of Type 2 Diabetes Variants. Diabetes. 2017 May 10;

50. Dimas AS, Lagou V, Barker A, Knowles JW, Mägi R, Hivert M-F, et al. Impact of type 2 diabetes susceptibility variants on quantitative glycemic traits reveals mechanistic heterogeneity. Diabetes. 2014 Jun;63(6):2158–71.

51. Roman TS, Cannon ME, Vadlamudi S, Buchkovich ML, Wolford BN, Welch RP, et al. A Type 2 Diabetes-Associated Functional Regulatory Variant in a Pancreatic Islet Enhancer at the Adcy5 Locus. Diabetes. 2017 Jul 6;

52. Thurner M, van de Bunt M, Torres JM, Mahajan A, Nylander V, Bennett AJ, et al. Integration of human pancreatic islet genomic data refines regulatory mechanisms at Type 2 Diabetes susceptibility loci. eLife. 2018 Feb 7;7.

53. Lyssenko V, Nagorny CLF, Erdos MR, Wierup N, Jonsson A, Spégel P, et al. Common variant in MTNR1B associated with increased risk of type 2 diabetes and impaired early insulin secretion. Nat Genet. 2009 Jan;41(1):82–8.

54. Banovich NE, Li YI, Raj A, Ward MC, Greenside P, Calderon D, et al. Impact of regulatory variation across human iPSCs and differentiated cells. Genome Res. 2018;28(1):122–31.

55. Nédélec Y, Sanz J, Baharian G, Szpiech ZA, Pacis A, Dumaine A, et al. Genetic Ancestry and Natural Selection Drive Population Differences in Immune Responses to Pathogens. Cell. 2016 Oct 20;167(3):657–669.e21.
Figure Legends

Figure 1. Human pancreatic islet chromatin accessibility profiles from 19 donors.

(A) UCSC Genome Browser view of ATAC-seq profiles at the NKK6.1 locus from 6 representative islet samples (both non-diabetic (ND) and T2D individuals), the lymphoblastoid cell line GM12878, CD4⁺ T cells, adipose tissue, and peripheral blood mononuclear cells (PBMC; data from 2 individuals). Orange and gray rectangles denote islet-specific or ubiquitously accessible regions, respectively, among cell types/tissues profiled. Green rectangles highlight regions showing variable accessibility between islet samples in the cohort. All chromatin accessibility profiles are normalized to their respective library size and have the same scale. Islet ChromHMM chromatin state annotations of these accessible sites (color code key found in Figure 1E), islet stretch enhancers (SEs), and RefSeq gene models are also shown. (B) Heatmap of Spearman correlation coefficients between ATAC-seq profiles from 19 islet samples and other cell types. Asterisks mark islet ATAC-seq samples from T2D donors (n=5). PBMC = peripheral blood mononuclear cells. (C) Transcription factor (TF) motif enrichments in open chromatin regions (OCRs) unique to islet samples (n=40,271) compared to islet OCVs that are also detected in skeletal muscle, adipose, GM12878, CD4T or PBMCs (n=41,639). TFs are clustered with respect to the similarity of their position weight matrices (PWMs). Motif logos are shown for TFs highlighted with maroon bars. (D) Histogram of the number of times an ATAC-seq OCR is detected in the cohort, ranging from individual-specific OCVRs (n=1) to shared OCVRs (n=19). (E) Stacked bar plot showing islet ChromHMM chromatin state annotations of OCVRs, binned according to the number of times an ATAC-seq OCR is detected in the cohort. Note that common OCVRs predominantly overlap promoter states, whereas individual-specific OCVRs overlap mostly unannotated (i.e., quiescent/low signal) regions.
Figure 2. Chromatin accessibility quantitative trait locus (caQTL) analysis identifies genetic variants affecting human islet cis-regulatory element use. (A) Schematic depicting genotype effects on chromatin accessibility detected by caQTL analyses. (B) Average ATAC-seq read counts of islet samples with ‘CC’ (blue), ‘CT’ (pink), or ‘TT’ (green) genotypes at rs488797, an islet caQTL overlapping an islet stretch enhancer (SE) within an intron of CELF4. The fraction of ATAC-seq reads overlapping rs488797 that contain the opening ‘T’ allele in ‘CT’ heterozygous islet samples (n=11) is plotted in the inset. The rs488797 ‘C’ allele is predicted to disrupt a FOXA2 binding motif (logo shown below the reference genome sequence), which is consistent with reduced chromatin accessibility observed for the ‘C’ allele. Average read counts for islet samples with the ‘TT’ genotype is 50.5 at the OCR summit. Islet samples with ‘CT’ or ‘CC’ genotype exhibited maximum average read counts of 32.36 and 6.5, respectively. Islet ChromHMM chromatin states, islet SEs, and RefSeq gene models are displayed as in Figure 1. hg19 chromosome coordinates: chr18:34969218-34972156. (C) UCSC genome browser view of FOXA2 ChIP-seq profiles (7) at the CELF4 locus for islets from two individuals (HI101, HI32). FOXA2 ChIP-seq read pileups are shown for the islet caQTL SNP (rs488797; gray rectangle) and a nearby SNP (rs648005; orange rectangle) in high LD ($r^2=0.99$), suggesting that the rs648005/rs488797 genotypes are ‘TC’/’CT’ for HI101 and ‘TT’/’CC’ for HI32. No FOXA2 binding is observed at rs488797 in HI32, whose ‘CC’ genotype is predicted to disrupt the FOXA2 binding motif on both parental chromosomes. In HI101, which is heterozygous at rs488797, all FOXA2 ChIP-seq reads contained the ‘T’ allele, supporting predictions that the ‘C’ allele disrupts FOXA2 binding. (D) TF motifs significantly enriched in islet caQTL OCRs. TFs are clustered based on their position weight matrix (PWM) similarity using hierarchical clustering, resulting in four major TF groups. Bar plots of p-values are color-coded according to this clustering. A representative motif logo is shown for each cluster. Asterisks mark the TF that
corresponds to depicted motif logos. (E) Luciferase reporter activity in MIN6 beta cells of sequences containing human islet caQTL alleles at selected loci. Plots show the ratio of luciferase activity of the more accessible, ‘open’ allele relative to the less accessible, ‘closed’ allele. Dashed red line indicates balanced activity of caQTL alleles. Error bars represent standard error of the mean (SEM). **** and *** indicate p<0.0001 and p<0.001, respectively, according to two-sided Mann-Whitney test; n.s. = not significant. Three plasmid preparations were tested for each sequence on three separate occasions. (F) QQ plot of observed (y-axis) vs. expected (x-axis) islet eQTL (eQTLs from 19 individuals in this study) association p-values for islet caQTL SNPs (black) or randomly selected non-caQTL SNPs (blue). Higher enrichment of eQTLs among statistically significant caQTLs links regulation of chromatin accessibility to gene expression.

Figure 3. Chromatin accessibility differences between T2D and non-diabetic (ND) islet samples. (A) T2D disease state-associated chromatin accessibility changes. Heatmap represents normalized chromatin accessibility levels at differentially accessible sites (FDR 10%). (B) UCSC Genome Browser view around the BHLHE41 locus highlighting an example closing OCR in T2D islet samples. (C) Islet ChromHMM chromatin state annotations of all islet OCRs (n=52,387) and differentially accessible OCRs (n=1515), further separated into closing (n=801) or opening (n=714) OCRs. Note that closing and opening OCRs predominantly overlap islet enhancer and promoter states, respectively. (D) Plot showing the fraction of non-diabetic (ND; x-axis) and T2D (y-axis) islet samples that have OCRs detected at differentially accessible regions, demonstrating that the majority of accessibility changes in T2D islet samples are quantitative in nature. The size of each pie represents the number of differential OCRs for that category. Pie sizes are listed for the rightmost column. Pink wedges indicate the proportion of T2D disease state-associated differential OCRs that are also islet caQTLs.
Asterisk denotes the group that contains the “opening” OCR shown in panel E. (E) T2D “opening” OCR that is also an islet caQTL. Average chromatin accessibility of all 19 islet samples at the T2D-associated TSPAN8 locus, stratified by rs1463768 genotype. Average read counts for islet samples with the ‘GG’ genotype is 97.33 at the OCR summit. Islet samples with ‘TG’ or ‘TT’ genotypes exhibited maximum average read counts of 67.75 and 29.125, respectively. Left inset shows the fraction of ATAC-seq reads containing the ‘G’ allele for each of the heterozygous islet samples (n=8). Right inset shows chromatin accessibility of the 5 ND and 5 T2D islet samples used in the differential OCR analysis, stratified by rs1463768 genotype. ChromHMM=chromatin states. hg19 coordinates: chr12:71586245-71591030.

**Figure 4. GWAS SNP enrichment in islet caQTLs.** (A) Disease- and trait-associated GWAS SNP enrichment in islet caQTLs. Enrichment (x-axis; observed/expected number of disease SNPs) and significance (y-axis) of GWAS SNP-islet caQTL overlaps are plotted. Red dots indicate significantly enriched diseases/traits at FDR 5% (after correcting for multiple hypothesis testing; 184 GWAS catalog diseases and traits tested). (B) Table showing the T2D-associated GWAS index or linked (r^2>0.8) SNP overlapping islet caQTLs. Asterisks mark sequence variants tested for allelic effects on luciferase activity shown in Figure 2E. The ‘eQTL allele’ refers to the allele linked to higher gene expression in islets (8,11,12). Reported pairwise SNP correlations (r^2 values) are based on European populations (EUR). (C) Average chromatin accessibility in islet samples stratified by genotype at rs11708067 in the ADCY5 locus. The inset shows the fraction of ATAC-seq reads containing the rs11708067 ‘G’ allele in each of the heterozygous islet samples (n=5). This is a putative loss-of-function T2D-associated caQTL, in which the T2D risk allele ‘A’ at rs11708067 is associated with lower chromatin accessibility in islets, as well as lower gene expression levels. Average read counts for islet samples
with the ‘GG’ genotype is 44 at the OCR summit. Islet samples with ‘AG’ or ‘AA’ genotypes exhibit maximum average read counts of 24.4 or 10.08, respectively. hg19 coordinates: chr3:123062482-123067947. (D) Average chromatin accessibility in islet samples stratified by genotype at rs6937795 in the *IL20RA* locus. The fraction of ATAC-seq reads containing the ‘C’ allele in each of the heterozygous islet samples (n=11) is plotted in the inset. This is an example of a gain-of-function T2D-associated caQTL, in which the T2D risk allele is associated with higher chromatin accessibility at this OCR. Average read counts for islet samples with the ‘AA’ genotype is 40.33 at the OCR summit. Islet samples with ‘AC’ or ‘CC’ genotypes exhibited maximum average read counts of 22.81 or 19.6, respectively. hg19 coordinates for zoomed-in view of ATAC-seq average read counts: chr6:137289071-137292315; hg19 coordinates for ChromHMM chromatin state, islet SE, and RefSeq Gene models: chr6:137277485-137324778.
A. Example locus around *NKH6.1*

B. ATAC-seq library correlations

C. TF motif enrichment in islet-specific OCRs

D. Frequency of ATAC-seq OCR calls in the cohort

E. Annotation of ATAC-seq OCRs in the cohort
Diabetes

A

SNP in ATAC-seq OCR
T
T
T
T
C
C
C
C

Chromatin Accessibility QTL (caQTL)

B

Example islet caQTL

TCA[GTAAATACT]TAC
rs488797

FOXA_Disc2

TT (n=2)
CT (n=11)
CC (n=6)

Fraction of reads with 'T' allele in heterozygous individuals

C

FOXA2 ChIP-seq

HI101
FOXA2

5 kb

hg19

rs648005 (T:16/C:16)
rs488797 (C:0/T:11)

HI32
FOXA2

rs648005 (T:52/C:0)
rs488797 (C:0/T:0)

CAQTl
Islet OCR
ChromHMM
Islet SEs

C/CELF

D

TF motif enrichments in caQTL OCRs

PWM clustering

E

Fold change in Luciferase Activity

Fold Change of Luciferase Activity

Open/Closed Allele

n.s.

F

QQ plot for 19 islet cohort caQTL-eQTL overlaps

Observed -log10(pvalue)

Expected -log10(pvalue)

caQTL SNPs
Non-caQTL SNPs
A. T2D disease state-associated changes

Chromatin accessibility changes at the TSPAN8 locus

B. Example closing OCR

Chromatin state annotations

C. Quantitative changes at differential OCRs

D. Fraction of ND islet samples with OCRs detected

E. Fraction of T2D islet samples with OCRs detected
### A GWAS Disease SNP enrichments at caQTLs

![Graph showing enrichment of GWAS Disease SNPs at caQTLs.](image)

### B caQTLs linked to T2D-associated SNPs

| Gene          | Index SNP   | caQTL SNP     | $r^2$ | caQTL Alleles | Risk Allele | Accessible Allele | eQTL Allele |
|---------------|-------------|---------------|------|---------------|-------------|------------------|-------------|
| ADCY5         | rs11717195  | rs11708067    | 0.94 | A/G           | A           | G                | G           |
| PLEKHA1*      | rs10510110  | rs2421016     | 0.99 | C/T           | T           | C                | NA          |
| IGF2BP2       | rs1470579   | rs10428126    | 1.0  | T/C           | T           | T                | NA          |
| LOC100507477  | rs642858    | rs9376483     | 0.92 | C/T           | C           | T                | NA          |
| ROR2*         | rs1873747   | rs7855529     | 1.0  | T/C           | C           | T                | NA          |
| EVADR*        | rs1048886   | rs16869158    | 0.81 | A/T           | A           | T                | NA          |
| IL20RA*       | rs6937795   | rs6937795     | 1.0  | A/C           | A           | A                | NA          |
| INS*          | rs3842770   | rs1154020895  | 1.0  | G/A           | A           | A                | NA          |
| RNF6          | rs10507349  | rs34584161    | 0.93 | A/G           | A           | A                | A           |
| CDKN2B*       | rs2383208   | rs10811661    | 0.95 | T/C           | T           | T                | NA          |
| ZMIZ1*        | rs12571751  | rs703977      | 0.98 | T/G           | T           | T                | T           |
| MTNR1B        | rs10830963  | rs10830963    | 1.0  | C/G           | G           | G                | G           |
| IRS1          | rs2943640   | rs2943656     | 0.85 | A/G           | G           | G                | NA          |

### C Loss-of-function T2D caQTL

![Bar chart showing fraction of reads with G allele in heterozygous individuals.](image)

### D Gain-of-function T2D caQTL

![Bar chart showing fraction of reads with C allele in heterozyzous individuals.](image)
**Supplementary Materials:**

**Supplementary Table List** (All supplementary tables are available via link below)

https://github.com/UcarLab/DiabetesSubmission/tree/master/SupplementaryTables

Table S1. Human islet donor demographic characteristics and islet metadata

Table S2. ATAC-seq quality control metrics for the 19 islets

Table S3. Meta data and associated statistics of the 10 islets used for differential analyses

Table S4. RNA-seq quality control metrics for the 19 islets

Table S5. Constructs for luciferase assay

Table S6. 154,437 OCRs in 19 islets

Table S7. Statistically significant islet caQTLs

Table S8. HaploReg Predictions for TF motifs disrupted by islet caQTLs

Table S9. Differentially accessible ATAC-seq peaks in T2D islets

Table S10. Differentially expressed genes in T2D islets
Figure S1. Chromatin accessibility profiles of human islet samples. (A) Insert size distributions of six representative islet samples (4 ND, 2 T2D). ATAC-seq libraries capture nucleosome free and mono-, di-nucleosomal regions. (B) Pairwise Spearman’s correlations between ATAC-seq read distributions of islet samples and other cell types ((1–3); Kursawe and Stitzel, unpublished data). ‘Ackerman_Alpha’ and ‘Ackerman_Beta’
represent ATAC-seq profiles of FACS-enriched islet alpha and beta cell populations (4). PBMC = peripheral blood mononuclear cells. (C) Stacked bar plot showing percent overlap of islet OCRs (n=154,437) with ChromHMM chromatin states in islets and 30 other cell types. Tissues are ordered based on the overlap of OCRs with the enhancer states. NHLF = normal human lung fibroblasts; HMEC = human mammary epithelial cells; hASC = human adipose derived stem cells; t1-4 denote adipogenesis stages as reported (5); HUVEC = human umbilical vein cells; NHEK = normal human epidermal keratinocytes; HSMM = human skeletal muscle cells; ES-HUES = human embryonic stem cells; CD34-PB = peripheral blood (PB) CD34+ cells; H1 = human embryonic stem cells; K562 = immortalized myelogenous leukemia. TSS = Transcription Start Site; Rep = repressed. (D) Fraction of unique tissue-specific stretch enhancers (SEs) overlapping one or more islet OCRs. Fisher’s exact test p-values are shown; n.s. = not significant. Cell type abbreviations are as described in S1C above.
Figure S2. Islet chromatin accessibility QTLs. (A) Quantile-quantile (QQ) plot of observed (y-axis) vs. expected (x-axis) association p-values between islet chromatin accessibility and genotype. (B) Location of significant caQTLs (marked in green) across
the autosomal chromosomes. (C) Density scatter plot showing the distance (x-axis; kb=kilobases) between 2949 islet caQTLs (y-axis; FDR 10%) and the transcription start site (TSS) of the nearest islet-expressed gene. The majority (>98.5%) of caQTLs are within the flanking 200 kb of the TSS of the nearest expressed gene. (D) Chromatin state annotations of caQTLs in islets vs. other tissues. Tissues are sorted from lowest to highest overlap between OCRs and the 'Quiescent/Low Signal' state. Note the enrichment of islet enhancer states in caQTLs. (E) Percent of tissue-specific SEs overlapping one or more islet caQTL. Fisher’s exact test p-values are shown; n.s. = not significant. Cell type abbreviations are as described in Figure S1C. (F) Table showing the overlaps of islet TF binding sites (i.e., ChIP-seq peaks) with all islet OCRs and caQTL-containing islet OCRs. Fisher’s exact test was used to calculate the significance of overlaps with the caQTL-containing islet OCRs. (G) QQ plot of observed (y-axis) vs. expected (x-axis) islet eQTL (eQTL from 112 individuals; (6)) association p-values for islet caQTL SNPs (black) or randomly selected non-caQTL SNPs (blue). Higher enrichment of eQTLs among statistically significant caQTLs links regulation of chromatin accessibility to gene expression. (H) Direction-of-effect between caQTL-eQTL pairs in 19 islet cohort. 37/44 islet caQTL-eQTL pairs (84%) show concordant changes (Q1 and Q3) in direction-of-effect on chromatin accessibility (x-axis) and gene expression (y-axis) (Pearson coefficient r=0.691; p=1.992e-07). Effect size < 0.5 signifies that the reference allele is more accessible/expressed than the alternate allele.
Figure S3. T2D disease state-associated chromatin accessibility changes. (A) Schematic of comparative analyses of five T2D and five ND islet samples. (B) The weighted average proportion variance explained by meta-variables available from the 10 islet samples before and after batch correction using Surrogate Variable Analysis (SVA) (7). Note that SVA reduces the variance attributed to all meta-variables, except for the disease state (ND or T2D). (C) MA plot for differential OCR analyses (at FDR 10%). Every dot represents an OCR considered in the analyses (n=52,387). Positive log fold change (logFC) indicates that the chromatin is opening in T2D islet samples, and...
negative logFC indicates that the chromatin is closing in T2D islet samples (CPM=counts per million). Differential OCRs are red; non-differential ones are shown in blue. (D) Transcription factor (TF) motif enrichment p-values and q-values for opening and closing OCRs. Mutually exclusive lists of TFs are enriched in closing and opening OCRs. (E) Gene expression changes (measured in log fold change of read counts from RNA-seq profiles) at promoters with OCRs (n=7499). Genes with opening OCRs exhibit positive fold change, i.e., increased expression in T2D islet samples; genes with closing OCRs show negative fold change, i.e., decreased expression in T2D islet samples (p=3.186e-06, Welch’s Two Sample two-sided t-test). (F) Venn diagram showing the number of differentially accessible OCRs overlapping islet caQTLs. 90/1515 (5.9%) of differential OCRs overlap islet caQTLs.
Figure S4. Chromatin accessibility QTLs in islets. (A) Lymphoblastoid cell line (LCL) DNase I sensitivity QTLs (dsQTLs; (8)) are enriched for GWAS SNPs associated with diabetes.
immune-related diseases (e.g., Inflammatory Bowel Disease, Crohn’s Disease, lupus, Type 1 Diabetes) and not for those associated with T2D or fasting glucose as observed with islet caQTLs. Diseases/traits with GWAS SNPs significantly overlapping LCL dsQTLs at FDR 5% are highlighted red. (B) Allelic imbalance of ATAC-seq reads for T2D-associated islet caQTL SNPs in heterozygous islet donors. Dot-and-box plots show the fraction of total ATAC-seq reads containing the ‘opening’ allele for each heterozygous individual. caQTL SNP rsID for each corresponding locus is indicated in the table below. The dashed red line at y=0.5 denotes equal representation of each allele. (C-D) Chromatin accessibility of 19 islet samples stratified by genotypes at caQTL SNP rs11708067 in the ADCY5 locus (C) and at rs6937795 in the IL20RA locus (D). Note that T2D islet samples (highlighted in red) are distributed across all three genotypes at each locus.

Supplemental References Cited:

1. Scott LJ, Erdos MR, Huyghe JR, Welch RP, Beck AT, Wolford BN, et al. The genetic regulatory signature of type 2 diabetes in human skeletal muscle. Nat Commun. 2016;7:11764.

2. Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat Methods. 2013 Dec;10(12):1213–8.

3. Ucar D, Márquez EJ, Chung C-H, Marches R, Rossi RJ, Uyar A, et al. The chromatin accessibility signature of human immune aging stems from CD8(+) T cells. J Exp Med. 2017 Oct 2;214(10):3123–44.

4. Ackermann AM, Wang Z, Schug J, Naji A, Kaestner KH. Integration of ATAC-seq and RNA-seq identifies human alpha cell and beta cell signature genes. Mol Metab. 2016 Mar;5(3):233–44.

5. Mikkelsen TS, Xu Z, Zhang X, Wang L, Gimble JM, Lander ES, et al. Comparative epigenomic analysis of murine and human adipogenesis. Cell. 2010 Oct 1;143(1):156–69.
6. Varshney A, Scott LJ, Welch RP, Erdos MR, Chines PS, Narisu N, et al. Genetic regulatory signatures underlying islet gene expression and type 2 diabetes. Proc Natl Acad Sci U S A. 2017 Feb 13;

7. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. Bioinforma Oxf Engl. 2012 Mar 15;28(6):882–3.

8. Degner JF, Pai AA, Pique-Regi R, Veyrieras J-B, Gaffney DJ, Pickrell JK, et al. DNase I sensitivity QTLs are a major determinant of human expression variation. Nature. 2012 Feb 16;482(7385):390–4.