Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

HHS Public Access

Author manuscript

Nat Genet. Author manuscript; available in PMC 2023 October 05.

Published in final edited form as:

Nat Genet. 2023 June ; 55(6): 984–994. doi:10.1038/s41588-023-01397-9.

Integrating genetics with single-cell multiomic measurements across disease states identifies mechanisms of beta cell dysfunction in type 2 diabetes

Gaowei Wang[1,2,16], **Joshua Chiou**[1,2,3,16], **Chun Zeng**[1,2,16], **Michael Miller**[4], **Ileana Matta**[1,2], **Jee Yun Han**[4], **Nikita Kadakia**[1,2], **Mei-Lin Okino**[1,2], **Elisha Beebe**[1,2], **Medhavi Mallick**[1,2], **Joan Camunas-Soler**[5], **Theodore dos Santos**[6,7], **Xiao-Qing Dai**[6,7], **Cara Ellis**[6,7], **Yan Hang**[8,9], **Seung K. Kim**[8,9,10], **Patrick E. MacDonald**[6,7], **Fouad R. Kandeel**[11], **Sebastian Preissl**[4,12,17], **Kyle J. Gaulton**[1,2,13,17], **Maike Sander**[1,2,13,14,15,17]

[1]Department of Pediatrics, University of California San Diego, La Jolla, CA, USA.

[2]Pediatric Diabetes Research Center, University of California San Diego, La Jolla, CA, USA.

[3]Biomedical Graduate Studies Program, University of California San Diego, La Jolla, CA, USA.

[4]Center for Epigenomics, University of California San Diego, La Jolla, CA, USA.

[5]Department of Bioengineering, Stanford University, Stanford, CA, USA.

[6]Department of Pharmacology, University of Alberta, Edmonton, Alberta, Canada.

[7]Alberta Diabetes Institute, University of Alberta, Edmonton, Alberta, Canada.

[8]Department of Developmental Biology, Stanford University School of Medicine, Stanford, CA, USA.

[9]Departments of Medicine and of Pediatrics, Stanford University School of Medicine, Stanford, CA, USA.

[10]Stanford Diabetes Research Center, Stanford University School of Medicine, Stanford, CA, USA.

Correspondence and requests for materials should be addressed to Sebastian Preissl, Kyle J. Gaulton or Maike, Sander.sebastian.preissl@pharmakol.uni-freiburg.de.

Author contributions

M.S., K.J.G. and S.P. conceived and supervised the research in the study; M.S., K.J.G., G.W. and J.C. wrote the manuscript; G.W. and J.C. performed analyses of single-cell and genetic data; C.Z., I.M., N.K., J.Y.H. and M.-L.O. performed experiments; M. Miller performed 10x single-cell assays; E.B. and M. Mallick contributed to data analyses; F.R.K. provided human islets; J.C.-S., T.d.S., X.-Q.D., C.E., Y.H., S.K.K. and P.E.M. provided Patch-seq data.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41588-023-01397-9.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41588-023-01397-9.

Peer review information *Nature Genetics* thanks Vivek Swarup, Judith Zaugg, and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permissions information is available at www.nature.com/reprints.

Competing interests

K.J.G. holds stock in Vertex Pharmaceuticals and Neurocrine Biosciences. J.C. is now employed by and holds stock in Pfizer Inc. The other authors declare no competing interests.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

[11]Department of Clinical Diabetes, Endocrinology & Metabolism, City of Hope, Duarte, CA, USA.

[12]Institute of Experimental and Clinical Pharmacology and Toxicology, Faculty of Medicine, University of Freiburg, Freiburg, Germany.

[13]Institute for Genomic Medicine, University of California San Diego, La Jolla, CA, USA.

[14]Department of Cellular and Molecular Medicine, University of California San Diego, La Jolla, CA, USA.

[15]Max Delbrück Center for Molecular Medicine in the Helmholtz Association, Berlin, Germany.

[16]These authors contributed equally: Gaowei Wang, Joshua Chiou, Chun Zeng.

[17]These authors jointly supervised this work: Sebastian Preissl, Kyle J. Gaulton, Maike Sander.

Abstract

Dysfunctional pancreatic islet beta cells are a hallmark of type 2 diabetes (T2D), but a comprehensive understanding of the underlying mechanisms, including gene dysregulation, is lacking. Here we integrate information from measurements of chromatin accessibility, gene expression and function in single beta cells with genetic association data to nominate disease-causal gene regulatory changes in T2D. Using machine learning on chromatin accessibility data from 34 nondiabetic, pre-T2D and T2D donors, we identify two transcriptionally and functionally distinct beta cell subtypes that undergo an abundance shift during T2D progression. Subtype-defining accessible chromatin is enriched for T2D risk variants, suggesting a causal contribution of subtype identity to T2D. Both beta cell subtypes exhibit activation of a stress-response transcriptional program and functional impairment in T2D, which is probably induced by the T2D-associated metabolic environment. Our findings demonstrate the power of multimodal single-cell measurements combined with machine learning for characterizing mechanisms of complex diseases.

Pancreatic islets are comprised of multiple endocrine cell types with distinct functions in the regulation of glucose homeostasis and metabolism[1]. Islet endocrine cell types, in particular the insulin-producing beta cells, exhibit functional and molecular heterogeneity[2–5]. We recently showed that beta cell subtypes can be distinguished by chromatin accessibility in nondiabetic individuals[6]. However, the relationship between subtype-specific chromatin and transcriptomic features, and cellular function, remains unclear.

Type 2 diabetes (T2D) results from the interplay of genetic and environmental factors. A decline in glucose-stimulated insulin secretion of beta cells is a hallmark of pre-T2D[7,8], culminating in beta cell failure and loss in T2D. Studies have compared islet gene expression in nondiabetic and T2D individuals at bulk[9,10] and single-cell levels[5,11–13]. However, findings are not robust across studies[14], likely owing to small sample sizes and confounding factors unrelated to disease obscuring disease-characteristic patterns with current analysis methods. One study observed a beta cell subtype shift in T2D[15], yet the gene regulatory programs driving this shift are not understood.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

In this Article, we measured chromatin accessibility and gene expression at the single-cell level in human islet preparations from 34 nondiabetic, pre-T2D and T2D donors. We developed a classifier based on machine learning and identified two beta cell subtypes that change in abundance in T2D in independent cohorts. Using Patch sequencing (Patch-seq)[16,17], we show that the beta cell subtypes are functionally distinct in nondiabetic donors and functionally impaired in T2D donors. Through gene regulatory network (GRN) analysis, we distinguish gene regulatory programs driving beta cell subtype identity from subtype-independent, T2D-associated changes. Finally, we describe the relationship of these gene regulatory programs to T2D genetic risk, suggesting a causal contribution of beta cell subtype identity to T2D.

Results

T2D affects chromatin state in beta cells

We collected islets from 11 nondiabetic, 8 pre-T2D and 15 T2D donors (Supplementary Table 1a,b) and profiled chromatin accessibility of individual cells by single-nucleus assay for transposase-accessible chromatin using sequencing (snATAC-seq) (Fig. 1a). After quality control (Methods and Extended Data Fig. 1a–g), we annotated cell type identities based on promoter chromatin accessibility at known marker genes (Fig. 1b, Extended Data Fig. 1h,i and Supplementary Table 1a,c) and identified 412,113 nonoverlapping candidate *cis*-regulatory elements (cCREs) (Supplementary Table 2).

Long-term T2D leads to beta cell loss[18]; therefore, we assessed changes in cell type composition among islets from nondiabetic, pre-T2D and T2D donors. Cell type composition showed substantial donor heterogeneity (Fig. 1c), consistent with previous reports[10]. Relative beta cell numbers were significantly reduced in T2D compared with nondiabetic donor islets ($P = 0.006$, analysis of variance (ANOVA) test), whereas relative alpha cell numbers were increased ($P = 0.007$, ANOVA test; Fig. 1d). By contrast, relative delta or gamma cell numbers were similar between groups (Fig. 1d).

Characterization of cell type-resolved changes in chromatin accessibility during T2D progression can reveal gene regulatory mechanisms leading to T2D. Considering biological and technical covariates (Methods and Extended Data Fig. 2), we identified cCREs with differential accessibility among nondiabetic, pre-T2D and T2D donors in pseudobulk beta cells[19]. A total of 3,097 and 3,614 cCREs gained and lost accessibility in T2D, respectively (false discovery rate (FDR) < 0.1, *P* values adjusted with the Benjamini–Hochberg method; Fig. 1e and Supplementary Table 3a). Of the 6,711 differential cCREs associated with T2D in our cohort, 78.8% (5,291/6,711) showed consistent changes in an independent cohort (the Human Pancreas Analysis Program (HPAP)[20]; $P < 2.2 \times 10^{-16}$, binominal test; Extended Data Fig. 3), demonstrating robustness of our findings. We further confirmed disease specificity of the identified beta cell differential cCREs by shuffling donor disease status and could not detect any differential cCREs (FDR < 0.1, *P* values adjusted with the Benjamini–Hochberg method).

No beta cell differential cCREs were identified between nondiabetic and pre-T2D donors, and only a few were identified between pre-T2D and T2D donors (Supplementary Table 3b),

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

also after adjusting for donor numbers among groups (Supplementary Table 3c–h). However, 92% of cCREs that gained (2,855/3,097, $P < 2.2 \times 10^{-16}$, binominal test) and 88% of cCREs that lost (3,175/3,614, $P < 2.2 \times 10^{-16}$, binominal test) accessibility in T2D exhibited directionally concordant changes in pre-T2D and T2D (Fig. 1e), suggesting intermediate chromatin accessibility changes in pre-T2D.

To test effects of T2D on chromatin in nonbeta islet cell types, we analyzed cCREs for differential accessibility in alpha, delta and gamma cells. We found no or very few regulated cCREs (Supplementary Table 4), including after downsampling to account for cell number differences among cell types (FDR < 0.1; Extended Data Fig. 4). These findings suggest more subtle effects of T2D on chromatin accessibility in nonbeta islet cell types compared with beta cells.

Machine learning identifies two beta cell subtypes based on chromatin accessibility

T2D-associated chromatin accessibility changes in pseudobulk beta cells could be due to a shift in beta cell subpopulations, a shift in chromatin accessibility in individual beta cells, or both (Fig. 2a). To distinguish between these possibilities, we reclustered beta cells and identified three clusters (Extended Data Fig. 5a). However, no cluster was enriched for beta cells from pre-T2D or T2D donors (Extended Data Fig. 5b). A shortcoming of clustering and dimensionality reduction is that technical factors unrelated to disease can drive subtype identity and obscure disease-relevant shifts. To circumvent these limitations, we applied machine learning[21] by training a classifier on individual beta cells and testing its ability to distinguish beta cell chromatin profiles from non-diabetic, pre-T2D and T2D donors. To eliminate donor-specific effects during training and to test whether beta cells between disease groups can be distinguished, we removed beta cells from one donor at a time in the testing group while using remaining donors as a training group. We then compared predictions of the classifier to each donor's annotated disease state. In the case of gradual chromatin activity changes during T2D progression (scenario 2, Extended Data Fig. 5c), the classifier should exhibit high prediction accuracy in all three disease states. By contrast, in the case of a T2D-associated beta cell subtype shift (scenario 3, Extended Data Fig. 5c), prediction accuracy will depend on the prevalence of the dominant beta cell subtype. The classifier predicted beta cells from nondiabetic and T2D donors with ~60% accuracy, whereas prediction accuracy of pre-T2D beta cells was only ~5% (Extended Data Fig. 5d,e). This indicates the presence of two beta cell subtypes, one enriched in nondiabetic donors and one enriched in T2D donors (scenario 3).

The same analysis for alpha and delta cells showed prediction accuracies for nondiabetic, pre-T2D and T2D close to randomness (Extended Data Fig. 5f–i, scenario 1, Extended Data Fig. 5c), suggesting that alpha and delta cells from nondiabetic, pre-T2D and T2D donors are indistinguishable.

Through reiterative training and testing on beta cells from only nondiabetic and T2D donors (Extended Data Fig. 5j), we next established a classifier capable of distinguishing beta cell subtypes enriched in nondiabetic (beta-1) and T2D (beta-2) donors (Supplementary Table 5) and calculated their relative abundance in each donor (Fig. 2b). Whereas beta-1 cells were the predominant subtype in the nondiabetic state (67.2 ± 2.8% beta-1 versus 32.7 ±

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

2.8% beta-2), beta-2 cells were the more abundant subtype in T2D (28.3 ± 3.7% beta-1 versus 71.7 ± 3.8% beta-2; Fig. 2c). Comparison with samples from pre-T2D donors showed that the subtype shift mostly occurred between pre-T2D and T2D (Fig. 2c). In individual donors, beta-2 cell abundance positively correlated with hemoglobin A1c (HbA1c; Pearson's $R = 0.78$; Fig. 2d), which is an index for long-term glycemic control. The percentage of beta-2 cells was unrelated to sex, body mass index (BMI) or the islet index as a technical confounding factor but showed a nominal but small positive correlation with age (Extended Data Fig. 6a–d).

Next, we validated our findings using independent datasets and analysis methods. Testing of our classifier on islet snATAC-seq data from HPAP[20] revealed similar proportions of beta-1 and beta-2 cells in nondiabetic and T2D donors as in our cohort (Extended Data Fig. 6e,f), showing that our classifier can successfully identify beta cell subtypes and T2D-associated changes and is not overfit to our data. To demonstrate detection of the two beta cell subtypes by an independent method, we clustered beta cells based on cCREs with differential activity in pseudobulk beta cells from T2D donors (Fig. 1e). Confirming our findings using machine learning, we identified two beta cell clusters with differential abundance in T2D (Extended Data Fig. 6g–k) and found that beta cells in clusters 1 and 2 overlapped significantly with beta-1 and beta-2 cells, respectively, identified by machine learning ($P < 2.2 \times 10^{-16}$, exact binomial test; Extended Data Fig. 6l).

The two beta cell subtypes are transcriptionally and functionally distinct

To understand the gene expression programs that distinguish the two beta cell subtypes, we profiled gene expression and chromatin accessibility jointly in single nuclei in a subset of donors (six nondiabetic, eight pre-T2D and eight T2D; Supplementary Table 1a). We then isolated beta cells (Extended Data Fig. 7a,b), showed that clustering beta cells based on genes linked to cCREs with differential accessibility in T2D (Fig. 1e) separate beta-1 and beta-2 subtypes (Extended Data Fig. 7c,d), and identified differential cCREs and differentially expressed genes between beta-1 and beta-2 cells (Fig. 3a,b). Changes in distal and promoter cCRE accessibility positively correlated with changes in gene expression (Extended Data Fig. 7e,f). Genes with higher expression and chromatin accessibility in beta-2 cells compared with beta-1 cells included insulin (*INS*) and positive regulators of insulin secretion, such as synaptotagmin 1 (*SYT1*) and glucokinase (*GCK*), as well as the transcription factor (TF) *PAX6*, which promotes insulin gene transcription[22] (Fig. 3b,c, Extended Data Fig. 7g and Supplementary Table 6a). Beta-1 cells expressed higher levels of the TFs *HNF1A* and *HNF4A* (Fig. 3b,c and Supplementary Table 6b). Accordingly, HNF1A and HNF4A motifs were enriched at cCREs with higher accessibility in beta-1 cells than in beta-2 cells, whereas NEUROD1, E2A and NF1 motifs were enriched at cCREs more accessible in beta-2 cells (Fig. 3d and Supplementary Table 7). Together, this analysis identified concordant gene regulatory and transcriptomic features that distinguish the two beta cell subtypes. We further validated the beta cell subtypes using human islet single-cell RNA sequencing (scRNA-seq) data from three independent cohorts[5,11,23] (Extended Data Fig. 8a–i).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

The higher expression of genes associated with insulin secretion in beta-2 cells indicates functional differences between the beta cell subtypes. To test this, we leveraged human islet Patch-seq data (electrophysiological measurements and scRNA-seq) in which we confirmed the two beta cell subtypes (Extended Data Fig. 8j–l). Comparison of exocytosis in beta-1 and beta-2 cells from nondiabetic donors revealed higher exocytosis in beta-2 cells than in beta-1 cells in high glucose (Fig. 3e–g).

In sum, we derived a classifier based on machine learning of chromatin profiles that can discern beta cell subtypes with distinct transcriptomic and functional features. In nondiabetic donors, the minority beta cell subtype expresses higher levels of insulin and exocytotic genes and exhibits increased exocytosis in high glucose compared with the majority subtype.

A bistable transcriptional circuit maintains the two beta cell subtypes

To uncover transcriptional mechanisms of beta cell subtype maintenance, we inferred beta cell GRNs, linking TFs to cCREs and their target genes using our multiome data (Fig. 4a). For each TF, we assigned target genes based on positive or negative expression correlation with the TF (Supplementary Table 8).

Next, we isolated TF–gene modules with differential regulation between beta-1 and beta-2 cells. First, we conducted gene set analysis[24–26] to identify modules containing genes with differential expression between beta-1 and beta-2 cells ($P < 0.05$; Methods). Second, we filtered TF–gene modules based on TF motifs enriched at cCREs with differential accessibility between beta-1 and beta-2 cells (Fig. 3d and Supplementary Table 7). This analysis revealed gene modules positively and negatively regulated by HNF1A, HNF4A and HNF4G with higher and lower expression, respectively, in beta-1 cells and beta-2 cells, and, conversely, gene modules positively and negatively regulated by NEUROD1, NFIA and TCF4 with higher and lower expression, respectively, in beta-2 cells and beta-1 cells (Fig. 4b,c). Among the genes positively regulated by HNF1A, HNF4A and HNF4G were known regulators of insulin secretion, including the glucose transporter *SLC2A2*[27], the suppressor of cytokine signaling *SOCS6*[28], the calcium-binding protein *S100A10*[29], and the ligand-gated calcium channel *ITPR1*[30] (Fig. 4b, Extended Data Fig. 9a and Supplementary Table 8). Likewise, positively regulated targets of NEUROD1, NFIA and TCF4 included many genes with established roles in beta cell function (*SLC30A8, RFX6, ABCC8, INS, GCK* and *PCSK1*) (Fig. 4b, Extended Data Fig. 9b and Supplementary Table 8).

To identify mechanisms reinforcing beta cell subtype identity, we analyzed regulation of beta-1 and beta-2 subtype-defining TFs. For *HNF1A, HNF4A* and *HNF4G*, promoter chromatin accessibility and expression were higher in beta-1 cells than in beta-2 cells (Extended Data Fig. 9c,d). Conversely, *TCF4* and *NFIA* exhibited higher promoter accessibility and expression in beta-2 cells (Extended Data Fig. 9e,f). We identified autoregulatory and crossregulatory feedback loops between these TFs that we predict to reinforce beta cell subtypes. For example, beta-1 versus beta-2 differentially accessible cCREs at *HNF1A*, *HNF4A* and *HNF4G* contained binding motifs for HNF1A, HNF4A and HNF4G (Extended Data Fig. 9g), and beta cell expression of these TFs positively correlated across donors (Fig. 4d). Similar positive feedback loops were identified among NEUROD1, NFIA and TCF4 (Extended Data Fig. 9h and Fig. 4e), whereas HNF1A and

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

TCF4 showed negative feedback (Extended Data Fig. 9i and Fig. 4f), suggesting that a bistable transcriptional switch between HNF1A and TCF4 maintains beta cell subtype identity (Fig. 4g). Together, this analysis proposes a core network of TFs and their target genes that control beta cell subtype identity.

T2D-related functional and gene regulatory changes in beta cells

Beta-2 cells exhibit higher insulin exocytosis than beta-1 cells in nondiabetic donors (Fig. 3e–g); however, beta-2 cells increase in abundance in T2D (Fig. 2c), which is difficult to reconcile with a T2D-associated decline in beta cell function[7,16]. To determine whether beta-1 and/or beta-2 cells undergo functional change during T2D progression, we compared insulin exocytosis in beta-1 and beta-2 cells from nondiabetic, pre-T2D and T2D donors using Patch-seq. There was no difference in exocytosis at stimulatory glucose between nondiabetic and pre-T2D donors. By contrast, beta-1 and beta-2 cells both exhibited decreased exocytosis in T2D compared with pre-T2D donors (Fig. 5a,b), consistent with the T2D-associated decline in beta cell function.

To understand the molecular basis of these functional changes in T2D, we identified differentially accessible cCREs in beta-1 and beta-2 cells between nondiabetic, pre-T2D and T2D donors (Extended Data Fig. 10a,b and Supplementary Table 9) and inferred T2D-regulated GRNs by identifying TF–gene modules in beta-1 and beta-2 cells with changes in T2D (Fig. 5c,d, Extended Data Fig. 10c,d and Supplementary Table 10). TFs predicted to regulate T2D-associated gene expression changes in both subtypes included the signal-dependent TFs DBP, ELF3, XBP1, TFEB, ETV6 and ATF6 (Fig. 5c,d). These TFs are regulated by cell extrinsic stimuli, including nutrients and circadian cues, and are known mediators of the cellular stress response[31–33]. This suggests that T2D-associated changes in the extracellular environment, such as elevated glucose, affect gene expression in both beta cell subtypes.

T2D-regulated genes in both beta cell subtypes associated with processes regulating beta cell function[34], including protein translation and protein quality control, cAMP signaling, oxidative phosphorylation, vesicle trafficking and lipid metabolism (Fig. 5c,d, Extended Data Fig. 10c,d and Supplementary Table 11). For example, T2D-downregulated modules included genes encoding mitochondrial electron transport chain proteins (*NDUFS6, NDUFS8* and *ATP5G2*), syntaxins (*STX5*), and ribosomal proteins important for protein translation (*RPL3, EEF2* and *EIF3I*) (Fig. 5c–e and Extended Data Fig. 10c,d). These gene expression changes are predicted to reduce insulin production and secretion. Gene modules with increased expression in T2D included negative regulators of cAMP signaling (*PDE4B* and *PDE7A*) (Fig. 5c,d,f and Extended Data Fig. 10c,d), known to dampen glucose-stimulated insulin secretion[35]. Furthermore, we observed upregulation of regulators of insulin secretion, including K_{ATP} channel subunits (*ABCC9*)[36] and P4-ATPases (*ATP8A1* and *ATP8A2*)[37], as well as lipogenic enzymes (*ELOVL6* and *ELOVL7*), which modulate the stress response[38] and inhibit insulin secretion[39] (Fig. 5c,d,f and Extended Data Fig. 10c,d). Thus, our analysis identified a core T2D-associated gene regulatory program comprised of signal-dependent TFs regulating genes involved in beta cell function.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

We observed positive correlation in expression among TFs downregulated (*XBP1* and *ELF3*) and upregulated (*ETV6, TFEB* and *ATF6*) in T2D (Extended Data Fig. 10e,f), as well as negative correlation between TFs changing in the opposite direction in T2D (Extended Data Fig. 10g). This suggests reinforcement of T2D-related gene expression changes via positive and negative feedback loops between TFs (Fig. 5g).

Genetic risk of T2D affects beta cell subtype regulation

Having identified and characterized two distinct beta cell subtypes correlated with T2D progression, we next asked whether T2D risk variants are associated with subtype-specific gene regulatory programs, which would suggest a causal role for the beta cell subtype shift in disease. To this end, we leveraged the highly polygenic inheritance of T2D[40] to test for enrichment of fine-mapped T2D risk variants in cCREs with increased activity in the beta-1 or beta-2 subtype compared with a background of permuted cCREs comprising all beta cell cCREs. We observed enrichment of T2D risk variants in cCREs with increased activity for both beta-1 and beta-2 subtypes compared with background cCREs (beta-1 log(odds ratio, OR) = 1.33, $P = 1.8 \times 10^{-3}$; beta-2 log(OR) = 1.75, $P = 1.5 \times 10^{-4}$; Fig. 6a). Next, we tested for T2D variant enrichment in cCREs with increased or decreased activity within beta-1 and beta-2 subtypes across the T2D disease state. We did not observe significant enrichment of these cCREs for T2D risk variants, although there was nominal evidence ($P < 0.05$) for enrichment of beta-2 cCREs with higher activity in T2D.

Given enrichment of T2D risk in cCREs defining the beta-1 and beta-2 subtypes, we next determined whether TFs that maintain subtype identity mediate this risk. Of the six TFs that we suggest maintain beta-1 and beta-2 identity, genes encoding *HNF1A, HNF4A, NEUROD1* and *TCF4* harbor mutations known to cause maturity-onset diabetes of the young (MODY), a monogenic form of diabetes[41], and *HNF1A, HNF4A* and *TCF4* additionally map to known T2D risk loci[40]. We determined whether subtype-defining binding sites for these TFs were enriched for T2D risk variants. There was significant enrichment for cCREs defining beta-1 identity bound by HNF4A and HNF4G (log(OR) = 1.32, $P = 8.1 \times 10^{-3}$; log(OR) = 1.32, $P = 8.0 \times 10^{-3}$), as well as nominal enrichment for cCREs bound by HNF1A (log(OR) = 1.07, $P = 0.033$; Fig. 6b). Similarly, there was significant enrichment for cCREs defining beta-2 identity bound by TCF4, NEUROD1 and NFIA (log(OR) = 1.86, $P = 1.6 \times 10^{-4}$; log(OR) = 1.81, $P = 3.8 \times 10^{-4}$; log(OR) = 1.97, $P = 5.9 \times 10^{-4}$). There was no evidence for enrichment ($P > 0.05$) in subtype-defining cCREs not bound by these TFs (Fig. 6b).

In total, there were 43 fine-mapped T2D variants that overlapped a cCRE defining beta-1 or beta-2 identity, including high-probability variants at the *GLIS3, RASGRP1, ZFPM1, SLC12A8, FAIM2* and *SIX2/3* loci (Supplementary Table 12). We determined whether the risk alleles of T2D variants in cCREs defining beta-1 or beta-2 identity were correlated with increased or decreased chromatin accessibility using allelic imbalance mapping (Supplementary Table 12). Among fine-mapped T2D variants in cCREs defining beta-1 identity, T2D risk alleles were significantly more likely to reduce beta-1 accessibility than expected (observed = 0.86, expected = 0.50, binomial $P = 0.013$). We observed the same pattern among T2D-associated variants genome-wide in cCREs defining beta-1

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

identity (observed = 0.59, expected = 0.50, binomial $P = 0.043$). For example, at the 12p24 locus, rs1617434 overlapped a cCRE defining beta-1 identity where the T2D risk allele significantly (FDR < 0.10) decreased beta-1 accessibility (beta-1 allelic effect (π) = 0.27; 95% confidence interval = 0.12, 0.46; q value = 0.048) and was predicted to disrupt a HNF4A motif (Fig. 6c). Furthermore, the same allele was associated with reduced expression of the nearby genes *ABCB9* ($P = 1.46 \times 10^{-7}$), *RILPL2* ($P = 1.23 \times 10^{-6}$) and *MPHOSPH9* ($P = 1.87 \times 10^{-3}$), as well as other genes in islet expression quantitative trait locus data[42]. By comparison, T2D risk alleles of variants in cCREs defining beta-2 identity were more likely to increase beta-2 accessibility than expected (fine-mapped variants, observed = 0.67, expected = 0.50, binomial $P = 0.51$; genome-wide variants, observed = 0.69, expected = 0.50, binomial $P = 0.011$).

Finally, we identified T2D variants with heterogeneity in allelic effects on beta cell subtype activity ($P < 0.05$) that may modulate subtype identity (Supplementary Table 13). For example, at the 4q31 locus, fine-mapped T2D variant rs6813195 had heterogeneous effects on beta cell subtype chromatin accessibility (beta-1 $\pi = 0.56$, beta-2 $\pi = 0.64$, $P = 0.024$), where the risk allele had increased accessibility in beta-2 cells compared with beta-1 cells (Fig. 6d). The risk allele was also predicted to create a binding site for PAX6 and was associated with increased islet expression of *FBXW7* ($P = 7.49 \times 10^{-4}$). At the 14q32 locus, fine-mapped T2D variant rs56330734 had heterogeneity in effects on beta cell subtype chromatin (beta-1 $\pi = 0$, beta-2 $\pi = 0.91$, $P = 5.2 \times 10^{-5}$). The risk allele had increased accessibility in beta-2 cells compared with beta-1 cells and was predicted to create a NKX2–2 motif. In each of these examples, both the TFs and target genes affected by variant activity were involved in the NEUROD1-related GRN, suggesting that these variants may affect T2D risk by promoting beta-2 subtype identity.

Together, our analysis identified two functionally distinct beta cell subtypes in human islets that we suggest are maintained by six core TFs (Fig. 6e). We provide genetic evidence that the transcriptional programs maintaining beta cell subtype identity are causal in T2D. In T2D, there is an abundance shift between the two beta cell subtypes. Both subtypes are functionally impaired in T2D, and these functional changes are putatively driven by signal-dependent TFs implicated in the cellular stress response.

Discussion

Despite efforts to define the molecular events underlying T2D pathogenesis, we still lack a detailed understanding of the gene regulatory programs driving T2D progression. Our study demonstrates the power of combining single-cell multiome data throughout T2D progression with machine learning, genetic association data, and single-cell functional measurements to define islet cell type and subtype gene regulatory programs involved in T2D pathogenesis. The machine learning approach overcomes limitations of unsupervised dimensionality reduction methods for identifying disease-associated patterns in single-cell data from tissues of heterogenous donors and could have broad applications for interpreting single-cell maps of human tissues.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Machine learning identified two beta cell subtypes in nondiabetic donors that are functionally distinct and undergo an abundance shift in T2D. Genetic data suggest a causal role for this subtype shift in T2D, supported by enrichment of T2D variants in accessible chromatin distinguishing beta cell subtypes and these variants favoring a transition toward the T2D-enriched beta cell subtype. We identified HNF4A and HNF1A as central to the GRN maintaining the beta cell subtype enriched in nondiabetic donors and found enrichment of T2D variants at HNF4A and HNF1A binding sites. *HNF4A* and *HNF1A* loss-of-function mutations cause MODY1[41]. Together, this suggests that reduced HNF4A and HNF1A activity could be a causal event in T2D pathogenesis by triggering a shift in beta cell subtype identity.

Additional T2D-related gene expression changes occur across both beta cell subtypes and are driven by signal-dependent TFs (for example, XBP1, ATF6, TFEB and DBP). These TFs are activated by elevated glucose and fatty acids[34] and regulate cellular stress response genes and beta cell function[31–33]. Therefore, these gene expression changes could be a consequence of T2D-associated metabolic abnormalities, and targeting this GRN could help reverse beta cell dysfunction in T2D.

It is possible that the beta cell subtype-defining GRN and the T2D-induced ‘stress response GRN’ are linked and that the changes are interdependent. This view is supported by evidence showing that loss of HNF1A, which maintains the beta-1 subtype, reduces XBP1 and sensitizes beta cells to endoplasmic reticulum (ER) stress[43]. Conversely, genetic reduction of insulin dosage—akin to forcing beta cells into a beta-2 subtype—alleviates beta cell ER stress[44]. Whether the more highly exocytotic beta-2 subtype is ultimately more vulnerable and prone to fail in the face of metabolic stress remains to be studied.

Analyzing and interpreting omics data from human primary islets has been challenging because of sample heterogeneity due to biological factors as well as technical variables from the islet isolation procedure. Although our findings will require validation on additional human islet data, this study suggests that machine learning can overcome limitations of widely used unsupervised dimensionality reduction methods for identifying disease-relevant patterns in heterogenous islet samples. The approach described here could prove generally useful for pinpointing disease-associated changes in single-cell data from human samples.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41588–023-01397–9.

Methods

Ethical approval

This research complies with all relevant ethical regulations. Studies were given exempt status by the Institutional Review Board of the University of California San Diego (UCSD). Informed consent was obtained from participants. Participants did not receive compensation.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Human islets

We obtained islet preparations for 34 donors from four resource centers (22 from City of Hope National Medical Center, 9 from Scharp-Lacy Research Institute, 2 from the University of Pennsylvania and 1 from the University of Wisconsin). Characteristics (that is, age, sex, BMI, HbA1c and ethnicity) and available clinical information for individual donors are listed in Supplementary Table 1a. The mean age, BMI and HbA1c, as well as number of donors by sex and ethnicity in each disease group, are summarized in Supplementary Table 1b. Classification of donors as nondiabetic, pre-T2D or T2D was based on the person's medical record or postmortem HbA1c value. Donors with prior T2D diagnosis per medical record or HbA1c ≥ 6.5 were classified as T2D, donors without prior T2D diagnosis and 5.7 ≤ HbA1c ≤ 6.4 were classified as pre-T2D, and donors without prior T2D diagnosis and HbA1c ≤ 5.6 (or HbA1c unavailable) were classified as nondiabetic. Islet preparations were further enriched using zinc-dithizone staining followed by handpicking and were snap-frozen with liquid nitrogen or dry ice.

Generation of snATAC-seq data

Approximately 1,000 islet equivalents (IEQs) (~1,000 cells per IEQ) were resuspended in 1 ml of nuclei permeabilization buffer (10 mM Tris-HCl (pH 7.5), 10 mM NaCl, 3 mM $MgCl_2$, 0.1% Tween-20 (Sigma), 0.1% IGEPAL-CA630 (Sigma), 0.01% digitonin (Promega) and 1% fatty acid-free BSA (Proliant, 68700) in molecular biology-grade water) and homogenized using a 1-ml glass Dounce homogenizer with a tight-fitting pestle (Wheaton, EF24835AA) for 10–20 strokes until the solution was homogeneous. Homogenized islets were filtered with a 30-μm filter (CellTrics, Sysmex) and then incubated for 10 min at 4 °C on a rotator. Nuclei were pelleted with a swinging-bucket centrifuge (500 × *g*, 5 min, 4 °C; Eppendorf, 5920 R) and washed with wash buffer (10 mM Tris-HCl (pH 7.5), 10 mM NaCl, 3 mM $MgCl_2$, 0.1% Tween-20, 1% BSA (Proliant, 68700) in molecular biology-grade water). Nuclei were pelleted and resuspended in 30 μl of 1x Nuclei Buffer (10x Genomics). Nuclei were counted using a hemocytometer, and 15,360 nuclei were used for tagmentation. snATAC-seq libraries were generated using the Chromium Single Cell ATAC Library & Gel Bead Kit (10x Genomics, 1000110), Chromium Chip E Single Cell ATAC Kit (10x Genomics, 1000086) and indexes (Chromium i7 Multiplex Kit N, Set A, 10x Genomics, 1000084) following manufacturer instructions. Final libraries were quantified using a Qubit fluorometer (Life Technologies), and the nucleosomal pattern was verified using a TapeStation (High Sensitivity D1000, Agilent). Libraries were sequenced on NextSeq 500, HiSeq 4000 and NovaSeq 6000 sequencers (Illumina) with the following read lengths (Read1 + Index1 + Index2 + Read2): 50 + 8 + 16 + 50.

Generation of joint single-nucleus RNA and ATAC-seq data

Islets were resuspended in 1 ml of wash buffer (10 mM Tris-HCl (pH 7.4), 10 mM NaCl, 3 mM $MgCl_2$, 0.1% Tween-20 (Sigma), 1% fatty acid-free BSA (Proliant, 68700), 1 mM DTT (Sigma), 1x protease inhibitors (Thermo Fisher Scientific, PIA32965), 1 U μl^{-1} RNasin (Promega, N2515) in molecular biology-grade water) and homogenized using a 1-ml glass Dounce homogenizer with a tight-fitting pestle (Wheaton, EF24835AA) for 10–20 strokes until the solution was homogeneous. Homogenized islets were filtered with 30-μm filter

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

(CellTrics, Sysmex) and pelleted with a swinging-bucket centrifuge (500 × *g*, 5 min, 4 °C; Eppendorf, 5920 R). Nuclei were resuspended in 400 µl of sort buffer (1% fatty acid-free BSA, 1x protease inhibitors (Thermo Fisher Scientific, PIA32965), 1 U μl^{-1} RNasin (Promega, N2515) in PBS) and stained with 7-aminoactinomycin D (7-AAD; 1 µM; Thermo Fisher Scientific, A1310). A total of 120,000 nuclei were sorted using an SH800 sorter (Sony) into 87.5 µl of collection buffer (1 U μl^{-1} RNasin (Promega, N2515), 5% fatty acid-free BSA (Proliant, 68700) in PBS). Nuclei suspension was mixed in a ratio of 4:1 with 5x permeabilization buffer (50 mM Tris-HCl (pH 7.4), 50 mM NaCl, 15 mM $MgCl_2$, 0.5% Tween-20 (Sigma), 0.5% IGEPAL-CA630 (Sigma), 0.05% digitonin (Promega), 5% fatty acid-free BSA (Proliant, 68700), 5 mM DTT (Sigma), 5x protease inhibitors (Thermo Fisher Scientific, PIA32965), 1 U μl^{-1} RNasin (Promega, N2515) in molecular biology-grade water) and incubated on ice for 1 min before pelleting with a swinging-bucket centrifuge (500 × *g*, 5 min, 4 °C; Eppendorf, 5920 R). Supernatant was gently removed, and ~50 µl were left behind to increase nuclei recovery. A total of 650 µl of wash buffer (10 mM Tris-HCl (pH 7.4), 10 mM NaCl, 3 mM $MgCl_2$, 0.1% Tween-20 (Sigma), 1% fatty acid-free BSA (Proliant, 68700), 1 mM DTT (Sigma), 1x protease inhibitors (Thermo Fisher Scientific, PIA32965), 1 U μl^{-1} RNasin (Promega, N2515) in molecular biology-grade water) was added with minimal disturbance of the pellet, and samples were centrifuged again with a swinging-bucket centrifuge (500 × *g*, 5 min, 4 °C; Eppendorf, 5920 R). Supernatant was gently removed without disturbing the pellet, leaving ~2–3 µl behind. Approximately 7–10 µl of 1x Nuclei Buffer (10x Genomics) were added, and nuclei were gently resuspended. Nuclei were counted using a hemocytometer, and 16,550–18,000 nuclei were used as input for tagmentation. Single-cell multiome ATAC and gene expression libraries were generated following manufacturer instructions (Chromium Next GEM Single Cell Multiome ATAC + Gene Expression Reagent Bundle, 10x Genomics, 1000283; Chromium Next GEM Chip J Single Cell Kit, 10x Genomics, 1000234; Dual Index Kit TT Set A, 10x Genomics, 1000215; Single Index Kit N Set A, 10x Genomics, 1000212) with the following PCR cycles: 7 cycles for ATAC index PCR, 7 cycles for complementary DNA (cDNA) amplification, and 13–16 cycles for RNA index PCR. Final libraries were quantified using a Qubit fluorometer (Life Technologies), and the size distribution was checked using a TapeStation (High Sensitivity D1000, Agilent). Libraries were sequenced on NextSeq 500 and NovaSeq 6000 sequencers (Illumina) with the following read lengths (Read1 + Inde x1 + Index2 + Read2): ATAC (NovaSeq 6000), 50 + 8 + 24 + 50; ATAC (NextSeq 500 with custom recipe), 50 + 8 + 16 + 50; RNA (NextSeq 500, NovaSeq 6000), 28 + 10 + 10 + 90.

Raw data processing and quality control

Data processing.—Alignment to the hg19 genome and initial processing were performed using the 10x Genomics Cell Ranger ATAC (v.1.1.0) and multiome ARC (v.2.0.0) pipelines. We filtered reads with MAPQ < 30, secondary or unmapped reads, and duplicate reads from the resulting bam files using bedtools (v.2.26.0), picard tools (v.2.0.4), and samtools[45] (v.1.6). Sample information and a summary of the Cell Ranger ATAC-seq and multiome quality metrics are provided in Supplementary Table 1a.

Filtering barcode doublets and low-quality cells for each individual donor.—Cell barcodes from the 10x Chromium snATAC-seq assay may have barcode

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

multiplets that have more than one oligonucleotide sequence[46]. We used the 'clean_barcode_multiplets_1.1.py' script from 10x to identify barcode multiplets for each donor and excluded these barcodes from further analysis. We then filtered low-quality snATAC-seq profiles by total unique molecular identifiers (UMIs; <1,000), fraction of reads overlapping transcription start site (TSS; <15%), fraction of reads overlapping called peaks (<30%), and fraction of reads overlapping mitochondrial DNA (>10%) according to the distribution of these metrics for all barcodes. We also excluded profiles that had extremely high unique nuclear reads (top 1%), fraction of reads overlapping TSS (top 1%), and called peaks (top 1%) to minimize the contribution of these barcodes to our analysis. Representative cell filtering from donor JYH809 is shown in Extended Data Fig. 1b. For multiome data, we used identical cutoffs to filter cells with low-quality ATAC profiles and used total UMIs (<1,000) and fraction of reads overlapping mitochondrial DNA (>10%) to filter cells with low-quality RNA profiles.

Cell clustering.—After filtering low-quality cells, we checked data quality from each sample by performing an initial clustering using Scanpy (v.1.6.0)[47]. We partitioned the hg19 genome into 5-kilobase (kb) sliding windows and removing windows overlapping blacklisted regions from ENCODE[48,49] (https://www.encodeproject.org/annotations/ENCSR636HFF). Using 5-kb sliding windows as features, we produced a barcode-by-feature count matrix consisting of the counts of reads within each feature region for each barcode. We normalized each barcode to a uniform read depth and extracted highly variable windows (using thresholds: minimal mean expression > 0.1 and dispersion > 0.2). Then, we regressed out the total read depth for each cell, performed principal component analysis (PCA), and extracted the top 50 principal components (PCs) to calculate the 30 nearest neighbors using the cosine metric, which were subsequently used for uniform manifold approximation and projection (UMAP) dimensionality reduction with the parameter 'min_dist=0.3' and Leiden[50] (v.0.7.0) clustering with the parameter 'resolution=0.8'. Representative cell clustering and marker gene promoter accessibility from donor JYH809 are shown in Extended Data Fig. 1c,d.

We then performed initial cell clustering for 255,598 cells from all donors using similar methods to cluster cells for each donor. Of note, we extracted highly variable windows across cells from all experiments. As read depth was a technical covariate specific to each experiment, we regressed this out on a per-experiment basis. We also used Harmony[51] (v.0.1.0) to adjust for batch effects across experiments.

We identified clusters and subclusters ('resolution' = 1.5) with significantly different total UMIs, fraction of reads overlapping TSS, or fraction of reads overlapping called peaks compared with other clusters and subclusters. We excluded these clusters and subclusters from further analysis, exemplified by cluster 14 and subcluster 1 from cluster 6 in Extended Data Fig. 1f. We also used marker hormones for alpha (*GCG*), beta (*INS-IGF2*) and delta (*SST*) cells to identify and remove potential doublets that have chromatin accessibility in more than one marker gene promoter. We retained 218,973 barcodes after excluding 22,929 cells in low-quality clusters and subclusters (8.9%) and 13,696 potential doublets (5.3%) and used identical methods to cluster these retained barcodes. UMAPs for cell clustering and marker gene promoter accessibility are shown in Extended Data Fig. 1g,h.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

We aggregated reads within each cluster (Extended Data Fig. 1e) and called peaks for each cluster using the MACS2 (v.2.2.7) call peak command with parameters '–nomodel –extsize 200 –shift 0 –keep-dup all -q 0.05' and filtered these peaks by the ENCODE hg19 blacklist. Then, we merged peaks from all clusters to get a union peak set containing the peaks observed across all clusters. We used these union peaks as features to generate a barcode-by-feature count matrix consisting of the counts of reads within each feature region for each barcode. We performed cell clustering using identical methods for initial clustering of all cells and identified 13 cell clusters (Fig. 1b). We determined the cell type represented by each cluster by examining chromatin accessibility at the promoter regions of known marker genes for alpha (*GCG*), beta (*INS-IGF2*), delta (*SST*), gamma (*PPY*), acinar (*REG1A*), ductal (*CFTR*), stellate (*PDGFRB*), endothelial (*CLEC14A*) and immune cells (*CCL3*).

Generating fixed-width and nonoverlapping peaks that represent open chromatin sites across all cell types

We called peaks for each cell type in Fig. 1b using the MACS2 call peak command with parameters '–nomodel –extsize 200 –shift 0 –keep-dup all -q 0.05' and filtered these peaks by the ENCODE hg19 blacklist. For each cell type, we generated fixed-width peaks (summits of these peaks from MACS2 were extended by 250 base pairs (bp) on either side to a final width of 501 bp)[52]. We quantified the significance of these fixed-width peaks in each cell type by converting the MACS2 peak scores ($-\log_{10}(q\text{ value})$) to a 'score quantile'. Then, fixed-width peaks for each cell type were combined into a cumulative peak set. As there are overlapping peaks across cell types, we retained the most significant peak, and any peak that directly overlapped with that significant peak was removed. This process was iterated to the next most significant peak and so on until all peaks were either kept or removed because of direct overlap with a more significant peak. In total, we retained 412,113 fixed-width (501 bp) and nonoverlapping peaks.

Identification of beta cell subtypes using machine learning

Train and test classifier to distinguish beta cells from different disease states.

—We used chromatin accessibility of 224,563 beta cell autosomal cCREs to characterize individual beta cells. A total of 90,290 beta cells (35,103 beta cells from 11 nondiabetic, 19,682 beta cells from pre-T2D, and 35,505 beta cells from T2D donors) was retained after excluding beta cells with less than 1,000 reads within beta cell autosomal cCREs. We used beta cells from one donor at a time as a testing group while using beta cells from remaining donors as a training group (Extended Data Fig. 5c). Using the chromatin accessibility profiles of training beta cells and their disease-state annotation, we trained a classifier using XGBoost[21] (v.0.80.1) to distinguish beta cells from nondiabetic, pre-T2D and T2D donors. We then predicted the disease state of beta cells from donors in the testing group using the trained classifier and compared predictions to the annotated disease state of testing donors to calculate the prediction accuracy. We used each donor as a testing group and obtained prediction accuracies for each donor. We downsampled beta cells from nondiabetic and T2D donors to numbers from pre-T2D donors and repeated the training and testing steps to test the effect of cell numbers.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Train classifier to predict two beta cell subtypes.—After recognizing two major beta cell subtypes enriched in either nondiabetic (beta-1 subtype) or T2D (beta-2 subtype) donors, we used reiterative training and testing steps to obtain a classifier distinguishing the two beta cell subtypes (Extended Data Fig. 5j). Using beta cells from nondiabetic and T2D donors, we trained and tested the classifier as described above. As beta-1 and beta-2 cells coexisted in each donor, we used reiterative model training and testing to identify the dominant beta cell subtype in nondiabetic (beta-1) and T2D (beta-2) donors. For each round of training and testing, we used beta cells whose disease state was correctly predicted for the next round of training and testing until the disease state of all selected beta cells was correctly predicted. Using this methodology, we obtained the final classifier to distinguish beta-1 and beta-2 cells and used the classifier to predict subtype identity of beta cells from pre-T2D donors in our snATAC-seq data and in an independent islet snATAC-seq dataset from nondiabetic and T2D donors from HPAP[20] (below).

Computing coaccessibility

For each endocrine cell type, we used Cicero[53] (v.1.3.4.10) to calculate coaccessibility scores for pairs of peaks for alpha, beta, delta and gamma cells. We started from the merged peak by cell sparse binary matrix, extracted alpha cells, and filtered out peaks that were not present in alpha cells. We used the 'make_cicero_cds' function to aggregate cells based on the 50 nearest neighbors. We then used Cicero to calculate coaccessibility scores using a window size of 1 megabase (Mb) and a distance constraint of 250 kb. We then repeated the same procedure for beta, delta and gamma cells. We used a coaccessibility threshold of 0.05 to define pairs of peaks as coaccessible. Peaks within and outside ±5 kb of a TSS in GENCODE (v.19) were considered proximal and distal, respectively. Peaks within ±500 bp of a TSS in GENCODE (v.19) were defined as promoter. Coaccessible pairs were assigned to one of three groups: distal-to-distal, distal-to-proximal or proximal-to-proximal. Distal-to-proximal coaccessible pairs were defined as potential enhancer–promoter connections. Genes linked to proximal or distal cCREs were identified.

Differential peak and gene expression analysis

Identification of independent confounding factors in snATAC-seq data.—To determine the factors that account for sample variability in our data, we conducted PCA on cell type-specific pseudobulk profiles generated from each of the 34 donors. Here, features were fixed-width peaks for each cell type and donor. Next, we calculated total-count normalized matrices, applied PCA to the normalized matrices using prcomp in R, and visualized the position of each donor using the autoplot function in R. In addition to disease status, we considered HbA1c, age, BMI and sex as biological covariates, as well as islet index, islet purity, sequencing depth, total read counts, and the fraction of reads overlapping TSS as technical covariates. We calculated the absolute Spearman correlation coefficient between the first six PCs and each biological or technical variable. We used an absolute Spearman correlation threshold of 0.4 as a cutoff to identify factors that have high correlation with each PC. We further identified independent confounding factors by calculating the pairwise Spearman correlation coefficients between factors. As high pairwise association (Spearman's $\rho > 0.9$) represents dependencies between factors such as disease status and HAb1c level, we retained only one of them.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Identification of differential peaks in cell type pseudobulk data.—For each cell type, we called differential peaks between disease groups using DESeq2[19] (v.1.22.1) in the R package. We used the cell type-specific pseudobulk feature-by-donor matrix as input and major biological and technical confounding factors (age, BMI, sex, islet index, fraction of reads overlapping TSS, and total reads) as covariates. FDR < 0.1 (*P* values adjustment with the Benjamini–Hochberg method) was used as the cutoff to identify differential peaks. We also identified differential peaks based on age, sex and BMI. We used CEAS[54] (v.1.0.2) to annotate differential sites. Of note, we found very few (0–301) differential peaks in each islet cell type based on sex, age and BMI, suggesting no consistent effect on chromatin accessibility in our data. We performed downsampling to match cell numbers for alpha, beta and delta cells. We downsampled alpha, beta and delta cells by randomly selecting 15,000 and 5,000 cells. Then, we called differential cCREs using downsampled cells. We also performed downsampling to match donor numbers in the nondiabetic, pre-T2D and T2D groups. We downsampled nondiabetic and T2D donors by randomly selecting eight donors from all nondiabetic and T2D donors. Then, we called beta cell differential cCREs with identical sample size ($n = 8$) for nondiabetic, pre-T2D and T2D groups. We repeated this process by randomly selecting six different combinations of eight nondiabetic and T2D donors.

Identification of differential peaks and genes between beta cell subtypes.—We generated beta-1 and beta-2 pseudobulk accessibility profiles (34 total) from snATAC-seq data and gene expression profiles from multiome data (20 total). Using these pseudobulk profiles, we performed paired *t*-tests to identify differential cCREs (FDR < 0.05, *P* values adjusted with the Benjamini–Hochberg method) and genes (FDR < 0.15, *P* values adjusted with the Benjamini–Hochberg method) between beta cell subtypes. We calculated the Pearson correction between $\log_2$ differences (beta-2/beta-1) in chromatin accessibility at differential cCREs and $\log_2$ differences (beta-2/beta-1) in gene expression of cCRE target genes with differential expression. To identify high-confidence differentially expressed genes between beta cell subtypes, we focused only on differentially expressed genes that also have significant changes in proximal (within ±5 kb of a TSS in GENCODE v.19) or distal cCREs accessibility (defined in Computing coaccessibility) between beta cell subtypes.

TF motif enrichment analysis

Using the barcode-by-peaks (501-bp fixed-width peaks) count matrix as input, we inferred enrichment of TF motifs for each barcode using chromVAR[55] (v.1.4.1). We filtered cells with reads less than 1,500 (min_ depth = 1,500) and peaks with fraction of reads less than 0.15 (min_in_ peaks = 0.15) by using the 'filterSamplesPlot' function from chromVAR. We also corrected GC bias based on 'BSgenome.Hsapiens.UCSC.hg19' using the 'addGCBias' function. Then, we used the TF binding profiles database JASPAR 2020 motifs[56] and calculated the deviation *z*-scores for each TF motif in each cell by using the 'computeDeviations' function. High-variance TF motifs across all cell types were selected using the 'computeVariability' function with the cutoff 1.15 ($n = 255$). For each of these variable motifs, we calculated the mean *z*-score for each cell type and normalized the values to 0 (minimal) and 1 (maximal).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

We performed both de novo and known motif enrichment analysis using the HOMER[57] (v.4.11) command 'findMotifsGenome.pl'. We focused on significantly enriched de novo motifs and assigned the best matched known TF motifs to de novo motifs.

Gene Ontology enrichment analysis

We performed Gene Ontology enrichment analysis using the R package Enrichr[58] (v.1.0). The library 'GO_Biological_Process_2018' was used with default parameters.

Inferring GRNs from multiome data

We first used a position frequency matrix (PFMatrixList object) of TF DNA-binding preferences from the JASPAR 2020 database[56] and width-fixed peaks as input to perform TF binding motif analysis. We used the 'matchMotifs' function in the R package motifmatchr (v.1.21.0) to infer beta cell cCREs occupied by 264 TFs expressed in beta cells (mean transcripts per million (TPM) across donors >4). We linked beta cell cCREs occupied by each TF to target genes based on proximity to the gene promoter (within ±5 kb of a TSS in GENCODE v.19) or coaccessibility between the distal cCRE and gene promoter across single beta cells (defined in Computing coaccessibility). We further calculated gene expression correlations between each TF and its target genes in pseudobulk beta-1 and beta-2 cells for each donor from multiome data (n = 20 donors). For each TF, we identified target genes that have significant positive and negative gene expression Pearson correlation with the TF (FDR < 0.05, P values adjusted with the Benjamini–Hochberg method) and defined positively correlated TF–gene modules and negatively correlated TF–gene modules.

Identification of differential TF–gene modules

We performed gene set analysis using R package GSA[24] (v.1.3.1) to evaluate changes of individual TF–gene modules (using all genes in the TF–gene module) between beta cell subtypes and during T2D progression (20 donors, each donor has beta-1 and beta-2 pseudobulk gene expression profiles). We used a P value < 0.05 and enrichment score to identify significantly upregulated (enrichment score > 0.6) or downregulated (enrichment score < −0.6) TF–gene modules between beta cell subtypes. We further filtered these TF–gene modules by intersecting with enriched TF motifs in cCREs with higher accessibility in beta-1 or beta-2. For each beta cell subtype, we used a P value < 0.05 and enrichment score to identify significantly upregulated (enrichment score > 1.3) or downregulated (enrichment score < −1.3) TF–gene modules during T2D progression. We further filtered the TFs by intersecting with enriched TF motifs in cCREs with significant changes in beta-1 or beta-2 during T2D progression.

Public human islet snATAC-seq and scRNA-seq data

We downloaded public human islet snATAC-seq data from HPAP[20] (https://hpap.pmacs.upenn.edu; v.2.0.0). We processed and analyzed the data using the pipeline described above. After quality control, the snATAC-seq data were used to validate results from our snATAC-seq data. Donor characteristics are summarized in Supplementary Table 14a. More information about these donors is available at https://hpap.pmacs.upenn.edu/explore/donor?by_donor.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

We downloaded scRNA-seq data and metadata of donors from three public islet scRNA-seq datasets[5,11,23]. We processed and analyzed the data using the pipeline described above. Donor characteristics are available in the original publications and are summarized in Supplementary Table 14b–d.

To classify donors from public islet datasets analyzed in this study as nondiabetic, pre-T2D or T2D, we applied the same classification criteria as used for classifying the 34 donors from the cohort profiled in this study ('Human islets'). In some cases, our classification criteria differed from the criteria used in the original studies leading to reclassification of select donors (Supplementary Table 14).

Genome-wide association studies (GWAS) enrichment analysis

We tested for enrichment of fine-mapped T2D risk variants from the DIAMANTE Consortium for beta cell cCREs defining the beta-1 and beta-2 subtype, as well as cCREs with differential activity in T2D. For each set of cCREs, we calculated the cumulative posterior probability of association (cPPA) of all fine-mapped variants overlapping cCREs. We then generated a null distribution of cPPA by randomly selecting the same number of cCREs from the set of all beta cell cCREs across 100,000 permutations. We calculated a *P* value as the number of permutations with a higher cPPA than for the observed set of cCREs. We further computed an odds ratio as $cPPA_{obs} \times (cPPA_{max} - cPPA_{mean})/cPPA_{mean} \times (cPPA_{max} - cPPA_{obs})$, where $cPPA_{obs}$ is the observed cPPA, $cPPA_{max}$ is the maximum possible cPPA for that number of sites, and $cPPA_{mean}$ is the average cPPA from the null distribution and took the natural log of the odds ratio.

Genotyping and imputation

Approximately 1,000–3,000 IEQ human islet pellets were resuspended in 200 μl of PBS and treated with 20 μl of 10 mg ml^{-1} RNase A (Invitrogen) and 20 μl of proteinase K (QIAGEN) for 30 min at room temperature, followed by the steps as described in the protocol of the DNeasy Blood & Tissue Kit (QIAGEN). Approximately 200–500 ng of DNA were used for genotyping using the Infinium Omni2.5–8 v1–4 and Omni2.5–8 v1–5 Genotyping BeadChip (Illumina) at the UCSD Institute for Genomic Medicine (IGM) core. We called genotypes with GenomeStudio (v.2.0.4) using default settings. For genotypes that passed quality filters (missing < 0.05, minor allele frequency (MAF) > 0.01, nonambiguous alleles defined by AT/GC variants with MAF > 40%), we imputed genotypes into the TOPMed r2 reference panel[59] using the TOPMed Imputation Server[60]. After imputation, we removed genotypes with low imputation quality ($R^2 < 0.3$) and used liftOver[61] to map the coordinates back to hg19.

Allelic imbalance analysis

To estimate cell type-specific chromatin accessibility allelic imbalance, we modified the WASP[62] pipeline for single-cell analysis by remapping reads using phase information and removing duplicate reads within each cell. For each sample, we aggregated remapped reads for cells from each beta cell subtype. We assessed allelic imbalance at each heterozygous variant using a binomial test, assuming a null hypothesis of equal proportions of reads for each allele. We meta-analyzed *z*-scores across all samples using Stouffer's *z*-score method

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

with remapped read depth as a weight. We used allelic imbalance *z*-scores to calculate two-sided *P* values. We annotated fine-mapped T2D variants in 99% credible sets from DIAMANTE[63] overlapping cCREs defining beta cell subtype identity with allelic imbalance *z*-scores, and calculated *q* values for these variants using Storey's method (R package qvalue v.2.16.0). For each subtype, we identified the most probable fine-mapped variant per T2D signal overlapping cCREs defining identity of that subtype. We then determined whether the proportion of T2D risk alleles for these variants with decreased subtype accessibility differed from the expected proportion of 0.50 using a binomial test. We further identified all variants with $P < 0.0001$ in DIAMANTE[63] overlapping cCREs defining beta cell subtype identity and again determined whether the proportion of T2D risk alleles for these variants with decreased accessibility differed from the expected proportion using a binomial test.

For the analyses comparing allelic imbalance between beta cell subtypes, we retained variants tested for allelic imbalance in at least two samples for each subtype and used two-sided binomial proportion tests to compare allelic imbalance *z*-scores between subtypes. We obtained islet eQTL data from the TIGER database (https://tiger.bsc.es).

Statistics and reproducibility

Statistical tests used are indicated in the figure legends. The correlation values and *P* values were calculated using packages of R (v.3.6.3) or Python (v.3.6.6). All of the code and data to reproduce the analyses were deposited in public repositories (listed in the Data availability and Code availability sections). All box plots were generated using the 'ggpaired' function in R. The minimum, median, maximum and interquartile range of the data are shown. No statistical method was used to predetermine sample size, no data were excluded from the analyses, and the experiments were not randomized.

Reporting summary

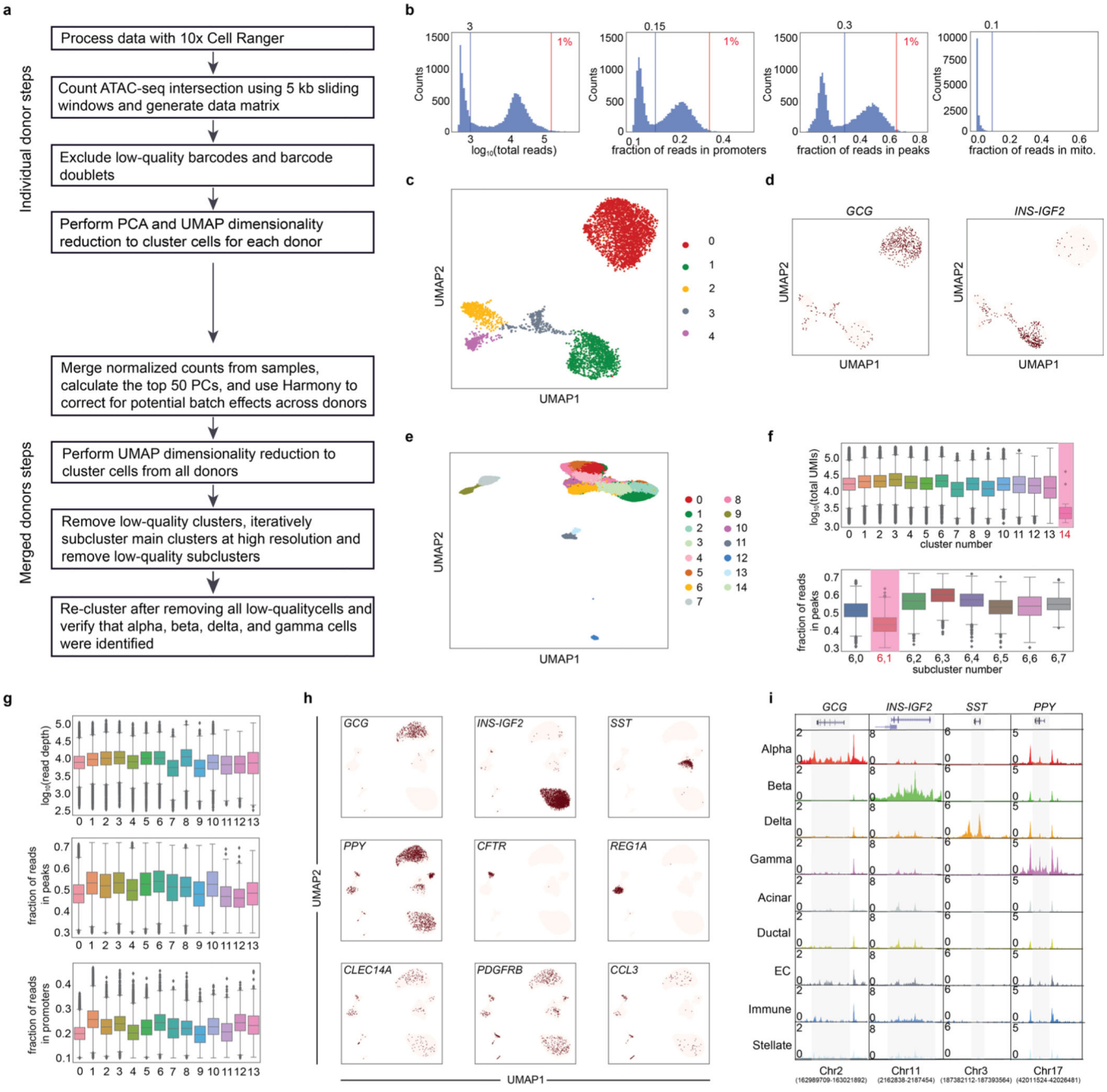
Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Code availability

Custom codes for the main analysis used in this study have been deposited on GitHub at https://github.com/gaoweiwang/Islet_snATACseq.

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

Extended Data



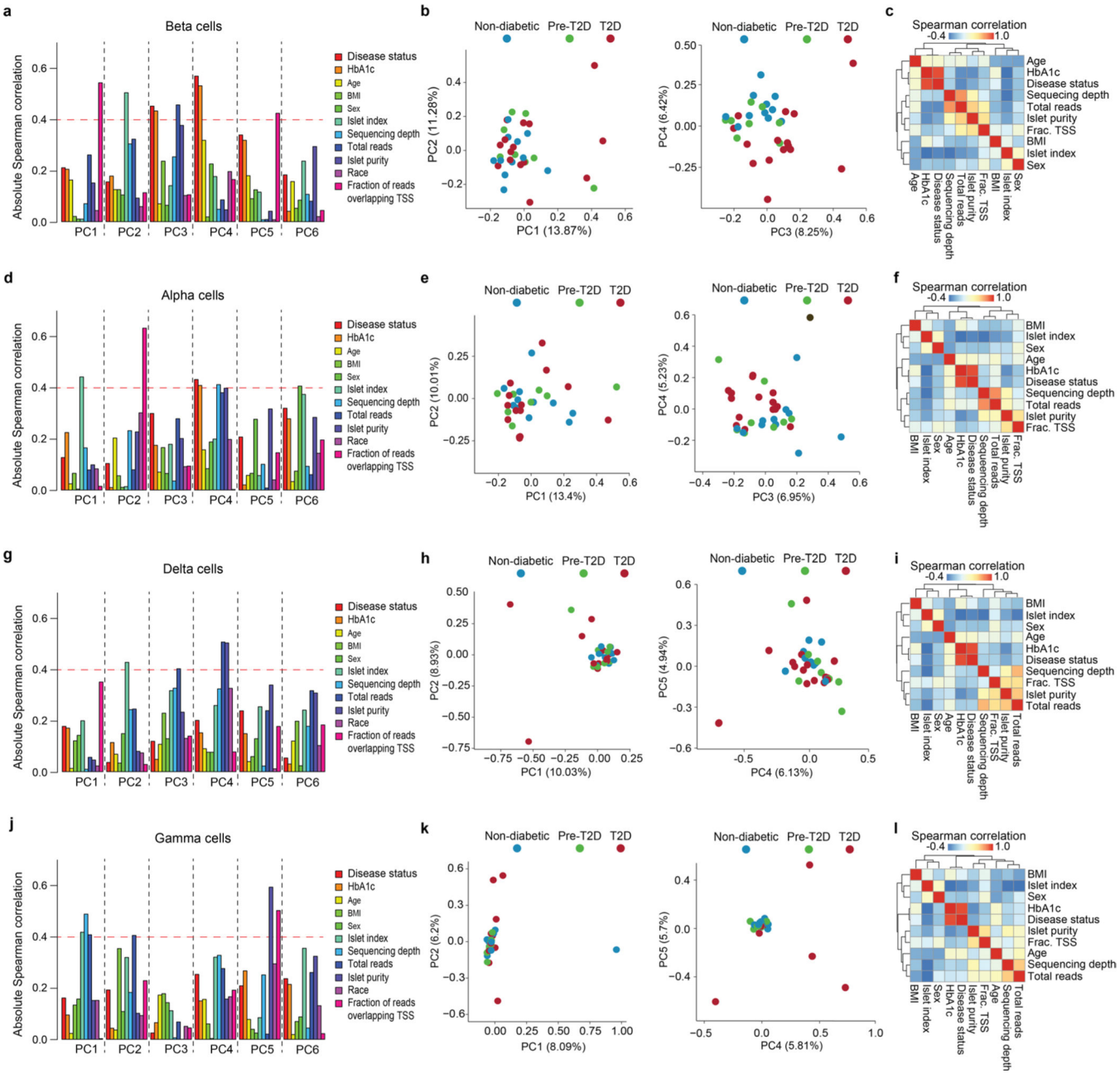
Extended Data Fig. 1 |. Quality control of snATAC-seq data.

(a) Steps for snATAC-seq data processing and quality control. **(b)** Representative quality control metrics for each donor. Log_{10} total reads, fraction of reads overlapping promoters, fraction of reads overlapping peaks, and fraction of reads overlapping mitochondria DNA distribution of cells from library 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.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

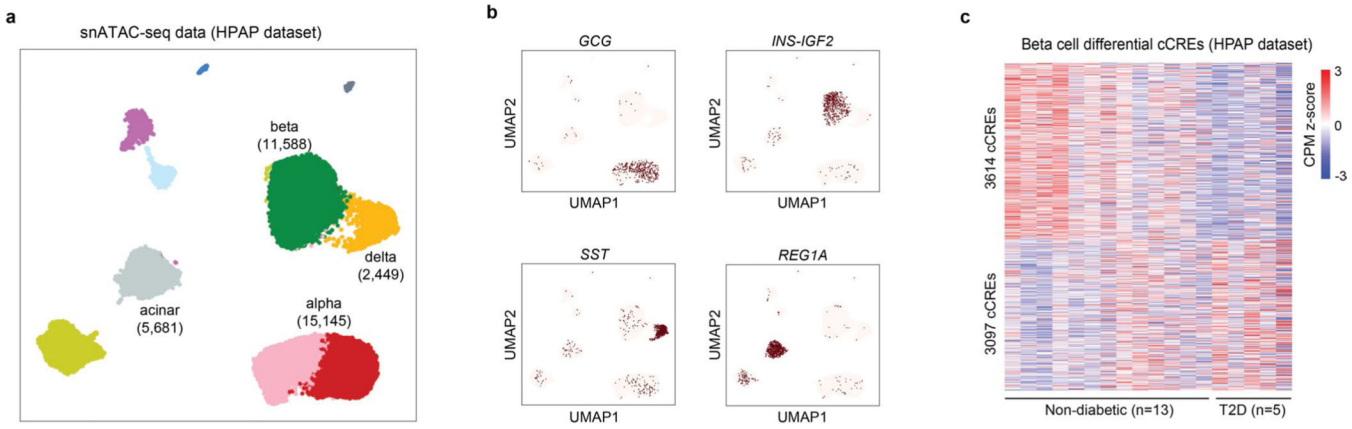
Red vertical lines denote thresholds to identify top 1% barcodes with extremely high total fragment number and fraction of reads overlapping promoters and peaks, respectively. **(c)** Representative cell clustering from library JYH809. **(d)** Promoter chromatin accessibility in a 5 kb window around TSS for endocrine marker genes in individual nuclei library JYH809. Total counts normalization and log-transformation were applied. **(e)** Cell clustering of chromatin accessibility profiles from all donors. **(f)** Representative low-quality cluster and subcluster. Cells in cluster 14 (top, highlighted in red) have significantly lower unique fragment than cells in other clusters (p = 2.3e-9, n = 255,598 cells). Cells in subcluster 1 (bottom, highlighted in red) have significantly lower fraction of reads overlapping peaks than cells in other clusters (p = 4.8e-5, n = 16,296 cells). Data are shown as mean ± S.E.M., ANOVA test with sex, age, BMI, disease status as covariates. **(g)** Log_{10} total reads, fraction of reads overlapping peaks and fraction of reads in promoters of cells from each cluster in Fig. 1b. Data are shown as mean ± S.E.M. **(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*). The UMAP projection is the same as in the main Fig. 1b. **(i)** Genome browser tracks showing aggregate read density (scaled to uniform 1×10^6 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.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript



Extended Data Fig. 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 (n = 11), pre-T2D (n = 8), and 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.

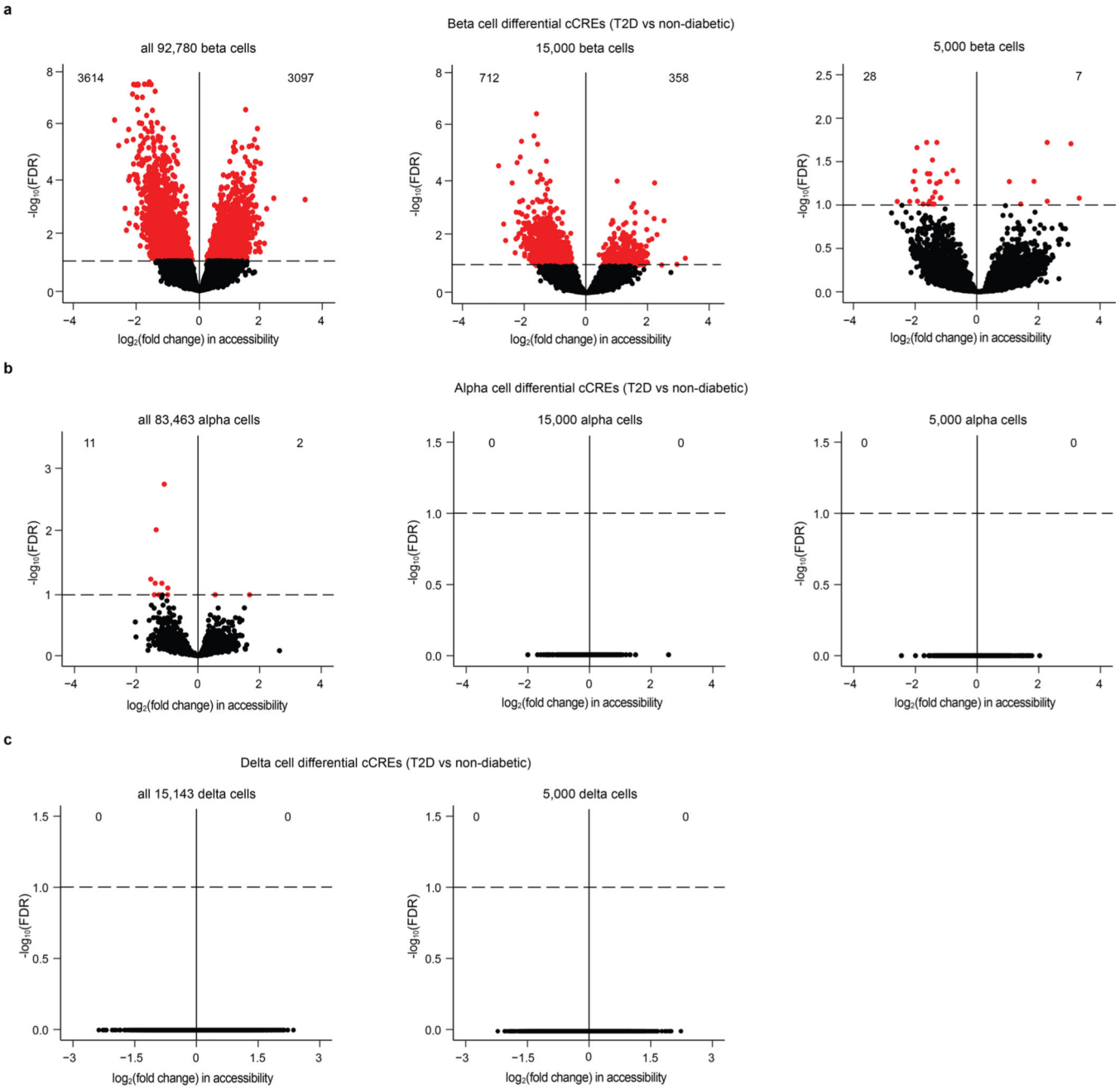


Extended Data Fig. 3 |. Validation of beta cell T2D-differential cCREs in snATAC-seq data from an independent cohort of donor islets.

(a) Clustering of chromatin accessibility profiles from HPAP human islet snATAC-seq data (non-diabetic, $n = 13$; pre-T2D, $n = 2$; T2D, $n = 5$). Nuclei are plotted using the first two UMAP components. Clusters are assigned cell type identities based on promoter accessibility of known marker genes (see Extended Data Fig. 3b). The number of nuclei 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 nucleus (alpha: *GCG*, beta: *INS-IGF2*, delta: *SST*, acinar: *REG1A*). **(c)** Heatmap showing chromatin accessibility at differential cCREs identified in Fig. 1e in HPAP snATAC-seq data. Columns represent beta cells from each donor (non-diabetic, $n = 13$; T2D, $n = 5$) and all non-diabetic and T2D donors with accessibility of peaks normalized by CPM (counts per million).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

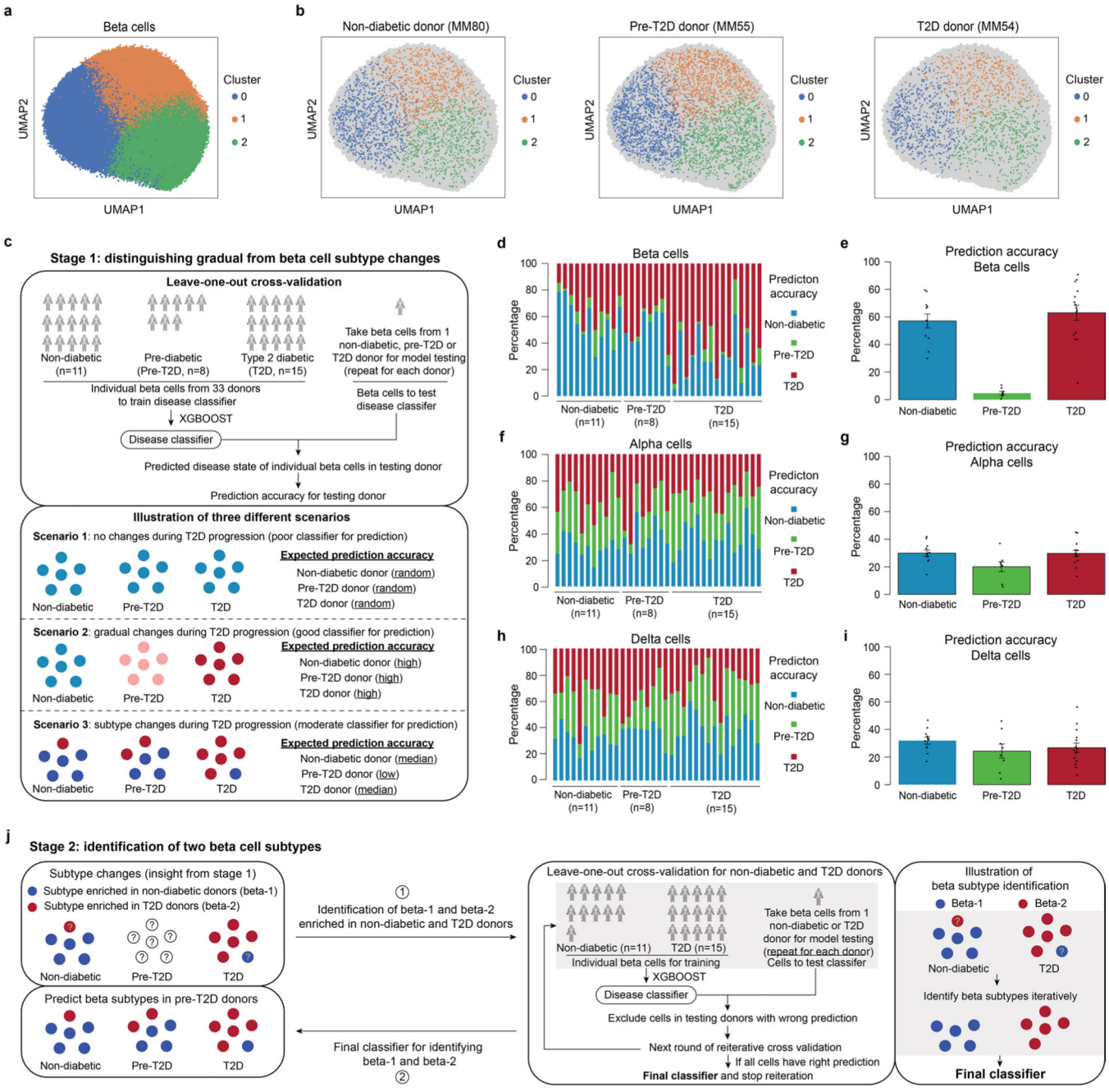
Author Manuscript Author Manuscript Author Manuscript Author Manuscript



Extended Data Fig. 4 |. T2D affects chromatin accessibility 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 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 non-diabetic 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 non-diabetic 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.



Extended Data Fig. 5 |. Machine learning identifies two beta cell subtypes.

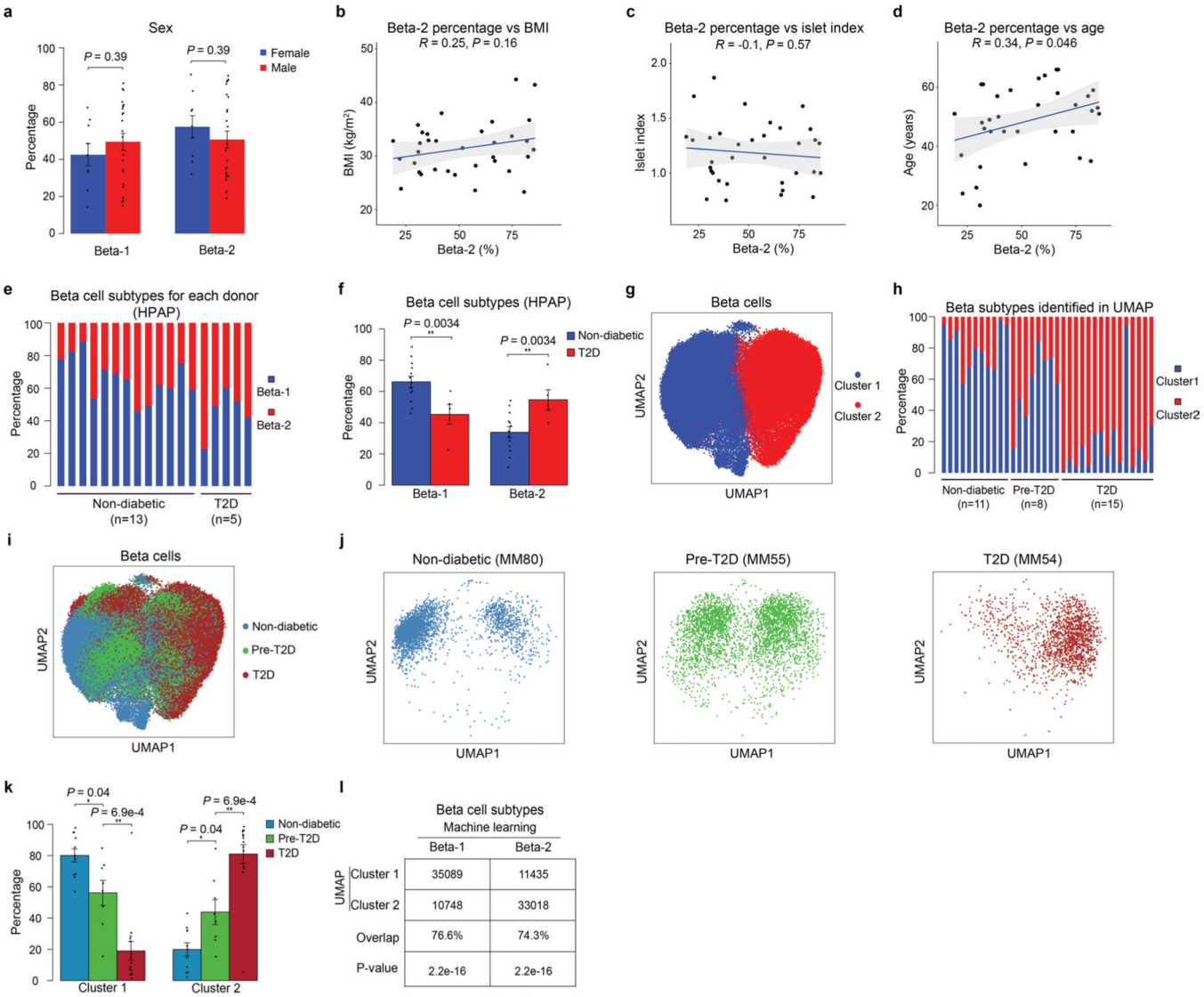
(a) Clustering of chromatin accessibility profiles from 92,780 beta cells from non-diabetic, pre-T2D and T2D donor islets using Scanpy (resolution=0.5). Nuclei are plotted using the first two UMAP components. **(b)** Position of beta cells from representative non-diabetic

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

(MM80), pre-T2D (MM55), and T2D (MM54) donors on the UMPA. **(c)** Illustration of the strategy 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 nuclei from one donor. **(e, g, i)** Relative abundance of predicted disease state among beta (**e**), alpha (**g**), and delta (**i**) cells in non-diabetic, pre-T2D and T2D donor islets. Data are shown as mean ± S.E.M. (n = 11 non-diabetic, n = 8 pre-T2D, n = 15 T2D donors), dots denote data points from individual donors. (**j**) Illustration of the strategy for identifying a classifier capable of distinguishing the two beta cell subtypes.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

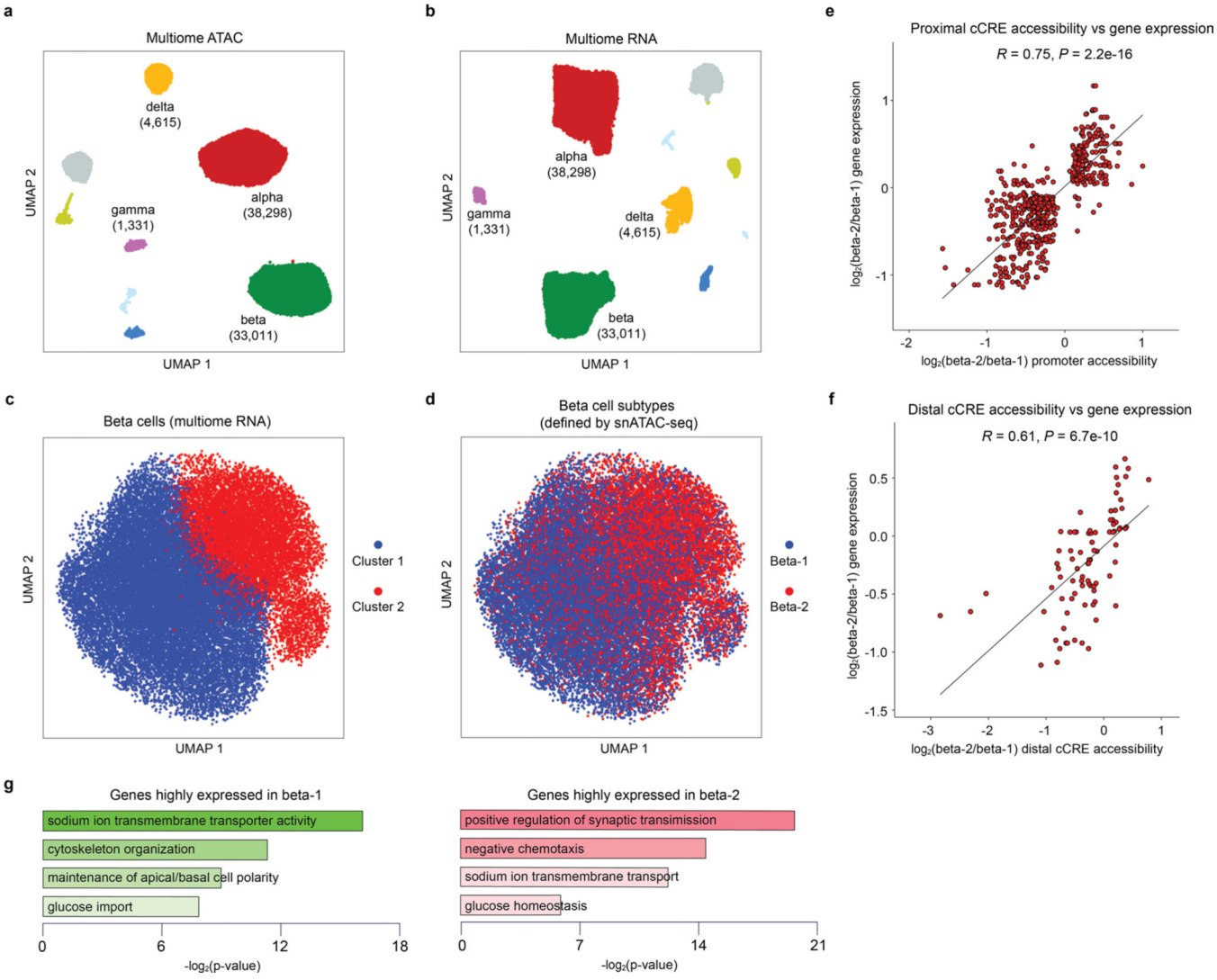


Extended Data Fig. 6 |. Validation of beta cell subtypes using independent data and computational methods.

(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-d)** Pearson correlation between relative abundance of beta-2 cells and BMI (**b**), islet index (**c**), age (**d**) across donors (n = 34 donors). The bands around the linear regression line represent the range of the 95% confidence interval, two-sided Pearson test. **(e)** Relative abundance of beta-1 and beta-2 cells in islet snATAC-seq data from an independent cohort (n = 13 non-diabetic, n = 5 T2D donors). Each column represents cells from one donor. **(f)** Relative abundance of each beta cell subtype in non-diabetic and T2D donor islets. Data are shown as mean ± S.E.M (n = 13 non-diabetic, 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 non-diabetic, pre-T2D and T2D donors using beta cell differential cCREs between non-diabetic and T2D donors from Fig. 1e. Cells are plotted using the

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

first two UMAP components. **(h)** Relative abundance of each beta cell cluster based on UMAP annotation. Each column represents cells from one donor. **(i)** Position of beta cells from non-diabetic, pre-T2D and T2D donors on the UMAP. **(j)** Position of beta cells from representative non-diabetic (MM80), pre-T2D (MM55) and T2D (MM54) donors on the UMPA. **(k)** Relative abundance of each beta cell cluster in non-diabetic, pre-T2D and T2D donor islets. Data are shown as mean ± S.E.M. (n = 11 non-diabetic, 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. The overlap is 76.6% between cluster 1 and beta-1 and 74.3% between cluster 2 and beta-2. P = 2.2e-16 (Two-sided Binominal test).



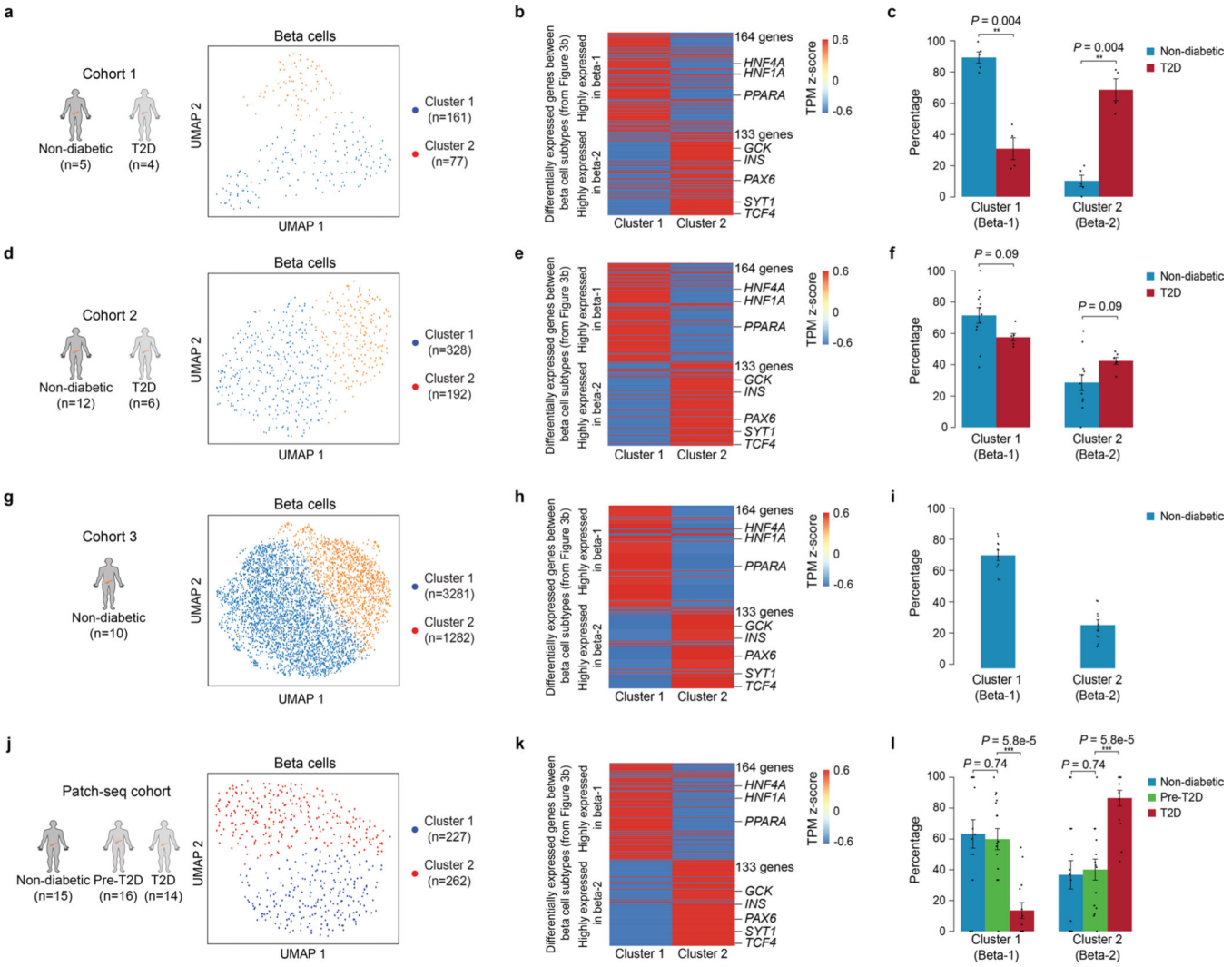
Extended Data Fig. 7 |. Validation and characterization of beta cell subtypes using multiome data.

(a, **b)** Clustering of chromatin accessibility profiles of nuclei from multiome data (**a**). **(b)** Clustering of gene expression profiles of cells from multiome data (**b**). Nuclei are plotted using the first two UMAP components. Clusters are assigned cell type identities based on

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

expression levels or promoter chromatin accessibility of known marker genes (alpha: *GCG*, beta: *INS*, delta: *SST*, gamma: *PPY*). The number of nuclei for each cell type cluster is shown in parentheses. $n = 20$ donors **(c)** Clustering of gene expression profiles of beta cells from multiome data using genes linked to differential proximal (within ± 5 kb of a TSS in GENCODE V19) and distal (based on potential distal cCRE-promoter connections inferred from Cicero, see Methods) cCREs between non-diabetic and T2D beta cells from Fig. 1e. Nuclei 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-f)** Correlation between changes in proximal cCRE (within ± 5 kb of a TSS in GENCODE V19) accessibility and gene expression differences between beta-1 and beta-2 cells for differentially expressed genes from Fig. 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 (**e**). 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 Fig. 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 (**f**). Two-sided Pearson test. **(g)** Enriched gene ontology terms among genes (see Fig. 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).

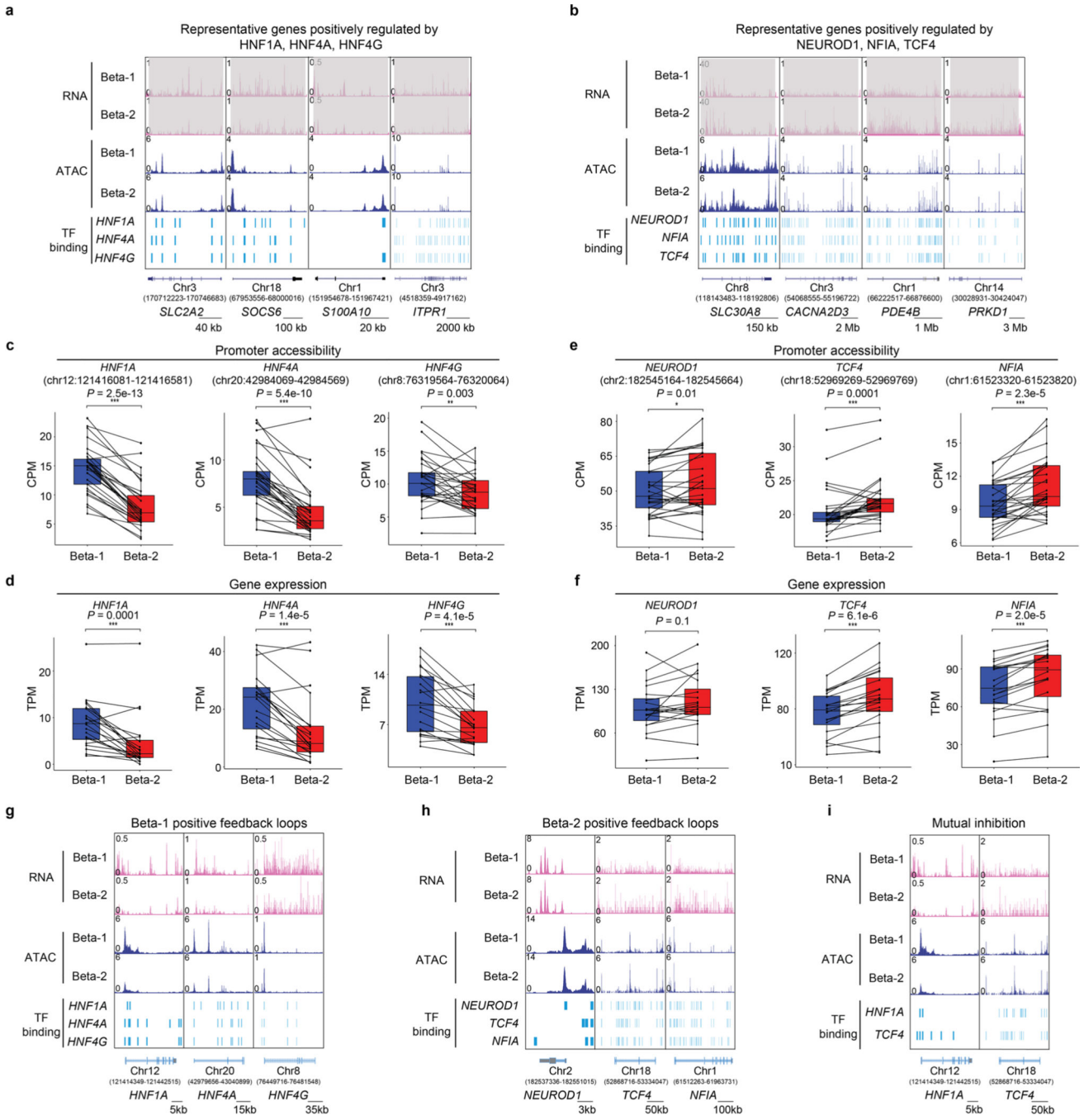
Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript



Extended Data Fig. 8 |. Beta-1 and beta-2 cell classification 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 Fig. 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 Fig. 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 non-diabetic and T2D donor islets in cohort 1[5] (n = 5 nondiabetic, n = 4 T2D) cohort 2[12] (n = 12 non-diabetic, n = 6 T2D), cohort 3[22] (n = 10 non-diabetic), and Patch-seq cohort (n = 15 non-diabetic, n = 16 pre-T2D, n = 14 T2D). 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.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

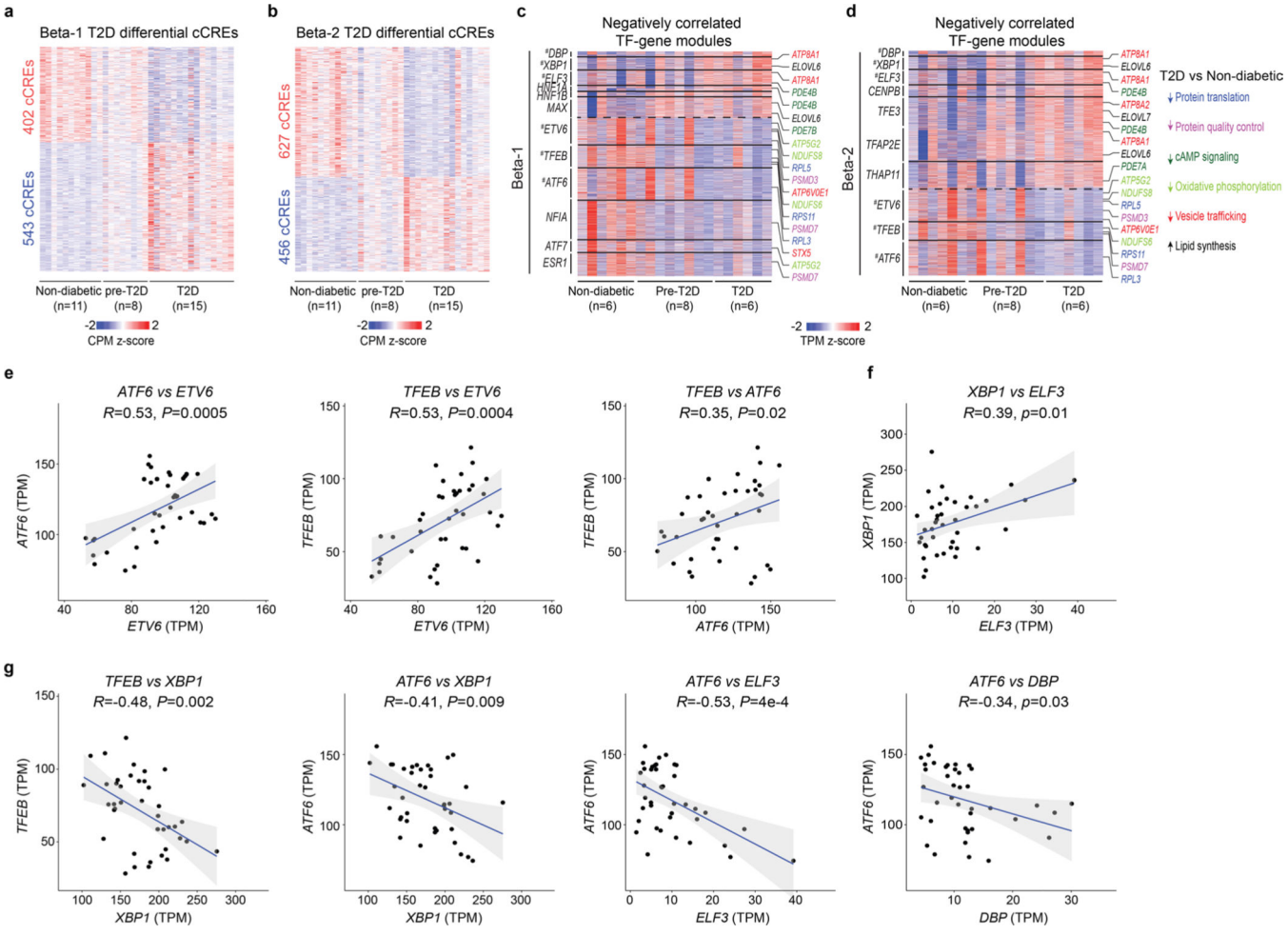


Extended Data Fig. 9 |. Transcriptional programs distinguishing the two beta cell subtypes.

(a) Genome browser tracks showing aggregate RNA and ATAC read density at representative genes positively regulated by HNF1A, HNF4A or HNF4G. Beta cell cCREs with binding sites for HNF1A, HNF4A and HNF4G are shown. **(b)** Genome browser tracks showing aggregate RNA and ATAC read density at representative genes positively regulated by NEUROD1, NFIA or TCF4. Beta cell cCREs with binding sites for NEUROD1, NFIA and TCF4 are shown.(**a**, **b**) All tracks are scaled to uniform 1×10^6 read depth, differential expressed genes between beta-1 and beta-2 are indicated by grey shaded boxes. **(c)** Box

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

plots showing accessibility at *HNF1A*, *HNF4A* and *HNF4G* proximal cCREs in beta-1 and beta-2 cells. Genomic coordinates of promotor regions are shown in parentheses. (**d**) Bar plots showing expression of *HNF1A*, *HNF4*A and *HNF4G* in beta-1 and beta-2 cells. (**e**) Bar plots showing accessibility at *NEUROD1*, *NFIA* and *TCF4* proximal cCREs in beta-1 and beta-2 cells. Proximal region of genes were shown in parentheses. (**f**) Bar plots showing expression of *NEUROD1*, *NFI*A, and *TCF4* in beta-1 and beta-2. (**c-f**) Accessibility of peaks is normalized by CPM, gene expression is normalized by TPM. Data are shown as mean ± S.E.M., n = 20 donors, Two-sided 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. Beta cell cCREs with binding sites for HNF1A, HNF4A and HNF4G are shown. (**h**) Genome browser tracks showing aggregate RNA and ATAC read density at *NEUROD1, NFIA* and *TCF4* in beta-1 and beta-2 cells. Beta cell cCREs with binding sites for NEUROD1, NFIA and TCF4 are shown. (**i**) Genome browser tracks showing aggregate RNA and ATAC read density at *HNF1A* and *TCF4* in beta-1 and beta-2 cells. Beta cell cCREs with binding sites for HNF1A and TCF4 are shown. All tracks are scaled to uniform 1×10^6 read depth.



Extended Data Fig. 10 |. Transcriptional programs changed in both beta cell subtypes in T2D.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

(a) Heatmap showing chromatin accessibility at cCREs with differential accessibility in beta-1 cells from non-diabetic and T2D donors. Columns represent beta cells from each donor (non-diabetic, $n = 11$; 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 non-diabetic and T2D donors. Columns represent beta cells from each donor (non-diabetic, $n = 11$; 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 non-diabetic compared to T2D beta-1 cells (see Methods) and TFs (red) with lower activity in non-diabetic compared to T2D beta-1 cells ($n = 6$ non-diabetic, $n = 8$ 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 non-diabetic compared to T2D beta-2 cells (see Methods) and TFs (red) with lower activity in non-diabetic compared to T2D beta-2 cells ($n = 6$ non-diabetic, $n = 8$ 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-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 $n = 6$ non-diabetic, $n = 8$ pre-T2D, $n = 6$ T2D). The bands around the linear regression line represent the range of the 95% confidence interval. Two-sided Pearson test.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

This publication includes data generated at the UCSD IGM Genomics Center using an Illumina NovaSeq 6000 that was purchased with funding from a National Institutes of Health (NIH) Shared Instrument Grant (#S10 OD026929). This work was supported by NIH U01DK105541 and R01DK122607 to M.S. and K.J.G.; R01DK114650 to K.J.G.; R01DK068471 to M.S.; U01DK120447 to P.E.M.; U01DK123716 to S.K.K. and P.E.M.; and UC4-DK112217, UC4-DK112232 and P30 DK116074 to S.K.K. Work at the UCSD Center for Epigenomics was supported by the UCSD School of Medicine. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript. We thank H. Gao, Y. Shi, K. A. Frazer, B. Ren, and members of Sander lab for scientific discussions and input on the project. We also thank the organ donors and their families for their contributions to make this study possible.

Data availability

snATAC-seq data and processed data are available through the Gene Expression Omnibus (GEO) under accession number GSE169453, single-nucleus multiome data are available under accession number GSE200044, and genotyping data are available under accession number GSE170763. UCSC Genome Browser Sessions of aggregated snATAC-seq data are available at https://genome.ucsc.edu/s/gaowei/hg19_cell_type and https://genome.ucsc.edu/s/gaowei/hg19_beta_cell. Previously published[16,17] Patch-seq data are available as raw sequencing reads in GEO under accession numbers GSE124742

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

and GSE164875. Additional Patch-seq data are accessible at the HPAP database (https://hpap.pmacs.upenn.edu). Blacklisted regions from ENCODE are found at https://www.encodeproject.org/annotations/ENCSR636HFF. Source data are provided with this paper.

References

1. Noguchi GM & Huising MO Integrating the inputs that shape pancreatic islet hormone release. Nat. Metab 1, 1189–1201 (2019). [PubMed: 32694675]
2. Wojtusciszyn A, Armanet M, Morel P, Berney T. & Bosco D. Insulin secretion from human beta cells is heterogeneous and dependent on cell-to-cell contacts. Diabetologia 51, 1843–1852 (2008). [PubMed: 18665347]
3. Dominguez-Gutierrez G, Xin Y. & Gromada J. Heterogeneity of human pancreatic β-cells. Mol. Metab 27, S7–S14 (2019).
4. Benninger RKP & Kravets V. The physiological role of β-cell heterogeneity in pancreatic islet function. Nat. Rev. Endocrinol 18, 9–22 (2022). [PubMed: 34667280]
5. Segerstolpe Å et al. Single-cell transcriptome profiling of human pancreatic islets in health and type 2 diabetes. Cell Metab. 24, 593–607 (2016). [PubMed: 27667667]
6. Chiou J. et al. Single-cell chromatin accessibility identifies pancreatic islet cell type– and state-specific regulatory programs of diabetes risk. Nat. Genet 53, 455–466 (2021). [PubMed: 33795864]
7. Cohrs CM et al. Dysfunction of persisting β cells is a key feature of early type 2 diabetes pathogenesis. Cell Rep. 31, 107469 (2020).
8. Chen C, Cohrs CM, Stertmann J, Bozsak R. & Speier S. Human beta cell mass and function in diabetes: recent advances in knowledge and technologies to understand disease pathogenesis. Mol. Metab 6, 943–957 (2017). [PubMed: 28951820]
9. Fadista J. et al. Global genomic and transcriptomic analysis of human pancreatic islets reveals novel genes influencing glucose metabolism. Proc. Natl Acad. Sci. USA 111, 13924–13929 (2014). [PubMed: 25201977]
10. Wigger L. et al. Multi-omics profiling of living human pancreatic islet donors reveals heterogeneous beta cell trajectories towards type 2 diabetes. Nat. Metab 3, 1017–1031 (2021). [PubMed: 34183850]
11. Xin Y. et al. RNA sequencing of single human islet cells reveals type 2 diabetes genes. Cell Metab. 24, 608–615 (2016). [PubMed: 27667665]
12. 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). [PubMed: 27864352]
13. Fang Z. et al. Single-cell heterogeneity analysis and CRISPR screen identify key β-cell-specific disease genes. Cell Rep. 26, 3132–3144.e7 (2019). [PubMed: 30865899]
14. Wang YJ & Kaestner KH Single-cell RNA-seq of the pancreatic islets—a promise not yet fulfilled? Cell Metab. 29, 539–544 (2019). [PubMed: 30581120]
15. Dorrell C. et al. Human islets contain four distinct subtypes of β cells. Nat. Commun 7, 11756 (2016). [PubMed: 27399229]
16. Camunas-Soler J. et al. Patch-seq links single-cell transcriptomes to human islet dysfunction in diabetes. Cell Metab. 31, 1017–1031. e4 (2020). [PubMed: 32302527]
17. Dai X-Q et al. Heterogenous impairment of α cell function in type 2 diabetes is linked to cell maturation state. Cell Metab. 34, 256–268.e5 (2022). [PubMed: 35108513]
18. Kahn SE, Cooper ME & Del Prato S. Pathophysiology and treatment of type 2 diabetes: perspectives on the past, present, and future. Lancet 383, 1068–1083 (2014). [PubMed: 24315620]
19. Love MI, Huber W. & Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15, 550 (2014). [PubMed: 25516281]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

20. Shapira SN, Naji A, Atkinson MA, Powers AC & Kaestner KH Understanding islet dysfunction in type 2 diabetes through multidimensional pancreatic phenotyping: the Human Pancreas Analysis Program. Cell Metab. 34, 1906–1913 (2022). [PubMed: 36206763]
21. Chen T. & Guestrin C. XGBoost: a scalable tree boosting system. In Proc. 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining 785–794 (Association for Computing Machinery, 2016).
22. Sander M. et al. Genetic analysis reveals that PAX6 is required for normal transcription of pancreatic hormone genes and islet development. Genes Dev. 11, 1662–1673 (1997). [PubMed: 9224716]
23. Xin Y. et al. Pseudotime ordering of single human β-cells reveals states of insulin production and unfolded protein response. Diabetes 67, 1783–1794 (2018). [PubMed: 29950394]
24. Efron B. & Tibshirani R. On the testing of significance of sets of genes. Ann. Appl. Stat 1, 107–129 (2007).
25. Cahan P. et al. CellNet: network biology applied to stem cell engineering. Cell 158, 903–915 (2014). [PubMed: 25126793]
26. Wang G. et al. A tumorigenic index for quantitative analysis of liver cancer initiation and progression. Proc. Natl Acad. Sci. USA 116, 26873–26880 (2019). [PubMed: 31843886]
27. Sansbury FH et al. *SLC2A2* mutations can cause neonatal diabetes, suggesting GLUT2 may have a role in human insulin secretion. Diabetologia 55, 2381–2385 (2012). [PubMed: 22660720]
28. Vlacich G, Nawijn MC, Webb GC & Steiner DF Pim3 negatively regulates glucose-stimulated insulin secretion. Islets 2, 308–317 (2010). [PubMed: 21099329]
29. Stancill JS et al. Chronic β-cell depolarization impairs β-cell identity by disrupting a network of Ca^{2+}-regulated genes. Diabetes 66, 2175–2187 (2017). [PubMed: 28550109]
30. Ye R. et al. Inositol 1,4,5-trisphosphate receptor 1 mutation perturbs glucose homeostasis and enhances susceptibility to diet-induced diabetes. J. Endocrinol 210, 209–217 (2011). [PubMed: 21565852]
31. Martina JA, Diab HI, Brady OA & Puertollano R. TFEB and TFE3 are novel components of the integrated stress response. EMBO J. 35, 479–495 (2016). [PubMed: 26813791]
32. Ohta Y. et al. Clock gene dysregulation induced by chronic ER stress disrupts β-cell function. eBioMedicine 18, 146–156 (2017). [PubMed: 28389215]
33. Eizirik DL, Pasquali L. & Cnop M. Pancreatic β-cells in type 1 and type 2 diabetes mellitus: different pathways to failure. Nat. Rev. Endocrinol 16, 349–362 (2020). [PubMed: 32398822]
34. Lytrivi M, Castell A-L, Poitout V. & Cnop M. Recent insights into mechanisms of β-cell lipo- and glucolipotoxicity in type 2 diabetes. J. Mol. Biol 432, 1514–1534 (2020). [PubMed: 31628942]
35. Pratt EPS, Harvey KE, Salyer AE & Hockerman GH Regulation of cAMP accumulation and activity by distinct phosphodiesterase subtypes in INS-1 cells and human pancreatic β-cells. PLoS ONE 14, e0215188 (2019).
36. Bryan J. et al. ABCC8 and ABCC9: ABC transporters that regulate K^+ channels. Pflugers Arch. 453, 703–718 (2007). [PubMed: 16897043]
37. Yang Y. et al. The phosphatidylserine flippase β-subunit *Tmem30a* is essential for normal insulin maturation and secretion. Mol. Ther 29, 2854–2872 (2021). [PubMed: 33895325]
38. Palu RAS & Chow CY Baldspot/ELOVL6 is a conserved modifier of disease and the ER stress response. PLoS Genet. 14, e1007557 (2018).
39. Tang N. et al. Ablation of *Elovl6* protects pancreatic islets from high-fat diet-induced impairment of insulin secretion. Biochem. Biophys. Res. Commun 450, 318–323 (2014). [PubMed: 24938128]
40. Gaulton KJ Mechanisms of type 2 diabetes risk loci. Curr. Diab. Rep 17, 72 (2017). [PubMed: 28741265]
41. Nkonge KM, Nkonge DK & Nkonge TN The epidemiology, molecular pathogenesis, diagnosis, and treatment of maturity-onset diabetes of the young (MODY). Clin. Diabetes Endocrinol 6, 20 (2020). [PubMed: 33292863]
42. Alonso L. et al. TIGER: the gene expression regulatory variation landscape of human pancreatic islets. Cell Rep. 37, 109807 (2021).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

43. Kirkpatrick CL et al. Hepatic nuclear factor 1α (HNF1α) dysfunction down-regulates X-box-binding protein 1 (XBP1) and sensitizes β-cells to endoplasmic reticulum stress. J. Biol. Chem 286, 32300–32312 (2011). [PubMed: 21784843]
44. Szabat M. et al. Reduced insulin production relieves endoplasmic reticulum stress and induces β cell proliferation. Cell Metab. 23, 179–193 (2016). [PubMed: 26626461]
45. Li H. et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078–2079 (2009). [PubMed: 19505943]
46. Lareau CA, Ma S, Duarte FM & Buenrostro JD Inference and effects of barcode multiplets in droplet-based single-cell assays. Nat. Commun 11, 866 (2020). [PubMed: 32054859]
47. Wolf FA, Angerer P. & Theis FJ SCANPY: large-scale single-cell gene expression data analysis. Genome Biol. 19, 15 (2018). [PubMed: 29409532]
48. Amemiya HM, Kundaje A. & Boyle AP The ENCODE blacklist: identification of problematic regions of the genome. Sci. Rep 9, 9354 (2019). [PubMed: 31249361]
49. The ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. Nature 489, 57–74 (2012). [PubMed: 22955616]
50. Traag VA, Waltman L. & van Eck NJ From Louvain to Leiden: guaranteeing well-connected communities. Sci. Rep 9, 5233 (2019). [PubMed: 30914743]
51. Korsunsky I. et al. Fast, sensitive and accurate integration of single-cell data with Harmony. Nat. Methods 16, 1289–1296 (2019). [PubMed: 31740819]
52. Satpathy AT et al. Massively parallel single-cell chromatin landscapes of human immune cell development and intratumoral T cell exhaustion. Nat. Biotechnol 37, 925–936 (2019). [PubMed: 31375813]
53. Pliner HA et al. Cicero predicts *cis*-regulatory DNA interactions from single-cell chromatin accessibility data. Mol. Cell 71, 858–871.e8 (2018). [PubMed: 30078726]
54. Ji X, Li W, Song J, Wei L. & Liu XS CEAS: *cis*-regulatory element annotation system. Nucleic Acids Res. 34, W551–W554 (2006). [PubMed: 16845068]
55. Schep AN, Wu B, Buenrostro JD & Greenleaf WJ chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. Nat. Methods 14, 975–978 (2017). [PubMed: 28825706]
56. Fornes O. et al. JASPAR 2020: update of the open-access database of transcription factor binding profiles. Nucleic Acids Res. 48, D87–D92 (2020). [PubMed: 31701148]
57. Heinz S. et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol. Cell 38, 576–589 (2010). [PubMed: 20513432]
58. Kuleshov MV et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res. 44, W90–W97 (2016). [PubMed: 27141961]
59. Taliun D. et al. Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. Nature 590, 290–299 (2021). [PubMed: 33568819]
60. Das S. et al. Next-generation genotype imputation service and methods. Nat. Genet 48, 1284–1287 (2016). [PubMed: 27571263]
61. Hinrichs AS et al. The UCSC Genome Browser Database: update 2006. Nucleic Acids Res. 34, D590–D598 (2006). [PubMed: 16381938]
62. van de Geijn B, McVicker G, Gilad Y. & Pritchard JK WASP: allele-specific software for robust molecular quantitative trait locus discovery. Nat. Methods 12, 1061–1063 (2015). [PubMed: 26366987]
63. Mahajan A. et al. Multi-ancestry genetic study of type 2 diabetes highlights the power of diverse populations for discovery and translation. Nat. Genet 54, 560–572 (2022). [PubMed: 35551307]

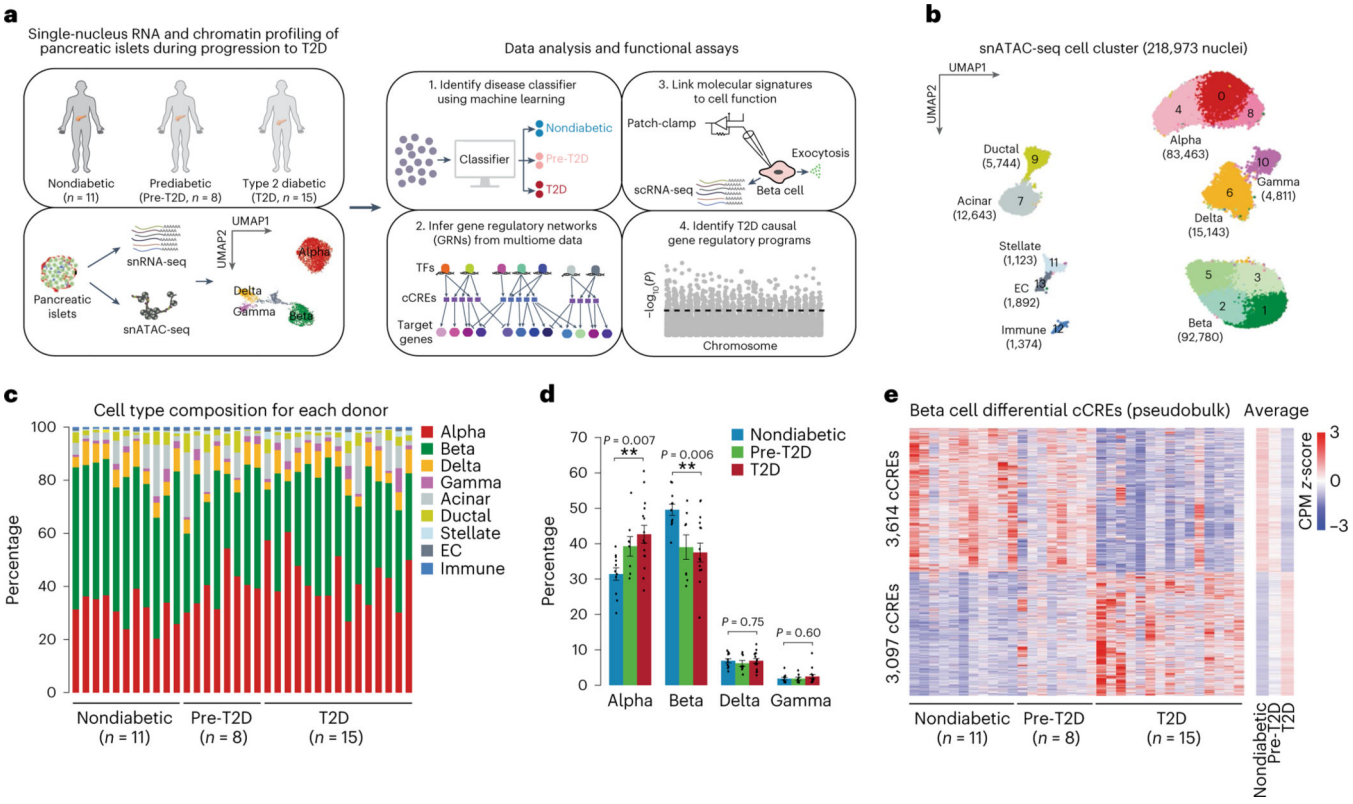


Fig. 1 |. Beta cells exhibit changes in chromatin activity in T2D.

a, Schematic of the study design. snATAC-seq was performed on nuclei from pancreatic islets from 11 nondiabetic, 8 pre-T2D and 15 T2D human donors. Single-nucleus multiome (ATAC + RNA) analysis was performed on a subset of donors (n = 6 nondiabetic, n = 8 pre-T2D, n = 6 T2D). We used machine learning to identify classifiers for beta cells in nondiabetic, pre-T2D and T2D donors; inferred GRNs; linked molecular signatures to beta cell function using Patch-seq; and identified T2D causal gene regulatory programs. **b**, Clustering of chromatin accessibility profiles from 218,973 nuclei from nondiabetic, pre-T2D and T2D donor islets. Nuclei are plotted using the first two UMAP components. Clusters are assigned cell type identities based on promoter accessibility of known marker genes. The number of nuclei for each cell type cluster is shown in parentheses. EC, endothelial cells. **c**, Relative abundance of each cell type based on UMAP annotation in **b**. Each column represents cells from one donor. **d**, Relative abundance of each islet endocrine cell type in nondiabetic, pre-T2D and T2D donor islets. Data are shown as mean ± s.e.m. (n = 11 nondiabetic, n = 8 pre-T2D, n = 15 T2D); dots denote data points from individual donors. ***P < 0.001; **P < 0.01; *P < 0.05; ANOVA test with age, sex, BMI and islet index as covariates. **e**, Heatmap showing chromatin accessibility at cCREs with differential accessibility in beta cells from nondiabetic and T2D donors. Columns represent beta cells from each donor and all nondiabetic, pre-T2D and T2D donors with accessibility of peaks normalized by counts per million (CPM).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

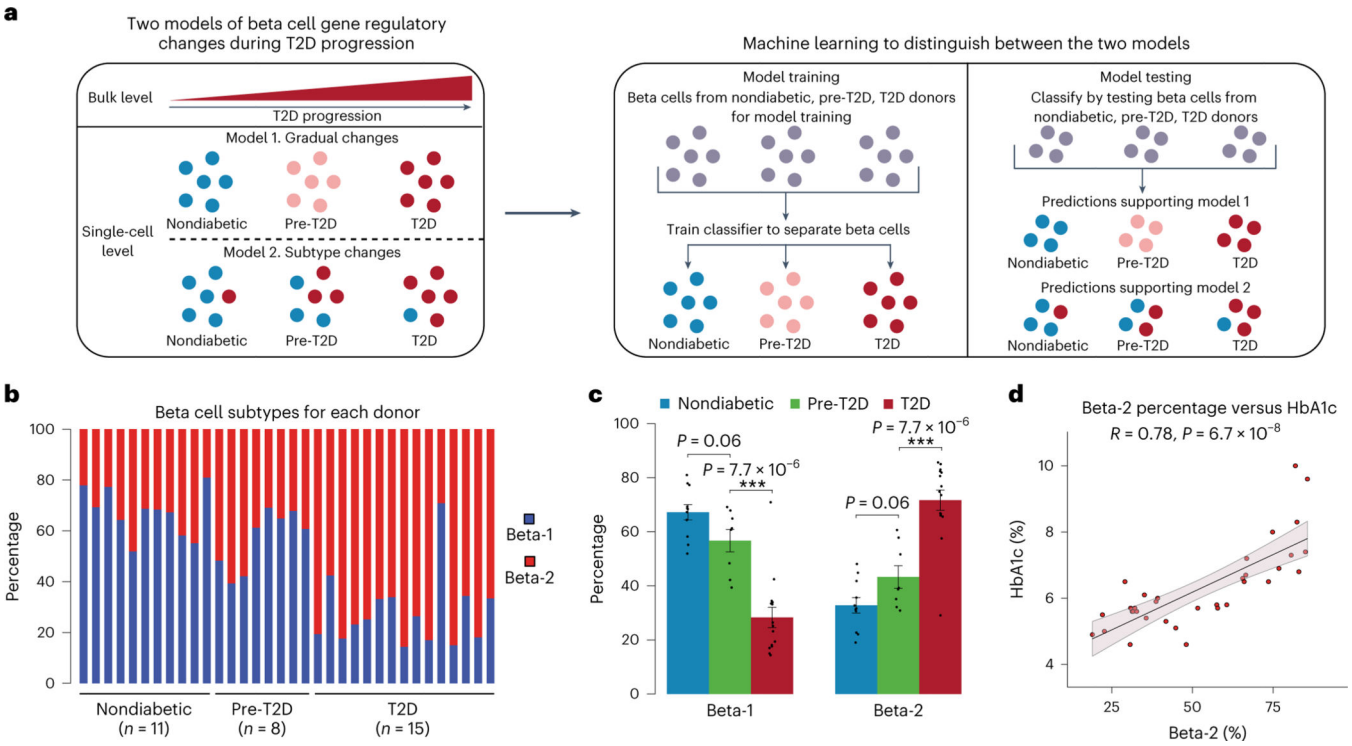


Fig. 2 |. Machine learning identifies two beta cell subtypes with differential abundance in T2D.

a, Schematic outlining the machine learning-based approach to distinguish two models that could account for gene regulatory changes in beta cells in T2D. **b**, Relative abundance of beta-1 (blue) and beta-2 (red) cells identified by machine learning. Each column represents nuclei from one donor. **c**, Relative abundance of each beta cell subtype in nondiabetic, pre-T2D and T2D donor islets. Data are shown as mean ± s.e.m. (n = 11 nondiabetic, n = 8 pre-T2D, n = 15 T2D); dots denote data points from individual donors. ***P < .001; ANOVA test with age, sex, BMI and islet index as covariates. **d**, Pearson correlation between relative abundance of beta-2 cells and HbA1c across donors (n = 34 donors). The bands around the linear regression line represent the range of 95% confidence interval; two-sided Pearson test.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

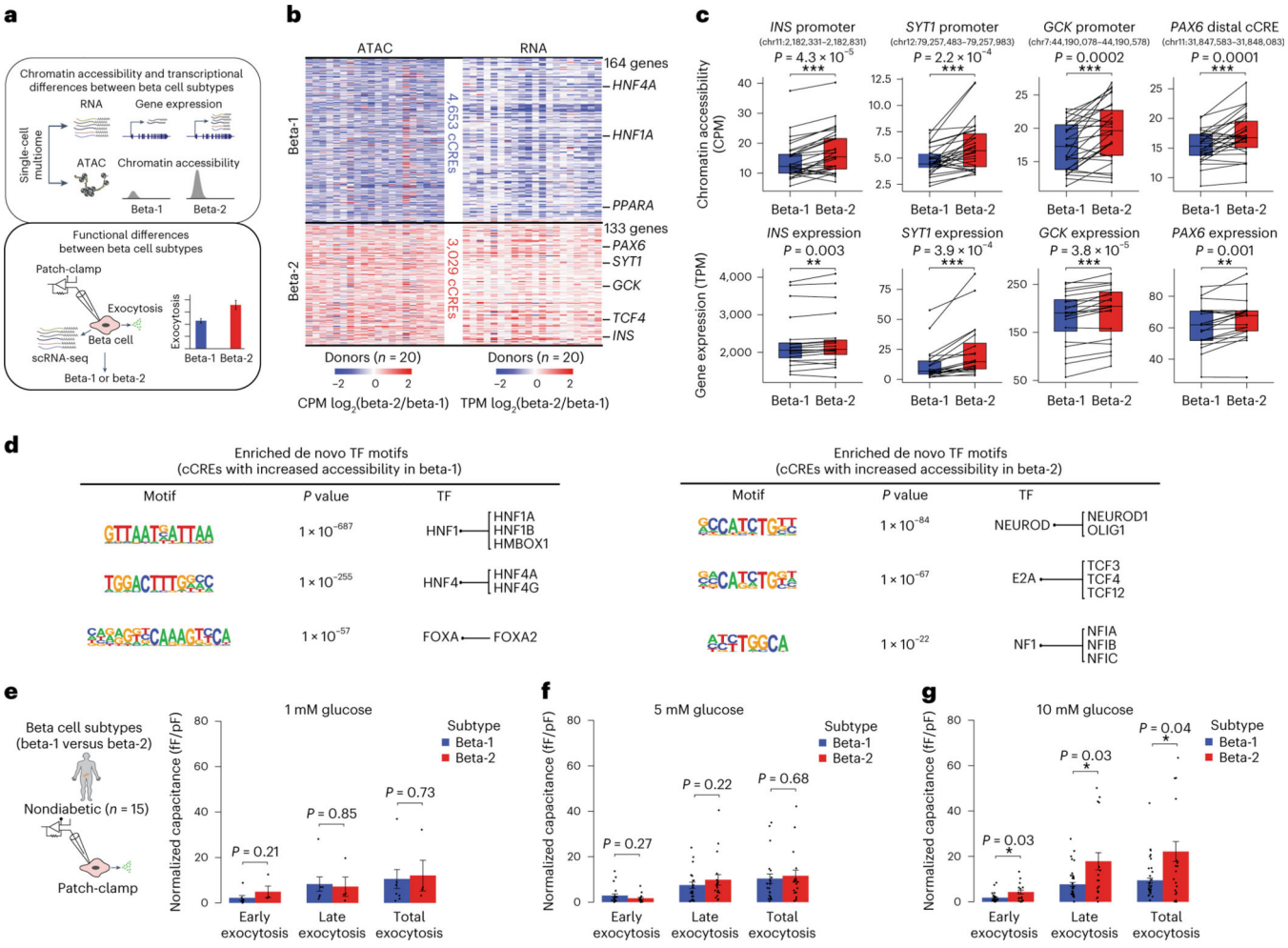


Fig. 3 |. The two beta cell subtypes are distinguished by chromatin accessibility, gene expression and function.

a, Workflow to link beta cell subtype chromatin accessibility to gene expression using islet single-nucleus multiome (ATAC + RNA) data and gene expression to function using Patch-seq. **b**, Heatmap showing $\log_2$ differences (beta-2/beta-1) in chromatin accessibility at cCREs with differential accessibility between beta cell subtypes (left, FDR < 0.05) and $\log_2$ differences (beta-2/beta-1) in gene expression of cCRE target genes with differential expression between beta cell subtypes (right, FDR < 0.15). Rows represent differential cCREs or genes, columns represent donors (n = 20 donors). Representative genes are highlighted. Accessibility of cCREs is normalized by CPM, and gene expression is normalized by TPM. Two-sided paired t-test, FDR, P values adjusted with the Benjamini–Hochberg method. **c**, Box plots showing cCRE accessibility (top) and gene expression (bottom) of representative genes in beta-1 and beta-2 cells. Proximal regions of genes are shown in parentheses. Accessibility of peaks is normalized by CPM, and gene expression is normalized by TPM. Data are shown as mean ± s.e.m., n = 20 donors, two-sided paired t-test. **d**, TF motif enrichment at cCREs with higher accessibility in beta-1 cells than in beta-2 cells (left) or higher accessibility in beta-2 cells than in beta-1 cells (right) against a background of all cCREs in beta cells using HOMER. The top three enriched de novo

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

motifs, their *P* values, and best matched known TF motif are shown. **e–g**, Bar plots from Patch-seq analysis showing early, late and total exocytosis in beta-1 cells (10 cells from 4 nondiabetic donors) and beta-2 cells (4 cells from 4 nondiabetic donors) stimulated with 1 mM glucose (**e**), in beta-1 cells (26 cells from 10 nondiabetic donors) and beta-2 cells (20 cells from 9 nondiabetic donors) stimulated with 5 mM glucose (**f**), and in beta-1 cells (42 cells from 5 nondiabetic donors) and beta-2 cells (23 cells from 6 nondiabetic donors) stimulated with 10 mM glucose (**g**). Data are shown as mean ± s.e.m.; $*P < 0.05$; ANOVA test with age, sex and BMI as covariates. pF, picofarad; fF, femtofarad.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

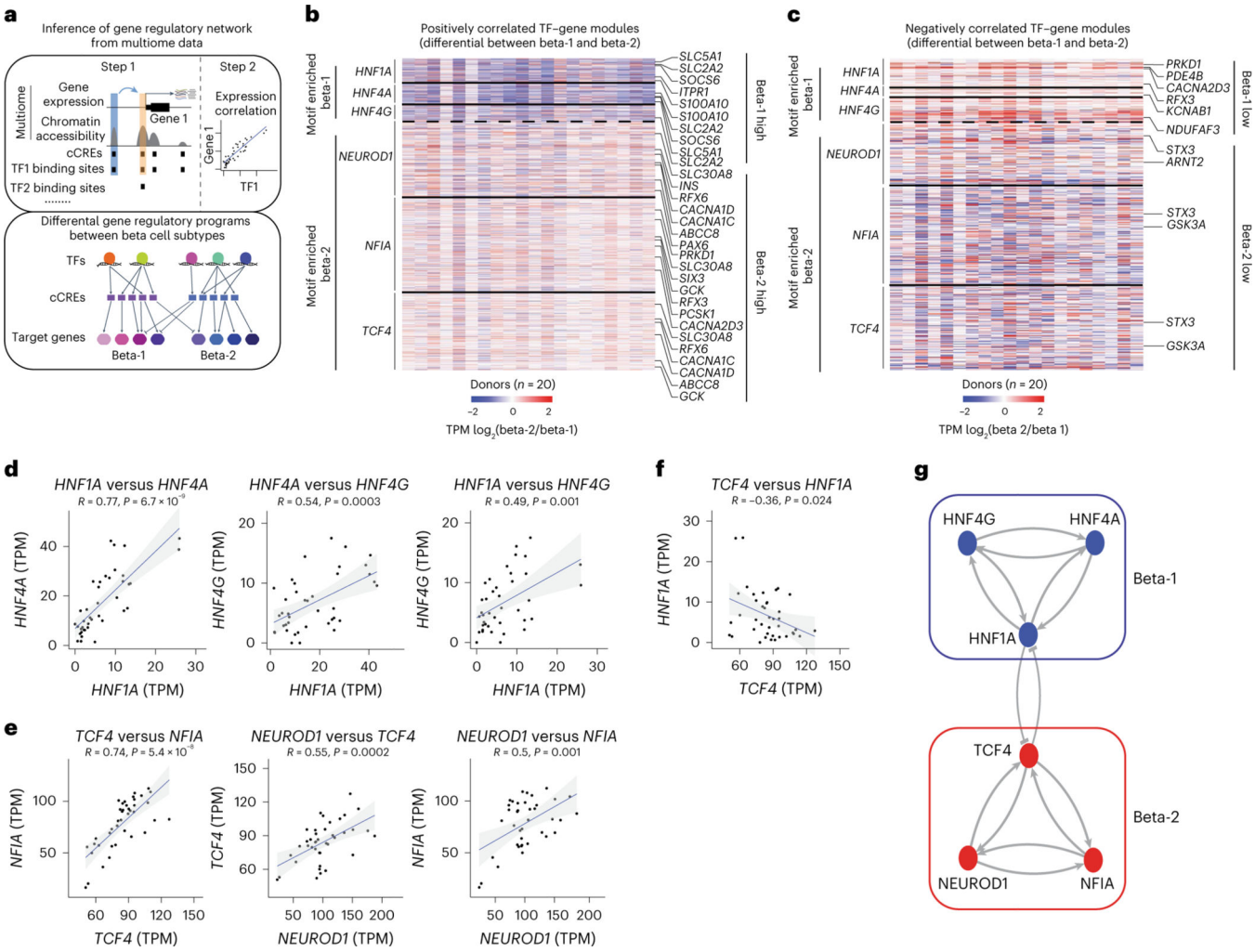


Fig. 4 |. GRNs defining the two beta cell subtypes.

a, Schematic outlining the inference of beta cell GRNs and differential gene regulatory programs (TF–gene modules) between beta cell subtypes. **b**, Heatmap showing $\log_2$ differences (beta-2/beta-1) in expression for genes positively regulated by TFs (HNF1A, HNF4A and HNF4G) with higher accessibility in beta-1 cells than in beta-2 cells and TFs (NEUROD1, NFIA and TCF4) with higher accessibility in beta-2 cells than in beta-1 cells (Methods). Representative target genes of individual TFs are highlighted. Gene expression is normalized by TPM. **c**, Heatmap showing $\log_2$ differences (beta-2/beta-1) in expression for genes negatively regulated by TFs (HNF1A, HNF4A and HNF4G) with higher accessibility in beta-1 cells than in beta-2 cells and TFs (NEUROD1, NFIA and TCF4) with higher accessibility in beta-2 cells than in beta-1 cells (Methods). Representative target genes of individual TFs are highlighted. Gene expression is normalized by TPM. **d**–**f**, Pearson correlation of expression levels between indicated TFs across pseudobulk RNA profiles from each beta cell subtype (20 donors, 40 dots in total). The bands around the linear regression lines represent the range of 95% confidence interval, two-sided Pearson test. **g**, A bistable circuit established by positive feedback among HNF1A, HNF4A and HNF4G; positive

feedback among NEUROD1, NFIA and TCF4; and mutual repression between HNF1A and TCF4.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

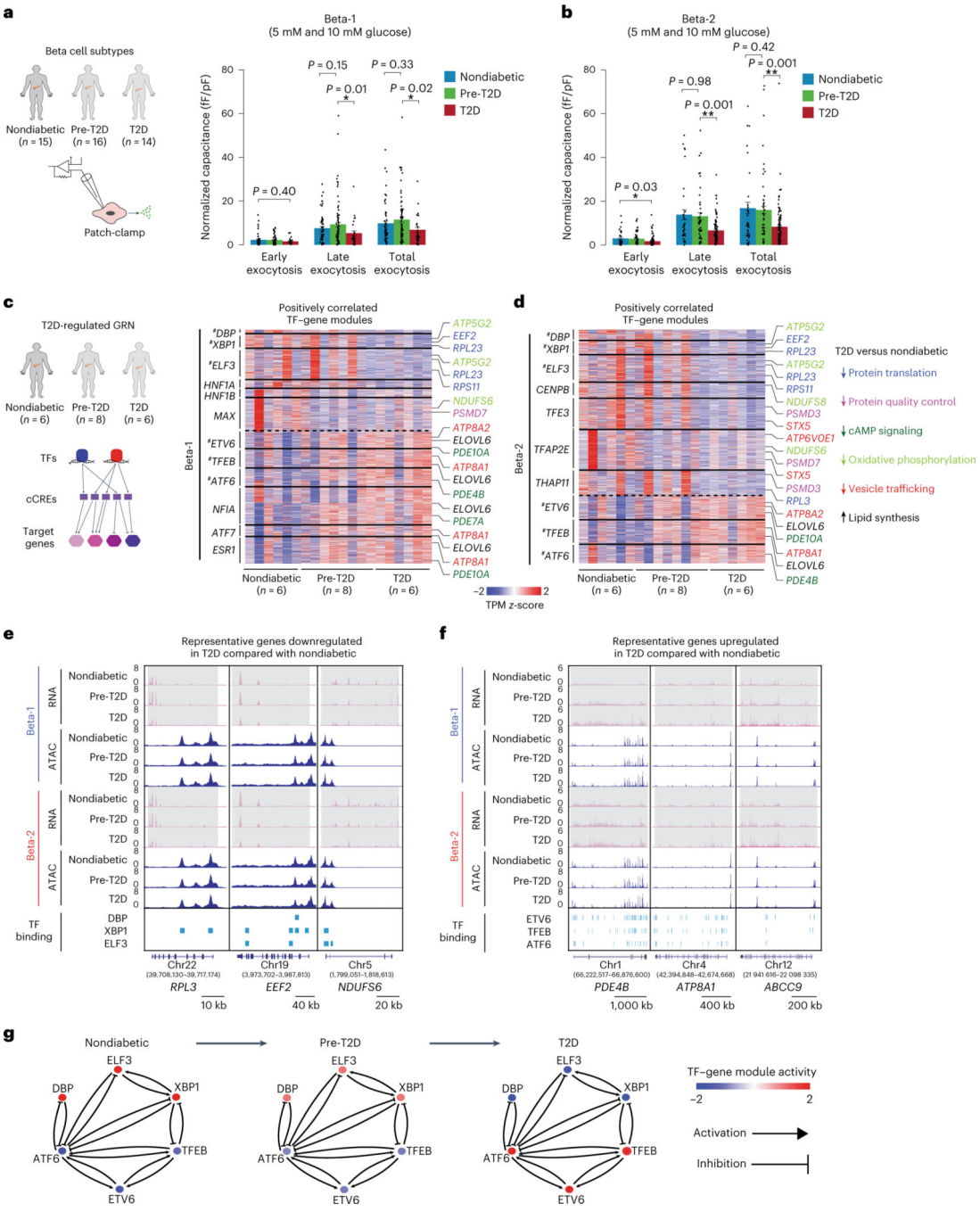


Fig. 5 |. Beta cell functional and gene regulatory changes in T2D.

a,**b**, Bar plots from Patch-seq analysis showing early, late and total exocytosis in beta-1 cells from nondiabetic (68 cells from 11 donors), pre-T2D (91 cells from 14 donors) and T2D (35 cells from 7 donors) donors (**a**) and in beta-2 cells from nondiabetic (43 cells from 10 donors), pre-T2D (57 cells from 14 donors) and T2D (131 cells from 14 donors) donors (**b**) stimulated with 5 mM or 10 mM glucose. Data are shown as mean ± s.e.m.; $*P < 0.05$; $**P < 0.01$; ANOVA test with age, sex and BMI as covariates. **c**,**d**, Heatmap showing expression of genes positively regulated by TFs with higher (green) or lower (red) activity

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

in nondiabetic compared with T2D beta-1 cells (**c**) and beta-2 cells (**d**). Representative target genes of individual TFs are highlighted and classified by biological processes. Gene expression is normalized by TPM. # denotes TFs with decreased or increased expression in T2D in both beta-1 and beta-2 cells, $n = 6$ nondiabetic, $n = 8$ pre-T2D, $n = 6$ T2D. **e**,**f**, Genome browser tracks showing aggregate RNA and ATAC read density at representative genes (*RPL3*, *EEF2* and *NDUFS6*) downregulated in T2D relative to nondiabetic for both beta-1 and beta-2 cells (**e**). Genome browser tracks showing aggregate RNA and ATAC read density at representative genes (*PDE4B*, *ATP8A1* and *ABCC9*) upregulated in T2D relative to nondiabetic for both beta-1 and beta-2 cells (**f**). Beta cell cCREs with binding sites for differential TFs in both beta-1 and beta-2 cells are shown. Differentially expressed genes in T2D beta cells are indicated by gray shaded boxes. All tracks are scaled to uniform 1×10^6 read depth. **g**, Crossregulation between TFs with activity change in T2D in both beta cell subtypes (derived from Fig. 5c,d). The color code for TFs in nondiabetic, pre-T2D and T2D donors reflects their expression change during T2D progression.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

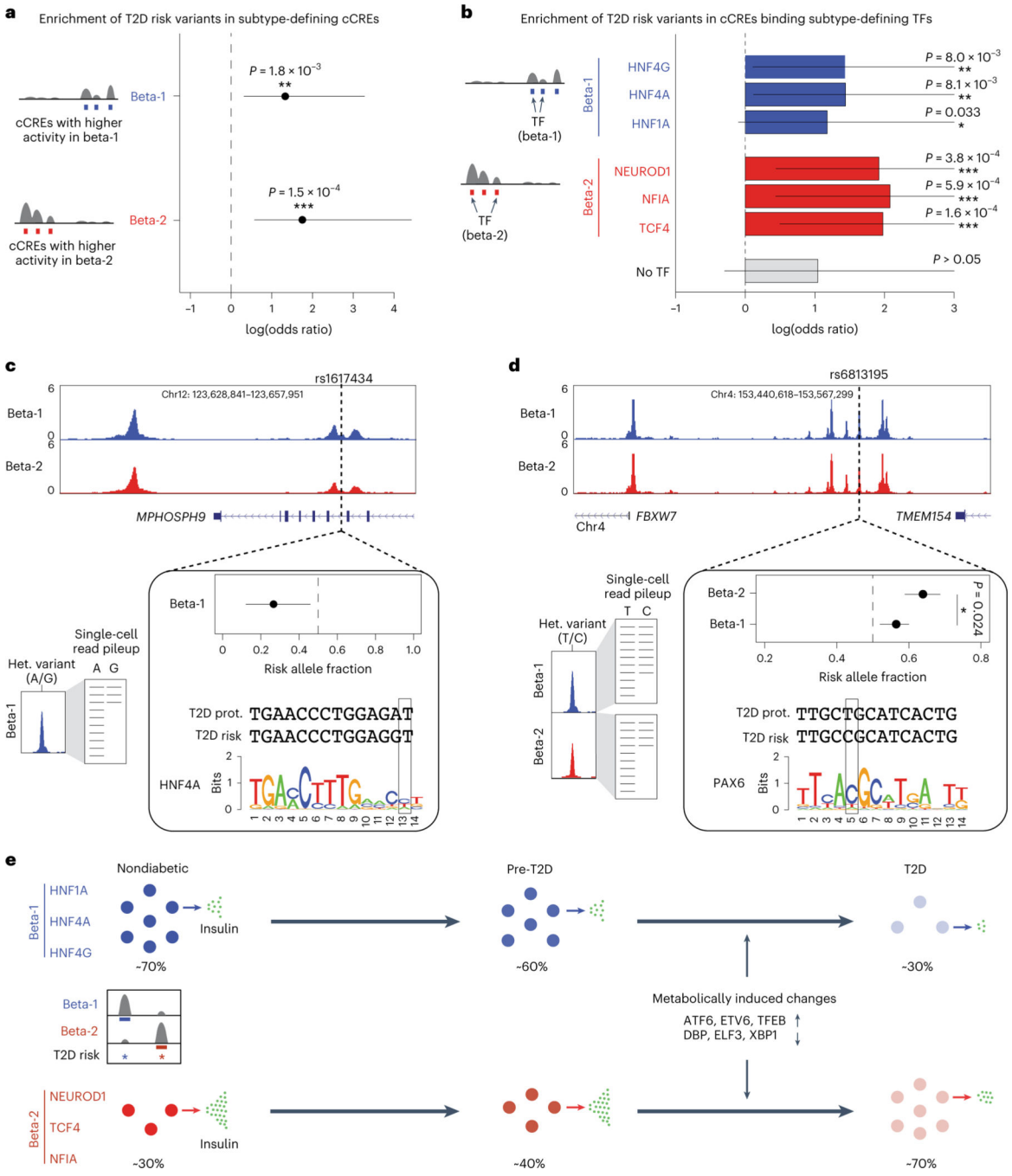


Fig. 6 |. T2D risk variants affect beta cell subtype chromatin accessibility.

a, Enrichment of fine-mapped T2D risk variants for cCREs defining the beta-1 and beta-2 subtypes. n = 4,653 and 3,029 differential cCREs for beta-1 and beta-2, respectively. Values represent log odds ratios and 95% confidence intervals. **b**, Enrichment of fine-mapped T2D risk variants for cCREs defining the beta-1 and beta-2 subtypes bound by each TF or not bound by any of the listed TFs ('no TF'). n = 16,023, 16,129, 15,743, 26,478, 13,636 and 35,694 cCREs bound by HNF4G, HNF4A, HNF1A, NEUROD1, NFIA and TCF4, respectively. Permutation test, values represent log odds ratios and 95% confidence intervals.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

c,**d**, Fine-mapped T2D risk variant rs1617434 at the *MPHOSPH9* locus overlaps a cCRE defining the beta-1 subtype. The T2D risk allele of this variant decreases beta-1 chromatin accessibility and disrupts a predicted binding site for HNF4A (**c**). Fine-mapped T2D risk variant rs6813185 at the *TMEM154/FBXW7* locus overlaps a cCRE active in both the beta-1 and beta-2 subtypes. This variant has significant heterogeneity in allelic imbalance in beta-2 and beta-2 chromatin accessibility, where the T2D risk allele has larger effect in beta-2 cells than in beta-1 cells (**d**) ($P = 0.024$). The values for allelic imbalance represent the fraction of reads from the risk allele and the 95% confidence interval. Two-sided binomial proportion tests. On the left is a schematic describing allelic imbalance mapping in reads from the beta-1 and/ or beta-2 subtype. *$P < 0.05$. **e**, Schematic showing abundance and functional changes of beta cell subtypes during T2D progression. The TFs maintaining beta cell subtype identity as well as TFs mediating T2D-associated changes are shown.