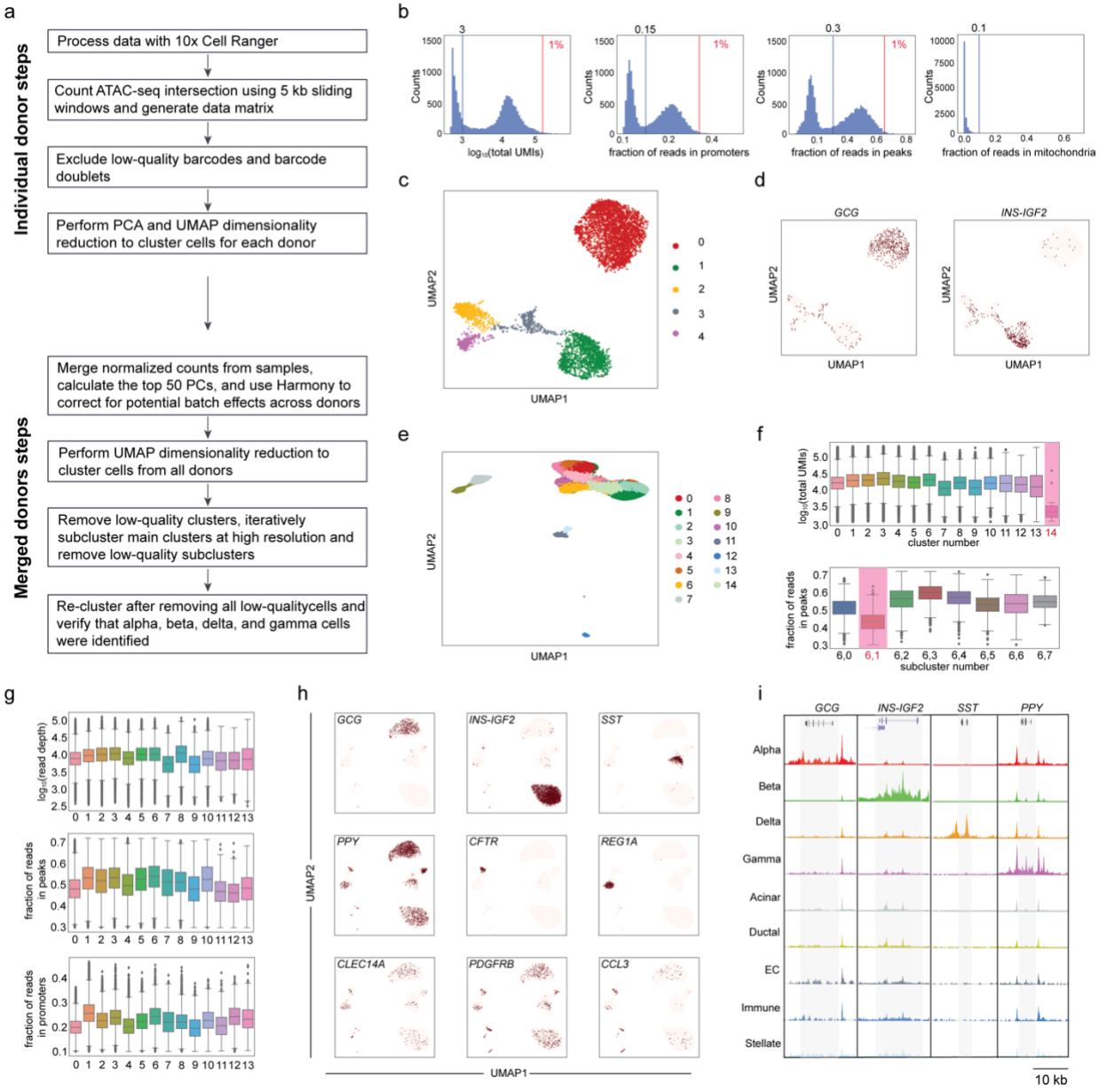
Supplementary Figures

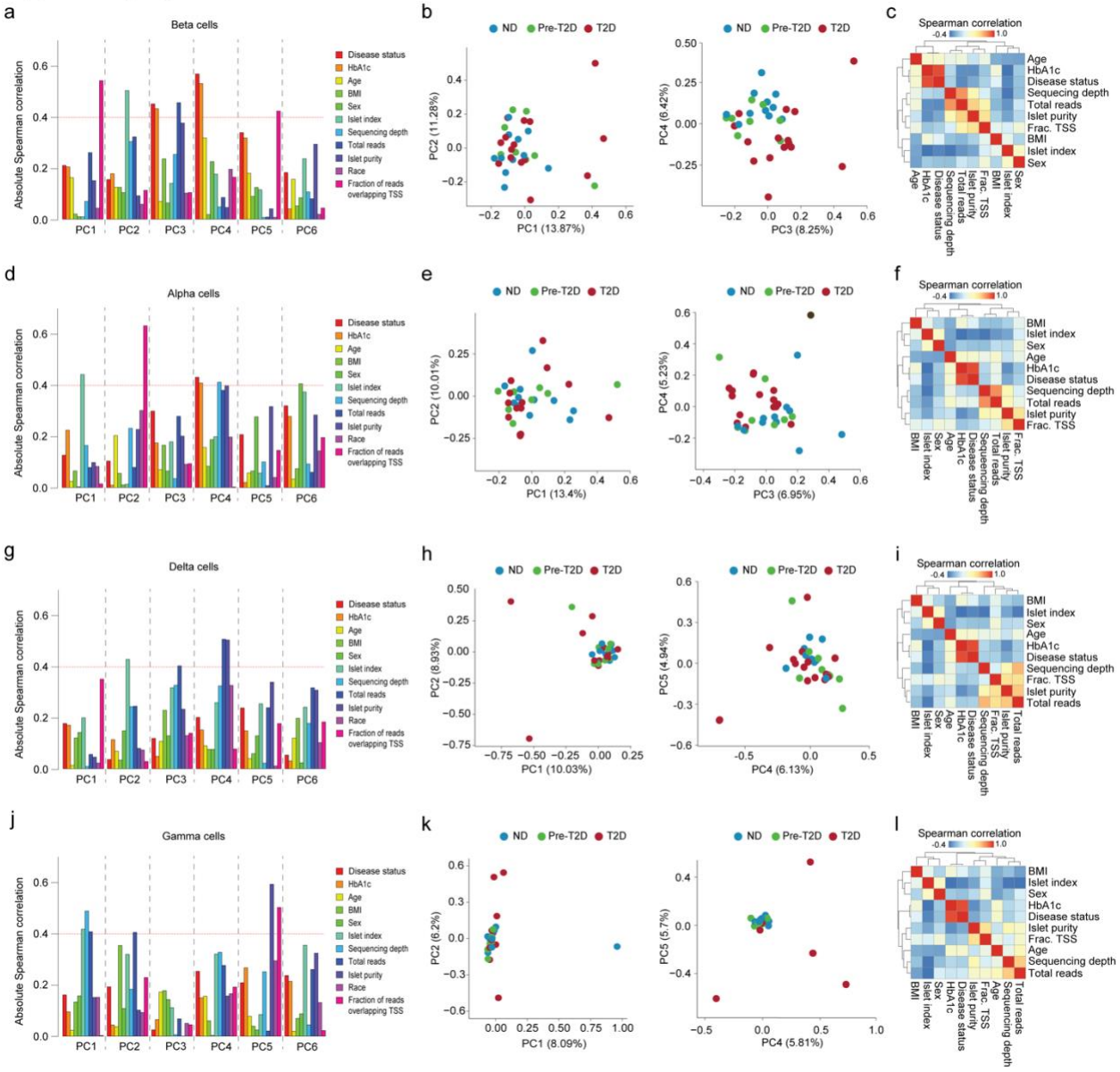
Supplementary Figure 1



Supplementary Figure 1. Quality control of snATAC-seq data.
(a) Steps for snATAC-seq data processing and quality control. **(b)** Representative quality control (QC) metrics for each donor. Log_{10} total UMIs, fraction of reads overlapping promoters, fraction of reads overlapping peaks, and fraction of reads overlapping mitochondria DNA distribution of cells from T2D donor JYH809 as example. Blue vertical lines denote thresholds of 1000 minimal fragment number, 15% fragments overlapping promoters, 30% fragments overlapping peaks, and 10% fraction of reads overlapping mitochondria DNA, respectively. Red vertical lines denote thresholds to identify top 1% barcodes with extremely high total fragment number and fraction of reads overlapping

promoters and peaks, respectively. **(c)** Representative cell clustering from donor JYH809 conducted for each donor. Cells are plotted using the first two UMAP components. **(d)** Promoter chromatin accessibility in a 5 kb window around TSS for endocrine marker genes for each profiled cell from donor JYH809. Total counts normalization and log-transformation were applied. **(e)** Cell clustering of chromatin accessibility profiles from all donors. Cells are plotted using the first two UMAP components. **(f)** Representative low-quality cluster and subcluster. Log_{10} total UMIs distribution of cells from each cluster. Cells in cluster 14 (top, highlighted in red) have significantly lower unique fragment than cells in other clusters. Fraction of reads overlapping peaks distribution of cells from each subcluster of main cluster 6. Cells in subcluster 1 (bottom, highlighted in red) have significantly lower fraction of reads overlapping peaks than cells in other clusters. **(g)** Log_{10} total UMIs, fraction of reads overlapping peaks and fraction of reads in promoters of cells from each cluster in Figure 1b, showing that these metrics do not drive single-cell grouping in UMAP space. **(h)** Promoter chromatin accessibility in a 5 kb window around TSS for selected endocrine and non-endocrine marker genes for each profiled cell (alpha: *GCG*, beta: *INS-IGF2*, delta: *SST*, gamma: *PPY*, acinar: *REG1A*, ductal: *CFTR*, stellate: *PDGFRB*, endothelial: *CLEC14A*, immune: *CCL3*). Total counts normalization and log-transformation were applied. **(i)** Genome browser tracks showing aggregate read density (scaled to uniform $1x10^6$ read depth) for cells within each cell type cluster at hormone gene loci for endocrine islet cell types. The gene body of each gene is highlighted.

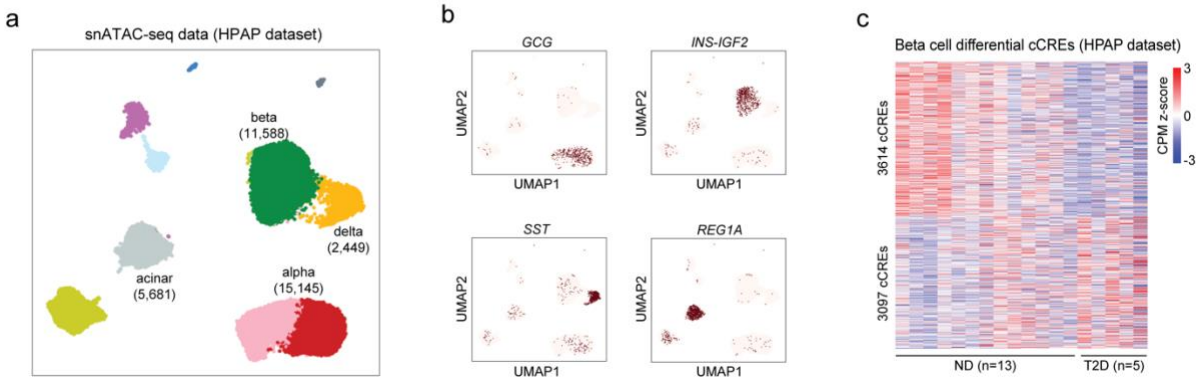
Supplementary Figure 2



Supplementary Figure 2. Identification of factors explaining donor variability in snATAC-seq data.

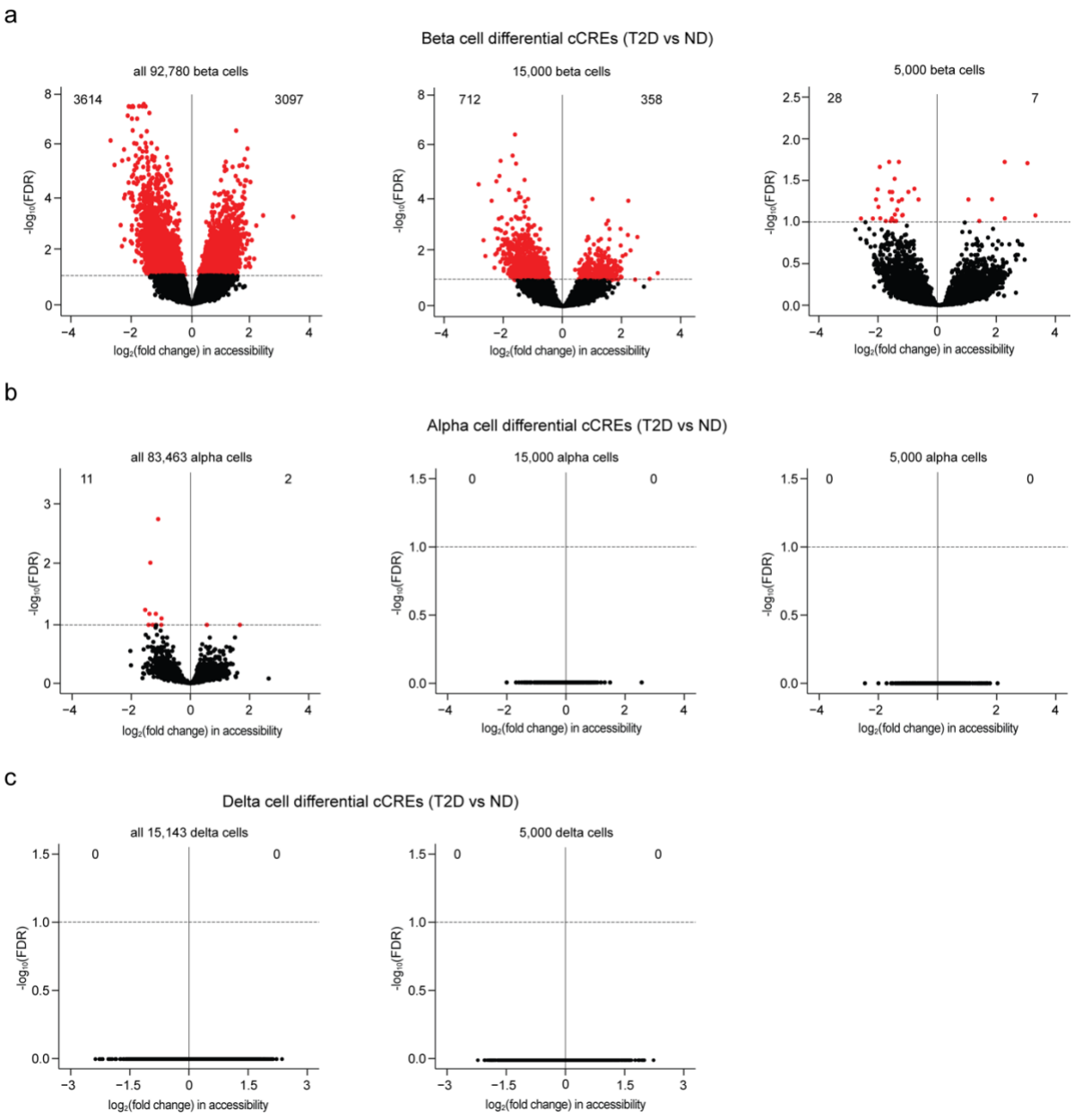
(a,d,g,j) Absolute Spearman correlation coefficient between the first 6 principle components (PCs) and each biological or technical variable in beta (a), alpha (d), delta (g), and gamma (j) cells. An absolute Spearman correlation threshold of 0.4 was used to identify factors having a high correlation with each PC. **(b,e,h,k)** Principal component analysis (PCA) based on cCREs in beta (b), alpha (e), delta (h), and gamma (k) cells from individual non-diabetic (ND, *n*=11), pre-diabetic (pre-T2D, *n*=8), and type 2 diabetic (T2D, *n*=15) donors which are color-coded by disease status. Each donor in the space is defined by the first two principal components (left) and the two principal components (right) that show highest correlation with disease status. **(c,f,i,l)** Pairwise Spearman correlation coefficients between biological or technical variables in beta (c), alpha (f), delta (i), and gamma (l) cells.

Supplementary Figure 3



Supplementary Figure 3. Validation of beta cell T2D-differential cCREs in snATAC-seq data from an independent cohort of donor islets.
(a) Clustering of chromatin accessibility profiles from HPAP human islet snATAC-seq data (non-diabetic (ND), *n*=13; pre-T2D, *n*=2; T2D, *n*=5). Cells are plotted using the first two UMAP components. Clusters are assigned cell type identities based on promoter accessibility of known marker genes (see Supplementary Figure 3b). The number of cells for each cell type cluster is shown in parentheses. **(b)** Promoter chromatin accessibility in a 5 kb window around TSS for selected endocrine and non-endocrine marker genes for each profiled cell (alpha: *GCG*, beta: *INS-IGF2*, delta: *SST*, acinar: *REG1A*). Total counts normalization and log-transformation were applied. **(c)** Heatmap showing chromatin accessibility at differential cCREs identified in Figure 1e in HPAP snATAC-seq data. Columns represent beta cells from each donor (ND, *n*=13; T2D, *n*=5) and all ND and T2D donors with accessibility of peaks normalized by CPM (counts per million).

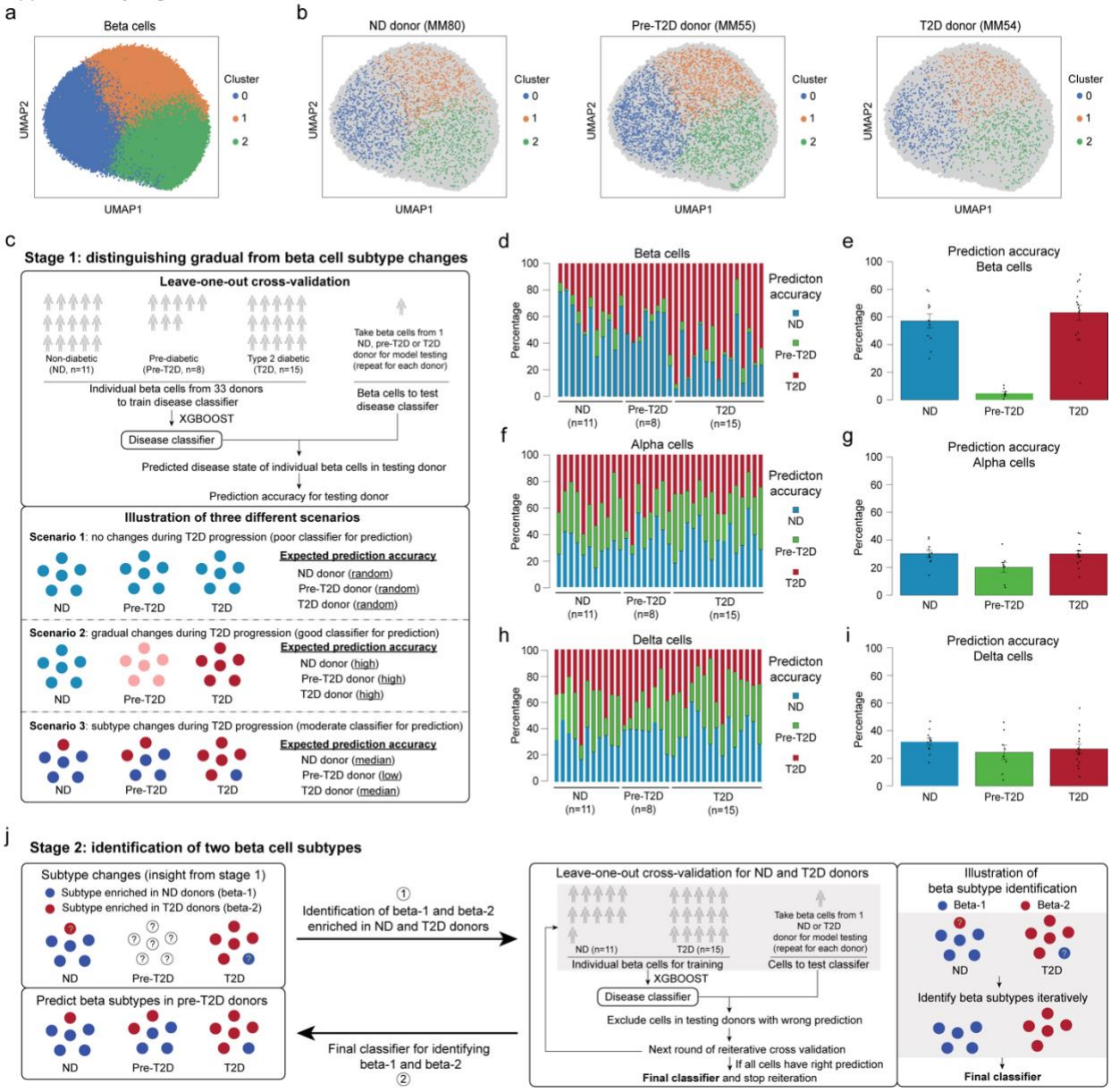
Supplementary Figure 4



Supplementary Figure 4. T2D affects chromatin activity more profoundly in beta cells than in other endocrine cell types.
(a) Volcano plot showing differential cCREs in beta cells between type 2 diabetic (T2D) and non-diabetic (ND) donors. Panels show all beta cells (left), beta cells down-sampled to 15,000 (middle), and 5,000 cells (right). Each dot represents a cCRE. cCREs with FDR < .1 after Benjamini-Hochberg correction (red dots) were considered differentially accessible. **(b)** Volcano plot showing differential cCREs in alpha cells between T2D and ND donors. Panels show all alpha cells (left), alpha cells down-sampled to 15,000 (middle), and 5,000 cells (right). Each dot represents a chromatin accessible cCRE. cCREs with FDR < .1 after Benjamini-Hochberg correction (red dots) were considered

differentially accessible. **(c)** Volcano plot showing differential cCREs in delta cells between T2D and ND donors. Panels show all delta cells (left) and delta cells down-sampled to 5,000 cells (right). Each dot represents a chromatin accessible cCRE. cCREs with FDR < .1 after Benjamini-Hochberg correction (red dots) were considered differentially accessible.

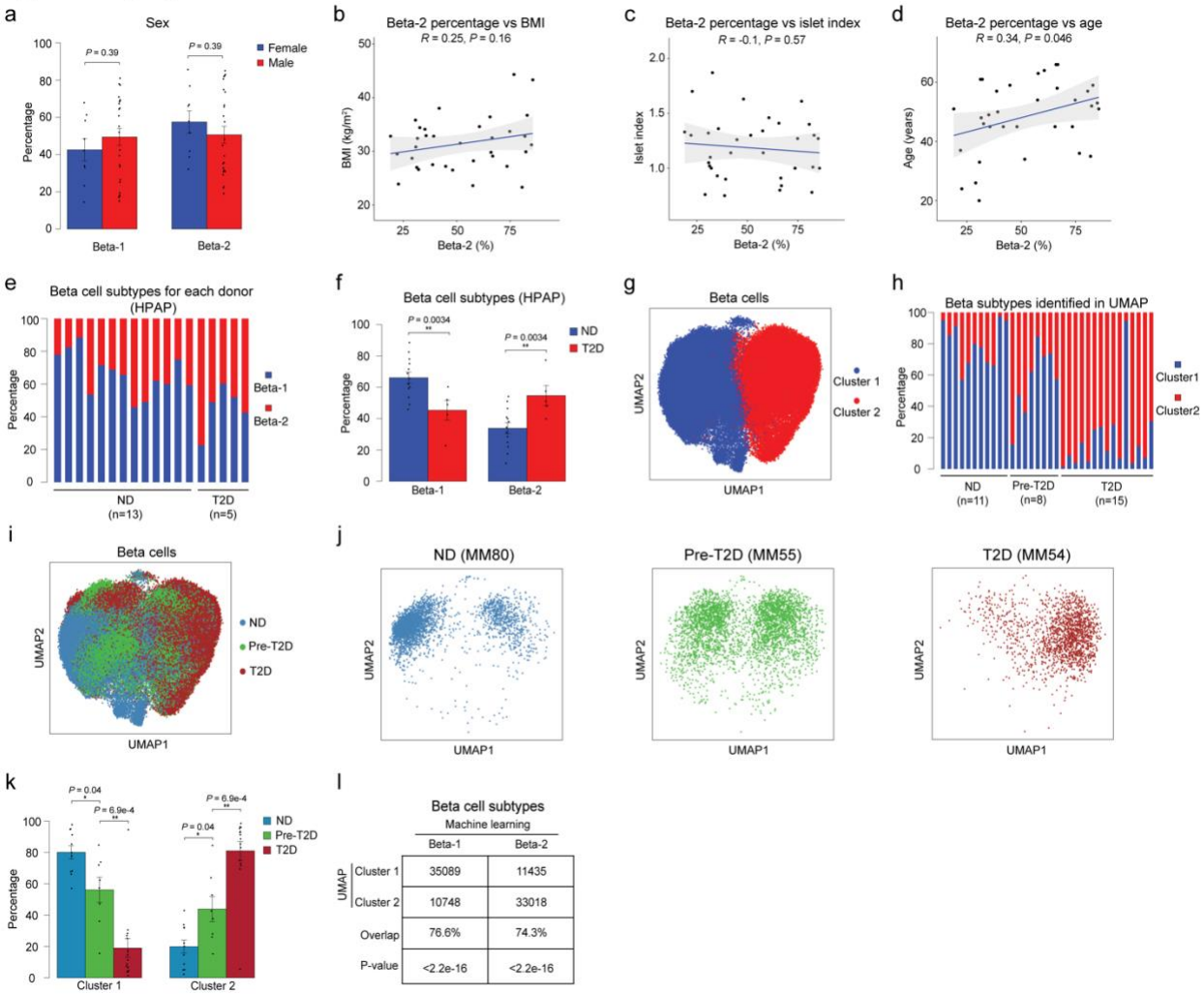
Supplementary Figure 5



Supplementary Figure 5. Machine learning undercovers two beta cell subtypes. **(a)** Clustering of chromatin accessibility profiles from 92,780 beta cells from non-diabetic (ND), prediabetic (pre-T2D) and type 2 diabetic (T2D) donor islets using Scanpy (resolution=0.5). Cells are plotted using the first two UMAP components. **(b)** Position of beta cells from representative ND (MM80), pre-T2D (MM55), and T2D (MM54) donors on the UMPA in panel a. **(c)** Illustration of process for distinguishing gradual from subtype changes in beta cells using machine learning. Possible scenarios for cell changes during T2D progression and expected disease state prediction accuracies for each scenario. In the case of no T2D-associated changes, the prediction accuracy for each disease state would be random (scenario 1), gradual cell state changes would be reflected by high prediction accuracy in each disease state (scenario 2), and subtype changes would be

reflected by median prediction accuracies (scenario 3, here shown for two cell subtypes). **(d, f, h)** Relative abundance of predicted disease state among beta (d), alpha (f), and delta (h) cells from each donor using XGBOOST. Each column represents cells from one donor. **(e, g, i)** Relative abundance of predicted disease state among beta (e), alpha (g), and delta (i) cells in ND, pre-T2D and T2D donor islets. Data are shown as mean ± S.E.M. (*n* = 11 ND, *n* = 8 pre-T2D, *n* = 15 T2D donors), dots denote data points from individual donors. (**j**) Illustration of process for identifying a classifier capable of distinguishing the two beta cell subtypes.

Supplementary Figure 6

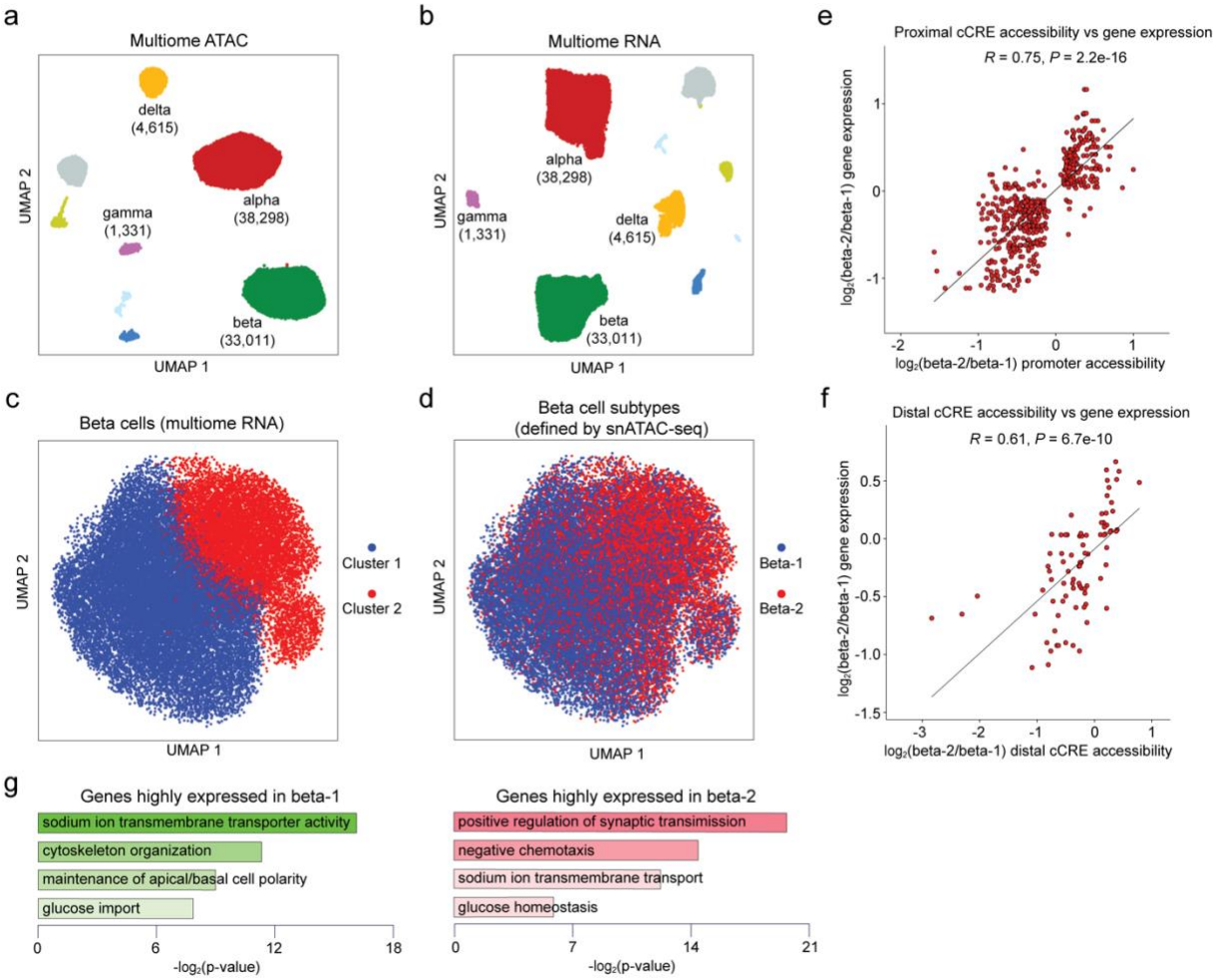


Supplementary Figure 6. Validation of beta cell subtypes using independent data and computational methods.

(a) Relative abundance of beta-1 and beta-2 cells in male and female donor islets. Data are shown as mean ± S.E.M. (*n* = 9 females, *n* = 25 males), dots denote data points from individual donors. ANOVA test with age, disease, BMI, and islet index as covariates. **(b)** Pearson correlation between relative abundance of beta-2 cells and BMI across donors (*n* = 11 ND, *n* = 8 pre-T2D, *n* = 15 T2D donors). **(c)** Pearson correlation between relative abundance of beta-2 cells and islet index across donors. **(d)** Pearson correlation between relative abundance of beta-2 cells and age across donors. **(e)** Relative abundance of beta-1 and beta-2 cells in islet snATAC-seq data from an independent cohort (*n* = 13 ND, *n* = 5 T2D donors). Each column represents cells from one donor. **(f)** Relative abundance of each beta cell subtype in ND and T2D donor islets. Data are shown as mean ± S.E.M (*n* = 13 ND, *n* = 5 T2D donors). ***P* < .01; ANOVA test with age, sex, and BMI as covariates. **(g)** Clustering of chromatin accessibility profiles from 92,780 beta cells from ND, pre-T2D and T2D donors using beta cell differential cCREs between ND and T2D donors from Figure 1e. Cells are plotted using the first two UMAP components. **(h)**

Relative abundance of each beta cell cluster based on UMAP annotation in panel g. Each column represents cells from one donor. **(i)** Position of beta cells from ND, pre-T2D and T2D donors on the UMPA in panel g. **(j)** Position of beta cells from representative ND (MM80), pre-T2D (MM55) and T2D (MM54) donors on the UMPA in panel g. **(k)** Relative abundance of each beta cell cluster in ND, pre-T2D and T2D donor islets. Data are shown as mean ± S.E.M. ($n = 11$ ND, $n = 8$ pre-T2D, $n = 15$ T2D donors). $^{**}P < .01$, $^{*}P < .05$; ANOVA test with age, sex, BMI, and islet index as covariates. **(l)** Overlap between beta cell subtypes identified using machine learning and beta cell clusters from UMPA in panel g. The overlap is 76.6% between cluster 1 and beta-1 and 74.3% between cluster 2 and beta-2. $P < 2.2e\text{-}16$ (Binominal test).

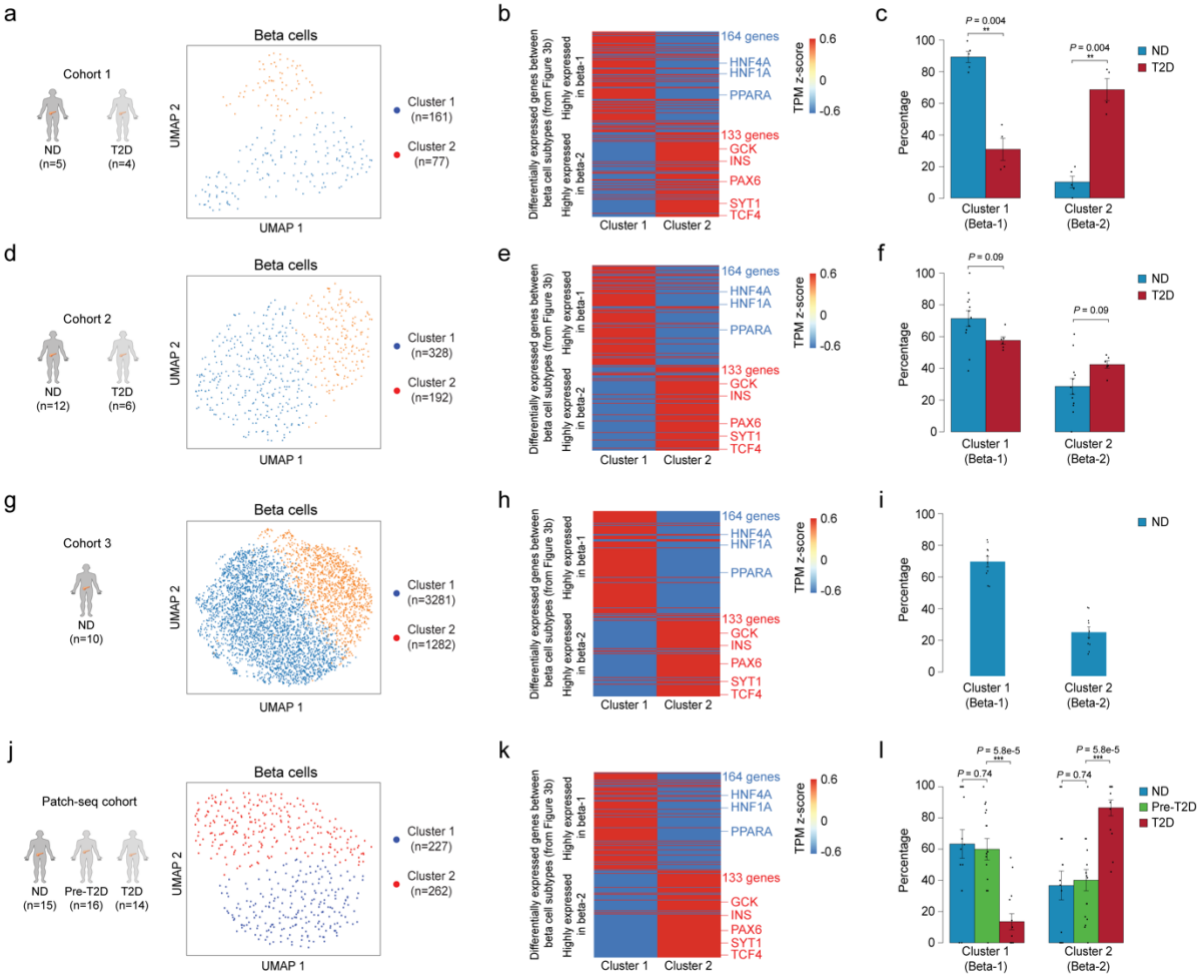
Supplementary Figure 7



Supplementary Figure 7. Validation and characterization of beta cell subtypes using multiome data.
(a) Clustering of chromatin accessibility profiles of cells from multiome data ($n = 6$ ND, $n = 8$ pre-T2D, $n = 6$ T2D). Cells are plotted using the first two UMAP components. Clusters are assigned cell type identities based on promoter accessibility of known marker genes (alpha: *GCG*, beta: *INS-IGF2*, delta: *SST*, gamma: *PPY*). The number of cells for each cell type cluster is shown in parentheses. **(b)** Clustering of gene expression profiles of cells from multiome data ($n = 6$ ND, $n = 8$ pre-T2D, $n = 6$ T2D). Cells are plotted using the first two UMAP components. Clusters are assigned cell type identities based on expression levels of known marker genes (alpha: *GCG*, beta: *INS*, delta: *SST*, gamma: *PPY*). The number of cells for each cell type cluster is shown in parentheses. **(c)** Clustering of gene expression profiles of beta cells from multiome data using genes linked to differential proximal (within ± 5kb of a TSS in GENCODE V19) and distal (based on potential distal cCRE-promoter connections inferred from cicero, see Methods) cCREs between ND and T2D beta cells from Figure 1e. Cells are plotted using the first two UMAP components. **(d)** Plots of beta cell subtypes predicted from chromatin accessibility profiles of beta cells from multiome data by machine learning. **(e)** Correlation between changes

in proximal cCRE (within ± 5kb of a TSS in GENCODE V19) accessibility and gene expression differences between beta-1 and beta-2 cells for differentially expressed genes from Figure 3b. There are 544 proximal cCREs and target gene pairs in total, of which 511 have consistent changes between proximal cCRE accessibility and gene expression. **(f)** Correlation between changes in distal cCRE (potential distal cCRE-promoter connections inferred from cicero, see Methods) accessibility and gene expression differences between beta-1 and beta-2 cells for differentially expressed genes from Figure 3b. There are 85 distal cCREs and target gene pairs in total, of which 72 have consistent changes between distal cCRE accessibility and gene expression. **(g)** Enriched gene ontology terms among genes (see Figure 3b) with higher (proximal or distal) cCRE accessibility and expression in beta-1 compared to beta-2 cells (left) and higher (proximal or distal) cCRE accessibility and expression in beta-2 compared to beta-1 cells (right).

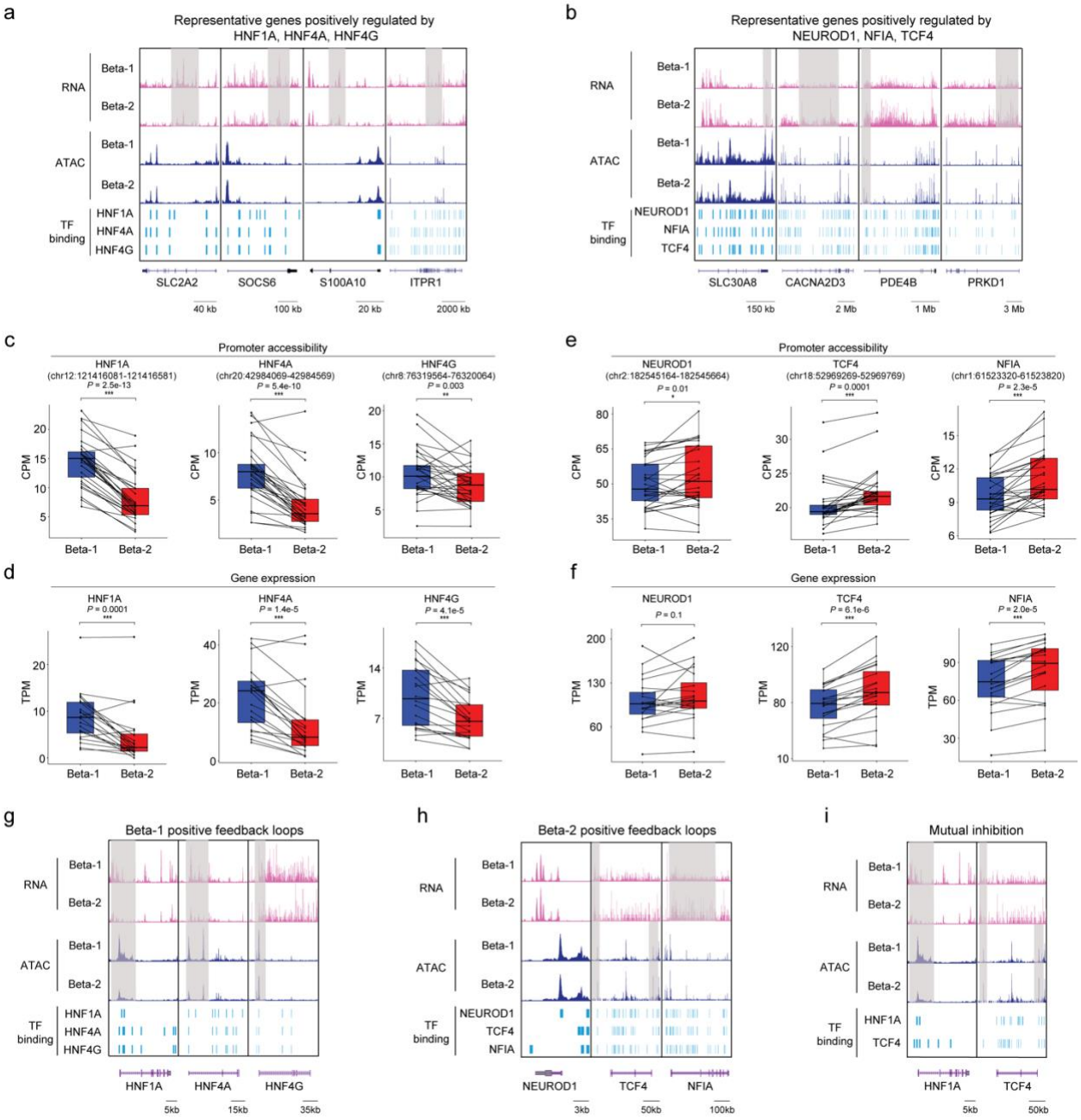
Supplementary Figure 8



Supplementary Figure 8. Beta-1 and beta-2 cell classification analysis in scRNA-seq data from independent cohorts.

(a, d, g, j) Clustering of gene expression profiles of beta cells from cohort 1[5], cohort 2[12], cohort 3[22], and Patch-seq cohort using differentially expressed genes between beta-1 and beta-2 from Figure 3b. Cells are plotted using the first two UMAP components. The number of donors for each cohort and cells for each cell cluster is shown in parentheses. **(b, e, h, k)** Heatmap showing pseudo-bulk expression levels of differentially expressed genes between beta-1 and beta-2 (see Figure 3b) in beta cells from cluster 1 and cluster 2 of cohort 1[5], cohort 2[12], cohort 3[22], and Patch-seq cohort. Expression levels of genes are normalized by TPM (transcripts per million). **(c, f, i, l)** Relative abundance of each beta cell subtype in ND and T2D donor islets in cohort 1[5], cohort 2[12], cohort 3[22], and Patch-seq cohort. Data are shown as mean ± S.E.M., dots denote data points from individual donors. ***P* < .01, ****P* < .001; ANOVA test with age, sex, and BMI as covariates.

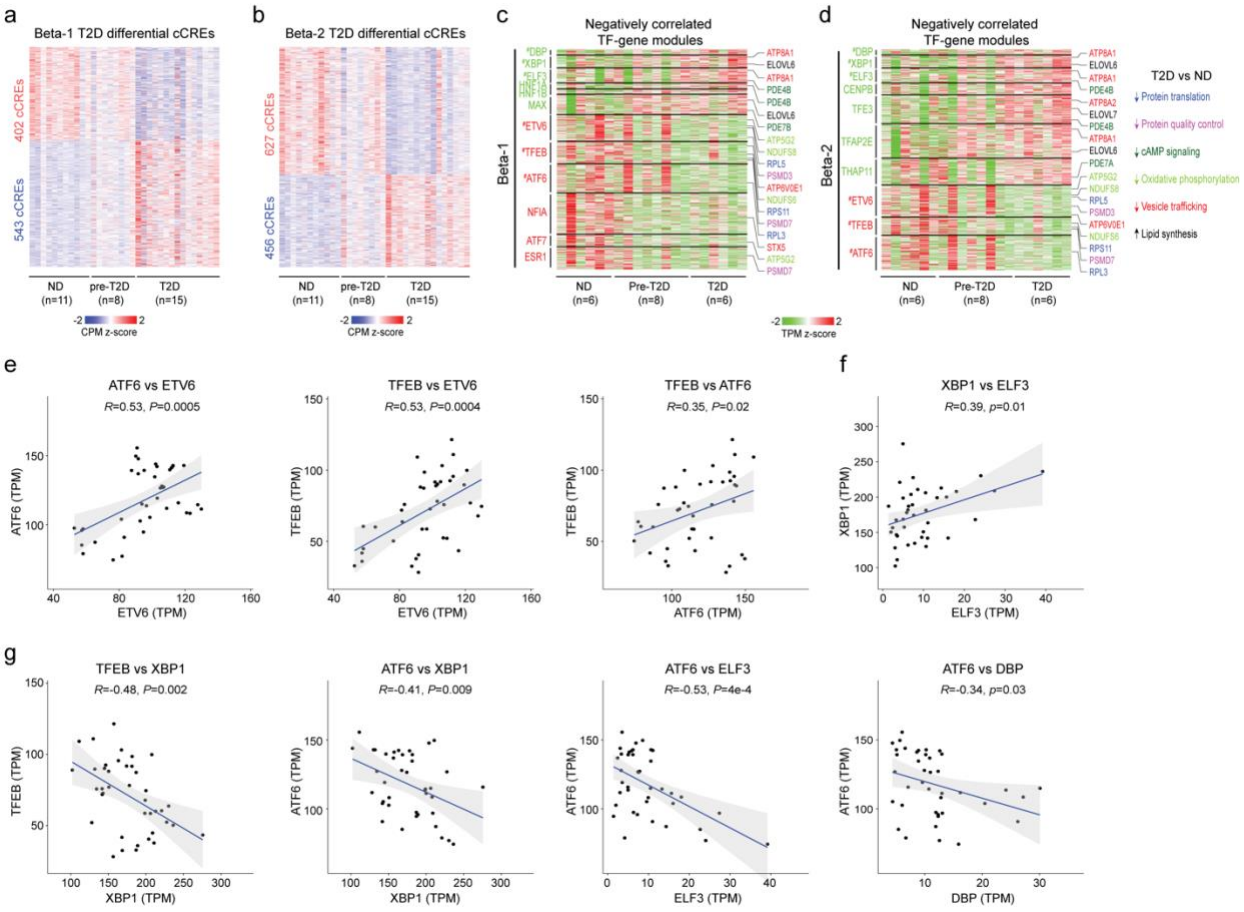
Supplementary Figure 9



Supplementary Figure 9. Transcriptional programs distinguishing the two beta cell subtypes.
(a) Genome browser tracks showing aggregate RNA and ATAC read density at representative genes (*SLC2A2, SOCS6, S100A10, ITPR1*) positively regulated by HNF1A, HNF4A or HNF4G. Differential regions between beta-1 and beta-2 are indicated by grey shaded boxes. Beta cell cCREs with binding sites for HNF1A, HNF4A and HNF4G are shown. All tracks are scaled to uniform $1x10^6$ read depth. **(b)** Genome browser tracks showing aggregate RNA and ATAC read density at representative genes (*SLC30A8, CACNA2D3, PDE4B, PRKD1*) positively regulated by NEUROD1, NFIA or TCF4.

Differential regions between beta-1 and beta-2 are indicated by grey shaded boxes. Beta cell cCREs with binding sites for NEUROD1, NFIA and TCF4 are shown. All tracks are scaled to uniform $1x10^6$ read depth. **(c)** Bar plots showing accessibility at *HNF1A*, *HNF4A* and *HNF4G* proximal cCREs in beta-1 and beta-2 cells. Proximal region of *HNF1A* (chr12:121416081-121416581), *HNF4A* (chr20:42984069-42984569), *HNF4G* (chr8:76319564-76320064). Accessibility of peaks is normalized by CPM (counts per million). Paired t-test. **(d)** Bar plots showing expression of *HNF1A*, *HNF4*A and *HNF4G* in beta-1 and beta-2 cells. Gene expression is normalized by TPM (transcripts per million). Paired t-test. **(e)** Bar plots showing accessibility at *NEUROD1*, *NFIA* and *TCF4* proximal cCREs in beta-1 and beta-2 cells. Proximal region of *NEUROD1* (chr2:182545164-182545664), *NFIA* (chr1:61523320-61523820), *TCF4* (chr18:52969269-52969769). Accessibility of peaks is normalized by CPM. Paired t-test. **(f)** Bar plots showing expression of *NEUROD1*, *NFI*A, and *TCF4* in beta-1 and beta-2. Gene expression is normalized by TPM. Paired t-test. **(g)** Genome browser tracks showing aggregate RNA and ATAC read density at *HNF1A, HNF4A* and *HNF4G* in beta-1 and beta-2 cells. Differential regions between beta-1 and beta-2 are indicated by grey shaded boxes. Beta cell cCREs with binding sites for HNF1A, HNF4A and HNF4G are shown. All tracks are scaled to uniform $1x10^6$ read depth. **(h)** Genome browser tracks showing aggregate RNA and ATAC read density at *NEUROD1, NFIA* and *TCF4* in beta-1 and beta-2 cells. Differential regions between beta-1 and beta-2 are indicated by grey shaded boxes. Beta cell cCREs with binding sites for NEUROD1, NFIA and TCF4 are shown. All tracks are scaled to uniform $1x10^6$ read depth. **(i)** Genome browser tracks showing aggregate RNA and ATAC read density at *HNF1A* and *TCF4* in beta-1 and beta-2 cells. Differential regions between beta-1 and beta-2 cells are indicated by grey shaded boxes. Beta cell cCREs with binding sites for HNF1A and TCF4 are shown. All tracks are scaled to uniform $1x10^6$ read depth.

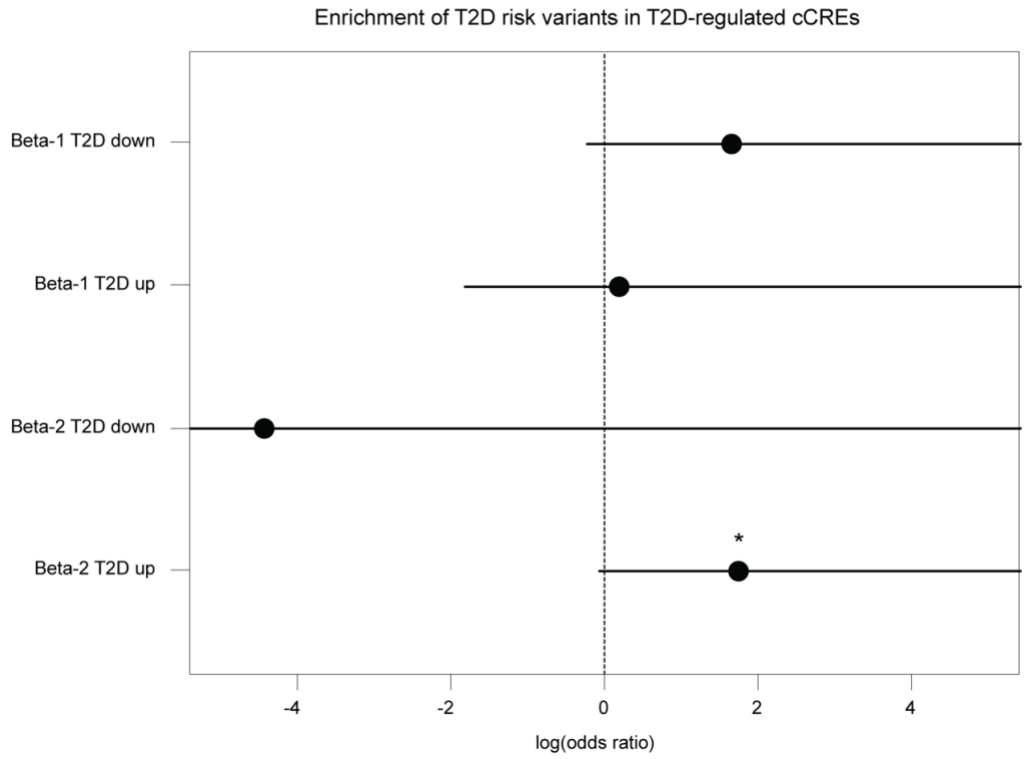
Supplementary Figure 10



Supplementary Figure 10. Transcriptional programs changed in both beta cell subtypes in T2D.
(a) Heatmap showing chromatin accessibility at cCREs with differential accessibility in beta-1 cells from ND and T2D donors. Columns represent beta cells from each donor (ND, *n*=11; pre-diabetic, pre-T2D, *n*=8; T2D, *n*=15) with accessibility of peaks normalized by CPM (counts per million). **(b)** Heatmap showing chromatin accessibility at cCREs with differential accessibility in beta-2 cells from ND and T2D donors. Columns represent beta cells from each donor (ND, *n*=11; pre-diabetic, pre-T2D, *n*=8; T2D, *n*=15) with accessibility of peaks normalized by CPM. **(c)** Heatmap showing expression of genes negatively regulated by TFs (green) with higher activity in ND compared to T2D beta-1 cells (see Methods) and TFs (red) with lower activity in ND compared to T2D beta-1 cells (*n*=6 ND, *n*=8 pre-T2D, *n*=6 T2D donors). Representative target genes of individual TFs are highlighted and classified by biological processes. Gene expression is normalized by TPM (transcripts per million). # denotes TFs with decreased or increased expression in T2D in both beta-1 and beta-2 cells. **(d)** Heatmap showing expression of genes negatively regulated by TFs (green) with higher activity in ND compared to T2D beta-2 cells (see Methods) and TFs (red) with lower activity in ND compared to T2D beta-2 cells (*n*=6 ND, *n*=8 pre-T2D, *n*=6 T2D donors). Representative target genes of individual TFs are highlighted and classified by biological processes. Gene expression is normalized by

TPM (transcripts per million). # denotes TFs with decreased or increased expression in T2D in both beta-1 and beta-2 cells. **(e,f,g)** Pearson correlation of expression levels between indicated TFs across pseudo-bulk RNA profiles from each beta cell subtype (40 dots in total: 20 donors including $n = 6$ ND, $n = 8$ pre-T2D, $n = 6$ T2D).

Supplementary Figure 11



Supplementary Figure 11. T2D risk variant enrichment for cCREs with T2D-dependent changes in the beta-1 and beta-2 subtype.
Enrichment of fine-mapped T2D risk variants for cCREs active in the beta-1 and beta-2 subtype with increased or decreased activity in T2D. Values represent log odds ratios and 95% confidence intervals. * $P < .05$