**Single cell chromatin accessibility reveals pancreatic islet cell type- and state-specific regulatory programs of diabetes risk**

# Joshua Chiou\*,1, Chun Zeng\*,2,3, Zhang Cheng4, Jee Yun Han4, Michael Schlichting2,3, Serina Huang2, Jinzhao Wang2,3, Yinghui Sui2,3, Allison Deogaygay2, Mei-Lin Okino2, Yunjiang Qiu3, Ying Sun2, Parul Kudtarkar2, Rongxin Fang3, Sebastian Preissl4, Maike Sander\*\*,#,2,3,5, David Gorkin\*\*,#,4, Kyle J Gaulton\*\*,#,2,5

1. Biomedical Graduate Studies Program, University of California San Diego, La Jolla CA
2. Department of Pediatrics, Pediatric Diabetes Research Center, University of California San Diego, La Jolla CA
3. Department of Cellular and Molecular Medicine, University of California San Diego, La Jolla CA
4. Center for Epigenomics, University of California San Diego, La Jolla CA
5. Institute for Genomic Medicine, University of California San Diego, La Jolla CA

\* Authors contributed equally to this work

\*\* Authors jointly supervised this work

# Corresponding author:

Kyle Gaulton

9500 Gilman Drive, #0746

Department of Pediatrics

University of California San Diego

858-822-3640

kgaulton@ucsd.edu

David Gorkin

9500 Gilman Drive, #0767

Center for Epigenomics

University of California San Diego

858-534-7581

dgorkin@ucsd.edu

Maike Sander

9500 Gilman Drive, #0695

Pediatric Diabetes Research Center

Departments of Pediatrics & Cellular and Molecular Medicine

University of California San Diego

858-246-0843

masander@ucsd.edu

# Abstract

Genetic risk variants for complex, multifactorial diseases are enriched in *cis*-regulatory elements. Single cell epigenomic technologies create new opportunities to dissect cell type-specific mechanisms of risk variants, but data from disease-relevant human tissues are lacking. Here, we report accessible chromatin profiles from 14.2k pancreatic islet cells revealing 13 cell clusters, including multiple alpha, beta and delta cell clusters representing different hormone-producing and signal-responsive states. We catalog 244,236 islet cell type accessible chromatin sites, and transcription factors that define both lineage- and state-specific regulation. To integrate these data with GWAS, we developed a framework to measure the enrichment of genetic association within accessible chromatin profiles of single cells, which revealed heterogeneity within different beta cell states for fasting glucose level. We further used machine learning to predict the cell type-specific regulatory function of genetic variants, and used single cell co-accessibility to link sites to putative cell type-specific target genes. Through integrative analyses we localized 239 T2D risk signals to islet accessible chromatin sites and prioritized fine-mapped variants at these signals, using both predicted regulatory function and co-accessibility with putative target genes. At the *KCNQ1* locus, causal T2D variant rs231361 had predicted effects on an enhancer which had beta cell-specific, long-range co-accessibility with the insulin promoter, and deletion of this enhancer in human embryonic stem cell-derived beta cells reduced insulin gene and protein expression. To facilitate querying T2D risk variants for their islet cell type regulatory programs, we created a web resource available at www.t2depigenome.org. Our findings provide a cell type- and state-resolved map of gene regulation in human islet cells and demonstrate the power of single cell epigenomics in dissecting the mechanisms of complex disease risk.

**Introduction**

Gene regulatory programs are largely orchestrated by *cis* regulatory elements that direct the expression of genes in response to specific developmental and environmental cues. Genetic variants associated with disease by genome-wide association studies (GWAS) are highly enriched within putative *cis* regulatory elements1, highlighting the importance of regulatory sequence in mediating disease risk. The Assay for Transposase-Accessible Chromatin using Sequencing (ATAC-seq) provides a powerful tool to map regulatory elements by profiling regions “open” or “accessible” chromatin that are relatively devoid of packaging nucleosome particles2. However, the activity of regulatory elements is often restricted to specific cell types and/or cell states, limiting the ability of ATAC-seq and other “ensemble” (or “bulk”) epigenomic technologies to map regulatory elements in individual cell types within disease-relevant tissues. To overcome this limitation, approaches to obtain ATAC-seq profiles from individual nuclei (referred to below as single nucleus ATAC-seq, or snATAC-seq) have been developed that allow for the disaggregation of ensemble open chromatin maps from heterogenous samples into component cell types and subtypes3–5. These developments create new opportunities to dissect the molecular mechanisms that underlie genetic risk of disease. However, to date there are no reported snATAC-seq data from disease-relevant human tissues outside of hematopoietic lineages and *in vitro* models6–8.

Defining the regulatory programs of individual pancreatic islet cell types is critical for understanding islet function and the pathophysiology of diabetes. Islets are comprised of endocrine alpha, beta, delta and gamma cells that each secrete a different hormone to regulate metabolism and glucose homeostasis, in addition to regulating cross-talk between endocrine cell types9–11. Other cell types in the islets such as endothelial, glial, and immune cells also contribute to the modulation of islet cell function12–14. Importantly, no single endocrine cell type typically comprises >50% of the total islet mass15, limiting insights that can be gained from molecular assays of ‘ensemble’ islet tissue. Furthermore, single cell gene expression analysis has revealed different states within alpha and beta cells16–18 with gradients in expression patterns across different states17. Different cell states have unique molecular signatures and likely reflect distinct functional properties16,19,20. While several studies have generated epigenomic profiles of alpha and beta cells isolated using fluorescence-activated cell sorting (FACS)21,22, these studies have not profiled the complete breadth of endocrine and other cell types nor captured state-specific heterogeneity within cell types. Furthermore, these studies do not enable analyses that leverage the large number of independent observations provided by single cell assays, such as co-accessibility between chromatin sites across individual cells6.

Type 2 diabetes (T2D) is a multifactorial disease with a highly polygenic inheritance, where GWAS have identified over 400 risk loci that explain 18% of the estimated disease heritability23. Pancreatic islets are central to genetic risk of T2D, as evidenced by shared association between many T2D risk variants and quantitative measures of islet function24–26 and enrichment of T2D risk variants in ensemble islet regulatory sites27–31. However, resolution of islet regulatory programs at the level of individual cell types as well as different states within these cell types is necessary to understand the molecular mechanisms of T2D pathogenesis. In the present study, we map accessible chromatin profiles of individual islet cells by snATAC-seq to define the regulatory programs of islet cell types and cell states. By relating these profiles to genetic association for T2D and fasting glycemia, we identify heterogeneity in the effects of endocrine cells on glucose levels and T2D risk and transcription factors (TFs) that drive these effects. We annotate hundreds of fine-mapped T2D risk signals with islet cell type-specific regulatory programs, including variants with predicted regulatory effects and co-accessibility linking distal sites to gene promoters. At the *KCNQ1* locus, causal T2D variant rs231361 was predicted to affect an enhancer with beta cell-specific, long-range co-accessibility to the insulin promoter, and deletion of this enhancer hESC-derived beta cells affected insulin expression and protein levels. These data provide unprecedented resolution of regulatory programs in human islet cells and the molecular mechanisms of T2D, and provide a roadmap for using single cell epigenomic data to understand the genetic basis of complex disease.

# Results

**Islet snATAC-seq reveals 13 cell clusters with distinct regulatory landscapes**

To map the chromatin landscapes of single islet cells, we performed snATAC-seq on human pancreatic islets from three donors. We used a combinatorial barcoding snATAC-seq approach previously optimized by our group for use on tissues3,5 (**Supplementary Table 1, see Methods**). To confirm the quality of data from islet snATAC-seq assays we first analyzed the resulting sequence data as ensemble ATAC-seq by aggregating all high-quality mapped reads irrespective of barcode. Ensemble snATAC-seq data from all three islet samples showed the expected insert size distribution (**Supplementary Figure 1A**), strong enrichment of signal at transcription start sites (TSS) (**Supplementary Figure 1B**), and peak calls concordant with published ATAC-seq data from islet samples27,32–34 (**Supplementary Figure 1C**).

To obtain a collection of high-quality single cell profiles, we first filtered out individual cells with fewer than 1,000 reads (**Supplementary Figure 1D**), resulting in a total of 17,995 cells across the three samples. We then clustered accessible chromatin profiles from these cells, making key modifications to previous approaches which are briefly described here (**see Methods** for full details)5. First, as the inputs to clustering we used normalized read counts in 5kb sliding windows genome-wide rather than read counts within peak calls, reasoning that peak calls would be biased towards more common cell types. Second, we performed an initial round of clustering and quality control on a per-sample basis, which removed 2,709 cells in low read depth clusters. Third, prior to clustering cells across samples, we used the mutual nearest neighbors method35 to correct for variability across donors. Finally, we clustered all cells together and performed additional quality control, which removed a cluster without substantial representation from all donors (694 cells) and a cluster with low fraction of reads in peaks, aberrant read depth, and low intra-cluster similarity (192 cells). After clustering and filtering steps, we retained 14,239 cells which mapped in 13 clusters, all of which had consistent representation across samples, read depth profiles and other quality metrics (**Figure 1A, Supplementary Figure 2A-C**).

To determine the cell type represented by each cluster, following established approaches for assigning cluster identity5, we examined chromatin accessibility at the promoter region of the cognate hormone genes for endocrine cells and known marker genes for non-endocrine cell types. Based on these marker genes, we identified clusters representing beta (*INS-IGF2/*insulin), alpha (*GCG/*glucagon), delta (*SST/*somatostatin) and gamma (*PPY*/pancreatic polypeptide) cells (**Figure 1B,C**). We also identified clusters representing other pancreatic cell types including acinar (*REG1A*)36, immune (*PTPN22*)37, stellate (*PDGFRB*)37, glial (*CDH19*)38, and endothelial (*CD93*)39 cells (**Figure 1B**). We defined a broader set of marker genes for each cluster by identifying gene promoters with differential accessibility across clusters and retaining the top 100 differential gene promoters for each cluster (**see Methods, Supplementary Table 2**). To confirm the cell type identity of each cluster, we then compared promoter-accessibility marker genes to marker genes that we derived from re-analyzing a published islet scRNA-seq dataset17 (**see Methods, Supplementary Figure 3**). We observed highly specific correlations between marker genes of endocrine cell types in snATAC-seq and scRNA-seq (alpha 1/alpha: Spearman ρ=.43 and alpha 2/div. alpha ρ=.35; beta 1/beta ρ=.45 and beta 2/beta ρ=.47; delta 1/delta ρ=.47 and delta 2/delta ρ=.23; gamma/gamma ρ=.34). Of note, the multiple sub-clusters of alpha, beta, and delta cells we identified in snATAC-seq were each highly correlated with their respective cell type. We also observed correlations between exocrine marker genes in snATAC-seq and marker genes for both acinar and ductal cells in scRNA-seq, confirming that this cluster represents a mixture of both cell types (exocrine/acinar ρ=.49; exocrine/ductal ρ=.46), and correlations between other corresponding pancreatic cell types (endothelial 1/endothelial ρ=.52; endothelial 2/endothelial ρ=.51; immune/immune ρ=.43; stellate/ act. stellate ρ=.50; stellate/quies. stellate ρ=.59; glial/glial ρ=.26) (**Figure 1D**).

In order to characterize the regulatory programs of each cell type, we aggregated cells within each cluster and then called accessible chromatin sites using MACS2 (**see Methods**). In total we identified 244,236 accessible chromatin sites across the 13 clusters (**Supplementary Data 1).** For each cluster, both promoter-proximal (+/-500bp) and distal sites were generally well represented in ensemble islet ATAC-seq data (proximal: 97.32% (stellate) - 99.4% (delta 1); distal: 56.67% (immune) - 94.92% (alpha 2)), particularly for the more common cell types (**Supplementary Figure 2D,E**). Notably, accessible chromatin of alpha and beta cells from each donor was highly concordant with published bulk ATAC-seq of the corresponding FACS-sorted populations21,22 (**Supplementary Figure 2F**).

To understand the regulatory logic underlying cell type accessible chromatin, we next used chromVAR40 to identify TF sequence motifs from JASPAR41 enriched within the accessible chromatin regions of each individual cell. We then focused on 111 TF motifs with evidence for variability across islet cells (**see Methods, Supplementary Table 3)**. Analysis of TF enrichments averaged across cells in each cell type revealed distinct patterns of motif enrichment, many consistent with the known functions of specific TFs in islet cell types (**Figure 1E, Supplementary Table 3**). For example, the PDX1 motif was enriched specifically in beta (normalized enrichment=0.93) and delta (1.0) cells, in line with the role of PDX1 in the regulation of both cell types42. In contrast, MAF motifs were enriched in alpha (1.0) and beta cells (0.93), reflecting the role of MAF TFs in regulating both cell types43–45 (**Figure 1E**). We also identified motifs enriched across all endocrine cell types such as FOXA, as well as in non-endocrine cell types like IRF motifs for immune cells46 (1.0) and ETS motifs for endothelial cells47 (1.0) (**Figure 1E**). We then further performed hierarchical clustering of cell types based on their TF motif enrichment patterns. Consistent with analyses of single cell expression data36,37,48, the regulatory programs of beta and delta cells were closely related as were the programs of alpha and gamma cells (**Figure 1E**).

**Heterogeneity in islet endocrine cell accessible chromatin and regulatory programs**

A major strength of single-cell approaches is the ability to reveal heterogeneity within the same cell type. Indeed, our initial clustering showed that endocrine cell types (beta, alpha, and delta) appeared to further segregate into sub-clusters. To examine these sub-clusters, we determined gene promoter accessibility in each sub-cluster, and identified promoters with variable accessibility between sub-clusters (**see Methods, Supplementary Data 2**). We focused on alpha and beta cells, where cell numbers allowed for robust calculations. Notably, we found *INS* among the genes with most variable promoter accessibility between beta cell sub-clusters (*INS-IGF2* beta OR=5.05, Fisher’s exact P=3.98x10-37), leading us to rename these clusters for downstream analysis as INS-high beta and INS-low beta, respectively (**Figure 1B,C; Figure 2A**). Similarly, *GCG* promoter accessibility is strikingly variable between alpha 1 and alpha 2 (*GCG* alpha OR=3.30, P=4.68x10-25). Hence we renamed those sub-clusters GCG-high alpha and GCG-low alpha, respectively (**Figure 2A)**. Apart from *INS* and *GCG*, we found substantial overlap in the genes with both increased and decreased promoter accessibility in INS-high beta and GCG-high alpha states (**Figure 2A**). INS-high beta and GCG-high alpha cells both had increased promoter accessibility at genes involved in hormone secretion and glucose response such as *GCK, ABCC8, G6PC2* and *SLC30A8* (**Figure 2A**). Conversely, INS-low beta and GCG-low alpha cells both had increased promoter accessibility at genes involved in stress-induced signaling responses including *ATF3*, *FOSL1*, and *FOSL2* (**Figure 2A**). We note that similar sub-states and patterns were also evident in delta cells from our data, although the low overall cell number impedes deeper analysis in the current study (**Supplementary Figure 4A**). These data suggest differences between endocrine cell sub-clusters among genes involved in hormone production and stress-induced signaling responses, and point to an underlying commonality in the genes that govern state-specific functions across different endocrine cell types.

Cellular heterogeneity of islet gene expression and function has been reported previously16,17,49, and we therefore sought to explore whether the alpha and beta cell states observed here shared commonalities with these previously described patterns. To this end, we compared genes with significantly different promoter accessibility between sub-clusters (from **Figure 2A**) to gene sets describing beta cell heterogeneity from a previous scRNA-seq study17. That study also identified a beta cell sub-cluster (“β-sub.4”) with lower INS expression, and genes expressed in this β-sub.4 sub-cluster were associated with GO terms related to protein folding and ER stress. We found that β-sub.4 genes were significantly enriched among those with increased promoter accessibility in INS-low beta cells (**Figure 2B**). Interestingly β-sub.4 genes were also enriched among those with increased promoter accessibility in GCG-low alpha cells, reinforcing the similarity between endocrine cell sub-states (**Figure 2B**). We further performed gene set enrichment (GSEA) for different beta and alpha cell states in our data using biological terms from the Gene Ontology (GO). We observed enrichment of peptide hormone secretion and insulin secretion terms for INS-high beta cells as well as GCG-high alpha cells, whereas INS-low beta and GCG-low alpha cells were enriched for terms related to stress-response, insulin signaling, and cell cycle (**Figure 2B, Supplementary Table 4**).

The transcriptional programs driving heterogeneity in alpha and beta cell function are unknown, and therefore we determined TF sequence motifs differentially enriched across alpha and beta cell states. We again focused on 111 TF motifs with strongest evidence for variable enrichment across islet cells (**see Methods, Supplementary Figure 5A, Supplementary Table 5**). Among these 111 TF motifs there were clear patterns that distinguished both alpha and beta cells as well as different states within alpha and beta cells, which again revealed commonalities in state-specific heterogeneity across different endocrine cell types (**Figure 2C, Supplementary Figure 5B, Supplementary Figure 5C**). For example, motifs for multiple RFX family members were prominently enriched in both GCG-high alpha and INS-high beta cells, but not in GCG-low alpha or INS-low beta cells (e.g. RFX3 mean beta 1 enrich=.26, beta 2 enrich=-.62, P=3.5x10-158; mean alpha 1 enrich=.29, alpha 2 enrich=-.56, P=7.3x10-91) (**Figure 2C**). In contrast, motifs for FOS and JUN family members were prominently enriched in the hormone-low cells but not the hormone-high cells (e.g. FOS::JUN mean beta 1 enrich=-1.45, beta 2 enrich=4.50, P=4.7x10-307; mean alpha 1 enrich=-1.45, alpha 2 enrich=4.46, P=2.3x10-292) (**Figure 2C**). Again, we observed similar motif enrichment patterns between states in delta cells (**Supplementary Figure 5B, Supplementary Figure 5C**).

Analysis of single cells ordered along a trajectory has been used to examine gene regulatory programs as a continuum rather than as discrete or binary states6,17,50. To explore potential gradations in our data, we used Cicero6 to order alpha and beta cells along respective trajectories based on patterns of chromatin accessibility. We used cells with highest promoter accessibility at insulin (beta cells) or glucagon (alpha cells) as the root states for each trajectory (**see Methods**). We refer to the axis of these trajectories below as “pseudo-state” rather than the conventional “pseudo-time”, as the heterogeneity observed here describes ordered changes along different cell states as opposed to actual time. By this approach, we observed that cells existed along a gradient between the hormone-high and hormone-low states of alpha and beta cells, although we noted a discernable transition point within the trajectory (**Figure 2D, Supplementary Figure 4B**). These trajectories allowed us to examine both gene promoter accessibility and TF motif enrichment as a function of pseudo-state (**Figure 2D, Supplementary Figure 4C).** Here again, we found that enrichment of lineage marker genes and TF enrichments (e.g. RFX, NEUROD and NFATC) decreased along the trajectory from hormone-high to hormone-low cells, and enrichment for FOS/JUN family member motifs increased along the trajectory from hormone-high to hormone-low cells (**Figure 2D**). To identify specific TFs likely driving the enrichments for a given motif, we then correlated TF motif enrichments with promoter accessibility of TF genes in corresponding structural TF sub-families across cells along the trajectory. For example, FOS/JUN motif enrichment was correlated with promoter accessibility of *FOSL1, FOSL2* and *JUND* (**Figure 2E**), suggesting a role for these TFs in state-specific alpha and beta cell gene regulation.

Taken together, these results reveal striking patterns of heterogeneity in chromatin accessibility within endocrine cell types and provide transcriptional regulatory programs and factors driving this heterogeneity.

**Enrichment of islet cell type- and state-specific regulatory sequence for diabetes- and fasting glycemia-associated genetic variants**

Variants associated with complex diseases and physiological traits are enriched within *cis* regulatory sequences1,51. More specifically, genetic variants influencing diabetes and fasting glucose level are enriched in pancreatic islet regulatory elements28–30,52. However, these enrichments based on ensemble data obscure the potential role of islet cell type- and state-specific regulation in these traits. Using our islet cell type- and state-resolved accessible chromatin profiles, we sought to determine the enrichment of genetic variants associated with type 1 and 2 diabetes23,53 and diabetes-related quantitative phenotypes26,54–58 as well as other complex traits and disease for calibration59–66. We first determined the enrichment of variants in accessible chromatin sites for each islet cell type and state using stratified LD score regression67,68 (**see Methods**). We observed significant enrichment (FDR<.1) of fasting glucose (FG) level and T2D association for both INS-high and INS-low beta cell states (T2D beta 1 Z=4.45 q-value=.001, beta 2 Z=4.00 q=.004; FG beta 1 Z=3.93 q=.004, beta 2 Z=3.34 q=.027), as well as enrichment of body-mass index (BMI) for SST-high delta cells (delta 1 Z=3.50 q=.027) (**Figure 3A**). We also observed more nominally-significant enrichment (P<.01) of 2hr glucose level adjusted for BMI for both alpha cell states (alpha 1 Z=2.45 P=.007, alpha 2 Z=2.40 P=.008), and T2D and fasting proinsulin level for GCG-low alpha cells (PI: alpha 2 Z=2.64, P=.004; T2D: alpha 2, Z=2.40 P=.008), although these enrichments did not pass multiple test correction.

In these analyses, we again noted evidence for differences in enrichments between the high- and low-hormone states of endocrine cells. In order to further resolve the heterogeneity of genetic association enrichment patterns, we used a novel framework to test the enrichment of genetic association signal within accessible chromatin profiles of single cells (**see Methods**). We applied this approach to genetic association data for T2D and fasting glucose level, as well as major depressive disorder and systemic lupus erythematosus (lupus) to provide negative control traits not expected to show enrichment (**Figure 3B**). We observed marked heterogeneity among beta cells in enrichment estimates for fasting glucose-associated variants, whereby cells in the INS-high state had significantly stronger enrichment than cells in the INS-low state (INS-high median Z=2.42, INS-low median Z=1.13, P<2.2x10-16) (**Figure 3B**). We further examined beta cell heterogeneity by calculating the average enrichment estimates for cells binned across the beta cell ‘pseudo-state’ trajectory (see **Figure 2**). This revealed decreasing enrichment for fasting glucose-associated variation across pseudo-state from INS-high to INS-low beta cells (**Figure 3B**). Conversely, for T2D we observed enrichment for beta cells that was consistent across INS-high and INS-low beta cells, as well as across the beta cell pseudo-state trajectory (INS-high median Z=0.48, INS-low median Z=0.51, P=0.84) (**Figure 3B**). Interestingly, we also observed evidence for stronger T2D enrichment of cells in GCG-low alpha cells relative to GCG-high alpha cells (GCG-high median Z=.14, GCG-low median Z=.31, P=7.6x10-13) (**Figure 3B**), which was also evident in the LD-score regression results (see **Figure 3A**). By comparison, major depressive disorder and lupus showed no evidence for enrichment for beta cells (all median Z<.001) (**Figure 3B**). Illuminating the contribution of different cell states to specific phenotypes can then inform interpretation of association signals for these phenotypes. For example, at the *DGKB* locus associated with fasting glucose and T2D, fine-mapped variants overlapped a chromatin site with higher activity in INS-high beta cells and that decreases across the pseudo-state trajectory, suggesting that this site is mediating the association signal (**Figure 3C**).

Given our ability to map both genetic association and TF motif enrichments to single cells, we reasoned that joint analysis of these data could provide insights into the TFs and gene regulatory pathways through which genetic effects on these traits are mediated. We therefore correlated fasting glucose level and T2D enrichment Z-scores with TF motif enrichments from chromVAR40 (see **Figure 1E**),both across all 14.2k islet cells as well as just the 7.2k beta cells (**see Methods**). Across all 14.2k cells, we observed strong positive correlation between fasting glucose level and T2D enrichment and beta cell lineage-specifying TFs such as PDX1 (FG ρ=.22, P=5.27x10-151; T2D ρ=.13, P=4.26x10-50), PAX4 (FG ρ=.21, P=8.97x10-141; T2D ρ=.14, P=2.80x10-59), and NKX6-1 (FG ρ=.20, P=4.68x10-133; T2D ρ=.14, P=2.81x10-65), reflecting the enrichment of these traits for beta cell chromatin, and negative correlation with TFs regulating other islet cell types (**Figure 3D, Supplementary Table 6**). When then considering the 7.2k beta cells only, we observed strongest positive correlation between fasting glucose level enrichment and TF motifs enriched in INS-high beta cells (from **Figure 2**) such as NRL/MAF (ρ=.14, P=5.36x10-32), RFX5 (ρ=.12, P=2.58x10-24), FOXA1 (ρ=.11, P=5.41x10-19), and NEUROG2 (ρ=.09, P=4.89x10-15), and strongest negative correlation with low-INS beta cell enriched TF motifs such as JUND and ATF4 (JUND ρ=-.23, P=1.23x10-85, ATF4 ρ=-.12, P=1.18x10-23) (**Figure 3D, Supplementary Table 6**). For T2D, significant positive correlations included TF motifs that were also correlated with fasting glucose such as NKX6-1 (ρ=.050, P=2.16x10-5) as well as those without strong fasting glucose correlations such as CREB1 (ρ=.053, P=7.44x10-6) and ELK-, ETV- and NFY-family members (ELK1 ρ=.072, P=1.21x10-9, ETV1 ρ=.069, P=4.33x10-9, NFYA ρ=.073, P=1.72x10-9), while significant negative correlations included JUN/FOS-family motifs such as JUND (ρ=.069, P=5.35x10-6) (**Supplementary Figure 6, Supplementary Table 6**).

Together these data support that a beta cell state driven by both lineage- and state-determining TFs plays a prominent role in the regulation and genetic basis of fasting glucose, whereas beta cell states that are both shared and distinct from glucose level contribute to T2D risk.

**Islet cell type- and state-specific predictions of variant effects genome-wide on regulatory sequence**

Predicting the effects of non-coding genetic variants on regulatory activity remains a major challenge, in large part because the sequence vocabularies that encode regulatory function differ for a given cell type and/or state. Our cell type- and state-resolved accessible chromatin profiles provided an ideal opportunity to apply machine learning methods that model these regulatory vocabularies and use these models to predict the effects of genetic variants on putative regulatory sequences. We therefore used deltaSVM69 to predict the effects of genetic variants from the Haplotype Reference Consortium panel70 on chromatin accessibility in each endocrine cell type and cell state (**see Methods**). We identified 543,537 variants genome-wide in endocrine islet accessible chromatin sites with predicted allelic effects (FDR<.1) on at least one cell type or state, and between 128k-210k variants (9.1%-14.8% of tested variants) per cell type or state (**Figure 4A**). Among variants with predicted effects in alpha or beta cells, a small percentage (9.4%) had predicted effects at FDR<.1 in all alpha and beta cell states; however, there was substantial directional concordance between the effect allele of variants with significant predictions in one alpha/beta state and the effect allele in other states (93%-95.9%; all P<2.2x10-16).

To validate that our predictions captured true allelic effects on islet chromatin, we first compared alpha and beta cell predictions to allelic imbalance in chromatin accessibility measured directly from read count data at heterozygous variants in each sample (**see Methods**). We found positive and highly significant correlations between predicted allelic effects and allelic imbalance estimates for all alpha and beta cell states (alpha 1 Spearman .261, P=3.27x10-46, alpha 2 .225, P=4.38x10-10, beta 1 .285, P=1.13x10-53, beta 2 .297, P=2.28x10-40) (**Figure 4C**). We further validated five likely causal T2D variants identified in fine-mapping and predicted to have allelic effects on beta cell chromatin using gene reporter assays in the MIN6 beta cell line. In each case, the luciferase reporter assays showed significant allelic effects on beta cell enhancer activity that were also directionally consistent with the predicted effects (**Figure 4D**). We also compared our predictions to chromatin accessibility quantitative trait loci (caQTLs) previously identified in ensemble islet samples71. We observed highly significant enrichment of caQTLs among variants with predicted allelic effects on alpha or beta cells (obs.=38.8%, exp.=23.6%, Fisher’s exact P=1.64x10-66) (**Figure 4D).** When further sub-dividing predictions based on those with shared (all alpha and beta states), cell type-specific (alpha, beta) or state-specific (state 1, state 2) effects we observed significant enrichment of caQTLs among shared effect variants yet not for other categories (**Figure 4D**), suggesting that the islet caQTLs preferentially capture variants with shared effects across alpha and beta cells.

We sought to further characterize genetic variants predicted to have more cell type- and state-specific effects on islet chromatin. For each category of variants, we performed motif enrichment comparing sequences around the effect allele to the corresponding non-effect allele (**see Methods**). Variants with state-specific predicted effects tended to disrupt motifs for TF families such as NEUROD, FOXA, MAF and RFX for high-hormone states (-log10(P)=59.2, 56.0, 50.3, 20.6), and signaling-responsive TF families such as JUN/FOS and CREB for low-hormone states (-log10(P)=107.6, 46.8) (**Figure 4E**). Similarly, variants with alpha or beta cell-specific predicted effects tended to disrupt motifs for lineage-defining TF families including GATA for alpha cells (-log10(P)=24.8), and NKX6 and PDX1 for beta cells (-log10(P)=17.0, 13.0) (**Figure 4E**). Structurally-related TFs often have very similar sequence motifs, and thus in order to assign motifs to specific TFs we examined promoter-accessibility patterns of TFs within the structural TF family72 (**see Methods**). For example, among GATA family members only GATA6 had high promoter accessibility in alpha cells (alpha 1: 1.00, alpha 2: .97, beta 1: .21, beta 2: .13), suggesting that GATA6 binding is likely disrupted in alpha cells by regulatory variants affecting the GATA motif. Similarly, among NKX6 family members both NKX6-1 and NKX6-3had promoter accessibility in beta cells (NKX6-1 alpha 1: .78, alpha 2: .80, beta 1: .98, beta 2: 1.00; NKX6-3 alpha 1: 0, alpha 2: 0, beta 1: .18, beta 2: .19), and among RFX family members RFX6 had promoter accessibility in high-hormone state cells (alpha 1: .93, alpha 2: .68, beta 1: 0.88, beta 2: .85) (**Figure 4E**).

Predictions of genetic effects on islet regulatory programs are particularly important in interpreting the function of low frequency and rare non-coding variants, which are impractical to assay by standard approaches such as QTL and allelic imbalance mapping without very large sample sizes. We thus sought to demonstrate that our predictions could prioritize functional lower frequency (defined as minor allele frequency [MAF]<.05) variants involved in T2D risk. We first compared the T2D association at different p-value thresholds for lower frequency variants with significant effects for any endocrine cell type, as well as for each cell type individually, to background variants in endocrine sites without predicted effects. We observed enrichment of genome-wide significant T2D association among lower frequency variants with effects in any endocrine cell type compared to background variants (**Figure 4F**). When considering effects in each cell type, we observed enrichment of T2D association among lower frequency variants with effects in beta cells as well as delta cells, even down to more nominal p-value thresholds (**Figure 4F**). We then identified specific low frequency variants with predicted allelic effects that influence T2D risk. At the *IGF2BP3* locus, predictions pointed to a low frequency, T2D-associated variant rs78840640 (MAF=.02) having allelic effects on beta cell chromatin (INS-high beta q=.0015; INS-low beta q=.041) (**Figure 4G**). Fine-mapping data at this locus supported that rs78840640 is likely to have a causal role in T2D (posterior probability of association [PPA]=.33) (**Figure 4G**). We confirmed in gene reporter assays that this variant affected beta cell enhancer activity where the alternate (and T2D risk) allele G had reduced activity (**Figure 4C**). We also observed predicted beta cell effects for rare T2D risk variants including rs186384225 (MAF=.0037) at *TCF7L2* and rs571342427 (MAF=.0015) upstream of the *INS-IGF2* promoter (**Supplementary Figure 7**).

Together these results reveal cell type- and state-specific genetic effects on islet chromatin that inform functional interpretation of T2D risk variants across the allele frequency spectrum.

**Co-accessibility links distal regulatory variants to putative target genes in specific islet cell types and states**

Defining the genes affected by regulatory element activity remains a major challenge given that enhancers can regulate gene activity over large, non-adjacent distances73. A number of approaches have been developed to link regulatory elements to target genes including assays of 3D chromatin architecture and correlation of activity across multiple samples74,75. While these approaches have different strengths, a common weakness is reliance on ensemble data and non-cell type-resolved information32,76. Recently, a new approach was developed to link regulatory elements based on co-accessibility across single-cells6, which has the potential to provide evidence of cell-type resolved enhancer-promoter interactions. We thus sought to leverage accessible chromatin profiles across thousands of islets cells to define co-accessibility between regulatory regions in specific islet cell types. For these analyses we again focused on alpha and beta cells where cell numbers (5,594 and 7,170 cells, respectively) gave us the most power to effectively derive co-accessibility maps.

To calibrate the extent to which co-accessibility reflects true physical interactions between regulatory elements, we first performed a distance-matched comparison between co-accessible sites called at different thresholds to chromatin loops identified from Hi-C and promoter capture Hi-C (pcHi-C) assays in primary human islets32,76. We observed strong enrichment of sites with co-accessibility scores >.05 in both alpha and beta cells for islet chromatin loops identified from pcHi-C and Hi-C compared to sites that had no evidence for co-accessibility (**Figure 5A**, **Supplementary Figure 8A-C**). We therefore used this threshold (.05) to define co-accessibility, through which we identified 593,769 co-accessible sites in alpha cells and 487,549 co-accessible sites in beta cells (**Supplementary Data 3**). Of these sites, 252,816 (alpha) and 165,513 (beta) were between distal sites and gene promoters. There were a total of 64,045 (alpha) and 57,374 (beta) unique distal sites co-accessible with a gene promoter (median 2 promoters per site), and 19,872 (alpha) and 19,269 (beta) unique gene promoters co-accessible with a distal site (median 9 per gene in alpha cells, 6 in beta cells) (**Supplementary Figure 8D-G**).

Among co-accessible links to gene promoters, the majority (71.9%) were alpha- or beta-cell specific, highlighting the value of single cell-resolved data for identifying putative cell type-specific regulatory interactions. As an example of cell type-specific co-accessibility, the *PDX1* promoter had co-accessibility with 35 chromatin sites in beta cells, including a site over 500kb distal that directly coincided with an islet pcHi-C loop, only 7 of which were also found in alpha cells (**Figure 4B**). In another example, at the *ARX* locus 17 sites were co-accessible with the *ARX* promoter in alpha cells, none of which were co-accessible in beta cells (**Supplementary Figure 8F**). Conversely, as an example of shared co-accessibility across cell types, the *NEUROD1* promoter was co-accessible with 52 and 47 chromatin sites in alpha and beta cells, respectively, of which 26 were shared and several were over 500kb distal (**Supplementary Figure 8G**).

Given heterogeneity in alpha and beta cell regulatory programs, we next cataloged co-accessible links between distal alpha and beta cell sites and gene promoters that had differential activity across high- and low-hormone states (**see Methods, Supplementary Data 3**). We observed 25,012 (alpha) and 9,641 (beta) co-accessible links where both the distal site (alpha=10,926 unique distal sites, beta=7,958) and the gene promoter (alpha=1,951 unique promoters, beta=1,516) were differentially active between states in the same direction (**Supplementary Data 3**). State-dependent co-accessible links included both gene promoters active in the high-hormone state such as *INS*, *GCG, G6PC2* and *NEUROD1*, and gene promoters active in the low-hormone state such as *FOSL1, FOSL2, CREB1* and *CREB5*. We also identified a set of genes which were co-accessible with different distal sites that had high-hormone and low-hormone dependent activity such as *GLIS3*, suggesting these genes have distinct regulatory programs driving their activity across cell states.

Distal sites with co-accessibility to gene promoters harbored risk variants for T2D in many cases, suggesting this approach can prioritize target genes of T2D risk variants in islet cells. We observed one such example at the *KCNQ1* locus, where an islet chromatin site located in intron 3 of *KCNQ1* had beta cell-specific co-accessibility with the *INS* promoter over 500kb distaland harbored a causal T2D risk variant rs231361 (PPA=1)23. (**Figure 5C**). To confirm that co-accessibility represented a physical relationship between the distal site and the insulin promoter in beta cells, we utilized a published insulin promoter 4C dataset from the EndoC-βH1 beta cell line77. As expected, the distal site was in a region interacting with the insulin promoter (**Supplementary Figure 9A**). Interestingly, the site was more accessible in INS-high beta cells compared to INS-low beta cells, and rs231361 was predicted to have state-specific effects on beta cell chromatin accessibility (INS-high beta FDR q=.060; INS-low beta FDR q=.40). Furthermore, rs231361 disrupted a *RFX* family sequence motif, which itself was enriched in the INS-high beta cell state (**Figure 5C,** and see **Figure 2C**). The *KCNQ1* locus is associated with quantitative measures of insulin secretion78–81 and fasting glucose level82, demonstrating that the mechanism of action of this locus on T2D risk is likely mediated through beta cell function in a state-dependent manner.

To validate the effects of the chromatin site containing rs231361 on distal regulation of *INS* in beta cells, we deleted a 2.6kb region flanking the site in hESCs by CRISPR/Cas9-mediated genome editing generating three bi-allelic deletion clones (*KCNQ1*∆Enh) (**Figure 5C**, **Supplementary Figure 9B,C**). We then differentiated the three *KCNQ1*∆Enh as well as two unedited control clones into beta cells using an established protocol83 with minor modifications (**see Methods**). Analysis of beta cell stage cultures revealed similar numbers of INS+ cells in *KCNQ1*∆Enh and control (91.1±4.02% vs 94.6±2.11%) (**Supplementary Figure 9D)**, suggesting that the enhancer deletion had no effect on beta cell differentiation. Further supporting this conclusion, the numbers of cells expressing the beta cell marker NKX6-1 in *KCNQ1*∆Enh and control cultures were similar (**Supplementary Figure 9E**), as were *NKX6-1* (FDR=0.98) mRNA levels (**Supplementary Figure 9F**). Next, we determined the effects of the enhancer deletion on gene expression in *cis,* interrogating all genes within 2Mb of the enhancer. We observed a significant decrease in the expression of *INS* (P=3.02x10-4; FDR=0.066) and *CDKN1C* (P=1.96x10-4; FDR=0.059) in *KCNQ1*∆Enh compared to control cells, whereas other genes in the region showed no significant difference in expression (all P>.05; note *KCNQ1* itself is not expressed) (**Figure 5D**). Analysis of INS protein by immunofluorescence staining, flow cytometry and ELISA further revealed reduced INS protein abundance in *KCNQ1*∆Enh beta cells (**Figure 5E-G**). By contrast, beta cell NKX6-1 protein levels were not affected (**Supplementary Figure 9E**), confirming specific effects of the enhancer deletion on *INS* mRNA and protein expression in beta cells.

In addition to *INS* and *CDKN1C* expression, genome-wide differential expression analysis identified 89 additional mRNAs as regulated in *KCNQ1*∆Enh cells (**Supplementary Table 7**). We hypothesized that these genes could be regulated in *trans* downstream of *cis* regulatory effects on *INS* and/or *CDKN1C*. To explore this possibility, we calculated the directional concordance between expression changes in *KCNQ1*∆Enh cells and correlations with *INS* and *CDKN1C* expression across 5,958 beta cells from published scRNA-seq data17 (**see Methods**). We reasoned that genes with regulation tied to *INS* levels would show either: 1) *decreased* expression in *KCNQ1*∆Enh cells and *positive* correlation in expression with *INS* across single cells, or 2) *increased* expression in *KCNQ1*∆Enh cells and *negative* correlation in expression with *INS* across single cells. Across these 89 differentially expressed genes, we observed significant directional concordance in correlations with *INS* (binomial P=1.5x10-4) but not *CDKN1C* (binomial P=1) expression. These results indicate that the regulation of many genes in *KCNQ1*∆Enh cells is secondary to effects of the enhancer deletion on INS expression.

Together, these data demonstrate that single cell co-accessibility can link regulatory elements harboring disease-associated variants to *bona fide* distal target genes in a cell type-specific manner.

**A resource of islet cell type and state regulatory programs to annotate T2D risk variants**

Together our results provide a multi-tiered reference of islet cell type and cell state regulatory programs through which non-coding genetic variants can be comprehensively annotated. As the majority of genetic risk variants for diabetes are non-coding, this resource can be used to annotate diabetes risk variants and interpret their molecular mechanisms. We annotated the islet cell type regulatory programs of T2D risk variants using fine-mapping ‘credible sets’ of 402 risk signals combined from the DIAMANTE and Biobank Japan studies23,84. Fine-mapped ‘credible set’ variants (only those with PPA>.01) at 239 T2D risk signals mapped in an islet cell type chromatin site (**Supplementary Table 8**). Among these 239 risk signals, fine-mapped variants at 183 and 131 signals were in a site co-accessible with a gene promoter or had predicted effects on islet chromatin, respectively, and fine-mapped variants at 97 risk signals had both predicted allelic effects and co-accessibility with a gene promoter. For these 97 signals only a single candidate variant on average overlapped an islet cell type chromatin site, had predicted allelic effects on the site, and was co-accessible with a gene promoter (**Supplementary Table 8).**

Genes co-accessible with fine-mapped T2D variants in islet cell type chromatin sites were enriched for biological processes related to protein localization and transport, stress response, cell cycle, and signal transduction (**Supplementary Table 9**) Co-accessible genes also included numerous genes involved in monogenic diabetes such as *INS, KCNJ11, ABCC8, HNF1A, HNF4A, GCK, NKX2-2*, as well as TFs in structural families with lineage- and state-specific motif enrichments (from **Figure 1-2**) such as *NKX6-1, NFATC2,* and *RFX6*. At 22 T2D loci, fine-mapped variants at multiple independent risk signals were co-accessible with the same gene, providing independent support for the putative role of these genes in diabetes pathogenesis. For example, at the 11p15 locus fine-mapped variants at four independent T2D risk signals (including rs231361 described above) were in sites co-accessible with the *INS* promoter in beta cells (**Supplementary Figure 10A**), and at the 9p21 locus fine-mapped variants at five independent T2D risk signals were in sites co-accessible with the *CDKN2A, MTAP* and *DMRTA1* promoters in both alpha and beta cells (**Supplementary Figure 10B**). In other examples, at the 7p21 locus fine-mapped variants at two independent T2D signals were in sites co-accessible with the *DGKB* promoter in beta cells (**Supplementary Figure 10C)**, and at the 7p13 locus fine-mapped variants at two independent T2D signals were in sites co-accessible with the *GCK* promoter in alpha and beta cells (**Supplementary Figure 10D**).

In order to effectively provide these data to the community, we developed a publicly-accessible database and web portal (http://www.t2depigenome.org) which contains the islet cell type data generated in this study, visualization of cell type projections and chromatin accessibility patterns, and an interface to query non-coding genetic variants for their respective islet cell type annotations (**Supplementary Figure 11**).

# Discussion

Single cell accessible chromatin in primary human pancreatic islets provided transcriptional regulatory programs of alpha, beta, delta, gamma and other pancreatic cell types, as well as heterogeneity in the regulatory programs of endocrine cell types. Numerous studies have characterized heterogeneity in beta cells including cell surface markers, gene expression patterns, and physiological function16,85. Our results reveal an additional layer of beta cell heterogeneity at the epigenomic level, where the cellular states we identified closely corresponded to previously characterized beta cell states related to insulin production and stress-related signaling response. Furthermore, our results provide TF families likely driving heterogeneity in beta cell regulatory programs, such as RFX for hormone-producing states and FOS/JUN for signaling states. While previous studies have focused primarily on heterogeneity in beta cells, our findings reveal that a common regulatory logic underlies state-specific heterogeneity across all endocrine cell types. Based on pseudo-time ordering we also find evidence that accessible chromatin in endocrine cells exists on a gradient across the state axis, although the extent and rate at which cells cycle between these states is currently unclear.

Integrating single cell accessible chromatin with genetic association data revealed heterogeneity in the enrichment of individual beta cells for fasting glucose level, with stronger enrichment among cells in the high-insulin beta cell state. By comparison, T2D risk was more uniformly enriched across beta cell states. Genetic studies have identified numerous loci affecting both T2D risk and fasting glucose level25,55,86, where the mechanism of action for these loci in islets is likely through processes related to insulin production and secretion. Other T2D loci affect beta cell functions that are less related to insulin secretion such as the unfolded protein response, proliferation and apoptosis87,88. Our results suggest that within beta cells, fasting glucose levels are largely mediated through a specific insulin-producing state, whereas T2D risk is mediated through both this insulin-producing state as well as other functional states. We also identify TF motifs that potentially drive fasting glucose- and T2D-relevant regulatory programs in islets. For fasting glucose, this includes motifs for TF families such as RFX, FOXA and NEUROD, which are all involved in insulin secretion89–91, and for T2D this includes CREB1, NFYA and ETV1/5 that have been implicated in beta cell proliferation, stress and survival92,93.

Through machine learning we obtained cell type- and state-specific predictions of genetic variant effects on islet accessible chromatin, which were highly concordant with allelic effects in gene reporter assays and cell type-specific allelic imbalance. Comparing predictions to caQTLs from ensemble islet data34 revealed that caQTLs preferentially identified variants with shared effects across cell types, suggesting that cell type- and state-specific effects are under-represented in studies of ensemble islets in contemporary sample sizes. These findings argue for the value of cell type-resolved islet caQTL maps derived from single cell data in deconvoluting cell type-specific effects, in line with what has been recently shown in a QTL study of single cell gene expression in mononuclear blood cells94. However, even with cell type-resolved islet QTLs, accurate functional predictions of non-coding genetic variants remain important in interpreting the disease-relevance of rare variants. In our study, we identified and validated several lower frequency T2D risk variants that affected beta cell regulatory activity. Thus, our predictions enable functional interpretation of rare risk variant function, and can also inform non-coding variant burden tests, the development of personalized risk models, and the identification of non-coding mutations that contribute to Mendelian forms of diabetes.

Identifying the genes affected by non-coding disease risk variants is paramount for understanding the molecular pathways dysregulated in disease and can inform therapeutic target discovery. By using single cell co-accessibility, we derived candidate target genes of alpha and beta cell chromatin sites which were highly concordant with 3D chromatin loops identified in ensemble islets, yet also revealed an additional layer of cell type-specificity to these relationships. Candidate genes of T2D risk signals were strongly enriched for biological processes relevant to diabetes pathogenesis in islets such as protein secretion, stress response and cell cycle, suggesting that we identified *bona fide* T2D risk genes, many of which could be compelling targets for future mechanistic studies.

At the *KCNQ1* locus, our co-accessibility data and hESC-derived cell models revealed that an enhancer harboring a causal T2D variant affects insulin expression and protein levels in beta cells. We also identified evidence for effects of this enhancer on *CDKN1C* expression, which has been implicated in beta cell function and proliferation95. Genetic mutations of insulin cause monogenic diabetes and tandem repeats in insulin affect T1D risk96,97, but insulin has not been directly implicated in T2D risk. A previous study identified *trans* effects on insulin regulation in beta cells through inter-chromosomal interactions with the insulin promoter, and which included numerous other T2D risk loci77*.* Here, our findings suggest that T2D risk is mediated through insulin directly via long-range, beta cell-specific *cis* regulatory effects. The *KCNQ1* locus has a highly polygenic contribution to T2D with >10 signals in the region that confer independent risk23, four of which (including rs231361) had beta cell co-accessibility with the insulin promoter in our data. In support of this potential mechanism, variants at this locus are associated with quantitative measures of insulin secretion and beta cell function78–81. This locus is also imprinted in islets and several local genes show mono-allelic expression98. Based on these findings, we speculate that the *KCNQ1* locus mediates T2D risk through a complex combination of multiple, long-range regulatory effects on insulin, in addition to effects on *CDKN1C, KCNQ1* and other genes.

Over 400 known risk signals for T2D have been identified to date, yet only a handful have been characterized molecularly. The detailed resource of islet cell type and state regulatory programs provided by this study, in combination with genetic fine-mapping and genome sequencing data, will greatly enhance efforts to define the molecular mechanisms of T2D risk. More broadly, our findings provide a roadmap demonstrating how single cell accessible chromatin derived from disease-relevant primary tissue can be utilized to inform on the cell types, cell states, *cis* regulatory elements and genes involved in genetic basis of complex disease.

# Methods

## Islet processing and nuclei isolation

## We obtained islet preparations from three donors for the Integrated Islet Distribution Program (IIDP) (Supplementary Table 1). Islet preparations were further enriched using zinc-dithizone staining followed by hand picking. Studies were given exempt status by the Institutional Review Board (IRB) of the University of California San Diego.

## Generation of snATAC-seq libraries

Combinatorial barcoding single nuclear ATAC-seq was performed as described previously3,5 with several modifications as described below. For each donor (N=3), approximately 3,000 islet equivalents (IEQ, roughly 1,000 cells each) were resuspended in 1 ml nuclei permeabilization buffer (10mM Tris-HCL (pH 7.5), 10mM NaCl, 3mM MgCl2, 0.1% Tween-20 (Sigma), 0.1% IGEPAL-CA630 (Sigma) and 0.01% Digitonin (Promega) in water) and homogenized using 1ml glass dounce homogenizer with a tight-fitting pestle for 15 strokes. Homogenized islets were incubated for 10 min at 4°C and filtered with 30 µm filter (CellTrics). Nuclei were pelleted with a swinging bucket centrifuge (500 x g, 5 min, 4°C; 5920R, Eppendorf) and resuspended in 500 µL high salt tagmentation buffer (36.3 mM Tris-acetate (pH = 7.8), 72.6 mM potassium-acetate, 11 mM Mg-acetate, 17.6% DMF) and counted using a hemocytometer. Concentration was adjusted to 4500 nuclei/9 µl, and 4,500 nuclei were dispensed into each well of a 96-well plate. Glycerol was added to the leftover nuclei suspension for a final concentration of 25 % and nuclei were stored at -80°C. For tagmentation, 1 µL barcoded Tn5 transposomes5,99 were added using a BenchSmart™ 96 (Mettler Toledo), mixed five times and incubated for 60 min at 37 °C with shaking (500 rpm). To inhibit the Tn5 reaction, 10 µL of 40 mM EDTA were added to each well with a BenchSmart™ 96 (Mettler Toledo) and the plate was incubated at 37 °C for 15 min with shaking (500 rpm). Next, 20 µL 2 x sort buffer (2 % BSA, 2 mM EDTA in PBS) were added using a BenchSmart™ 96 (Mettler Toledo). All wells were combined into a FACS tube and stained with 3 µM Draq7 (Cell Signaling). Using a SH800 (Sony), 20 nuclei were sorted per well into eight 96-well plates (total of 768 wells) containing 10.5 µL EB (25 pmol primer i7, 25 pmol primer i5, 200 ng BSA (Sigma), PMID: 29434377). Preparation of sort plates and all downstream pipetting steps were performed on a Biomek i7 Automated Workstation (Beckman Coulter). After addition of 1 µL 0.2% SDS, samples were incubated at 55 °C for 7 min with shaking (500 rpm). We added 1 µL 12.5% Triton-X to each well to quench the SDS and 12.5 µL NEBNext High-Fidelity 2× PCR Master Mix (NEB). Samples were PCR-amplified (72 °C 5 min, 98 °C 30 s, (98 °C 10 s, 63 °C 30 s, 72 °C 60 s) × 12 cycles, held at 12 °C). After PCR, all wells were combined. Libraries were purified according to the MinElute PCR Purification Kit manual (Qiagen) using a vacuum manifold (QIAvac 24 plus, Qiagen) and size selection was performed with SPRI Beads (Beckmann Coulter, 0.55x and 1.5x). Libraries were purified one more time with SPRI Beads (Beckmann Coulter, 1.5x). Libraries were quantified using a Qubit fluorimeter (Life technologies) and the nucleosomal pattern was verified using a Tapestation (High Sensitivity D1000, Agilent). The library was sequenced on a HiSeq2500 sequencer (Illumina) using custom sequencing primers, 25% spike-in library and following read lengths: 50 + 43 + 40 + 50 (Read1 + Index1 + Index2 + Read2).

## Raw data processing and quality control

For each read, we first appended the cell barcode metadata to the read name. The cell barcode consisted of four pieces (P7, I7, I5, P5) which were derived from the index read files. We first corrected for sequencing errors by calculating the Levenshtein distance between each of the four pieces and a whitelist of possible sequences. If the piece did not perfectly match a whitelisted sequence, we took the best matching sequence if it was within 2 edits and the next matching sequence was at least 2 additional edits away. If none of these conditions were met, we discarded the read from further analyses.

We trimmed Nextera adapter sequences from sequence reads using the trim\_galore program with default parameters (https://github.com/FelixKrueger/TrimGalore). We used bwa mem100 to align reads to the hg19 reference genome with the options ‘-M -C’. We then used samtools101 to filter out reads that did not align to the autosomes or sex chromosomes and low mapping quality reads (MAPQ<30). We used samtools fixmate to perform additional checks for FR proper pairs and removed secondary or unmapped reads. We used the MarkDuplicates tool from picard (https://broadinstitute.github.io/picard/) to remove duplicates on a per-barcode basis with ‘BARCODE\_TAG’ option. For each experiment, we used a Gaussian mixture model on log-transformed read depths to separate barcodes with a 99% probability of belonging to the high read distribution, likely representing real cells, from those in the low read distribution, likely representing background reads. We then set an additional threshold of 1000 read depth, reasoning that low read cells would contribute additional noise to clustering.

## Cluster analysis (snATAC-seq)

We split the genome into 5kb windows and removed windows overlapping blacklisted regions from ENCODE. For each experiment, we then created a sparse *m* x *n* matrix containing read depth for *m* cells passing read depth thresholds at *n* windows. For further quality checks, we performed initial clustering for each experiment individually using scanpy102 . We extracted highly variable windows using mean read depths and normalized dispersion. After normalization to a uniform read depth and log-transformation of read depth, we regressed out the log-transformed total read depth for each cell. We then performed PCA and extracted the top 50 principal components. We used these components to calculate the nearest 30 neighbors using the cosine metric, which were subsequently used for UMAP dimensionality reduction with the parameters ‘min\_dist=0.3’ and Louvain clustering with the parameters ‘resolution=1.5’. For each experiment, we removed 2,709 cells that were in clusters corresponding to low read depth.

After removing these cells, we used similar methods to cluster cells from all experiments together with the following modifications. We extracted highly variable windows across cells from all experiments. Since read depth was a technical covariate specific to each experiment, we regressed this out on a per-experiment basis. We used mutual nearest neighbors (MNN) correction35 to adjust for batch effects across experiments with the parameters ‘k=10’. We then performed clustering as described above. We used chromatin accessibility at windows overlapping promoters for marker hormones (*GCG*, *INS-IGF2*, *SST*, and *PPY*) to assign cell types for the endocrine islet cell types (alpha, beta, delta, and gamma). We performed re-clustering on non-endocrine islet clusters and used chromatin accessibility at windows around marker genes from single cell RNA-seq to assign cluster labels. In our clustering results, we identified a cluster of 694 alpha cells that were mostly derived from a single donor (96% of cells from Islet 1). Because we were uncertain whether this represented technical or biological differences, we excluded this cluster from further analyses. We also excluded a cluster of 192 cells likely representing lower quality cells as it had low intra-cluster similarity and lower fraction of reads in peaks.

## Comparison to bulk and sorted islet ATAC-seq

We obtained raw sequence data of ATAC-seq for 42 bulk islet samples from four prior studies27,32,33,71 and 4 bulk pancreas samples from ENCODE. We re-processed all samples with a uniform pipeline: we aligned all reads to hg19 with bwa mem, identified and removed duplicate reads with picard MarkDuplicates, and called peaks with MACS2 with the parameters ‘—shift -100 –extsize 200 –keep-dup all’. For the three islet snATAC-seq samples, we used aggregated per-barcode deduplicated reads to call peaks. We defined all possible accessibility peaks by filtering out ENCODE blacklisted regions and retaining merged peaks on autosomal chromosomes found in more than one sample. We then calculated the read coverage at all possible accessibility peaks and TPM-normalized the counts. We calculated the Spearman correlation between normalized read coverages and used hierarchical clustering to assess similarity between bulk islet samples. To check peak call overlap between aggregated single cell ATAC and bulk ATAC data, we split peaks based into promoter proximal (+/-500bp from GENCODE transcript TSS) and distal peaks based on promoter overlap. For each cluster, we calculated the percentage of aggregate peaks that overlapped merged autosomal bulk peaks and individual sample-level autosomal bulk peaks.

We also obtained raw sequence data of ATAC-seq from flow-sorted pancreatic cells (alpha, beta, acinar, ductal) from two prior studies21,22 and re-processed all samples with the uniform pipeline described above. For alpha, beta, and exocrine cells from islet snATAC-seq, we split reads on a per-donor and per-cluster basis to obtain read files. Because total read depth was highly variable across sorted samples, we merged autosomal peaks after filtering out ENCODE blacklist regions. We calculated read coverage in each sample for each merged peak and TPM normalized count values. We then calculated the Spearman correlation between normalized read coverages and used hierarchical clustering to assess similarity between sorted and snATAC-seq islet samples.

## Identifying marker peaks of chromatin accessibility

To identify peaks for each cell type, we aggregated reads for all cells within a cluster or sub-cluster. We shifted reads aligning to the positive strand by +4bp and reads aligning to the negative strand by -5bp, extended reads to 200bp, and centered reads. We used MACS2103 to call peaks of chromatin accessibility for each aggregated read file with the following settings ‘--nomodel --keep-dup all’. We removed peaks that overlapped ENCODE blacklisted regions104. We then used bedtools105 to merge peaks from all clusters and sub-clusters to create a superset of islet regulatory peaks.

We generated a sparse *m* x *n* binary matrix containing binary overlap between *m* peaks in the superset of islet regulatory peaks and *n* cells. We then calculated t-statistics of peak specificity for each cluster or sub-cluster through linear regression models. We used binary encodings to specify which donor a given cell came from as covariates in the model. For each peak and cluster, we used binary encoding of read overlap with the peak as the predictor and whether a cell was in the cluster (1 if yes, -1 if no) as the outcome.

## Matching islet snATAC-seq with scRNA-seq clusters

To verify that clusters definitions and labels from single cell chromatin accessibility data matched those from single cell expression data, we obtained published single cell RNA-seq data from 12 non-diabetic islet donors17. Because cluster definitions for all cell types were not available, we re-analyzed the data and performed clustering analyses. Starting with the gene expression matrix, we first performed quality control steps to remove potential doublets. For each marker genes of different cell types within the pancreas, including *GCG* (alpha), *INS* (beta), *SST* (delta), *PPY* (gamma), *CTRB2* (acinar), *CFTR* (ductal)*, PLVAP* (endothelial)*, PDGFRB* (stellate)*,* and *C1QC* (immune) we used a Gaussian mixture model on log-transformed read depth for the gene to identify whether a cell expressed it (high distribution) or not (low distribution). We verified that cells expressing more than one marker gene had on average higher read depth and more genes expressed (**Figure SX**). We regressed out covariates including sex, BMI, and read depth, and separated cells by donor of origin. We then used MNN correction35 to adjust for batch effects. After scaling the data, we performed PCA and used the top 50 principal components to calculate the 10 nearest neighbors using the cosine metric. We used the nearest neighbor map for UMAP dimensionality reduction with the parameters ‘min\_dist=0.3’ and to perform Louvain clustering with the parameters ‘resolution=1’. We used a similar regression framework as the chromatin accessibility marker peaks to calculate t-statistics for gene specificity for each cluster with the following modifications. We included sex, BMI, and log-transformed read coverage as covariates and used log2 read counts for each gene instead of binary peak coverage as the predictor.

We used the Spearman correlation between t-statistics from islet snATAC-seq and scRNA-seq data to match up clusters. Specifically, we took the top 100 (sorted by descending t-statistic) most specific promoter peaks for each cluster or sub-cluster to define a list of genes for comparison. To facilitate one-to-one comparisons between the two datasets, for this analysis only we defined promoter peaks as peaks within +/-500bp of a GENCODE v19 gene TSS. This list contained 966 genes, which is less than 100x13 (number of clusters) because 1) marker genes were sometimes shared between sub-clusters and 2) not all genes were present in the expression dataset. For each cluster from accessible chromatin data, we then compared t-statistics of genes in the list with t-statistics for all clusters from single cell expression using the Spearman correlation, which is robust to very specific marker genes such as insulin which could otherwise bias these comparisons.

## Motif enrichment with chromVAR

We used chromVar40 to calculate TF motif-associated difference between cell populations. We first calculated counts per peak per cell matrix and then input it to chomVar. We filtered cells with minimal reads less than 1500 (min\_depth=1500) and peaks with fraction of reads less than 0.15 (min\_in\_peaks=0.15) by using ‘filterSamplesPlot’ function from chromVar. We also corrected GC bias based on ‘BSgenome.Hsapiens.UCSC.hg19’ using ‘addGCBias’ function. Then we used the Jaspar motifs from ‘getJasparMotifs’ function with default parameter and calculated the deviation Z scores for each TF motif in each cell by using ‘computeDeviations’ function. High-variance TF motifs across all cell types were selected by ‘computeVariability’ function using cut-off 1.2 (N=111). For each of these variable motifs, we calculated the mean Z score for each cell types and normalized the values to 0 (minimal) and 1 (maximal).

## Comparison of alpha and beta cell states

To identify TF motifs variable between alpha or beta cell states, we performed Student’s t-test on motifs’ Z scores between cells labeled with alpha1 and alpha 2 cells or beta 1 and beta 2. The raw p-value were then adjusted by the Benjamini-Hochberg procedure to obtain FDR. Motifs with FDR less than 0.05 and absolute delta Z score greater than 0.5 were defined as differential motifs (N=46 for beta cells, N=109 for alpha cells and 111 motifs combined). For these 111 motifs, we summarized the mean Z scores over alpha 1, alpha 2, beta 1 and beta 2 cells and plotted the normalized value. In order to check how these motifs usage changes along the pseudo cell states, we smoothed motif Z scores along pseudo-states for alpha and beta cells separately at step=0.05, using the shrinkage version of cubic regression spline (by ‘gam’ function from R package ‘mgcv’ with parameter bs =’cs’). We then smoothed motif enrichment profiles and normalized values for visualization.

To analyze differential promoter accessibility between alpha and beta cell states, we first calculated the binary promoter by cell matrix containing information about read overlap per cell in a promoter peak. Based on this matrix and cell cluster labels, we performed Fisher’s exact test between state 1 and state 2 of alpha, beta, and delta cells for each promoter against the null hypothesis that the promoter had similar accessibility across states. We used Bonferroni adjusted p-values (adjusted p-value<0.01) for alpha and beta cells with the sign of the log2 transformed odds ratio to identify genes whose promoter had either increased or decreased accessibility across states. Differentially-accessible promoters were further input into Enrichr106 to perform GO term enrichment analysis on biological processes terms (2018 version). To identify more specific processes, we filtered for gene ontology terms that contained less than 150 total genes.

To plot the profile of each promoter across pseudo-state, we first binned alpha cells or beta cells to 100 bins along the state trajectory. For each bin, we calculated the fraction of cells had a peak in the promoter region for each promoter. Then we smoothed these 100 fractions using the ‘loess’ function from R. The smoothed data were then normalized and clustered using k-medoids clustering, with k determined by optimum average silhouette width using the ‘pamk’ function from the R ‘fpc’ package. Genes attributed to the promoters in each cluster were then used to perform GO term enrichment analysis.

In order to compare with previous published data, we collected gene lists from Bader et al.19, Dorrell et al.16 and Xin et al.17. From Dorrell et al, we used four gene lists: genes significantly enriched in ST8SIA1+ beta cells (Supplement Figure 6 from Dorrell et al.), ST8SIA1- beta cells (Supplement Figure 7 from Dorrell et al.), CD9+ beta cells (Supplement Figure 8 from Dorrell et al.), and CD9- beta cells (Supplement Figure 9 from Dorrell et al.). From Bader et al, we obtained two gene lists from Supplementary Table 1 and converted the genes from mouse to human: genes enriched in FVR+ beta cells (i.e. mature beta cells) and FVR- beta cells (i.e. immature beta cells). From Xin et al, we obtained four gene lists for Beta 1-4 subpopulations (Supplementary Table S3 in Xin et al.). For each gene list, we performed gene set enrichment analysis107 on the log2 odds ratios from the differential promoter accessibility analysis.

## Ordering alpha and beta cells along a trajectory and finding dynamic peaks

We used cicero6 to order all alpha and beta cells along separate trajectories. We started with a sparse binary matrix encoding overlap between the superset of islet regulatory peaks and cells. We extracted all cells belonging to alpha cell sub-clusters and filtered out peaks that were not present in alpha cells. We used the aggregate\_nearby\_peaks function from cicero to find peaks within 10kb and merging their counts to make an aggregate matrix. We then chose peaks to define progress with the aggregated matrix by using the differentialGeneTest function from monocle250 to search for peaks that were different between the alpha 1 and alpha 2 sub-clusters (q-value<.1), while modeling total peaks in each cell as a covariate. We then used DDRTree to reduce dimensions and ordered cells along the trajectory, setting the root state as the state with the highest glucagon promoter accessibility. We grouped cells into 10 bins based on their trajectory values. Then we repeated the same procedure for beta cells, with the modification of ordering cells by insulin promoter accessibility.

## GWAS enrichment with aggregate peak annotations

We used cell type specific (CTS) LD score regression68,108 to calculate enrichment for GWAS traits. We obtained GWAS summary statistics for quantitative traits related to diabetes26,54–58, diabetes23, and control traits including psychiatric and autoimmune diseases59–66. We prepared summary statistics to the standard format for LD score regression. We used peaks from aggregated reads for each cluster as a binary annotation, and the superset of islet regulatory peaks as the background control. For each trait, we then used CTS LD score regression to estimate the enrichment coefficient of each annotation jointly with the background control.

## GWAS enrichment with single cell annotations

We determined genetic enrichment of accessible chromatin profiles in individual cells. We first split the genome into 5kb windows and removed windows overlapping blacklisted regions from ENCODE. We created a sparse *m* x *n* matrix containing read depth for *m* cells passing read depth thresholds at *n* windows, and extracted highly variable (HV) windows using mean read depths and normalized dispersion. We then retained genetic variants mapping in HV windows with minor allele frequency [MAF]>.05 mapping outside of the major histocompatibility complex region (MHC, defined by chr6:25,000,000-35,000,000 in hg19 coordinates).

As the accessible chromatin profiles from an individual cell are sparse, we used the bagging algorithm in the make\_cicero\_cds function from cicero6 to aggregate cells into groups of 10. For each aggregate cell group, we created a binary annotation based on mapped reads for cells in the aggregate. We also created baseline annotations consisting of pooled islet cell type accessible chromatin sites and the 53 baseline v1.1 annotations from LD score regression67. We then annotated all variants in HV windows with the aggregate cell and baseline annotations. We determined enrichment of HV variant annotations for fasting glucose level55, type 2 diabetes23, and two control traits, major depressive disorder65 and lupus62 GWAS data. In order to correct for the confounding effects of linkage disequilibrium (LD), we performed LD pruning of GWAS data for each trait by first sorting variants based on p-value and iteratively removing variants in LD (r2>.5, 1000 Genomes European subset) with a more significant variant. To then perform enrichment tests on pruned GWAS data we used a previously described method polyTest109 to jointly model the annotation for each aggregated cell group with the baseline pooled site and 53 annotations from LD score baseline v1.1. We then calculated a z-score for each aggregate cell based on the effects and standard error from the resulting model. As the grouping method for cicero uses bootstrap aggregation, a given cell was potentially assigned to multiple aggregates. We therefore calculated an enrichment z-score for each individual cell by averaging enrichment z-scores for each cell across its respective aggregates.

To identify TFs correlated with trait enrichments, we calculated the Spearman correlation coefficient between fasting glucose or type 2 diabetes single cell GWAS enrichment z-scores and chromVAR motif enrichment z-scores using data from all cells or within beta cells. Within each trait, we used Bonferroni correction to adjust correlation p-values for multiple tests.

## Mapping allelic imbalance within clusters

Genomic DNA for genotyping was extracted either from spare islet nuclei (donors 1 and 2), or acinar cells (donor 3). Genomic DNA was extracted using the DNeasy Blood & Tissue Kits (Qiagen) according to manufacturer’s protocol for purification of total DNA from animal blood or cells. Extracted genomic DNA was used for genotyping on the Illumina Infinium Omni2.5-8 v1.4 genotyping array. For genotypes that passed quality filters (non-missing, MAF>.01 in European or African populations in 1KGP), we then imputed genotypes into the HRC reference panel r1.1110 using the Michigan Imputation Server111. Post-imputation, we removed genotypes with low imputation quality (R2<.3). As an additional filter to remove potential false positive heterozygote genotype calls, we removed variants that had greater than 20 read coverage without reads for both alleles. Using cluster assignments for each cell, we split mapped reads for each donor into cluster-specific reads. For cluster-specific reads, we used the WASP pipeline112 to correct for reference mapping bias at heterozygous variants. We then used a binomial test to assess imbalance at heterozygous variants, assuming a null hypothesis where both alleles were equally likely to be observed. For each variant, we then calculated combined imbalance z-scores across donors using Stouffer’s z-score method and used sequencing depth to weight statistics from each sample.

## Predicting genetic variant effects on chromatin accessibility

We used deltaSVM69 to predict the effects of non-coding variants on chromatin accessibility in each cell type and cell state. We obtained sequences underlying promoter-distal (> +/-500bp from GENCODEv19 transcript TSS for protein coding and long non-coding RNA genes) peaks for each cluster, used ‘genNullSeqs’ to generate background sequences, and then trained a model for each cluster with ‘gkmtrain’ with default settings. For all possible combinations of 11mers, we then used ‘gkmpredict’ to predict the effects of 11mers based on the trained model for the cluster. For each SNP in the HRC reference panel r1.1110 overlapping an islet cell type accessible chromatin site, we created 19bp sequences around each allele (9bp flanking either side of the variant base). We then used the ‘deltasvm.pl’ script to calculate deltaSVM scores for differential chromatin accessibility between variant alleles. We built a null distribution by randomly permuting the effects of 11mers and re-calculating deltaSVM scores and using the parameters of this null distribution, we calculated z-scores for each variant. From variant z-scores we calculated p-values and then q-values and considered variants significant at FDR<.1.

For variants with predicted effects on chromatin accessibility in alpha or beta cells, we categorized them based on their effects across cell type and states. Variants with significant effects in both alpha cell states but neither beta cell state were classified as “alpha” (n=10,564) and vice versa for “beta” (n=12,833). Variants with significant effects in both alpha and beta state 1 but not alpha and beta state 2 were classified as “state 1” (n=15,769), and vice versa for “state 2” (n=12,471). Variants with significant effects in all four alpha and beta cell states were classified as “shared” (n=31,331). We also determined the concordance in the direction of effect for variants across alpha and beta cell states. For the set of variants with significant effects in each state, we calculated the fraction of variants where the allele with the higher predicted effect had a higher predicted effect in other states. We determined significance using a binomial test assuming an expected fraction of 50%. We assessed enrichment of predicted effect variants in alpha or beta cell states for islet caQTLs71 compared to any islet caQTL in alpha or beta cell sites using Fisher’s exact test. We then stratified variants with predicted effects by category (alpha, beta, state 1, or state 2) and assessed enrichment of caQTLs with predicted effects within each category with Fisher’s exact test.

## Luciferase gene reporter assays

We selected T2D risk variants to test for allelic differences in enhancer activity in beta cells. We cloned sequences containing alternate or reference alleles in the forward orientation upstream of the minimal promoter of firefly luciferase vector pGL4.23 (Promega) using KpnI and SacI restriction sites.

The primer sequences were the following:

*TH* locus

rs7482891left: AGAGGTCTGAGGAGCCCTTG

rs7482891right: TAGACCCTGCAGAGCCACAG

*RNF6* locus

rs34584161left: AAGCTGACAGACAGAGGGTCA

rs34584161right: GGGCTTCATAAACATCAGCA

*PROX-AS1* locus

rs17712208left: AAGCCCACCTTCGTAAACAT

rs17712208right: TGAAGTAGCTCCCAGTGAAGG

*IGF2BP3* locus

rs78840640left: CACAATGAAGCCATGTCCTTT

rs78840640right: TCAGCTTTCTATTTTGGGGAAA

*SLC12A8* locus

rs4679370left: TCAATGTCTACCTCAAAATTCTTTGT

rs4679370right: CACTGCAGCCTTAAACTCCTG

We seeded MIN6 beta cells into 6 (or 12)-well trays at 1 million cells per well. At 80% confluency, we co-transfected cells with 400ng\* of the experimental firefly luciferase vector pGL4.23 containing the alt or ref allele in either orientation or an empty vector and 50ng\* of the vector pRL-SV40 (Promega) using the Lipofectamine 3000 reagent. All transfections were done in triplicate. The media for each well was changed 6 hours following transfection under the following conditions: normal glucose (25mM) or high glucose (40mM). We lysed cells 48 hours after transfection and assayed for Firefly and Renilla luciferase activities using the Dual-Luciferase Reporter system (Promega). Firefly activity was normalized to Renilla activity and compared to the empty vector and normalized results were expressed as fold change compared to empty vector control per allele. We used a two-sided t-test to compare the luciferase activity between the two alleles in two orientations.

## TF motif enrichment within predicted effect variant categories

For each cell- or state-resolved category (state 1, state 2, alpha, beta) of variants with predicted effects, we extracted 29bp sequences (+/-14bp around each SNP) corresponding to the higher or lower predicted effect allele. Here, we reasoned that extracting sequences for a larger window around SNPs would alleviate bias for the analysis against motifs with longer PWMs. We then used AME from the MEME suite113 to predict motif enrichment, using position weight matrices from the latest non-redundant motif library JASPAR 201841. We used sequences from the higher effect allele as the test set and sequences from the lower effect allele as the background set. Because motif for TFs within the same structural family can potentially show similar enrichment, we used the TFClass database (http://tfclass.bioinf.med.uni-goettingen.de/) to group motifs by TF family. To determine which TF was most likely driving the enrichment, we used min-max normalized promoter accessibility within TF family members with a promoter peak in alpha or beta cells and highlighted corresponding cell type patterns of promoter accessibility.

## Enrichment of predicted variants for lower frequency variants

We obtained genome-wide summary statistics of T2D from the DIAMANTE consortium23. We estimated LD patterns for variants with MAF<.05 using HRC imputed genotype data from samples in the UK Biobank (UKB, March 2018 release). We randomly selected 10,000 non-related UKB samples of European ancestry and calculated LD between lower frequency variants using PLINK. We then LD-pruned variants with MAF<.05 in DIAMANTE T2D data by first sorting variants based on their p-values and then removing variants in r2>.5 with a more significant variant. Using the LD-pruned results, we then determined enrichment of variants with predicted effects on endocrine cell types. We created sets of variants that had significant effects (FDR<.1) in any endocrine cell type, as well as variants with FDR<1. for each cell type. For alpha, beta and delta cells, we considered variants with effects in either cell state. We then created a background set of variants as those without significant effects in any endocrine cell type (all FDR>.1). We set a series of p-value thresholds (5x10-8, 1x10-7, 1x10-6, 1x10-5, 1x10-4, 1x10-3), and at each threshold determined the fraction of variants in each category as well as background variants reaching that p-value threshold to calculate a fold-enrichment based on these fractions compared to background. We determined significance of the enrichments by using a binomial test of the counts for each category using the background fraction as the expected count.

## Chromatin co-accessibility with Cicero

We used cicero6 to calculate peak-peak co-accessibility scores for alpha and beta cells. Like the trajectory analysis, we started with a sparse binary matrix encoding overlap between the superset of islet regulatory peaks and cells. We extracted all cells belonging to alpha cell sub-clusters and filtered out peaks that were not present in alpha cells. We then used the make\_cicero\_cds function to aggregate cells based on the 50 nearest neighbors from the UMAP reduced dimensions. We then used cicero to calculate co-accessibility scores using a window size of 1Mb and a distance constraint of 500kb, leaving other parameters at the default setting. We then repeated the same procedure for beta cells. We used Fisher’s exact test to assess whether distal co-accessible sites had higher accessibility in either state 1 or state 2, and defined significance at FDR<.1. To compare promoter-distal co-accessibility links that had higher accessibility in the same direction (either both state 1 or state 2), we used differential promoters between states (from the previous analysis in Figure 2).

## Enrichment of islet Hi-C and pcHi-C loops in co-accessible peaks

We obtained sets of merged Hi-C loops32 and high-confidence promoter capture Hi-C (pc-HiC) loops76 from public datasets. For Hi-C loops, we used anchors directly from the loops. For pcHi-C loops, we used a 5kb window centered on the interaction point as the anchor. To compare alpha and beta cell co-accessibility with Hi-C, we then used direct overlap of alpha or beta cell peaks with anchors. For different binned thresholds of co-accessibility in .05 increments, we then calculated distance-matched odds ratios for co-accessible peaks containing Hi-C loops versus non-co-accessible peaks (co-accessible<0). We then used Fisher’s exact test to assess significance. We repeated the procedure for high confidence pcHi-C loops for both cell types.

## Annotating fine-mapped diabetes risk variants

We annotated risk signals in compiled fine-mapping data for (1) type 2 diabetes from the DIAMANTE consortium and Biobank Japan studies and (2) type 1 diabetes from a published study. For the Biobank Japan T2D GWAS, we constructed LD-based 99% genetic credible sets for main signals at 22 novel loci that were distinct from the DIAMANTE study. We used the East Asian subset of the 1000 Genomes Project to define credible set variants by taking all variants in at least low LD (r2>.1) with the index variant in a 5MB window. We used effect size and standard error estimates to calculate Bayes factors for each variant. For each signal, we then calculated the posterior probability causal probability (PPA) that each variant drives the association by dividing its Bayes factor by the sum of Bayes factors for all variants in the signal’s credible set. We then sorted each signal by descending PPA and retained variants that added up to a cumulative probability of .99 to derive 99% credible sets.

For each signal, we identified candidate casual variants that were both in the 99% credible set and had a posterior causal probability greater than .01. We intersected these candidate variants with accessible chromatin sites for each islet cell type and cell state, and then identified variants with predicted effects on the overlapping cell types/states. We finally annotated variants based on overlap with sites co-accessible to gene promoters. For target genes linked to diabetes risk variants we determined enriched gene sets using GSEA.

## Analysis of insulin promoter 4C data

We downloaded and re-analyzed published 4C data of the insulin promoter for the beta cell line EndoC-βH177 with the 4C-ker pipeline114. We first created a reduced genome using 25bp flanking sequences of BglII cutting sites. For each of the 3 replicates, we then aligned reads to this reduced genome using bowtie2115 v2.2.9 with the parameter “-N 0 -5 20”. We then extracted counts for each fragment from the SAM file after removed self-ligated and undigested fragments, and we used the bedGraph files as input to the R.4Cker package. We generated normalized counts and called high interaction regions using the ‘nearBaitAnalysis’ function with the parameter ‘k=10’.

## CRISPR/Cas9-mediated enhancer knockout

H1 hESCs (WA01; purchased from WiCell; NIH registration number: 0043) were seeded onto Matrigel®-coated six-well plates at a density of 50,000 cells/cm2 and maintained in mTeSR1 media (StemCell Technologies) for 3-4 days with media changed daily. hESC research was approved by the University of California, San Diego, Institutional Review Board and Embryonic Stem Cell Research Oversight Committee, and by the institutional review board at CCHMC.

To generate clonal homozygous *KCNQ1* enhancer knockout hESC lines, two sgRNAs targeting the enhancer were designed and cloned into Px333-GFP, a modified version of Px333 (#64073, Addgene). The plasmid was transfected into H1 hESCs with XtremeGene 9 (Roche). 24 hours later, 5000 GFP+ cells were sorted into a well of six-well plate. Individual colonies that emerged within 5-7 days after transfection were subsequently transferred manually into 48-well plates for expansion, genomic DNA extraction, PCR genotyping, and Sanger sequencing. sgRNA oligos and genotyping primers are listed below. For control clones, we transfected the Px333-GFP plasmid into H1 hESCs and subjected the cells to the same workflow as H1 hESCs transfected with sgRNAs.

sgRNA oligos:

KCNQ1\_sgRNA1-s: ACTGTCGGGCCCATCTGCCA  
KCNQ1\_sgRNA1-as: TGGTTGGATCTGTTGCGGGG

Genotyping primers:

Span-F: AGTGGGGCCATGAACAATAA  
Span-R: GCCTGAGTTTCCGTGACTGT

## Pancreatic differentiation of enhancer knockout hESCs

hESCs were differentiated in a suspension-based format using rotational culture with some modifications to a published protocol83. Undifferentiated hESCs were aggregated by preparing a single cell suspension in mTeSR media (STEMCELL Technologies) at 1 × 106 cells/mL and overnight culture in six-well ultra-low attachment plates (Costar) with 5.5ml per well on an orbital rotator (Innova2000, New Brunswick Scientific) at 100 rpm. The following day, undifferentiated aggregates were washed in DMEM/F12 (VWR) and differentiated using a multistep protocol with daily media changes and continued orbital rotation at either 100 rpm or at 108 rpm from days 8 to 28. In addition to 1% GlutaMAX™ (Gibco) and 15 mM (days 0-10) or 20 mM (days 11-28) glucose, MCDB 131 media (Life Technologies) was supplemented with 0.5% (days 0-5) or 2% (days 6-14) fatty acid-free BSA (Proliant), 1.5 g/L (days 0-5 and days 11-28) or 2.5 g/L (days 6-10) NaHCO3 (Sigma-Aldrich), and 0.25 mM (days 3-10) ascorbic acid (Sigma-Aldrich).

Human Activin A, mouse Wnt3a, and human KGF were purchased from R&D Systems. Other media components included ascorbic acid (Sigma-Aldrich), Insulin-Transferrin-Selenium-Ethanolamine (ITS-X; Thermo Fisher Scientific), ZnSO4 (Sigma-Aldrich), heparin (Sigma-Aldrich), retinoic acid (RA) (Sigma-Aldrich), SANT-1 (Sigma-Aldrich), 3,3′,5-Triiodo-L-thyronine (T3) (Sigma-Aldrich), the protein kinase C activator TPB (EMD Chemicals), the BMP type 1 receptor inhibitor LDN-193189 (Stemgent), the TGFβ type 1 activin like kinase receptor ALK5 inhibitor, ALK5 inhibitor II (Enzo Life Sciences), N-Acetyl-L-cysteine (Sigma), R428 (SelleckChem), Trolox (EMD Millipore), -secretase inhibitor XX (Calbiochem).

Day 0: MCDB 131, 100ng/mL Activin, 25ng/mL mouse Wnt3a

Day 1 – Day 2: MCDB 131, 100ng/mL Activin A

Day 3 – Day 5: MCDB 131, 50ng/mL KGF

Day 6 – Day 7: MCDB 131, 50ng/mL KGF, 0.25 µM SANT-1, 1 µM RA 100 nM LDN-193189, 200 nM TPB, 0.5% ITS-X

Day 8 – Day 10: MCDB 131, 2ng/mL KGF, 0.25 µM SANT-1, 0.1 µM RA, 200 nM LDN-193189, 100 nM TPB, 0.5% ITS-X

Day 11 – Day 13: MCDB 131, 0.25 µM SANT-1, 0.05 µM RA, 100 nM LDN-193189, 1 µM T3, 10 µM ALK5i II, 10 µM ZnSO4, 10 µg/mL heparin, 0.5% ITS-X

Day 14 – Day 21: MCDB 131, 100 nM LDN-193189, 1 µM T3, 10 µM ALK5i II, 10 µM ZnSO4, 10 µg/mL heparin, 100nM -secretase inhibitor XX, 0.5% ITS-X

Day 21 – Day 28: MCDB 131, 100 nM LDN-193189, 1 µM T3, 10 µM ALK5i II, 10 µM ZnSO4, 10 µg/mL heparin, 1mM N-Acetyl-L-cysteine, 10µM Trolox, 2µM R428 , 0.5% ITS-X

## Characterization of hESC-derived cultures at beta cell (day 28) stage

*Flow cytometry analysis*

hESC-derived cell aggregates were dissociated into a single-cell suspension with Accutase™ (Innovative Cell Technologies) at 37 °C for 5 min. Accutase™ was quenched with FACS buffer (0.2% (w/v) BSA in PBS). Cells were then pelleted, fixed, and permeabilized with Cytofix/Cytoperm Fixation/Permeabilization Solution (BD Biosciences) for 20 min at 4 °C, and washed twice in BD Perm/Wash™ Buffer. We incubated cells with APC-conjugated mouse anti-NKX6-1 (BD Biosciences) and PE-conjugated rabbit anti-INS (Cell Signaling Technology) antibody in 50 µl BD Perm/Wash™ Buffer for 1 hour at 4 °C. Following three washes in BD Perm/Wash™ Buffer, cells were analyzed on a FACSCanto II (BD Biosciences) cytometer.

*Immunofluorescence staining and quantification of immunofluorescence signal*

hESC-derived cell aggregates were washed twice with PBS and then fixed with 4% paraformaldehyde in PBS for 30 min at room temperature. Following three washes in PBS, aggregates were incubated in 30% sucrose at 4 °C overnight, frozen in Optimal Cutting Temperature Compound (Sakura Finetek USA), and sectioned at 10 µm with a CM3050S cryostat. Sections were washed with PBS, permeabilized, and blocked with Permeabilization/Blocking Buffer for 1 h at room temperature. Primary and secondary antibodies were diluted in Permeabilization/Blocking Buffer. We incubated sections overnight at 4°C with primary antibodies, and then secondary antibodies for 30 min at room temperature. The following primary antibodies were used: mouse anti-NKX6-1 (LifeSpan BioSciences, 1:250), guinea pig anti-INS (Dako, 1:1000). Secondary antibodies (1:1000) were Cy3-, Alexafluor488-conjugated antibodies raised in donkey against mouse and guinea pig (Jackson Immuno Research Laboratories). We acquired images on a Zeiss Axio-Observer-Z1 microscope with a Zeiss AxioCam digital camera.

*mRNA sequencing*

For each clone, we collected aggregates from two independent batches of differentiation and lysed them in RLT Buffer. We then extracted total RNA using the RNeasy Micro Kit (QIAGEN) following the manufacturer’s instructions. mRNA libraries were prepared using KAPA mRNA Hyper Prep kit (KAPA) and single-end 50 bp reads were sequenced using HiSeq4000 (Illumina). We used STAR to map reads to the hg19 genome, allowing for up to 10 mismatches. We retained reads aligned uniquely to one genomic location for subsequent analysis. We then created input count files for DESeq2 with htseq-count from the HTSeq python package and tested for differential gene expression using DESeq2 v1.10.1 with default parameters, incorporating differentiation batch as a technical covariate in our analysis. We considered genes with an FDR<.1 as significantly differentially expressed.

*Insulin content measurement*

We washed hESC-derived cell aggregates with PBS, resuspended in 50l of 0.1% SDS TE buffer and sonicated for 3 cycles of 30 sec on/ 30 sec off each using a Bioruptor on the high setting. We then immersed the lysate in a solution of 2% HCl and 80% ethanol overnight at 4°C and centrifuged at max speed for 10 min at 4°C. We collected the supernatant and measured insulin content using a human insulin ELISA kit (ALPCO). We resuspended the pellets in 50µl TE buffer and measured DNA content with Nanodrop. We normalized insulin content to DNA content.

# Figure Legends

## Main Figures

**Figure 1. Pancreatic islet cell type accessible chromatin defined using snATAC-seq.** (A) Clustering of accessible chromatin profiles from 14.2k pancreatic islet cells identifies 13 distinct clusters. Cells are plotted using the first two UMAP components, and clusters are are assigned cell type identities based on promoter accessibility of known marker genes for each cell type. (B) Promoter accessibility in a 1kb window around the TSS for selected endocrine and non-endocrine marker genes for each profiled cell. A cell is colored if it had promoter accessibility for a given marker gene, and otherwise is grey. (C) Genome browser plots showing loci around hormone genes for endocrine islet cell types, including *GCG* (alpha), *INS-IGF2* (beta), *SST* (delta), and *PPY* (gamma), and aggregate read density for cells within each cell type cluster. The promoter region for each gene is highlighted, and the number of cells for each cell type cluster is listed in parenthesis. (D) Spearman correlation between t-statistics of marker genes based on promoter accessibility (snATAC-seq) or gene expression (scRNA-seq) using the top 100 most specific promoters from each islet snATAC-seq cluster. (E) Normalized chromVAR motif enrichment values for 111 TF sequence motifs that have variable activity across clusters. We collapsed multiple clusters for each cell type into a single cluster. Position weight matrices and names are shown for sequence motifs for TF families enriched across different endocrine and non-endocrine cell types. Enrichment z-scores for FOXA and PDX in each cell projected onto UMAP coordinates.

**Figure 2. Heterogeneity in alpha and beta cell accessible chromatin and regulatory programs.** (A) Gene promoters with significantly different chromatin accessibility across sub-clusters of alpha cells (left) and beta cells (right). Among genes with increased promoter accessibility in ‘alpha 1’ and ‘beta 1’ sub-clusters include *GCG* (glucagon) for alpha cells and *INS* (insulin) for beta cells, as well as genes involved in glucose response and hormone secretion such as *G6PC2, GCK* and *SLC30A8.* Conversely, genes with increased promoter accessibility in ‘alpha 2’ and ‘beta 2’ sub-clusters included transcription factors such as *FOSL2* and *ATF3.* (B) Genes with differential promoter accessibility for alpha and beta cell sub-clusters were enriched for (left) genes in different beta cell sub-sets identified in an islet single cell gene expression study ( sub. 1-4), (middle) gene sets with differential promoter accessibility across alpha and beta cells, and (right) Gene Ontology terms for biological processes related to glucose response, hormone secretion, stress response, insulin signaling and cell cycle. (C) Normalized chromVAR enrichments for 111 TF motifs grouped by alpha and beta cell sub-cluster. We observed motifs enriched for different sub-clusters including RFX family members (RFX2-5) for alpha 1 and beta 1, and FOS/JUN family members for alpha 2 and beta 2. Enrichment z-scores of RFX3 and FOS::JUN motifs for each cell are plotted on UMAP coordinates, and the boxplots below each plot show enrichment values for each alpha and beta sub-cluster. (D) Ordering of alpha and beta cells on a trajectory using high *GCG/INS* as the anchor point with Cicero. Plots show cells binned across this trajectory from left to right, where the top shows the percentage of cells in state 1/state 2 in a given bin and the bottom shows chromVAR enrichments for motifs in a given bin. (E) Motifs in the FOS/JUN family show increasing enrichment across the alpha and beta cell trajectory. Specific genes in the FOS/JUN family with matching patterns of promoter accessibility (Spearman correlation>.9) are highlighted (in blue and starred), such as FOSL1, FOSL2, and JUND.

**Figure 3. Enrichment of islet single cell accessible chromatin for diabetes and related trait genetic association data.** (A) Cell type specific LD score regression enrichment Z-scores for diabetes-related quantitative endophenotypes (top), type 1 and 2 diabetes (middle), and control traits (bottom) for islet snATAC-seq clusters. \*\*FDR<.01 \*FDR<.1. (B) Single cell enrichment Z-scores from polyTest for fasting glucose, type 2 diabetes, major depressive disorder, and lupus projected onto UMAP coordinates (left panels), boxplot showing z-score enrichment distribution per cell type and state (middle panels), and z-score enrichment distribution split into 10 bins based on beta cell trajectory value (right panels). (C) Genome browser shot of the *DGKB* locus which is associated with both type 2 diabetes and fasting glucose level. Candidate causal variants fall in an enhancer with higher accessibility in beta 1 and with dynamic chromatin accessibility decreasing across the beta cell trajectory. (D) Correlation between single cell FG level enrichments and TF motif enrichments from chromVAR across all 14.2k cells (left) and 7.2k beta cells (right). Across all cells, FG level is positively correlated with beta cell TFs such as PDX1 and NKX6-1 and negatively correlated with alpha cell TFs such as GATA. Within beta cells, FG level is positively correlated with TF motifs enriched in the beta 1 state such as RFX, NRL/MAF, and FOXA1, and negatively correlated with TF motifs enriched in the beta 2 state such as JUND and NFE2.

**Figure 4. Genetic variants with islet cell type- and state-specific effects on chromatin accessibility.** (A) Percentage of variants that had significant deltaSVM predictions at FDR<.1 for the reference (ref) or alternate (alt) allele in different endocrine cell types and states. (B) Spearman correlation comparing deltaSVM score to chromatin accessibility allelic imbalance Z-scores using variants with significant deltaSVM predictions for alpha and beta subtypes. (C) Luciferase gene reporter assays of five fine-mapped T2D variants with predicted beta cell effects in MIN6 cells. All tested variants had significant effects in reporter assays and were directionally consistent with deltaSVM effects (highlighted with a circle around the predicted allele). 2-sided Student’s T-test \*P<.05 \*\*P<.01 \*\*\*P<.001. (D) Enrichment of ensemble islet caQTLs for SNPs with significant deltaSVM effects in alpha and beta cells (left) and categorized based on shared, cell type- and state-specific deltaSVM effects on alpha and beta cells (right). Fisher’s exact test. ns, not significant. (E) Variants with predicted cell type- and state-specific effects on alpha and beta cells, where size indicates magnitude of the Z-score and color indicates the effect allele. Ref=blue, alt=red (left). TF motif families enriched in sequences surrounding the effect allele compared to the non-effect allele for each variant category (middle). Promoter accessibility patterns of genes in in enriched TF motif families. TFs with promoter accessibility patterns that match TF motif enrichment patterns are highlighted in blue and starred (right). (F) Enrichment of low frequency and rare variants with significant effects on islet chromatin for T2D association at different p-value thresholds. \*P<.05. (G) Low-frequency T2D-associated variant rs78840640 at the *IGF2BP3* signal has a high causal probability (PPA=0.33), overlaps peaks in both beta cell states, and is predicted to have allelic effects in beta cells.

**Figure 5. Chromatin co-accessibility links cell type enhancers and diabetes risk variants to target genes.** (A) Distance-matched odds that beta cell co-accessibility links overlap islet pCHi-C chromatin loops at different co-accessibility threshold bins.(B) Beta cell (top) and alpha cell (middle) co-accessibility between pairs of accessible chromatin sites and high-confidence promoter capture Hi-C interactions from bulk islets (bottom) anchored at the *PDX1* promoter. (C) Beta cell co-accessibility anchored on an enhancer within *KCNQ1* harboring causal T2D variant rs231361 (PPA=1) shows distal links to the insulin promoter as well as other non-promoter sites. This enhancer has an accessible peak call in the INS-high beta cell state but not the INS-low state and has dynamic accessibility across the beta cell state trajectory. rs231361 disrupts a sequence motif for *RFX,* which itself is enriched in INS-high beta cells, has dynamic enrichment across the beta cell trajectory, and is predicted to have allelic effects on INS-high beta cells. We performed CRISPR/Cas9-mediated deletion of the 2.6kb genomic region flanking this enhancer (highlighted in grey) in hESCs (*KCNQ1*∆Enh). (D) Differential expression analysis of genes within 2Mb of the *KCNQ1* enhancer in beta cell stage cultures (day 28) from *KCNQ1*∆Enh (n=6; 3 clones each differentiated two times) and control (n=2; 1 clone differentiated two times) hESC clones. *INS* and *CDKN1C* expression is significantly reduced in *KCNQ1*∆Enh compared to control cells, while other genes in the region show no significant difference in expression. Data are shown as transcripts per million (TPM). (E) Representative immunofluorescence staining for INS (green), NKX6-1 (red), and DAPI staining (blue) on beta cell stage *KCNQ1*∆Enh and control aggregates. Scale bar, 50m. (F) Histogram showing insulin fluorescence intensity by flow cytometry (left panel) and quantification of insulin median fluorescence intensity (MFI, right panel) in beta cell stage cultures from *KCNQ1*∆Enh (n=9; 3 clones each differentiated three times) and control (n=6; 2 clones each differentiated three times) cells. (G) Insulin content in beta cell stage cultures from *KCNQ1*∆Enh (n=9; 3 clones each differentiated three times) and control (n=6; 2 clones differentiated three times) clones. Data are shown as mean ± SEM. \* p < 0.05, \*\*\* p<0.001, ns, not significant by T-test.

# Acknowledgements

Support for this work was provided by NIH funding R01DK114650 to K.G., U01DK120429 to K.G. and M.S., ...

# Conflict of Interest

The authors have no conflict of interest to disclose.

# Author Contributions

K.J.G., D.U.G, and M.S.conceived of and supervised the research in the study; K.J.G., D.U.G., M.S., J.C., C.Z, and Z.C.wrote the manuscript; J.C. performed analyses of single cell and genetic data; C.Z. performed hESC experiments; Z.C. performed analyses of single cell data; J.Y.H. performed single cell assays and genotyping; M.S. and J.W. performed hESC experiments; S.H., A.D. and M.O. performed gene reporter experiments; Y.S. contributed analyses of hESC data; Y.S. and P.K. developed and processed data for the epigenome database; R.F. contributed analyses of single cell data; S.P. contributed to the development of single cell assays;

# References

1. Maurano, M. T. *et al.* Systematic Localization of Common Disease-Associated Variation in Regulatory DNA. *Science* **337**, 1190–1195 (2012).

2. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218 (2013).

3. Cusanovich, D. A. *et al.* Multiplex Single Cell Profiling of Chromatin Accessibility by Combinatorial Cellular Indexing. *Science* **348**, 910–914 (2015).

4. Buenrostro, J. D. *et al.* Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* **523**, 486–490 (2015).

5. Preissl, S. *et al.* Single-nucleus analysis of accessible chromatin in developing mouse forebrain reveals cell-type-specific transcriptional regulation. *Nat. Neurosci.* **21**, 432–439 (2018).

6. Pliner, H. A. *et al.* Cicero Predicts cis-Regulatory DNA Interactions from Single-Cell Chromatin Accessibility Data. *Mol. Cell* **71**, 858–871.e8 (2018).

7. Litzenburger, U. M. *et al.* Single-cell epigenomic variability reveals functional cancer heterogeneity. *Genome Biol.* **18**, 15 (2017).

8. Buenrostro, J. D. *et al.* Integrated Single-Cell Analysis Maps the Continuous Regulatory Landscape of Human Hematopoietic Differentiation. *Cell* **173**, 1535–1548.e16 (2018).

9. van der Meulen, T. *et al.* Urocortin3 mediates somatostatin-dependent negative feedback control of insulin secretion. *Nat. Med.* **21**, 769–776 (2015).

10. Caicedo, A. Paracrine and autocrine interactions in the human islet: more than meets the eye. *Semin. Cell Dev. Biol.* **24**, 11–21 (2013).

11. DiGruccio, M. R. *et al.* Comprehensive alpha, beta and delta cell transcriptomes reveal that ghrelin selectively activates delta cells and promotes somatostatin release from pancreatic islets. *Mol. Metab.* **5**, 449–458 (2016).

12. Ranjan, A. K., Joglekar, M. V. & Hardikar, A. A. Endothelial cells in pancreatic islet development and function. *Islets* **1**, 2–9 (2009).

13. Tsui, H. *et al.* Islet glia, neurons, and beta cells. *Ann. N. Y. Acad. Sci.* **1150**, 32–42 (2008).

14. Radenkovic, M. *et al.* Characterization of resident lymphocytes in human pancreatic islets. *Clin. Exp. Immunol.* **187**, 418–427 (2017).

15. Steiner, D. J., Kim, A., Miller, K. & Hara, M. Pancreatic islet plasticity: interspecies comparison of islet architecture and composition. *Islets* **2**, 135–145 (2010).

16. Dorrell, C. *et al.* Human islets contain four distinct subtypes of β cells. *Nat. Commun.* **7**, ncomms11756 (2016).

17. Xin, Y. *et al.* Pseudotime Ordering of Single Human β-Cells Reveals States of Insulin Production and Unfolded Protein Response. *Diabetes* db180365 (2018). doi:10.2337/db18-0365

18. Korsunsky, I. *et al.* Fast, sensitive, and accurate integration of single cell data with Harmony. *bioRxiv* 461954 (2018). doi:10.1101/461954

19. Bader, E. *et al.* Identification of proliferative and mature β-cells in the islets of Langerhans. *Nature* **535**, 430–434 (2016).

20. Liu, J. S. E. & Hebrok, M. All mixed up: defining roles for β-cell subtypes in mature islets. *Genes Dev.* **31**, 228–240 (2017).

21. Ackermann, A. M., Wang, Z., Schug, J., Naji, A. & Kaestner, K. H. Integration of ATAC-seq and RNA-seq identifies human alpha cell and beta cell signature genes. *Mol. Metab.* **5**, 233–244 (2016).

22. Arda, H. E. *et al.* A Chromatin Basis for Cell Lineage and Disease Risk in the Human Pancreas. *Cell Syst.* **7**, 310–322.e4 (2018).

23. Mahajan, A. *et al.* Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. *Nat. Genet.* **50**, 1505–1513 (2018).

24. Wood, A. R. *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* **66**, 2296–2309 (2017).

25. Morris, A. P. *et al.* Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat. Genet.* **44**, 981–990 (2012).

26. Dupuis, J. *et al.* New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nat. Genet.* **42**, 105–116 (2010).

27. Thurner, M. *et al.* Integration of human pancreatic islet genomic data refines regulatory mechanisms at Type 2 Diabetes susceptibility loci. *eLife* **7**,

28. Fuchsberger, C. *et al.* The genetic architecture of type 2 diabetes. *Nature* **536**, 41–47 (2016).

29. Gaulton, K. J. *et al.* Genetic fine mapping and genomic annotation defines causal mechanisms at type 2 diabetes susceptibility loci. *Nat. Genet.* **47**, 1415–1425 (2015).

30. Pasquali, L. *et al.* Pancreatic islet enhancer clusters enriched in type 2 diabetes risk-associated variants. *Nat. Genet.* **46**, 136–143 (2014).

31. Gaulton, K. J. *et al.* A map of open chromatin in human pancreatic islets. *Nat. Genet.* **42**, 255–259 (2010).

32. Greenwald, W. W. *et al.* Pancreatic islet chromatin accessibility and conformation defines distal enhancer networks of type 2 diabetes risk. *bioRxiv* 299388 (2018). doi:10.1101/299388

33. Varshney, A. *et al.* Genetic regulatory signatures underlying islet gene expression and type 2 diabetes. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 2301–2306 (2017).

34. Khetan, S. *et al.* Type 2 Diabetes-Associated Genetic Variants Regulate Chromatin Accessibility in Human Islets. *Diabetes* **67**, 2466–2477 (2018).

35. Haghverdi, L., Lun, A. T. L., Morgan, M. D. & Marioni, J. C. Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. *Nat. Biotechnol.* **36**, 421–427 (2018).

36. Segerstolpe, Å. *et al.* Single-Cell Transcriptome Profiling of Human Pancreatic Islets in Health and Type 2 Diabetes. *Cell Metab.* **24**, 593–607 (2016).

37. Baron, M. *et al.* A Single-Cell Transcriptomic Map of the Human and Mouse Pancreas Reveals Inter- and Intra-cell Population Structure. *Cell Syst.* **3**, 346–360.e4 (2016).

38. George, D., Ahrens, P. & Lambert, S. Satellite glial cells represent a population of developmentally arrested Schwann cells. *Glia* **66**, 1496–1506 (2018).

39. Galvagni, F. *et al.* CD93 and dystroglycan cooperation in human endothelial cell adhesion and migration adhesion and migration. *Oncotarget* **7**, 10090–10103 (2016).

40. Schep, A. N., Wu, B., Buenrostro, J. D. & Greenleaf, W. J. chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. *Nat. Methods* **14**, 975–978 (2017).

41. Khan, A. *et al.* JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework. *Nucleic Acids Res.* **46**, D1284 (2018).

42. Wilson, M. E., Scheel, D. & German, M. S. Gene expression cascades in pancreatic development. *Mech. Dev.* **120**, 65–80 (2003).

43. Conrad, E. *et al.* The MAFB transcription factor impacts islet α-cell function in rodents and represents a unique signature of primate islet β-cells. *Am. J. Physiol. Endocrinol. Metab.* **310**, E91–E102 (2016).

44. Katoh, M. C. *et al.* MafB Is Critical for Glucagon Production and Secretion in Mouse Pancreatic α Cells In Vivo. *Mol. Cell. Biol.* **38**, (2018).

45. Nishimura, W., Takahashi, S. & Yasuda, K. MafA is critical for maintenance of the mature beta cell phenotype in mice. *Diabetologia* **58**, 566–574 (2015).

46. Ozato, K., Tailor, P. & Kubota, T. The interferon regulatory factor family in host defense: mechanism of action. *J. Biol. Chem.* **282**, 20065–20069 (2007).

47. De Val, S. & Black, B. L. Transcriptional Control of Endothelial Cell Development. *Dev. Cell* **16**, 180–195 (2009).

48. Lawlor, N. *et al.* Single-cell transcriptomes identify human islet cell signatures and reveal cell-type–specific expression changes in type 2 diabetes. *Genome Res.* **27**, 208–222 (2017).

49. Benninger, R. K. P. & Hodson, D. J. New Understanding of β-Cell Heterogeneity and In Situ Islet Function. *Diabetes* **67**, 537–547 (2018).

50. Trapnell, C. *et al.* Pseudo-temporal ordering of individual cells reveals dynamics and regulators of cell fate decisions. *Nat. Biotechnol.* **32**, 381 (2014).

51. Hindorff, L. A. *et al.* Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 9362–9367 (2009).

52. Parker, S. C. J. *et al.* Chromatin stretch enhancer states drive cell-specific gene regulation and harbor human disease risk variants. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 17921–17926 (2013).

53. Aylward, A., Chiou, J., Okino, M.-L., Kadakia, N. & Gaulton, K. J. Shared genetic risk contributes to type 1 and type 2 diabetes etiology. *Hum. Mol. Genet.* doi:10.1093/hmg/ddy314

54. Strawbridge, R. J. *et al.* Genome-wide association identifies nine common variants associated with fasting proinsulin levels and provides new insights into the pathophysiology of type 2 diabetes. *Diabetes* **60**, 2624–2634 (2011).

55. Manning, A. K. *et al.* A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. *Nat. Genet.* **44**, 659–669 (2012).

56. Locke, A. E. *et al.* Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518**, 197–206 (2015).

57. Saxena, R. *et al.* Genetic variation in GIPR influences the glucose and insulin responses to an oral glucose challenge. *Nat. Genet.* **42**, 142–148 (2010).

58. Wheeler, E. *et al.* Impact of common genetic determinants of Hemoglobin A1c on type 2 diabetes risk and diagnosis in ancestrally diverse populations: A transethnic genome-wide meta-analysis. *PLoS Med.* **14**, e1002383 (2017).

59. Hou, L. *et al.* Genome-wide association study of 40,000 individuals identifies two novel loci associated with bipolar disorder. *Hum. Mol. Genet.* **25**, 3383–3394 (2016).

60. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421–427 (2014).

61. de Lange, K. M. *et al.* Genome-wide association study implicates immune activation of multiple integrin genes in inflammatory bowel disease. *Nat. Genet.* **49**, 256–261 (2017).

62. Bentham, J. *et al.* Genetic association analyses implicate aberrant regulation of innate and adaptive immunity genes in the pathogenesis of systemic lupus erythematosus. *Nat. Genet.* **47**, 1457–1464 (2015).

63. Jin, Y. *et al.* Genome-wide association studies of autoimmune vitiligo identify 23 new risk loci and highlight key pathways and regulatory variants. *Nat. Genet.* **48**, 1418–1424 (2016).

64. Cordell, H. J. *et al.* International genome-wide meta-analysis identifies new primary biliary cirrhosis risk loci and targetable pathogenic pathways. *Nat. Commun.* **6**, 8019 (2015).

65. Wray, N. R. *et al.* Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat. Genet.* **50**, 668–681 (2018).

66. Lambert, J. C. *et al.* Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer’s disease. *Nat. Genet.* **45**, 1452–1458 (2013).

67. Finucane, H. K. *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015).

68. Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).

69. Lee, D. *et al.* A method to predict the impact of regulatory variants from DNA sequence. *Nat. Genet.* **47**, 955–961 (2015).

70. 1000 Genomes Project Consortium *et al.* A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).

71. Khetan, S. *et al.* Chromatin accessibility profiling uncovers genetic- and T2D disease state-associated changes in cis-regulatory element use in human islets. *bioRxiv* 192922 (2017). doi:10.1101/192922

72. Wingender, E., Schoeps, T., Haubrock, M. & Dönitz, J. TFClass: a classification of human transcription factors and their rodent orthologs. *Nucleic Acids Res.* **43**, D97-102 (2015).

73. Shlyueva, D., Stampfel, G. & Stark, A. Transcriptional enhancers: from properties to genome-wide predictions. *Nat. Rev. Genet.* **15**, 272–286 (2014).

74. Schmitt, A. D., Hu, M. & Ren, B. Genome-wide mapping and analysis of chromosome architecture. *Nat. Rev. Mol. Cell Biol.* **17**, 743–755 (2016).

75. Thurman, R. E. *et al.* The accessible chromatin landscape of the human genome. *Nature* **489**, 75–82 (2012).

76. Miguel-Escalada, I. *et al.* Human pancreatic islet 3D chromatin architecture provides insights into the genetics of type 2 diabetes. *bioRxiv* 400291 (2018). doi:10.1101/400291

77. Jian, X. & Felsenfeld, G. Insulin promoter in human pancreatic β cells contacts diabetes susceptibility loci and regulates genes affecting insulin metabolism. *Proc. Natl. Acad. Sci. U. S. A.* **115**, E4633–E4641 (2018).

78. Holmkvist, J. *et al.* The type 2 diabetes associated minor allele of rs2237895 KCNQ1 associates with reduced insulin release following an oral glucose load. *PloS One* **4**, e5872 (2009).

79. Jonsson, A. *et al.* A variant in the KCNQ1 gene predicts future type 2 diabetes and mediates impaired insulin secretion. *Diabetes* **58**, 2409–2413 (2009).

80. Tan, J. T. *et al.* Genetic Variation in KCNQ1 Associates With Fasting Glucose andβ-Cell Function. *Diabetes* **58**, 1445–1449 (2009).

81. Yasuda, K. *et al.* Variants in KCNQ1 are associated with susceptibility to type 2 diabetes mellitus. *Nat. Genet.* **40**, 1092–1097 (2008).

82. Kanai, M. *et al.* Genetic analysis of quantitative traits in the Japanese population links cell types to complex human diseases. *Nat. Genet.* **50**, 390–400 (2018).

83. Rezania, A. *et al.* Reversal of diabetes with insulin-producing cells derived in vitro from human pluripotent stem cells. *Nat. Biotechnol.* **32**, 1121–1133 (2014).

84. Suzuki, K. *et al.* Identification of 28 new susceptibility loci for type 2 diabetes in the Japanese population. *Nat. Genet.* **51**, 379–386 (2019).

85. Camunas-Soler, J. *et al.* Pancreas patch-seq links physiologic dysfunction in diabetes to single-cell transcriptomic phenotypes. *bioRxiv* 555110 (2019). doi:10.1101/555110

86. Scott, R. A. *et al.* Large-scale association analyses identify new loci influencing glycemic traits and provide insight into the underlying biological pathways. *Nat. Genet.* **44**, 991–1005 (2012).

87. Dooley, J. *et al.* Genetic predisposition for beta cell fragility underlies type 1 and type 2 diabetes. *Nat. Genet.* **48**, 519–527 (2016).

88. Kong, Y. *et al.* CDKN2A/B T2D Genome-Wide Association Study Risk SNPs Impact Locus Gene Expression and Proliferation in Human Islets. *Diabetes* **67**, 872–884 (2018).

89. Chandra, V. *et al.* RFX6 regulates insulin secretion by modulating Ca2+ homeostasis in human β cells. *Cell Rep.* **9**, 2206–2218 (2014).

90. Naya, F. J., Stellrecht, C. M. & Tsai, M. J. Tissue-specific regulation of the insulin gene by a novel basic helix-loop-helix transcription factor. *Genes Dev.* **9**, 1009–1019 (1995).

91. Gao, N. *et al.* Foxa1 and Foxa2 Maintain the Metabolic and Secretory Features of the Mature β-Cell. *Mol. Endocrinol.* **24**, 1594–1604 (2010).

92. Kennedy, J. *et al.* Protective Unfolded Protein Response in Human Pancreatic Beta Cells Transplanted into Mice. *PLoS ONE* **5**, (2010).

93. Jansson, D. *et al.* Glucose controls CREB activity in islet cells via regulated phosphorylation of TORC2. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 10161–10166 (2008).

94. van der Wijst, M. G. P. *et al.* Single-cell RNA sequencing identifies celltype-specific cis-eQTLs and co-expression QTLs. *Nat. Genet.* **50**, 493–497 (2018).

95. Ou, K. *et al.* Targeted demethylation at the CDKN1C/p57 locus induces human β cell replication. *J. Clin. Invest.* **129**, 209–214 (2019).

96. Steck, A. K. & Rewers, M. J. Genetics of Type 1 Diabetes. *Clin. Chem.* **57**, 176–185 (2011).

97. Yang, Y. & Chan, L. Monogenic Diabetes: What It Teaches Us on the Common Forms of Type 1 and Type 2 Diabetes. *Endocr. Rev.* **37**, 190–222 (2016).

98. Travers, M. E. *et al.* Insights into the molecular mechanism for type 2 diabetes susceptibility at the KCNQ1 locus from temporal changes in imprinting status in human islets. *Diabetes* **62**, 987–992 (2013).

99. Picelli, S. *et al.* Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. *Genome Res.* **24**, 2033–2040 (2014).

100. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinforma. Oxf. Engl.* **26**, 589–595 (2010).

101. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinforma. Oxf. Engl.* **25**, 2078–2079 (2009).

102. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15 (2018).

103. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).

104. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).

105. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).

106. Kuleshov, M. V. *et al.* Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **44**, W90-97 (2016).

107. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 15545–15550 (2005).

108. Finucane, H. K. *et al.* Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *Nat. Genet.* **50**, 621–629 (2018).

109. Li, Y. I. *et al.* RNA splicing is a primary link between genetic variation and disease. *Science* **352**, 600–604 (2016).

110. McCarthy, S. *et al.* A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.* **48**, 1279–1283 (2016).

111. Das, S. *et al.* Next-generation genotype imputation service and methods. *Nat. Genet.* **48**, 1284–1287 (2016).

112. van de Geijn, B., McVicker, G., Gilad, Y. & Pritchard, J. K. WASP: allele-specific software for robust molecular quantitative trait locus discovery. *Nat. Methods* **12**, 1061–1063 (2015).

113. Bailey, T. L. *et al.* MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* **37**, W202-208 (2009).

114. Raviram, R. *et al.* 4C-ker: A Method to Reproducibly Identify Genome-Wide Interactions Captured by 4C-Seq Experiments. *PLoS Comput. Biol.* **12**, e1004780 (2016).

115. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).