



# Type 2 Diabetes–Associated Genetic Variants Regulate Chromatin Accessibility in Human Islets

Shubham Khetan,<sup>1,2</sup> Romy Kursawe,<sup>1</sup> Ahrim Youn,<sup>1</sup> Nathan Lawlor,<sup>1</sup> Alexandria Jillette,<sup>1</sup> Eladio J. Marquez,<sup>1</sup> Duygu Ucar,<sup>1,2,3</sup> and Michael L. Stitzel<sup>1,2,3</sup>

*Diabetes* 2018;67:2466–2477 | <https://doi.org/10.2337/db18-0393>

**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 and 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 2,949 SNPs associated with in vivo cis-regulatory element 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.**

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 noncoding (~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 variants 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 (LD) with each other. QTL approaches have recently been applied in cell lines to link genetic variation to epigenomic changes, such as DNaseI sensitivity (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 without diabetes [ND] and 5 with 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 road map for

<sup>1</sup>The Jackson Laboratory for Genomic Medicine, Farmington, CT

<sup>2</sup>Department of Genetics and Genome Sciences, University of Connecticut, Farmington, CT

<sup>3</sup>Institute of Systems Genomics, University of Connecticut, Farmington, CT

Corresponding authors: Michael L. Stitzel, [michael.stitzel@jax.org](mailto:michael.stitzel@jax.org), and Duygu Ucar, [duygu.ucar@jax.org](mailto:duygu.ucar@jax.org).

Received 5 April 2018 and accepted 22 August 2018.

This article contains Supplementary Data online at <http://diabetes.diabetesjournals.org/lookup/suppl/doi:10.2337/db18-0393/-/DC1>.

© 2018 by the American Diabetes Association. Readers may use this article as long as the work is properly cited, the use is educational and not for profit, and the work is not altered. More information is available at <http://www.diabetesjournals.org/content/license>.

translating T2D-associated GWAS SNPs into functional molecular effects.

## RESEARCH DESIGN AND METHODS

### Study Subjects and Islet Culture

Fresh human cadaveric pancreatic islets were procured from Prodo Laboratories or the Integrated Islet Distribution Program (Supplementary Table 1) 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 Laboratories) and incubated overnight in a T-150 non-tissue culture–treated flask (VWR) at 37°C and 5% CO<sub>2</sub> 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 & 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 three-donor batches (corresponding to nine barcoded transposition reactions), and sequenced using 2 × 75 bp Illumina NextSeq 500 to an average depth of 62.6 (± 18.6) million paired-end reads per donor (Supplementary Table 2). 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 a false discovery rate (FDR) of 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 mappability regions (blacklisted regions available from <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 National Institutes of Health 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 transcription start

site (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 (SE) regions (6,8) were overlapped with islet OCRs and tested for enrichment using the Fisher exact test. For each tissue-specific test, the background set comprised SEs 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 (Supplementary Table 1). We mapped Illumina array probe sequences to the hg19 genome assembly using Burrows-Wheeler Aligner. 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 version 5 [27]) and phasing (Eagle version 2.3 [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 Kumasaka et al. (15) and used the first five principal components 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, and 2) then 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 five T2D and five ND individuals with the most comparable demographics (Supplementary Table 3). 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  $\geq 3$  islet donors and 2) present in  $\geq 1$  T2D and  $\geq 2$  ND (or  $\geq 1$  ND and  $\geq 2$  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 GWAS Catalog of GWAS index SNPs for 184 diseases/traits was retrieved on 19 January 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 ( $r^2 > 0.8$ , LD window size = 1 Mb, minimum neighbor number = 500) were enriched in islet caQTLs or differentially accessible OCRs.

### Transcription Factor Motif Enrichments

Homer (35) findMotifsGenome.pl script (parameters: hg19, -size given) was used to identify transcription factor (TF) motifs enriched in islet OCRs. We compared motifs in OCRs that are accessible only in islets ( $n = 40,271$  islet-specific OCRs) to motifs in OCRs that are also accessible in adipose, CD4<sup>+</sup> T, GM12878, or peripheral blood mononuclear cells (PBMCs) ( $n = 41,639$  shared/background OCRs) (Fig. 1C). Motifs enriched in caQTL-containing OCRs (Fig. 2D) were identified by comparing caQTL OCRs ( $n = 2,949$ ) 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 ( $n = 1,515$ ) 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; External RNA Controls Consortium (ERCC) Mix 1 or Mix 2 spike-ins were randomly added to each sample (Thermo Fisher, catalog #4456740) (Supplementary Table 4). RNA-seq from 10 islet samples used in differential analyses were sequenced together on an Illumina NextSeq 500 to minimize batch effects, whereas the remaining nine samples were sequenced on Illumina HiSeq 2500, each to an average sequencing depth of 87.2 ( $\pm 27.8$ ) million paired-end reads (Supplementary Table 4). 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 (Supplementary Table 3) were used for differential expression analysis. Expected read counts for autosomal genes with FPKM  $> 5$  in  $\geq 3$  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 surrogate variables ( $n = 1$ ) from SVA were used as covariates to minimize the impact of confounding factors on T2D disease state-associated gene expression changes.

### 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 kb flanking the gene body were tested. To minimize potential batch effects, we adopted the strategy described in Kumasaka et al. (15) and used the first four principal components, 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

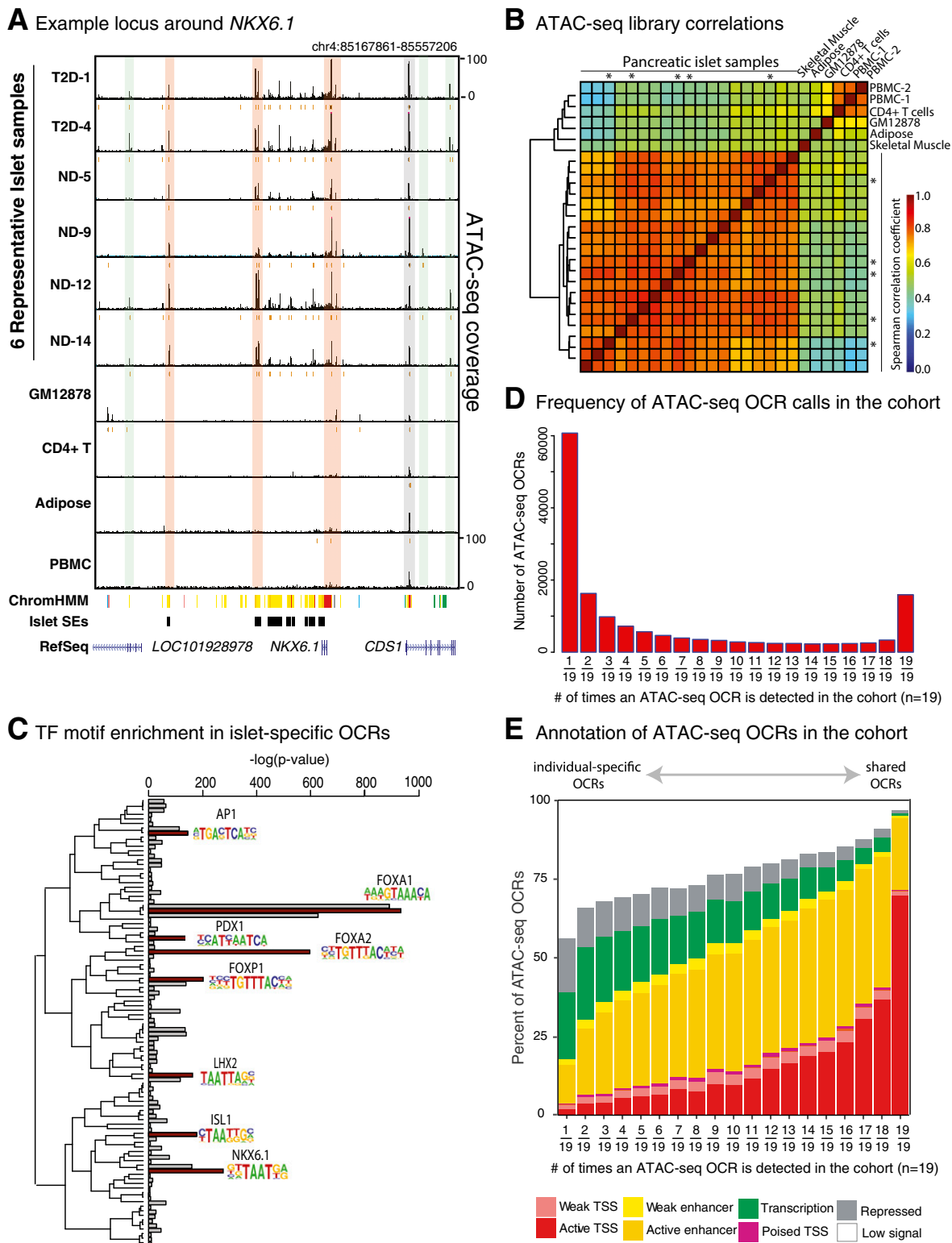
Quantile-quantile (QQ) plots for caQTL  $P$  values were generated against the expectation of a uniform  $P$  value distribution between 0 and 1. The QQ plot was generated for islet eQTL SNPs from 112 individuals (8) and caQTL SNPs from 19 individuals by conditioning on lead caQTL SNPs that were either statistically significant at FDR 10% or background sets of randomly selected nonsignificant ones. Random sets of nonsignificant SNPs ( $n = 2,545$ ) were generated 10 times to eliminate sampling bias; a representative result from one random set is shown in Fig. 2F and Supplementary Fig. 2G.

### 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 (Supplementary Table 5). 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 h prior to transfection as previously



**Figure 1**—Human pancreatic islet chromatin accessibility profiles from 19 donors. **A:** UCSC Genome Browser view of ATAC-seq profiles at the *NKX6.1* locus from six representative islet samples (ND and T2D individuals), the lymphoblastoid cell line GM12878, CD4<sup>+</sup> T cells, adipose tissue, and PBMCs (data from two 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 Fig. 1E), islet 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). **C:** TF motif enrichments in OCRs unique to islet samples (*n* = 40,271) compared with islet OCRs that are also detected in skeletal muscle, adipose tissue, GM12878, CD4<sup>+</sup> T cells, or PBMCs (*n* = 41,639). TFs are clustered with respect to the similarity of

described (39). Gateway-modified Firefly (0.072 pmol) (pGL4.23, Promega) plasmid containing each islet caQTL sequence (Supplementary Table 5) and 2 ng Renilla (pRL-TK, Promega) plasmid were cotransfected 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  $1 \times$  passive lysis buffer 38–40 h after transfection. Luminescence was measured on a Synergy 2 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*-REs in human islets, we generated high-quality ATAC-seq profiles from 19 islet donors (Supplementary Fig. 1A, Supplementary Tables 1 and 2). Investigating chromatin accessibility near the *NKX6.1* locus, a well-characterized  $\beta$ -cell-specific TF, revealed both OCRs unique to islet samples (Fig. 1A, orange and green rectangles) and OCRs shared with other cell types (22,40) (Fig. 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 (Fig. 1B and Supplementary Fig. 1B) (41). Notably, ATAC-seq profiles from T2D donors ( $n = 5$ ; Fig. 1B, asterisks) did not cluster separately from 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 (Supplementary Table 6). 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 (Supplementary Fig. 1C). Putative promoter OCRs were shared with several of 30 tissues profiled by the National Institutes of Health 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 SEs, which are 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 (Supplementary Fig. 1D), significantly greater than overlaps observed between islet OCRs and SEs in other tissues (Fisher exact test  $P < 2.2 \times 10^{-16}$ ). 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<sup>+</sup> T cells, skeletal muscle, or adipose tissues (Fig. 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 (Fig. 1D), which were overwhelmingly annotated as promoters (Fig. 1E, red bars). In contrast, 39.3% ( $n = 60,713$ ) of OCRs were detected in only 1 out of 19 individuals (Fig. 1D) and were found predominantly (45%) in quiescent/low signal chromatin states (Fig. 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, as references were based on data from 2–3 individuals. OCRs detected in 2–18 individuals (Fig. 1D) were mostly enhancers (Fig. 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

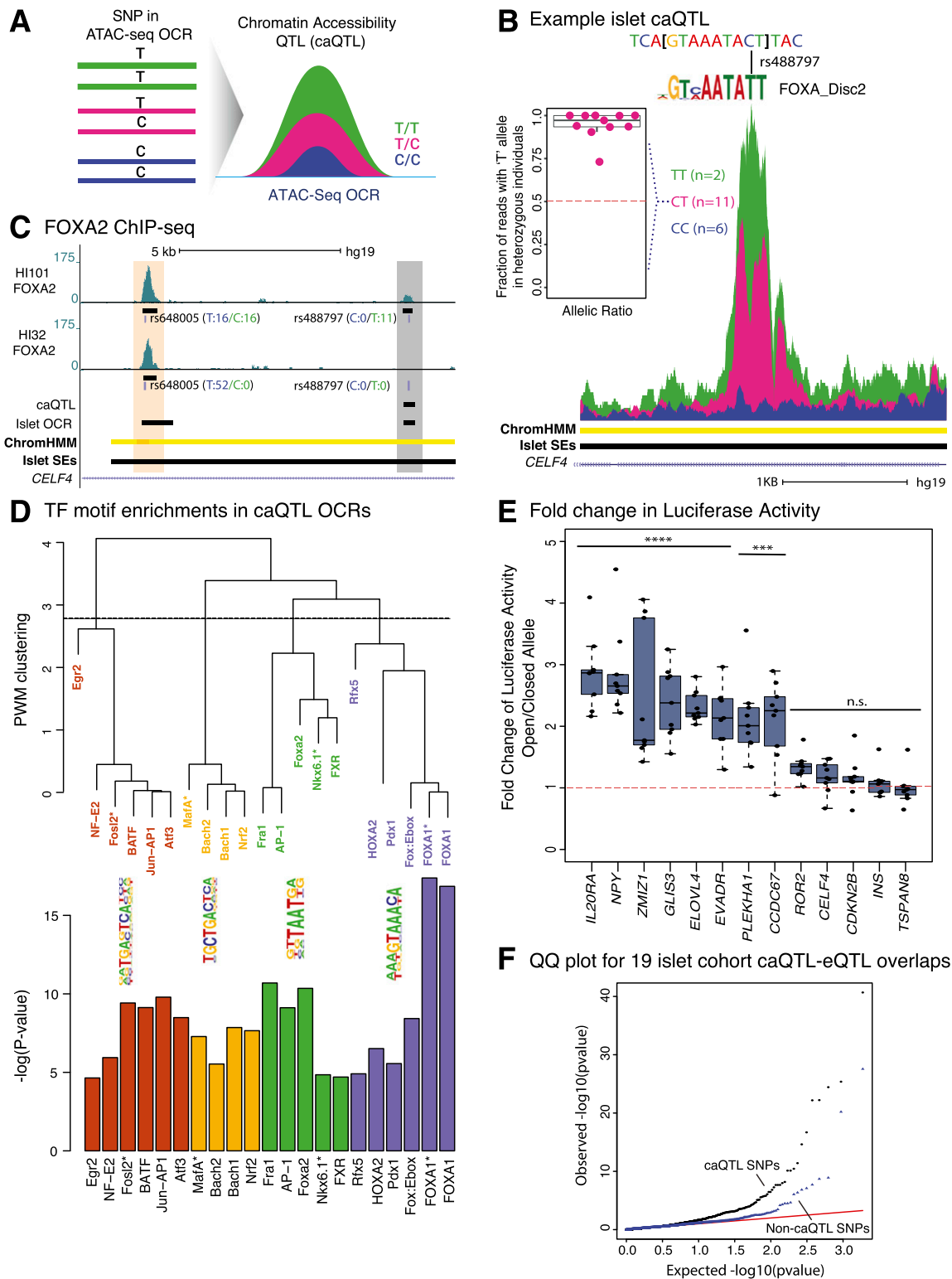
To identify genetic variants (SNPs and small insertions/deletions) that alter chromatin accessibility of islet OCRs in which they reside (Fig. 2A), we genotyped islet samples and conducted caQTL analysis using RASQUAL (15), a method that can discover QTLs from small sample sizes. Using RASQUAL, we identified 2,949 SNPs associated with increased or decreased chromatin accessibility at FDR 10% (Supplementary Fig. 2A, Supplementary Table 7) 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* (Fig. 2B), a gene selectively expressed in islets (8,40). CC homozygous islet donors exhibited dramatically lower accessibility than CT or TT genotypes (Fig. 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 (Fig. 2B, inset), reinforcing genetics as a strong determinant of chromatin accessibility at this OCR.

The rs488797 C allele is predicted to disrupt FOXA2 binding (Fig. 2B, compare top sequence between brackets to FOXA2 binding motif below). To test this, we analyzed FOXA2 ChIP-seq data from two islet donors (HI101 and HI32) (7) (Fig. 2C). We leveraged FOXA2 ChIP-seq reads and genetic LD to infer genotypes of these individuals in this region. As 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).

---

their 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 OCRs ( $n = 1$ ) to shared OCRs ( $n = 19$ ). E: Stacked bar plot showing islet ChromHMM chromatin state annotations of OCRs, binned according to the number of times an ATAC-seq OCR is detected in the cohort. Note that common OCRs predominantly overlap promoter states, whereas individual-specific OCRs overlap mostly unannotated (i.e., quiescent/low signal) regions.

---



**Figure 2**—caQTL analysis identifies genetic variants affecting human islet *cis*-RE 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 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 Fig. 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 overlaps a distinct OCR and a FOXA2 binding site 8,178 nucleotides away but is neither an islet caQTL nor is predicted to disrupt a FOXA2 motif. In HI101, FOXA2 ChIP-seq reads overlapping rs648005 contained both C and T alleles (Fig. 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. Supplementary Table 8 provides predicted motif disruptions from HaploReg (43) for all islet caQTL including rs488797.

Islet caQTLs were uniformly distributed across the autosomal chromosomes (Supplementary Fig. 2B), and the majority (>98%) were located within 200 kb flanking the TSS of the nearest islet-expressed gene (Supplementary Fig. 2C). Twelve percent of islet caQTLs were in promoters, whereas 30% overlapped enhancers (Supplementary Fig. 2D). Islet caQTLs were exclusively enriched in islet SEs compared with SEs in other tissues (Supplementary Fig. 2E). Finally, sequence motifs for islet-specific TFs, such as FOXA2, NKX6.1, and PDX1, were enriched in caQTL OCRs (Fig. 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 (Supplementary Fig. 2F), 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., Fig. 2B). We cloned DNA sequences containing each islet caQTL allele (Supplementary Table 5) and measured their enhancer activity using luciferase reporter assays in MIN6 mouse  $\beta$ -cells. We observed allelic effects on luciferase activity for 8 of the 13 caQTLs tested (Fig. 2E). Importantly, for all 8 caQTLs, the allele that increased *in vivo* chromatin accessibility also increased *in vitro* enhancer activity (Fig. 2E). Finally, we studied whether caQTL variants were also associated with variability in islet gene expression levels using islet eQTL data from this cohort ( $n = 19$ ). As shown in Fig. 2F, caQTL variants exhibited more significant allelic effects on islet gene expression than randomly selected variants in OCRs (noncaQTLs). We observed the same trend comparing these caQTLs to eQTLs detected in a larger independent cohort ( $n = 112$ ) (Supplementary Fig. 2G) (8). Importantly, for 84% of caQTL-eQTL pairs in our cohort (37/44) (Supplementary Fig. 2H), 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 (Supplementary Fig. 2H; Q1 and Q3), linking chromatin accessibility effects of these variants to downstream changes in islet gene expression.

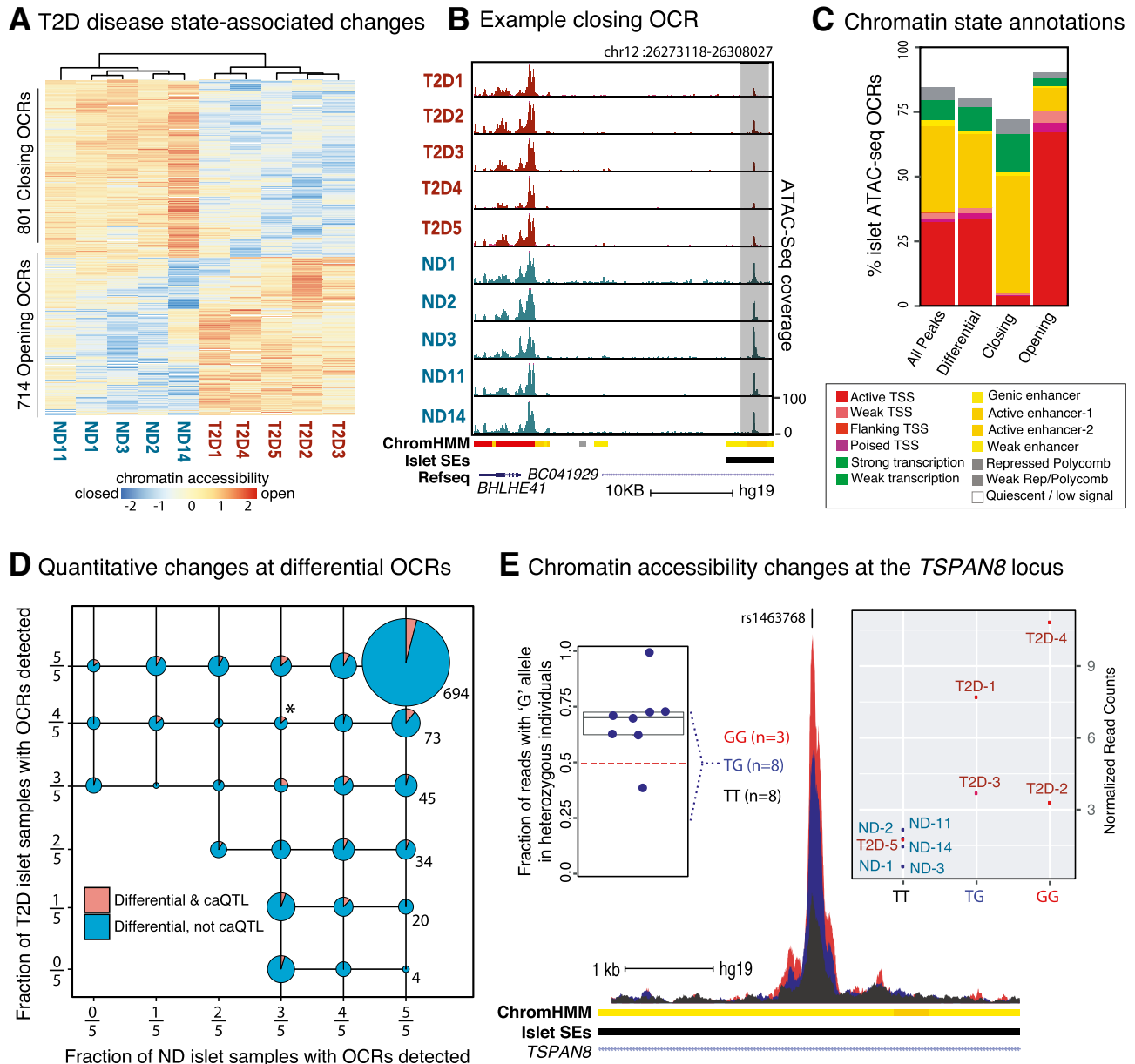
#### Chromatin Accessibility Changes in T2D Versus ND Islet Samples

To assess potential environmental effects of T2D disease state on the islet epigenome, we compared chromatin accessibility between five T2D donors and five ND donors with comparable demographics, e.g., age, race, sex) (Supplementary Fig. 3A, Supplementary Table 3). After completing SVA (31) to remove unwanted variation in the data, e.g., batch effect, sex, postmortem interval, drug treatment (Supplementary Fig. 3B), we identified 1,515 of 52,387 (2.8%) OCRs that were differentially accessible between T2D and ND islet samples at FDR 10% (see RESEARCH DESIGN AND METHODS, Fig. 3A, and Supplementary Fig. 3C; 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 with ND samples (Fig. 3A, Supplementary Table 9). There was a remarkable difference in the chromatin state annotation of opening and closing OCRs. Closing OCRs, e.g., the one highlighted near

---

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, who 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 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 are SEM. \*\*\*\* $P < 0.0001$ ; \*\*\* $P < 0.001$ , according to two-sided Mann-Whitney test. 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 noncaQTL SNPs (blue). Higher enrichment of eQTLs among statistically significant caQTLs links regulation of chromatin accessibility to gene expression. Red line denotes the line of equality ( $y = x$ ).

---



**Figure 3**—Chromatin accessibility differences between T2D and 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 of closing OCR in T2D islet samples. **C:** Islet ChromHMM chromatin state annotations of all islet OCRs ( $n = 52,387$ ) and differentially accessible OCRs ( $n = 1,515$ ), 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 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 five ND and five T2D islet samples used in the differential OCR analysis, stratified by rs1463768 genotype. hg19 coordinates: chr12:71586245–71591030.

*BHLHE41* (Fig. 3B, gray rectangle), mostly overlapped enhancers (48%), whereas opening OCRs extensively overlapped promoters (70%) (Fig. 3C). This difference was also reflected in TF motif enrichments, where opening and

closing OCRs were enriched for distinct motifs (Supplementary Fig. 3D). 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 (Fig. 3D). A total of 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 (Supplementary Fig. 3E). 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 (Supplementary Table 9), where only 120 and 54 genes were up- or downregulated, respectively, with T2D disease state at FDR 10% (Supplementary Table 10).

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) (Fig. 3D, Supplementary Fig. 3F), including the opening OCR that contains the caQTL variant rs1463768. Four of five T2D islet samples were heterozygous or homozygous for opening G allele for this variant, whereas all five ND donors were homozygous for the closing T allele (Fig. 3E). rs1463768-containing sequences did not show allelic differences in luciferase reporter activity in MIN6 cells (Fig. 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 noncoding 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 (see RESEARCH DESIGN 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 (Fig. 4A;  $P < 5.43 \times 10^{-04}$ , FDR 5%). In contrast, DNaseI sensitivity QTLs (13) in lymphoblastoid cell lines were enriched for mostly autoimmune disease-associated GWAS SNPs (Supplementary Fig. 4A), emphasizing the specificity of T2D-associated GWAS SNP enrichments in islet caQTLs.

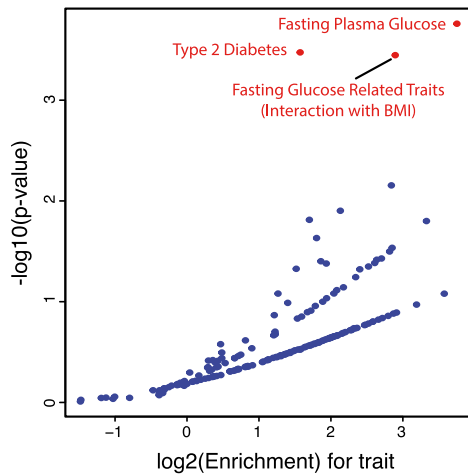
We identified SNPs in 13 T2D-associated loci that alter islet chromatin accessibility, thereby nominating these as putative causal/functional SNPs (Fig. 4B, Supplementary Fig. 4B). caQTL SNP alleles for 4 of 13 T2D-associated loci (*ADCY5*, *ZMIZ1*, *MTNR1B*, *RNF6*) were previously linked to altered 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 (Fig. 4B). For 6 of 13 T2D-associated caQTLs, the risk allele decreased chromatin accessibility, designated as loss of function (Fig. 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 (Fig. 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 ND human islets in vivo (12,51). The T2D risk allele was associated with increased chromatin accessibility for the remaining 7 of 13 islet caQTLs (Fig. 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 (Fig. 4D) and conferred 2.5-fold higher transcriptional reporter activity than the nonrisk C allele (Fig. 2E). Although 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 2,949 SNPs with altered in vivo chromatin accessibility. Allelic effects on in vivo chromatin accessibility correlated well with effects on in vitro enhancer activity; 8 of 13 caQTLs tested showed concordant allelic effects in luciferase reporter assays. Although we cannot eliminate the possibility of false-positive associations for the remaining 5 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.

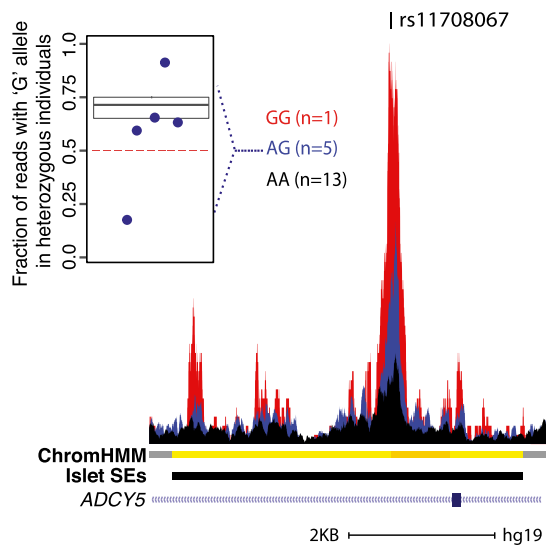
The data suggest that islet caQTL variants modulate regulatory programs important for islet cell identity and function. They were enriched in islet-specific TF motifs, TF ChIP-seq peaks (7), and islet SEs (6). They were specific to islets, as only 2.3% (68/2,949) of the islet caQTL variants altered chromatin accessibility in induced pluripotent stem cells 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., noncaQTL 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

**A** GWAS Disease SNP enrichments at caQTLs **B** caQTLs linked to T2D-associated SNPs

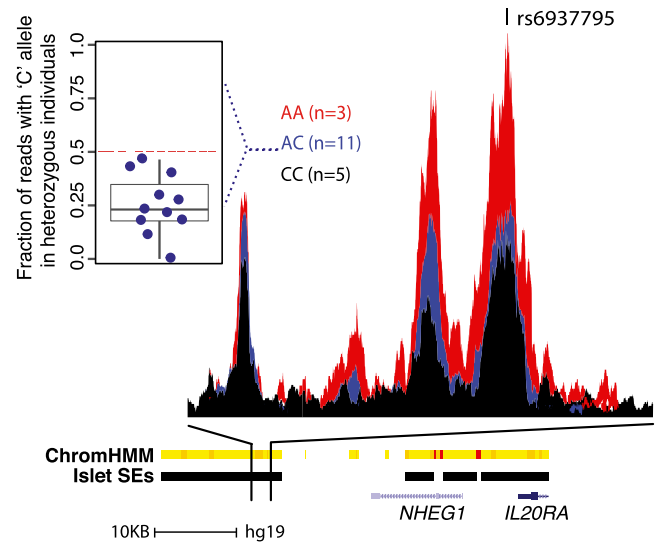


	Gene	Index SNP	caQTL SNP	r <sup>2</sup>	caQTL Alleles	Risk Allele	Accessible Allele	eQTL Allele
Loss-of-function	<i>ADCY5</i>	rs11717195	rs11708067	0.94	A/G	A	G	G
	<i>PLEKHA1*</i>	rs10510110	rs2421016	0.99	C/T	T	C	NA
	<i>IGF2BP2</i>	rs1470579	rs10428126	1	T/C	C	T	NA
	<i>LOC100507477</i>	rs642858	rs9376483	0.92	C/T	C	T	NA
	<i>ROR2*</i>	rs1873747	rs7855529	1	T/C	C	T	NA
	<i>EVADR*</i>	rs1048886	rs16869158	0.81	A/T	T	A	NA
Gain-of-function	<i>IL20RA*</i>	rs6937795	rs6937795	1	A/C	A	A	NA
	<i>INS*</i>	rs3842770	rs115420895	1	G/A	A	A	NA
	<i>RNF6</i>	rs10507349	rs34584161	0.93	A/G	A	A	A
	<i>CDKN2B*</i>	rs2383208	rs10811661	0.95	T/C	T	T	NA
	<i>ZMIZ1*</i>	rs12571751	rs703977	0.98	T/G	T	T	T
	<i>MTNR1B</i>	rs10830963	rs10830963	1	C/G	G	G	G
	<i>IRS1</i>	rs2943640	rs2943656	0.85	A/G	G	G	NA

**C** Loss-of-function T2D caQTL



**D** Gain-of-function T2D caQTL



**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 Fig. 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. **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 and 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.

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 HbA<sub>1c</sub> 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 and 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 linked 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.

**Acknowledgments.** The authors 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 National Institute of Diabetes and Digestive and Kidney Diseases–funded Integrated Islet Distribution Program at City of Hope (grant 2UC4DK098085). The authors thank Jane Cha, Jackson Laboratory for Genomic Medicine, for help generating artwork for the figures; members of the Stitzel and Ucar laboratories 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 the manuscript.

**Funding.** This study was made possible by generous financial support from The Jackson Laboratory startup funds (to M.L.S. and D.U.); the Doug Coleman Research Fund at The Jackson Laboratory; the National Institute of Diabetes and Digestive and Kidney Diseases under award number DK092251 (to M.L.S.); the Assistant Secretary of Defense for Health Affairs, through the Peer Reviewed Medical Research Program under award number W81XWH-16-1-0130 (to M.L.S.); and 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.** No other potential conflicts of interest relevant to this article were reported.

**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.J.M. 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.

**Data Availability.** The accession number for human islet ATAC-seq and RNA-seq data reported in this article is NCBI Sequence Read Archive: SRP117935.

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

## References

- Halban PA, Polonsky KS, Bowden DW, et al.  $\beta$ -Cell failure in type 2 diabetes: postulated mechanisms and prospects for prevention and treatment. *Diabetes Care* 2014;37:1751–1758
- Lawlor N, Khetan S, Ucar D, Stitzel ML. Genomics of islet (dys)function and type 2 diabetes. *Trends Genet* 2017;33:244–255
- Ashcroft FM, Rorsman P. Diabetes mellitus and the  $\beta$  cell: the last ten years. *Cell* 2012;148:1160–1171
- Fuchsberger C, Flannick J, Teslovich TM, et al. The genetic architecture of type 2 diabetes. *Nature* 2016;536:41–47
- Mohlke KL, Boehnke M. Recent advances in understanding the genetic architecture of type 2 diabetes. *Hum Mol Genet* 2015;24(R1):R85–R92
- Parker SCJ, Stitzel ML, Taylor DL, et al.; NISC Comparative Sequencing Program; National Institutes of Health Intramural Sequencing Center Comparative Sequencing Program Authors; NISC Comparative Sequencing Program Authors. Chromatin stretch enhancer states drive cell-specific gene regulation and harbor human disease risk variants. *Proc Natl Acad Sci U S A* 2013;110:17921–17926
- Pasquali L, Gaulton KJ, Rodríguez-Seguí SA, et al. Pancreatic islet enhancer clusters enriched in type 2 diabetes risk-associated variants. *Nat Genet* 2014;46:136–143
- Varshney A, Scott LJ, Welch RP, et al.; NISC Comparative Sequencing Program. Genetic regulatory signatures underlying islet gene expression and type 2 diabetes. *Proc Natl Acad Sci U S A* 2017;114:2301–2306
- Gaulton KJ, Ferreira T, Lee Y, et al.; DIABetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium. Genetic fine mapping and genomic annotation defines causal mechanisms at type 2 diabetes susceptibility loci. *Nat Genet* 2015;47:1415–1425
- GTEx Consortium. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* 2015;348:648–660
- Fadista J, Vikman P, Laakso EO, 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;111:13924–13929
- van de Bunt M, Manning Fox JE, Dai X, et al. Transcript expression data from human islets links regulatory signals from genome-wide association studies for type 2 diabetes and glycemic traits to their downstream effectors. *PLoS Genet* 2015;11:e1005694
- Degner JF, Pai AA, Pique-Regi R, et al. DNase I sensitivity QTLs are a major determinant of human expression variation. *Nature* 2012;482:390–394
- Gaffney DJ, McVicker G, Pai AA, et al. Controls of nucleosome positioning in the human genome. *PLoS Genet* 2012;8:e1003036
- Kumasaka N, Knights AJ, Gaffney DJ. Fine-mapping cellular QTLs with RASQUAL and ATAC-seq. *Nat Genet* 2016;48:206–213
- Alasoo K, Rodrigues J, Mukhopadhyay S, et al.; HIPSCI Consortium. Shared genetic effects on chromatin and gene expression indicate a role for enhancer priming in immune response. *Nat Genet* 2018;50:424–431
- McVicker G, van de Geijn B, Degner JF, et al. Identification of genetic variants that affect histone modifications in human cells. *Science* 2013;342:747–749
- 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;10:1213–1218
19. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 2014;30:2114–2120
20. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009;25:1754–1760
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;7:11959
22. Ucar D, Márquez EJ, Chung C-H, et al. The chromatin accessibility signature of human immune aging stems from CD8<sup>+</sup> T cells. *J Exp Med* 2017;214:3123–3144
23. Zhang Y, Liu T, Meyer CA, et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol* 2008;9:R137
24. Ross-Innes CS, Stark R, Teschendorff AE, et al. Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature* 2012;481:389–393
25. Kundaje A, Meuleman W, Ernst J, et al.; Roadmap Epigenomics Consortium. Integrative analysis of 111 reference human epigenomes. *Nature* 2015;518:317–330
26. Wickham H. *Ggplot2: Elegant Graphics for Data Analysis*. New York, Springer, 2009
27. Auton A, Brooks LD, Durbin RM, et al.; 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature* 2015;526:68–74
28. Loh P-R, Palamara PF, Price AL. Fast and accurate long-range phasing in a UK Biobank cohort. *Nat Genet* 2016;48:811–816
29. Das S, Forer L, Schönerr S, et al. Next-generation genotype imputation service and methods. *Nat Genet* 2016;48:1284–1287
30. Jun G, Flickinger M, Hetrick KN, et al. Detecting and estimating contamination of human DNA samples in sequencing and array-based genotype data. *Am J Hum Genet* 2012;91:839–848
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. *Bioinformatics* 2012;28:882–883
32. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 2010;26:139–140
33. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007;81:559–575
34. Schmidt EM, Zhang J, Zhou W, et al. GREGOR: evaluating global enrichment of trait-associated variants in epigenomic features using a systematic, data-driven approach. *Bioinformatics* 2015;31:2601–2606
35. Heinz S, Benner C, Spann N, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* 2010;38:576–589
36. Tan G, Lenhard B. TFBSTools: an R/bioconductor package for transcription factor binding site analysis. *Bioinformatics* 2016;32:1555–1556
37. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 2012;9:357–359
38. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 2011;12:323
39. Stitzel ML, Sethupathy P, Pearson DS, et al.; NISC Comparative Sequencing Program. Global epigenomic analysis of primary human pancreatic islets provides insights into type 2 diabetes susceptibility loci. *Cell Metab* 2010;12:443–455
40. Scott LJ, Erdos MR, Huyghe JR, 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;5:233–244
42. Heintzman ND, Hon GC, Hawkins RD, et al. Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* 2009;459:108–112
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;44:D877–D881
44. Guo S, Dai C, Guo M, et al. Inactivation of specific  $\beta$  cell transcription factors in type 2 diabetes. *J Clin Invest* 2013;123:3305–3316
45. Abebe T, Mahadevan J, Bogachus L, et al. Nrf2/antioxidant pathway mediates  $\beta$  cell self-repair after damage by high-fat diet-induced oxidative stress. *JCI Insight* 2017;2:92854
46. Kondo K, Ishigaki Y, Gao J, et al. Bach1 deficiency protects pancreatic  $\beta$ -cells from oxidative stress injury. *Am J Physiol Endocrinol Metab* 2013;305:E641–E648
47. Gurzov EN, Barthson J, Marhfour I, et al. Pancreatic  $\beta$ -cells activate a JunB/ATF3-dependent survival pathway during inflammation. *Oncogene* 2012;31:1723–1732
48. Scott RA, Scott LJ, Mägi R, et al. An expanded genome-wide association study of type 2 diabetes in Europeans. *Diabetes* 2017;66:2888–2902
49. Wood AR, Jonsson A, Jackson AU, 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;66:2296–2309
50. Dimas AS, Lagou V, Barker A, et al.; MAGIC Investigators. Impact of type 2 diabetes susceptibility variants on quantitative glycemic traits reveals mechanistic heterogeneity. *Diabetes* 2014;63:2158–2171
51. Roman TS, Cannon ME, Vadlamudi S, et al. A type 2 diabetes-associated functional regulatory variant in a pancreatic islet enhancer at the *Adcy5* locus. *Diabetes* 2017;66:2521–2530
52. Thurner M, van de Bunt M, Torres JM, et al. Integration of human pancreatic islet genomic data refines regulatory mechanisms at type 2 diabetes susceptibility loci. *eLife* 2018;7:7
53. Lyssenko V, Nagorny CLF, Erdos MR, et al. Common variant in *MTNR1B* associated with increased risk of type 2 diabetes and impaired early insulin secretion. *Nat Genet* 2009;41:82–88
54. Banovich NE, Li YI, Raj A, et al. Impact of regulatory variation across human iPSCs and differentiated cells. *Genome Res* 2018;28:122–131
55. Nédélec Y, Sanz J, Baharian G, et al. Genetic ancestry and natural selection drive population differences in immune responses to pathogens. *Cell* 2016;167:657–669.e21