

# SCIENTIFIC REPORTS



OPEN

## Alpha TC1 and Beta-TC-6 genomic profiling uncovers both shared and distinct transcriptional regulatory features with their primary islet counterparts

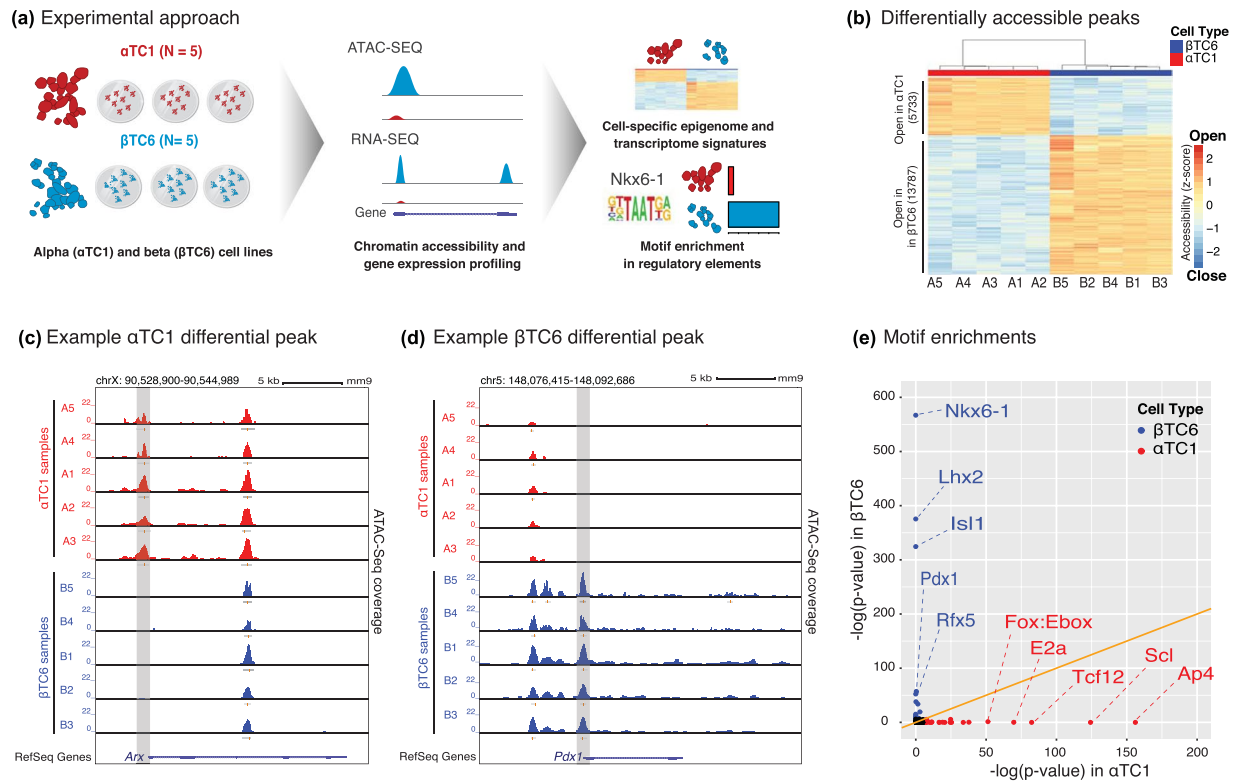
Nathan Lawlor<sup>1</sup>, Ahrim Youn<sup>1</sup>, Romy Kursawe<sup>1</sup>, Duygu Ucar<sup>1,2,3</sup> & Michael L. Stitzel<sup>1,2,3</sup>

Alpha TC1 ( $\alpha$ TC1) and Beta-TC-6 ( $\beta$ TC6) mouse islet cell lines are cellular models of islet (dys)function and type 2 diabetes (T2D). However, genomic characteristics of these cells, and their similarities to primary islet alpha and beta cells, are undefined. Here, we report the epigenomic (ATAC-seq) and transcriptomic (RNA-seq) landscapes of  $\alpha$ TC1 and  $\beta$ TC6 cells. Each cell type exhibits hallmarks of its primary islet cell counterpart including cell-specific expression of beta (e.g., *Pdx1*) and alpha (e.g., *Arx*) cell transcription factors (TFs), and enrichment of binding motifs for these TFs in  $\alpha$ TC1/ $\beta$ TC6 *cis*-regulatory elements.  $\alpha$ TC1/ $\beta$ TC6 transcriptomes overlap significantly with the transcriptomes of primary mouse/human alpha and beta cells. Our data further indicate that ATAC-seq detects cell-specific regulatory elements for cell types comprising  $\geq 20\%$  of a mixed cell population. We identified  $\alpha$ TC1/ $\beta$ TC6 *cis*-regulatory elements orthologous to those containing type 2 diabetes (T2D)-associated SNPs in human islets for 33 loci, suggesting these cells' utility to dissect T2D molecular genetics in these regions. Together, these maps provide important insights into the conserved regulatory architecture between  $\alpha$ TC1/ $\beta$ TC6 and primary islet cells that can be leveraged in functional (epi)genomic approaches to dissect the genetic and molecular factors controlling islet cell identity and function.

Pancreatic islets are heterogeneous clusters of endocrine cell types (alpha, beta, delta, gamma/PP, and epsilon) that secrete different hormones to control glucose homeostasis. Their (dys)function and/or death is central to the genetic etiology<sup>1</sup> and pathophysiology<sup>2</sup> of all forms of diabetes. Recent islet single cell<sup>3–9</sup> and enriched cell population<sup>10–13</sup> transcriptome analyses highlight the unique gene expression profiles that underlie each specific cell type's identity and function and have identified islet cell-specific expression differences between non-diabetic and Type 2 diabetic (T2D) individuals<sup>4–6</sup>. Epigenomic profiling of islet cells revealed that beta and alpha cells possess unique open chromatin landscapes that are differentially enriched for transcription factor (TF) binding sites and diabetes associated genetic variants<sup>14</sup>. Together these studies suggest that the perturbation of cell-specific transcriptomic and epigenomic programs contribute to both islet dysfunction and diabetes progression<sup>15</sup>. However, the specific functional consequences of disrupting these cell-specific regulatory programs have not been extensively elucidated.

The recent emergence of functional genomics and (epi)genome editing using the CRISPR/Cas9 platform offers a new and exciting opportunity to experimentally dissect these regulatory circuits that control islet cell identity and function and whose disruption by genetic variants and/or environmental risk factors may contribute to T2D pathogenesis. Many studies utilize islet-derived cell lines such as MIN6<sup>16–18</sup>, INS-1(832/13)<sup>19–21</sup>, beta-TC-6<sup>22–26</sup>, alpha TC1<sup>27–31</sup> and, most recently, human EndoC-BH1-3<sup>32–34</sup> to gain insights into the molecular processes governing islet cell identity and function. Genomic characterization of these cell lines is essential to

<sup>1</sup>The Jackson Laboratory for Genomic Medicine, Farmington, CT, 06032, USA. <sup>2</sup>Institute for Systems Genomics, University of Connecticut, Farmington, CT, 06032, USA. <sup>3</sup>Department of Genetics & Genome Sciences, University of Connecticut, Farmington, CT, 06032, USA. Nathan Lawlor and Ahrim Youn contributed equally to this work. Correspondence and requests for materials should be addressed to D.U. (email: [duygu.ucar@jax.org](mailto:duygu.ucar@jax.org)) or M.L.S. (email: [michael.stitzel@jax.org](mailto:michael.stitzel@jax.org))



**Figure 1.** Assay for transposase-accessible chromatin (ATAC-seq) profiling of  $\alpha$ TC1 and  $\beta$ TC6 identifies cell-type-specific open-chromatin regions. **(a)** Cartoon outline of experimental procedure.  $\alpha$ TC1 and  $\beta$ TC6 replicates were profiled using ATAC-seq and RNA-seq to characterize their transcriptomic and epigenomic landscapes. Further downstream analyses were performed including pathway and transcription factor motif enrichment analyses. **(b)** Differential analysis of open chromatin regions revealed 5,733 and 13,787 sites open in  $\alpha$ TC1 and  $\beta$ TC6 respectively. Values in heatmap reflect  $\log_2$  TMM normalized read counts after mean centering and scaling. **(c)** UCSC genome browser views of a chromatin site exclusively open in  $\alpha$ TC1 at *Arx* promoter (highlighted in grey) and **(d)** a similar site exclusively open in  $\beta$ TC6 at *Pdx1* promoter (highlighted in grey). **(e)** Sequences of differentially accessible chromatin regions demonstrate cell-type-specific binding of TF motifs. Colored points denote motifs significantly enriched (FDR < 1%) in a cell type (red =  $\alpha$ TC1, blue =  $\beta$ TC6) while black points represent motifs not enriched in either cell type. Note the cell-type-specificity of TF enrichments.

guide such studies and to interpret their findings. In the current study, we report the epigenome and transcriptome of alpha TC1 ( $\alpha$ TC1) and beta-TC-6 ( $\beta$ TC6) islet cell models with four major goals in mind: (1) to identify cell-specific epigenome and transcriptome signatures and their hallmark features; (2) to assess the sensitivity of epigenome profiling for detecting cell-specific signatures and for reflecting relative cell proportions in a mixed cell population; (3) to elucidate the extent to which these genomic features are shared with primary mouse and human alpha and beta cells; and (4) to define the T2D SNP-containing regulatory elements in human islets that are functionally conserved in these cell lines and therefore appealing targets for experimental manipulation using CRISPR/Cas9 (epi)genome editing to study their function. Together, these detailed maps reveal both important similarities and differences between these cell models and primary islet cells and provide an important resource to guide their use in future functional genomics experiments to dissect the genetic and molecular bases for islet (dys)function and diabetes.

## Results

**Alpha ( $\alpha$ TC1) and beta ( $\beta$ TC6) cell lines exhibit distinct regulatory landscapes.** To determine the transcriptional regulatory landscapes of  $\alpha$ TC1 and  $\beta$ TC6 cells, we profiled and compared their chromatin accessibility (ATAC-seq<sup>35</sup>) and expression (RNA-seq) patterns (Fig. 1a; Supplementary Table S1). Open chromatin profiling of  $\alpha$ TC1 (n = 5) and  $\beta$ TC6 (n = 5) replicates identified 65,053 consensus ATAC-seq peaks (Methods). Hierarchical clustering and principal component analysis (PCA) using these consensus peaks separated  $\alpha$ TC1 and  $\beta$ TC6 samples into two clusters (Supplementary Fig. S1), suggesting distinct and cell-specific epigenomic landscapes for these cells. Indeed, differential analyses revealed 13,787 and 5,733 differentially accessible (DA; FDR < 5%, absolute  $\log_2$  fold-change > 2) peaks that were significantly more open in  $\beta$ TC6 or  $\alpha$ TC1 cells, respectively; Fig. 1b; Supplementary Table S2). These included DA sites exclusively open in one cell type versus the other, such as the  $\alpha$ TC1-specific *Arx* (Fig. 1c) and  $\beta$ TC6-specific *Pdx1* (Fig. 1d) promoters. To identify the TFs that may modulate the observed cell-specific epigenomic landscapes, we conducted motif enrichment analysis using HOMER<sup>36</sup> (Supplementary Table S3). As expected,  $\beta$ TC6 DA peaks were enriched in motifs of TFs (Nkx6-1<sup>37-39</sup>, Isl1<sup>40,41</sup>, Pdx1<sup>42,43</sup>, and Rfx5<sup>44</sup>) necessary for islet beta cell development and survival and interestingly for

Lhx2, a LIM-HD factor linked to neurogenesis in the hippocampus<sup>45</sup> and expansion of multipotent progenitor cell populations<sup>46</sup> (Fig. 1e). In  $\alpha$ TC1-specific DA peaks, we observed motif enrichment for Fox family member TF (Fox:Ebox dimer, Foxa2, Foxo1, and Foxp1), which are involved in islet alpha cell function as suggested by *Foxp1/2/4* knockout mice developing hypoglycemia and having impaired glucagon secretion<sup>47</sup>, and *Tal1/Scl*, which targets *Ldb1*<sup>48</sup>, a coregulator of the Lin11-Isl1-Mec3 (LIM)-homeodomain (HD) complex implicated in islet alpha, beta, and delta cell development<sup>49,50</sup>. Other enriched TF motifs included Tcf12, which is involved in neural stem cell expansion<sup>51</sup>, and *Tfap4/Ap4*, a motif that interacts with *Igfbp2*<sup>52</sup>, a diagnostic and prognostic marker of pancreatic cancer<sup>53</sup>. These results highlight the cell-specific regulatory networks at work in  $\alpha$ TC1 and  $\beta$ TC6 to govern their distinct cell type identity and function and reflect those of primary alpha and beta cells.

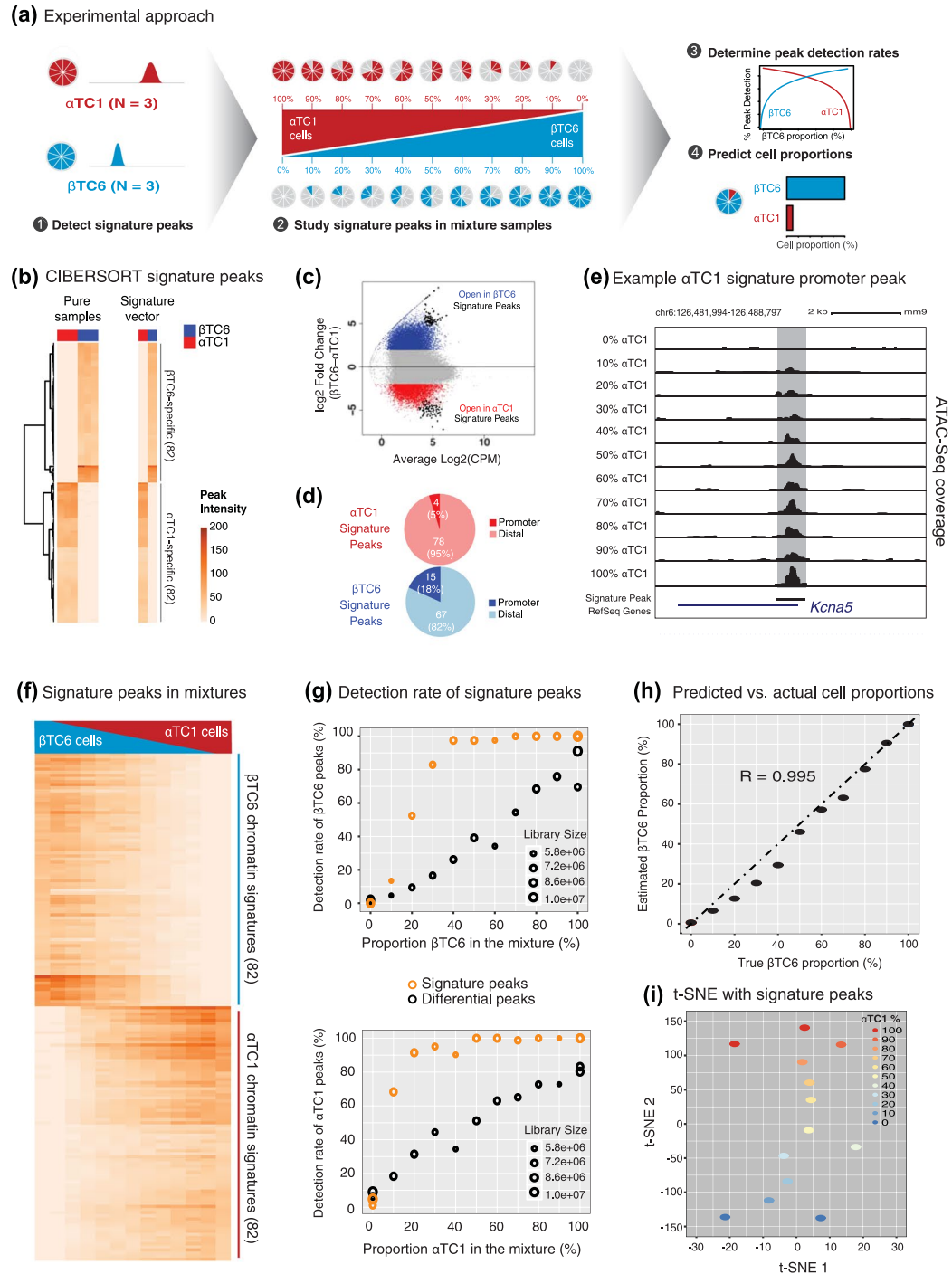
**ATAC-seq captures cell-specific patterns in heterogeneous  $\alpha$ TC1 and  $\beta$ TC6 mixtures.** Analyses of  $\alpha$ TC1 and  $\beta$ TC6 open chromatin profiles established major epigenomic differences between these homogeneous cell types. However, most genomic medicine studies profile tissues (e.g., pancreatic islets) that are composed of multiple cell types in different proportions. This cellular heterogeneity can impede the elucidation of cell-specific gene expression programs, especially those stemming from less abundant cell types<sup>3–9</sup>. To determine the sensitivity of the ATAC-seq technology to capture cell-specific epigenomic patterns within cell mixtures, we generated ATAC-seq maps from  $\alpha$ TC1/ $\beta$ TC6 mixtures ranging from 0–100% of each cell type in 10% intervals (Fig. 2a, Supplementary Fig. S2). First, we identified  $\alpha$ TC1/ $\beta$ TC6 cell-specific signature peaks using *CIBERSORT*<sup>54</sup> (Fig. 2a, Step 1). Next, we determined at what rate these signature peaks are detected in mixture samples containing variable proportions of  $\alpha$ TC1 and  $\beta$ TC6 cells (Fig. 2a, Step 2). Finally, we investigated whether the detection rates of these signature peaks reflect  $\alpha$ TC1/ $\beta$ TC6 proportions in a given mixture (Fig. 2a, Steps 3–4).

Among 65,053 consensus peaks, *CIBERSORT* selected 82 signature ATAC-seq peaks for  $\alpha$ TC1 cells ( $n = 3$ ) and 82 signature peaks for  $\beta$ TC6 cells ( $n = 3$ ) (Fig. 2b). Signature peaks (Fig. 2c, black points) exhibited the highest fold change among all  $\beta$ TC6 (blue points) and  $\alpha$ TC1 DA peaks (red points), respectively. 78/82 (95%) of  $\alpha$ TC1 and 67/82 (82%) of  $\beta$ TC6 signature peaks were distal (Fig. 2d), implying that distal regions of the genome contain more discriminative cell-specific patterns. As shown for the  $\alpha$ TC1 signature peak in the *Kcna5* promoter (Fig. 2e), we observed that read counts in signature peaks reflect the relative cell proportion in the mixture. This trend was consistent for all 164 signature peaks where read counts of  $\beta$ TC6 (Fig. 2f, top) and  $\alpha$ TC1 (Fig. 2f, bottom) signature peaks increased proportionately to their relative representation in the mixture. This demonstrates that ATAC-seq is sensitive enough to capture chromatin accessibility of a cell-specific regulatory element proportionately to that cell type's contribution to the mixture. However, we noted that the detection rates of signature peaks also depend on the number of raw read counts at these loci. Samples with lower sequence depth contained signature peaks at rates lower than expected, such as the sample that is composed of 40%  $\alpha$ TC1 (Fig. 2g, bottom plot; Supplementary Fig. S3). These results suggest that if ATAC-seq will be used to study cells that are less abundant in a population (i.e., represent a small proportion of the heterogeneous cell composite), the libraries should be sequenced more extensively to capture signatures stemming from this cell type of interest.

The detection rates of signature peaks in mixture samples were far greater than those of DA peaks, since signature peaks have more reads on the average (Fig. 2g, Supplementary Fig. S3). Notably, > 50% of signature peaks were detectable within mixtures when the corresponding cell type proportion was as low as 20% (Fig. 2g bottom panel). Finally, we studied whether ATAC-seq could be used to estimate/predict cell type proportions from these mixtures. Indeed, cell compositions of the mixtures were accurately predicted using signature peaks, where the correlation between estimated and true proportions is 0.995 (Fig. 2h). Moreover, t-SNE (t-distributed stochastic neighbor embedding)<sup>55</sup> arranged the 13 mixture samples with respect to their relative cell composition based on the signature peaks (Fig. 2i, t-SNE 2). Variability between pure  $\alpha$ TC1 and  $\beta$ TC6 replicates in the first t-SNE dimension reflects differences in their library size (Supplementary Fig. S4). The second t-SNE dimension reflects cell compositions and suggests that replicate epigenomes correlate well (Supplementary Fig. S2). Notably, similar t-SNE analysis using all  $\alpha$ TC1/ $\beta$ TC6 DA peaks, but not all consensus peaks (Supplementary Fig. S4), separated samples by cell type proportion, suggesting that cell-specific regulatory elements reflect the relative contribution of the corresponding cell to the mixture. Taken together, these analyses support the utility of ATAC-seq to profile cell mixtures and predict relative cell type composition in clinically relevant heterogeneous samples, as has been previously done with human blood samples<sup>56</sup>. Thus, ATAC-seq profiles of sorted human pancreatic islet cells (e.g. alpha, beta) would allow the estimation of relative cell type compositions in human pancreatic islets.

To test this hypothesis, we generated ATAC-seq libraries for human islets from seven individuals (Khetan *et al.*, in preparation), for which we also have estimated cell type proportions based on single cell transcriptome profiling<sup>4</sup> (Supplementary Table S4). We identified 28 alpha and 50 beta signature peaks from enriched human alpha and beta cell ATAC-seq profiles using the same data analysis pipeline<sup>14</sup>. Deconvolution of bulk islet ATAC-seq profiles with these signatures yielded alpha cell proportions that closely resembled counts for each cell type from single cell RNA-seq data (Supplementary Fig. S5;  $R = 0.9$ ,  $p = 0.005$ ). Beta cell proportions were less precisely estimated ( $R = 0.5$ ,  $p = 0.25$ ), potentially due to the contamination of other cell types (e.g., delta cells) during FACS-enrichment. These data provide encouraging proof-of-concept results suggesting the feasibility of computational deconvolution of islet cell composition using ATAC-seq of islet samples. We anticipate that improvements in cell sorting/enrichment protocols and resulting profiles of more and purer islet cell subpopulations will dramatically improve the predictive capacity of this algorithm.

**Cell-type-specific expression of genes involved in islet signaling, hormone secretion, and metabolism.** We completed RNA-seq and quantified gene expression of 24,531 protein coding genes and long intergenic non-coding RNAs (lincRNA) in  $\alpha$ TC1 and  $\beta$ TC6 replicates (Methods;  $n = 3$  each). Hierarchical clustering and PCA of  $\alpha$ TC1 and  $\beta$ TC6 transcriptomes using all detected genes (FPKM  $\geq 1$ ;  $N = 12,234$ ),



**Figure 2.** High sensitivity of ATAC-seq technology permits accurate open chromatin profiling of heterogeneous cell mixtures. **(a)** Cartoon representation of experimental workflow. Briefly, cell-specific “signature” peaks were defined for both  $\alpha$ TC1/ $\beta$ TC6 (Step 1). Next, the sensitivity of these cell-specific sites were compared in each heterogeneous mixture sample (Step 2) and used to assess detection rates of cell-specific chromatin sites (Step 3) and finally to predict each sample’s cellular composition (Step 4). **(b)** Signature peaks determined by CIBERSORT. Heatmap values represent TMM normalized read counts (peak intensity). Signature vector represents the median accessibility profile for these signature peaks. **(c)** MA plot highlighting DA peaks specific to  $\alpha$ TC1 (red) and  $\beta$ TC6 (blue). Signature peaks are colored in black (CPM = counts per million). **(d)** Genomic locations of signature peaks. Note most signature peaks are distal. **(e)** UCSC genome browser view of an  $\alpha$ TC1 signature peak at *Kcna5* promoter, that displays decreased accessibility as the  $\alpha$ TC1 proportions decreases in mixture samples. **(f)** Heatmap illustrating the peak intensity of the 82  $\alpha$ TC1 and 82  $\beta$ TC6 signature peaks in all mixture samples. **(g)** Scatterplots comparing the detection rate of the 13,787 differential (black) and 82 signature  $\beta$ TC6 peaks (top) and the 5,733 differential (black) and 82 signature  $\alpha$ TC1 peaks (bottom) in all mixture samples. Sizes of points in the scatterplot reflect respective library sizes (reads) for each sample. **(h)** Estimated cellular compositions of each mixture sample (y-axis), as determined

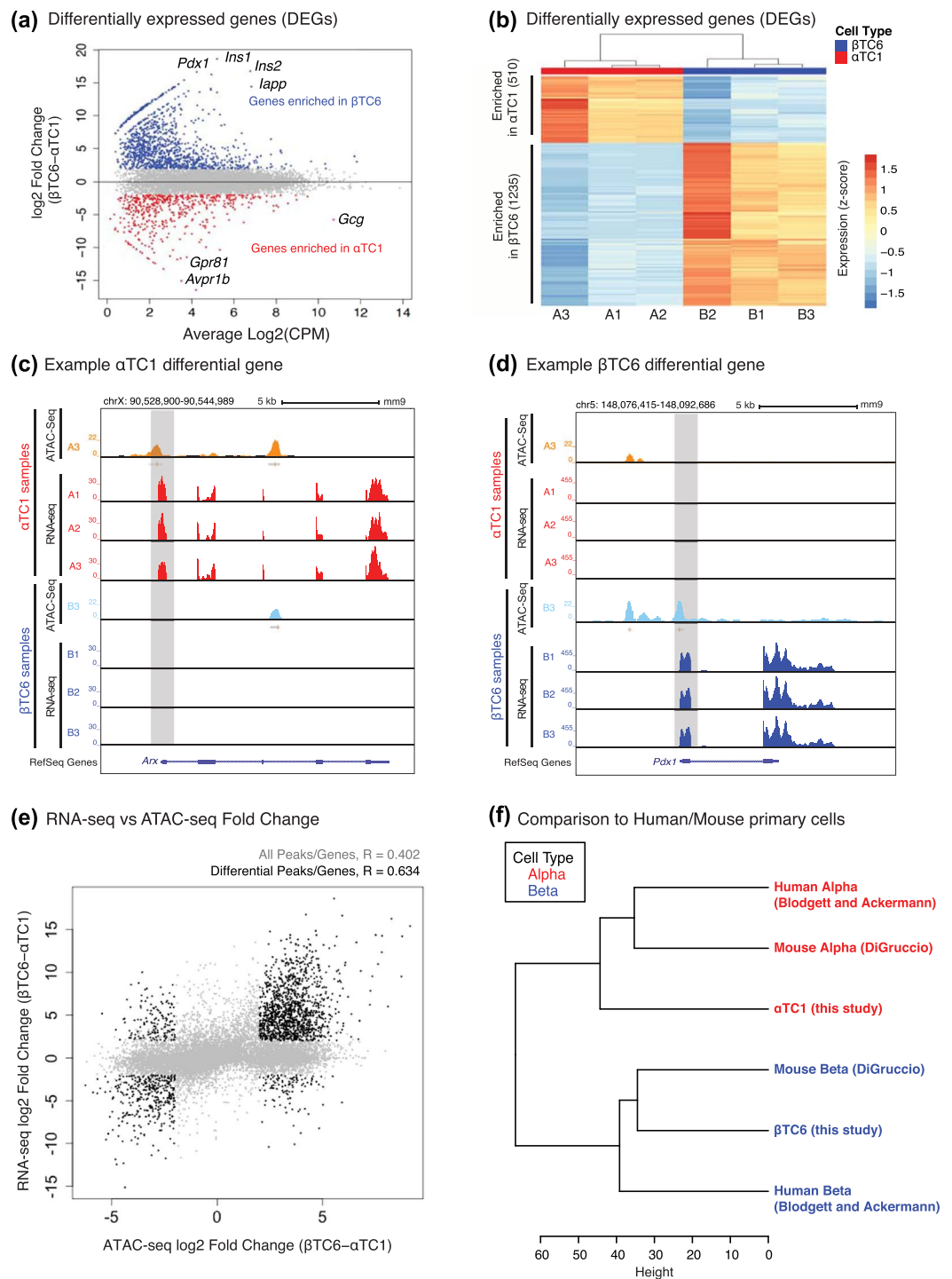
by *CIBERSORT*, closely matches that of true cellular compositions (x-axis). R represents Pearson's correlation coefficient. (i) t-SNE analyses of all mixture samples using the 164 *CIBERSORT*-defined signature peaks demonstrates clustering of samples based on their cellular composition.

separated samples into two cell-specific clusters (Supplementary Fig. S6) highlighting that the majority of the variability captured in the transcriptomic data is attributable to cell-type-specific gene expression patterns. Differential gene expression analysis revealed 510  $\alpha$ TC1-enriched ( $\log_2$  fold change  $< -2$ ; red points in Fig. 3a) and 1,235  $\beta$ TC6-enriched ( $\log_2$  fold change  $> 2$ ; blue points in Fig. 3a) genes at FDR  $< 5\%$  (Fig. 3a,b) (Methods). Genes specifically expressed in  $\alpha$ TC1 included classic alpha cell maturation TF genes such as *Irx1*<sup>12,57</sup> ( $\log_2$  FC -3.69), *Irx2*<sup>12,57</sup> ( $\log_2$  FC -3.70), and *Maftb*<sup>58</sup> ( $\log_2$  FC -5.95), while  $\beta$ TC6-specific transcripts included genes encoding established beta cell TFs *Nkx6-1* ( $\log_2$  FC 2.18), *Pax4*<sup>59,60</sup> ( $\log_2$  FC 9.48), and *Mafa*<sup>61,62</sup> ( $\log_2$  FC 7.46). In addition to the rodent insulin-encoding genes (*Ins1* and *Ins2*), top differentially expressed (DE) genes in  $\beta$ TC6 included a transmembrane receptor with unknown function (*Tmem215*), a regulator of beta cell mass (*Egfr*)<sup>63</sup>, and lipid transporter (*Abca5*). Top DE genes in  $\alpha$ TC1 included the glucagon-encoding gene *Gcg* and *Avpr1b*, a G-protein coupled receptor (GPCR) whose activation contributes to increased glucagon secretion (and indirectly insulin secretion) in islets<sup>64,65</sup>. Epigenomic patterns captured in these cells were concordant with the cell-specific changes in gene expression. For example, at the *Arx* locus, which encodes a TF regulating islet alpha cell development<sup>66,67</sup>, we noted  $\alpha$ TC1-specific chromatin accessibility and expression activity (Fig. 3c, grey box). Similarly,  $\beta$ TC6-specific chromatin accessibility pattern at the *Pdx1* promoter (Fig. 3d, grey box), an islet beta cell TF necessary for cell survival and function<sup>42,43</sup>, was accompanied by  $\beta$ TC6-specific expression of this gene.

KEGG pathway analyses (Supplementary Table S5) revealed significant enrichment (FDR  $< 1\%$ ) of genes (*Adcy1*, *Adcy2*, *Adcy3*, *Adcy9*, *Camk2b*) involved in insulin secretion, cAMP signaling, and calcium signaling in  $\beta$ TC6 cells, consistent with reported roles for these pathways in modulating beta cell insulin secretion. Similarly, we observed enrichment of genes associated with the molecular genetics of type 2 diabetes (T2D) and Maturity Onset Diabetes of the Young (MODY; e.g. *Hnf1a*, *Hnf4a*, *Slc2a2*, *Gck*) in  $\beta$ TC6. In the past decade, islet genomic studies have found that increased apoptosis and dysregulation of human beta cells are major contributing factors to diabetes development and susceptibility<sup>68–71</sup>. Consistent with recent studies highlighting the importance of circadian rhythmicity in beta cell function and insulin secretion<sup>23,72,73</sup>, we also observed enrichment of the circadian entrainment pathway in  $\beta$ TC6 cells.  $\beta$ TC6 transcriptomes were also enriched for genes involved in oxytocin signaling - a hormone and pathway that has been linked recently to mitigating metabolic stress in mouse pancreatic islets and a mouse beta cell line (MIN6) in a recent study<sup>74</sup>. Genes associated with gluconeogenesis and glycolysis (e.g. *Fbp1*, *Fbp2*, *Hk1*, *Hk2*) were enriched in  $\alpha$ TC1, reflecting the roles for islet alpha cells as a counterbalance to maintaining blood glucose and energy homeostasis. Ultimately, transcriptomics of  $\alpha$ TC1 and  $\beta$ TC6 suggests that these cell lines reflect pathways and TFs that are conserved in islet alpha and beta cells, and therefore may effectively model molecular and cellular mechanisms relevant to islet biology.

**Chromatin accessibility and gene expression are highly correlated in  $\alpha$ TC1 and  $\beta$ TC6.** To uncover the association between cell-specific epigenomic and transcriptomic patterns, we studied the correlation between RNA-seq and ATAC-seq profiles in  $\alpha$ TC1 and  $\beta$ TC6. 14,699/24,531 genes (~60%) considered in this study were associated with at least one open chromatin site. On average, four open chromatin sites mapped to a given gene, consistent with other studies indicating that multiple distal regulatory elements converge to govern transcription. Therefore, to consider a one-to-one relationship between chromatin accessibility and expression level of a gene, we assigned the open chromatin peak with the highest absolute fold change in accessibility between cell types to its corresponding gene. Overall correlation between gene expression changes ( $\beta$ TC6 vs.  $\alpha$ TC1  $\log_2$  fold changes) and chromatin accessibility changes for these 14,699 genes (Fig. 3e, grey points) was positive and significant (Pearson  $R = 0.402$ ,  $p < 2.2 \times 10^{-16}$ ). This correlation was more robust (Pearson  $R = 0.634$ ,  $p < 2.2 \times 10^{-16}$ ) for regions involving cell-specific chromatin accessibility and gene expression patterns ( $n = 1,833$  genes, black points), suggesting that transcriptional programs of cell-specific regulatory elements (both promoters and enhancers) are more tightly regulated than the rest of the genome by the epigenomic landscape<sup>75</sup>. These observations confirm significant associations between cell-type-specific gene expression programs and the chromatin accessibility patterns around these loci. Further investigation revealed that 17% of  $\alpha$ TC1- and 31% of  $\beta$ TC6-specific peaks corresponded to differential gene expression in that cell type (Supplementary Fig. S7;  $p < 1 \times 10^{-16}$ , Fisher's exact test). In contrast, shared peaks were not associated with differentially expressed genes ( $p = 1$ , Fisher's exact test). In primary human islet cells, only 12% of beta- and 5% of alpha-specific ATAC-seq peaks mapped to differentially expressed genes<sup>14</sup>. Together, these results suggest that alpha cells exhibit less congruence between cell-specific epigenomic and transcriptomic patterns, which may represent additional molecular hallmarks of alpha cell plasticity<sup>11</sup> and trans-differentiation potential/proensity<sup>76</sup>.

**Identification of shared and distinct gene expression programs between  $\alpha$ TC1/ $\beta$ TC6 and their primary cell counterparts.** To evaluate if these cell lines reflect transcriptional features of primary islet cells, we compared these profiles with those of human<sup>13,14</sup> and mouse<sup>77</sup> alpha/beta primary cells. After batch correction (Methods), primary and cell line sample transcriptomes clustered primarily by cell type (e.g. alpha or beta) (Fig. 3f), suggesting that suggesting that several alpha and beta cell-specific differences are preserved between primary and cell line samples.  $\beta$ TC6 were most similar to primary mouse beta cells. In contrast,  $\alpha$ TC1 samples clustered separately from primary human/mouse alpha cells, suggesting this cell line is less reflective of its primary alpha cell counterparts.



**Figure 3.** Transcriptome profiling (RNA-seq) of  $\alpha$ TC1 and  $\beta$ TC6 characterizes genes uniquely enriched in each cell type. **(a,b)** Differential gene expression analysis identifies 510 and 1,235 genes with enriched expression in  $\alpha$ TC1 and  $\beta$ TC6 respectively (CPM = counts per million). Values in heatmap reflect  $\log_2$  FPKM values after mean centering and scaling. **(c)** UCSC genome browser views of  $\alpha$ TC1-specific expression of *Arx* (highlighted in grey) and **(d)**  $\beta$ TC6-specific expression of *Pdx1* (highlighted in grey). For each view, representative ATAC-seq profiles for a single  $\beta$ TC6 (B3; light blue) and  $\alpha$ TC1 (A3; orange) sample were included. **(e)** Scatterplot of  $\beta$ TC6 vs.  $\alpha$ TC1 fold changes in gene expression (y-axis) and chromatin accessibility (x-axis) illustrates a positive and significant correlation at regions of the genome that display cell-specific chromatin accessibility and gene expression. R represents Pearson's correlation coefficient. **(f)** Hierarchical clustering of mouse<sup>77</sup> and human<sup>13,14</sup> primary islet cell and  $\alpha$ TC1/ $\beta$ TC6 cell line transcriptomes groups samples by cell type regardless of species.

Closer examination of these datasets identified 1,393 (and 1,188) genes that were upregulated (FDR < 5%) in both  $\beta$ TC6 and mouse beta cells (and in  $\alpha$ TC1 and mouse alpha cells) (Supplementary Table S6) ( $p < 2.2 \times 10^{-16}$ , Fisher's exact test). Overlaps between cell-specific genes in human primary alpha/beta cells and mouse cells (both cell lines and sorted cells) were less significant, potentially due to species-specific divergence in gene expression programs<sup>78,79</sup> (Supplementary Table S7). 111 genes were commonly enriched among human beta, mouse beta, and mouse  $\beta$ TC6 cells, whereas 83 genes were commonly enriched among human alpha, mouse alpha, and mouse  $\alpha$ TC1 transcriptomes (Supplementary Fig. S8). These 111 genes were enriched in KEGG pathways associated with insulin secretion, resistance, and diabetes (FDR < 10%) (Supplementary Table S6). In contrast, the 83 genes with alpha and  $\alpha$ TC1 activity were enriched for Rap1 and calcium signaling pathways - important activating components of GSIS<sup>80,81</sup> - and also the glucagon signaling pathway, a hallmark function of islet alpha cells (FDR ~ 15%). Genes uniquely enriched in  $\beta$ TC6 and  $\alpha$ TC1 were enriched for ribosome, cell development, and morphogenesis KEGG pathways and GO terms (FDR < 10%; Supplementary Table S8), possibly reflecting the increased energy demands and cycling nature of these cells compared to their primary cell counterparts. In contrast, significantly enriched GO terms for genes specific to primary mouse and human cells were more generic and included regulation of biological process and regulation of cellular process. Notably, we observed genes enriched in human/mouse primary alpha cells (e.g., *Gc*, *Prkar1a*, *Epcam*, *Igf1r*, *Dpp4*) that were not expressed (CPM = 0) in  $\alpha$ TC1 suggesting there are clear components of islet alpha cell biology not reflected in these cell lines. Expression of human-mouse orthologues in the primary and cell line datasets examined here are provided as a resource for investigators to determine the conservation of specific genes or pathways of interest (Supplementary Table S9).

Benner *et al.*<sup>82</sup> identified 9,474 genes shared between primary human and beta cell transcriptomes. Within primary human<sup>13</sup> and mouse<sup>77</sup> beta transcriptomes examined in our study, 7,828/9,474 genes (82%) (FPKM > 1 in both human/mouse beta cells) overlapped between Benner *et al.* and these datasets. We determined 7,828 genes detected both in the primary human beta<sup>13</sup> and primary mouse beta<sup>77</sup> transcriptomes (FPKM > 1) that overlapped these 9,474 genes. Minor discrepancies between these gene sets can be attributable to a variety of technical factors between the human alpha and beta data examined here<sup>13</sup> and in the other study<sup>10</sup>, including differences in FACS sorting/gating (Newport Green staining<sup>10</sup> vs. paraformaldehyde fixation<sup>13</sup> of cells) vs. sorting of live cells expressing cell-specific reporters<sup>82</sup>, library preparation, reference transcriptome builds, and pre-processing of sequencing data.

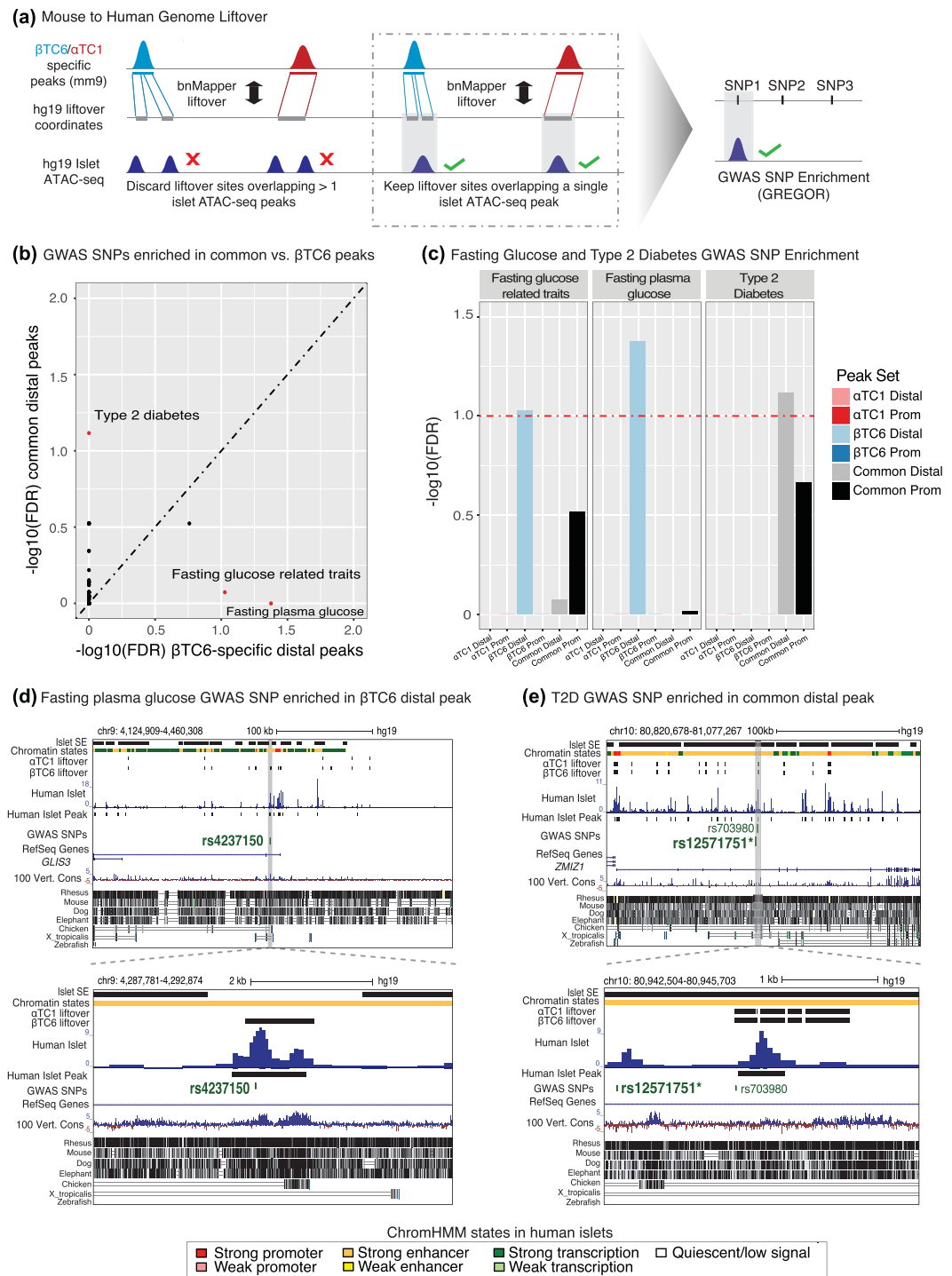
### Open chromatin sites common to $\alpha$ TC1 and $\beta$ TC6 overlap genetic variants associated with T2D.

Sequence variants identified by genome wide association studies (GWAS) are enriched in distal regulatory elements of disease relevant cell types<sup>83–86</sup>. We sought to determine which of these SNP-containing regulatory elements in human islets were evolutionarily functionally preserved in  $\alpha$ TC1/ $\beta$ TC6 cells. First, we identified the human-mouse orthologous sequences that overlapped ATAC-seq peaks in  $\alpha$ TC1/ $\beta$ TC6 cells and in human islets. Then, we tested if T2D GWAS SNPs were significantly enriched in these evolutionarily conserved and functionally preserved sequences. We used *bnMapper*<sup>87</sup> to identify orthologous human sequences corresponding to  $\alpha$ TC1/ $\beta$ TC6 ATAC-seq peaks (Fig. 4a; Methods). As expected, sequences overlapping mouse ATAC-seq peaks that lifted over from mm9 to hg19 genomes exhibited significantly higher sequence conservation than those that did not ( $p < 1e-16$ , Wilcoxon rank test)(Supplementary Fig. S9; Supplementary Table S2). Among these liftover human sequences, only those overlapping with human islet ATAC-seq peaks were retained. Liftover peaks overlapping bulk islet ATAC-seq peaks showed more extensive sequence conservation ( $p < 1e-16$ , Wilcoxon rank test) than non-overlapping peaks. Next, we retrieved all GWAS SNPs from the NHGRI-EBI Catalog (<https://www.ebi.ac.uk/gwas/>) and determined which disease-associated SNPs were enriched in these liftover/overlapping sites using GREGOR<sup>88</sup>. Among the 636 GWAS traits/diseases tested, only variants associated with 'fasting glucose related traits' ( $p = 2.96 e-04$ ), and 'fasting plasma glucose' ( $p = 6.62 e-05$ ) were significantly enriched in distal  $\beta$ TC6 peaks (Fig. 4b,c; Supplementary Table S10). For example, a SNP associated with fasting plasma glucose (rs4237150) overlapped a  $\beta$ TC6-specific peak occurring within the *GLIS3* locus (Fig. 4d). In a non-obese mouse model for type 1 diabetes, SNPs in *Glis3* have been associated with increased beta cell unfolded protein response and apoptosis<sup>89</sup>. These findings suggest that this genetic variant is contributing to impaired fasting glucose and pre-diabetic traits through altered beta cell activity, which is also reflected in the  $\beta$ TC6 cell line. Although not significant ( $p = 0.269$ , Wilcoxon rank test), overlapping ATAC-seq peaks mapping to GWAS SNPs exhibited a trend of higher sequence conservation compared to peaks not overlapping GWAS SNPs.

Distal peaks shared by both cell types were the most enriched in SNPs associated with T2D ( $p = 1.20 e-04$ ), suggesting that T2D SNPs may alter both alpha and beta cell transcriptional regulation through these loci. For example, an intronic region of *ZMIZ1* gene was accessible both in human islets and  $\alpha$ TC1/ $\beta$ TC6 cells and harbored the T2D index SNP (Fig. 4e, rs1251751). Perturbation of *ZMIZ1* activity has been recently associated with altered insulin secretion/exocytosis<sup>90</sup>. This locus also coincided with a human islet stretch enhancer (SE) suggesting that the SNP rs1251751 may alter activity of this islet-specific transcriptional enhancer. Furthermore, the risk allele of the overlapping SNP (rs703977) in LD with rs1251751 was conserved in  $\alpha$ TC1/ $\beta$ TC6 sequences (Supplementary Table S11), suggesting this may be a relevant target for experimental manipulation in these cell lines. Together, these findings suggest that a subset of T2D SNPs may impact both alpha and beta cell regulation and delineate the regulatory site(s) for which (epi)genome editing in  $\alpha$ TC1/ $\beta$ TC6 should provide important insights into the molecular genetics of T2D.

### Discussion

In this study, we performed the first integrative transcriptomic and epigenomic analysis of  $\alpha$ TC1 and  $\beta$ TC6 alpha and beta cell-derived cell lines to define their transcriptional regulatory elements and compare it to that of primary islet alpha and beta cell types. This approach revealed to what extent these profiles overlaps with their corresponding mouse and human primary islet cell types. The insights gleaned from these analyses highlight



**Figure 4.** Common open chromatin sites in  $\alpha$ TC1 and  $\beta$ TC6 demonstrate robust enrichment for T2D GWAS SNPs. **(a)** Cartoon schematic detailing the liftover analysis conducted from mouse genome (mm9) to human genome (hg19) using *bnMapper* and SNP enrichment analysis using *GREGOR* (Methods). **(b)** Scatterplot illustrating the false discovery rate (FDR) adjusted p-value enrichment scores for each category of GWAS SNPs in common distal peaks (y-axis) and  $\beta$ TC6-specific distal peaks (x-axis). Names of the GWAS categories that passed significance threshold (FDR < 10%) are displayed on the plot and the points for these categories are represented in red. **(c)** Barplots of FDR enrichment scores of GWAS SNPs for T2D and fasting glucose categories in each peak set. Horizontal dashed red line indicates an FDR threshold of 10%. **(d)** UCSC genome browser view of a  $\beta$ TC6-specific distal peak at *GLIS3* directly overlapping rs4237150, a GWAS SNP associated with fasting plasma glucose levels. **(e)** UCSC genome browser view of a common distal peak at the *ZMIZ1* locus overlapping a SNP that is in LD with a T2D associated SNP (rs12571751; indicated by bold and asterisk). Human islet stretch enhancer (SE) and chromatin state information were obtained from Parker et al.<sup>85</sup>. 100 Vertebrates Basewise Conservation by PhyloP and MultiZ Alignments of 100 Vertebrates tracks are provided to illustrate the sequence conservation for the highlighted ATAC-seq peaks between human, mouse, and other species.



important areas of utility and limitation of these cell models in studying the molecular genetics of islet cell function. Moreover, the epigenome and transcriptome maps generated represent an important resource to guide data-driven approaches for future experimentation in them.

ATAC-seq epigenome profiling suggests that the  $\beta$ TC6 epigenome is shaped by TFs such as Nkx6-1 and Pdx1 similar to primary beta cells<sup>14</sup> and islets<sup>91,92</sup> (Supplementary Fig. S10). We also identified a potentially novel TF for islet biology, Lhx2, which is conserved in  $\beta$ TC6 and human beta cell open chromatin landscapes<sup>14</sup>. However, the gene is expressed in  $\beta$ TC6 and not in human beta cells (Supplementary Table S9). Further investigation of the Lhx2 motif consensus sequence revealed strong resemblance to other homeobox factors (e.g. Pdx1, Nkx6-1, Isl1) (Supplementary Fig. S11) suggesting that enrichment of  $\beta$ TC6 specific peaks for the Lhx2 motif may be reflecting an enrichment of homeobox-associated factors. Further study of Lhx2 could elucidate whether this factor plays a regulatory role in islets. In human studies, binding sites for ISL1 and RFX5 were enriched in alpha cells<sup>14</sup>, whereas we observed these motifs to be more accessible in  $\beta$ TC6 cells. Notably, Isl1 has been implicated both in beta cell development via regulation of *Pdx1* and *Slc2a2*<sup>41</sup> and in alpha cell development via activation of *Arx* transcription<sup>93</sup>. Rfx5 was more enriched in  $\beta$ TC6 DA peaks, consistent with reported Rfx family factor roles in beta cell differentiation and insulin secretion<sup>44,94</sup>.

Although position weight matrices (PWM) are missing for well-established alpha cell TFs such as *Arx* and therefore could not be included in this study, ATAC-seq profiles of  $\alpha$ TC1, primary alpha, and islets were strikingly similar at this locus (Supplementary Fig. S10). Despite the missing PWM information, we validated the enrichment of Fox family TFs, (Fox:Ebox dimer, Foxa2, Foxo1, and Foxp1), which are implicated in regulation of alpha cell development and function<sup>47</sup>, in  $\alpha$ TC1-specific DA peaks. This contrasts with a recent report describing enrichment of TFAP4/AP4 and FOX sequence motifs in both human alpha/beta open chromatin regions<sup>14</sup>, and could represent mouse-human differences similar to those reported for *Six* and *Maf* transcription factor expression<sup>78,79</sup>. Indeed, expression of *Tfap4*/*Ap4* was higher in primary mouse and  $\beta$ TC6/ $\alpha$ TC1 cells (average log<sub>2</sub> CPM > 2) in comparison to that of human primary beta (average log<sub>2</sub> CPM ~ 1.5) and alpha (average log<sub>2</sub> CPM ~ 0.8) cells (Supplementary Table S9). Further study of these cell-specific TFs with unknown islet functions, and the potential species differences, may provide greater understanding of beta and alpha cell developmental programs.

Through  $\alpha$ TC1/ $\beta$ TC6 cell mixing experiments, we found that ATAC-seq profiles were able to capture the majority of (> 50%) cell signature open chromatin sites when  $\alpha$ TC1 or  $\beta$ TC6 comprised ~20% of the total mixture. Moreover, chromatin accessibility of these signature peaks in the mixtures was proportional to the corresponding cell's relative proportion in the mixture. Estimation of relative cell proportions appears to be robust over a range of sequencing depths in the mixture samples (Supplementary Fig. S12), as long as signature peaks are established in each purified population. Human islets are heterogeneous tissues that on average consist of ~55% beta, ~35% alpha, < 10% delta, and up to a few percent of PP/gamma cells<sup>13,95,96</sup>. Thus, our initial cell line mixture findings suggest that ATAC-seq profiles of whole human islets<sup>97</sup> should effectively detect open chromatin signatures for both beta and alpha cells, but are likely missing those of less abundant delta and PP/gamma cells. We also observed that by increasing the sequencing depth one can increase the proportion of cell-specific regulatory elements detected in ATAC-seq profiles from cell mixtures. Therefore, if the cell of interest constitutes a small proportion of the mixture that will be sampled, increasing the sequencing depth could enable detecting regulatory elements specific to this cell type.

Predicting cell type proportions in whole islets based on ATAC-seq signatures from primary human alpha and beta was moderately successful, considering the factors that may confound the data and analyses. First, ATAC-seq profiles of other islet endocrine cell types (e.g. delta, gamma/PP) have not been determined, limiting the ability to estimate their proportions in islets. Second, and perhaps related, the beta cell proportion was consistently overestimated (Supplementary Fig. S5). We hypothesize that this may be due to contributions of other cell types (e.g., delta cells) contaminating the FACS-enriched beta cells to the ATAC-seq profiles attributed as “pure beta”. In support of this, beta cell type composition was most overestimated in Islets P2 and P7, which have larger proportions of non-alpha/non-beta cell types. Third, actual cell proportions were determined based on counts from single cell transcriptomes from only a few hundred cells. These counts therefore may not accurately reflect the true cell proportions of each bulk islet sample consisting of millions of cells. Taken together, these results suggest that despite several caveats in data availability it is feasible to accurately deconvolute bulk islet cell compositions using purified cell ATAC-seq profiles. We expect that ATAC-seq profiles of additional purified islet endocrine and contaminating exocrine cell types from islet preparations will yield dramatically improved cell composition estimates from whole islet samples. These improvements, combined with sectioning and immunohistochemistry staining of multiple sections from distinct regions of the pancreas, may be necessary to obtain more accurate estimates of each islet cell type proportion.

Integrative analysis revealed that  $\beta$ TC6/ $\alpha$ TC1 transcriptomes largely resembled transcriptomes of corresponding mouse beta/alpha primary cells, respectively, but also highlighted important differences between cell lines and primary cells at multiple loci. Pathway analyses suggest that, unlike their primary cell counterparts, cell line transcriptomes were enriched for transcripts associated with cell proliferative processes. Closer inspection also revealed several enriched genes in human/mouse alpha cells that were otherwise undetected in  $\alpha$ TC1 transcriptomes. This suggests that although these cell line transcriptomes broadly resemble their primary cell counterparts, specific islet cell expression programs may not be adequately conserved in these cell lines.

GWAS have implicated SNPs in over 150 genomic loci that contribute to T2D risk and related quantitative measures of islet (dys)function<sup>1,98–100</sup>. Integrative analyses of islet epigenomic and GWAS SNPs have demonstrated that these SNPs are enriched in islet regulatory elements<sup>14,85,92</sup>. These studies have also suggested that genetic disruption of islet transcriptional regulatory control contribute to islet dysfunction and decreased insulin secretion, which are T2D hallmarks. Similar to observations in human alpha and beta cells<sup>14</sup>, we find that T2D SNP-containing (orthologous) sequences have regulatory potential in both  $\beta$ TC6 and  $\alpha$ TC1. Together, these datasets suggest that T2D-associated genetic variants may affect both alpha and beta cells and delineate the sites

for which CRISPR/Cas9 (epi)genomic manipulation in these cell lines may provide mechanistic insights into the molecular genetics of T2D that will translate between these cell models and human islets.

In summary, we have elucidated the transcriptional and epigenetic landscapes of  $\alpha$ TC1 and  $\beta$ TC6 mouse islet cell lines. Cell mixing experiments suggest that existing human islet epigenome profiles capture alpha and beta cell regulatory elements but may not represent those active in less abundant delta and PP/gamma cell types. Furthermore, these analyses suggest that estimation of bulk islet cell compositions using primary cell ATAC-seq profiles is feasible, particularly once the epigenomes of each islet cell type are defined. Overall, these analyses document important similarities between these cell lines and their primary islet counterparts, including evidence of common regulatory element use and putative TF binding motif enrichments, but also highlight significant differences at multiple loci. The data and analyses from this study should serve as a useful resource and tool for individual investigators to determine the utility of these cell lines to study their specific regulatory elements, genes, and pathways of interest and relevant to islet cell identity, function, and diabetes.

## Methods

**Cell Culture.**  $\alpha$ TC1 clone 6 ( $\alpha$ TC1) (ATCC<sup>®</sup> CRL-2934<sup>™</sup>) and beta-TC-6 ( $\beta$ TC6) (ATCC<sup>®</sup> CRL-11506<sup>™</sup>) were purchased from American Type Culture Collection (ATCC) Manassas, VA.  $\alpha$ TC1 cells were cultured in Dulbecco's Modified Eagle's Medium (Gibco 11885-076) supplemented with 10% heat-inactivated fetal bovine serum (Seradigm), 15 mM HEPES (Gibco), 0.1 mM non-essential amino acids (Gibco), 0.02% bovine serum albumin (Sigma), 2 g/L glucose at 37 C and 5% CO<sub>2</sub>.  $\beta$ TC6 cells were cultured in Dulbecco's Modified Eagle's Medium (Gibco 11965-084) supplemented with 15% heat-inactivated fetal bovine serum (Seradigm) and 10% sodium pyruvate (Gibco) at 37 C and 5% CO<sub>2</sub>.

**ATAC-seq.**  $\alpha$ TC1 and  $\beta$ TC6 cells were counted and mixed in 10% increments to create mixture samples consisting of 100,000 total cells (i.e. 90,000  $\beta$ TC6 and 10,000  $\alpha$ TC1). In addition, five replicates of 100,000 pure  $\alpha$ TC1 and  $\beta$ TC6 cells were processed as controls. ATAC-seq libraries for all samples were prepared as previously described<sup>35</sup> and sequenced on an Illumina NextSeq 500 with 2 × 75 bp cycles to a mean depth of ~100 million reads (Supplementary Table S1). Paired-end ATAC-seq reads were quality trimmed using *Trimmomatic* version 0.32<sup>101</sup> and parameters "TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36". Trimmed reads were aligned to mouse genome (mm9) using BWA version 0.7.9a<sup>102</sup>, specifically using the *bwa mem -M* option. Duplicate reads were removed using "MarkDuplicates" from *Picard-tools* version 1.95<sup>103</sup>. After preprocessing and quality filtering, peaks were called on alignments with *MACS* version 2.1.0<sup>104</sup> using the parameters "-g 'mm' -nomodel -f BAMPE -q 0.01". The peak sets from all samples were merged to generate one consensus peak set (N = 65,053) by using R package *DiffBind* 2.2.5<sup>105</sup>. Peaks only present in at least two samples were included in the analysis. Raw read counts were normalized using the effective library size (total number of reads in consensus peaks) and using the trimmed mean of M-values normalization method (TMM). Consensus peaks were annotated using *HOMER* version 4.6<sup>36</sup> and were classified into two groups i) distal peaks (peaks whose distance to a gene TSS is >2 kb) and ii) promoter peaks (peaks whose distance to TSS is <2 kb). Spearman rank-order correlation was calculated for all pure and mixture samples using consensus peaks with *deepTools* version 2.4.2<sup>106</sup>.

**RNA-seq.** Total RNA was extracted and purified from three  $\alpha$ TC1 and three  $\beta$ TC6 samples using Trizol (Life Technologies) according to the manufacturer's instructions, ERCC spike ins (Life Technologies) were added and library prepared using the Kapa Biosystems KAPA stranded mRNA-seq kit according to the manufacturer's instructions. All sequencing was performed on an Illumina NextSeq 500 with 2 × 100 bp cycles. RNA libraries were sequenced to an average depth of 60 million reads (Supplementary Table S1). Paired-end RNA-seq reads with Phred quality scores < 30 and adaptor sequences were removed using *Trim Galore!* version 0.40<sup>107</sup> and reads with < 50 bp after trimming were discarded. Trimmed reads were aligned to mouse genome (mm9) using *Bowtie* 2 version 2.23<sup>108</sup> with default parameters and expression levels of all genes were determined using *RSEM* version 1.2.12<sup>109</sup> with default parameters and reference transcript annotations (NCBI37/mm9, Ensembl v67). A total of 24,531 protein-coding genes and long intergenic non-coding RNAs (lincRNAs) were considered in the study.

**Differential Expression and Peak Analyses.** Differential gene expression analysis was performed using R package *edgeR* 3.16.5<sup>110</sup> to identify genes enriched in  $\alpha$ TC1 and  $\beta$ TC6 cells. Prior to analysis, gene expression counts were normalized using ERCC spike-in controls with R package *RUVSeq* 1.8.0<sup>111</sup>. A total of 12,234 genes with FPKM ≥ 1 in all three  $\beta$ TC6 or  $\alpha$ TC1 replicates were considered in the analysis. Genes with FDR < 5% and absolute log<sub>2</sub> fold change > 2 were considered differentially expressed. Differential analysis of consensus ATAC-seq peaks was also performed using *edgeR* 3.16.5 to identify those that are differentially accessible between  $\alpha$ TC1 and  $\beta$ TC6 cells. We considered peaks with FDR < 5% and absolute log<sub>2</sub> fold change of the normalized read counts > 2 as differentially accessible (DA).

**Pathway Analysis.** "findGo.pl" (*HOMER*) script was used to identify enrichment of mouse Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways in  $\alpha$ TC1 and  $\beta$ TC6-specific genes. *HOMER* enrichment p-values were adjusted using the Benjamini-Hochberg procedure and pathways with FDR < 1% were regarded as significant.

**Motif Enrichment Analysis.** "findMotifsGenome.pl" (*HOMER*) script with parameters "mm9 -size 200" was used to determine TF motifs enriched in  $\alpha$ TC1 and  $\beta$ TC6 DA peaks. In each analysis, non-cell-specific peaks were provided as background (e.g.  $\alpha$ TC1 DA peaks (target) vs. non-cell-specific peaks (background)). In order to identify motifs of expressed TFs enriched in each cell type, we solely considered motifs of expressed (FPKM > 1) TF genes.

**Determination of Chromatin Signatures and Estimation of Cellular Proportions.** We used the analytical tool, *CIBERSORT*<sup>54</sup>, available at <https://cibersort.stanford.edu>, to identify signature peaks and estimate the proportions of  $\alpha$ TC1 and  $\beta$ TC6 cells in the mixture samples. From consensus peaks of pure  $\alpha$ TC1 and  $\beta$ TC6 cells (3 replicates each), *CIBERSORT* selects signature peaks that are the most significantly differentially accessible between cell types using a two-sided unequal variance *t*-test. Then, a signature matrix is generated by calculating the median of the normalized read counts for signature peaks. With the signature matrix, *CIBERSORT* estimates the proportion of each cell type in the mixture samples using a linear model and assuming that the read counts of signature peaks in the mixture are the sums of signature peaks weighted by the unknown proportion of each cell type in the mixture. ATAC-seq cell mixtures were down-sampled to 25, 15, 5, and 1 million read depth intervals using *SAMtools*<sup>112</sup> version 1.5 and cell type proportions were estimated.

**Estimation of Bulk Islet Cell Type Proportions.** Raw sequence data for primary human beta and alpha ( $n = 3$  each) ATAC-seq profiles<sup>14</sup> were obtained and processed in the same manner as described in this study. *CIBERSORT* was used to derive alpha and beta signature peaks and quantify cell type proportions for seven islets (sample identifiers correspond to islet donor names in<sup>4</sup>; Supplementary Table S4).

**Comparison of Human and Mouse Primary Islet Cell Transcriptomes with those of  $\alpha$ TC1/ $\beta$ TC6.** Processed raw count gene expression data were obtained for primary mouse<sup>77</sup> and human<sup>13,14</sup> islet cells. Differential gene expression analyses were performed using R package *edgeR*\_3.16.5 to identify genes enriched in primary alpha and beta cells of each dataset at FDR < 5%. Prior to identifying overlapping genes between each dataset, human and mouse orthologues were identified using the Mouse Genome Informatics database (<http://www.informatics.jax.org>). Venn diagrams were constructed using the R package *Vennable*\_3.1.0.9000<sup>113</sup> and genes that existed in all three datasets (primary mouse islet, primary human islet, and mouse islet cell line). We tested the significance of overlap of genes enriched in  $\alpha$ TC1/ $\beta$ TC6 cells and mouse/human alpha/beta cells using a Fisher's exact test. Prior to hierarchical clustering, batch effects between datasets were removed using ComBat in the Bioconductor package *sva*\_3.24.4.

**Genome Lift-over, SNP LD-Pruning, and GWAS SNP Enrichment Analysis.**  $\alpha$ TC1 and  $\beta$ TC6 cell-specific promoter and distal peaks were converted from mm9 genome coordinates to hg19 genome coordinates using *bnMapper*<sup>87</sup>, a one-to-one nucleotide cross species mapper to identify evolutionary conserved sequences between two genomes. At times, the lift-over from mouse genomic coordinates would result in several non-contiguous human genomic coordinates (Supplementary Table S12). To merge non-contiguous coordinates and ensure that we were only using uniquely mapping and contiguous human lift-over coordinates, we overlapped the resulting lift-over coordinates with bulk islet open chromatin sites ( $n = 69,261$ ) (ATAC-seq) generated from 14 non-diabetic islets (Khetan *et al.* in preparation) (Supplementary Table S13). Only the coordinates that overlapped a unique human islet open chromatin site were used for downstream analysis. Lists of reference SNP identifiers were obtained from the NHGRI-EBI Catalog of SNPs (<https://www.ebi.ac.uk/gwas/>; accessed on January 19th, 2017) for 636 disease categories. For each disease category, GWAS SNPs were pruned using PLINK version 1.9<sup>114</sup> and parameters “-maf 0.05-clump-clump-p1 0.0001-clump-p2 0.01-clump-r2 0.2-clump-kb 1000” to ensure that each variant haplotype was tested only once during the enrichment analysis. For each SNP pair in linkage disequilibrium (LD) ( $R^2 > 0.2$ ) the SNP with the least significant p-value was discarded. Enrichment of LD-pruned GWAS SNPs within the unique overlapping islet ATAC-seq peaks was performed using *GREGOR* version 1.4.0<sup>88</sup>. PhastCons sequence conservation scores were obtained from <ftp://hgdownload.cse.ucsc.edu/goldenPath/mm9/phastCons30way/vertebrate> and average scores were calculated in all mouse cell line consensus peaks.

**Data Availability.** Raw sequence data generated and analyzed during this study are available in NCBI Sequence Read Archive repository (SRA; <http://www.ncbi.nlm.nih.gov/sra>) under accession number SRP108440 and BioProject repository (<https://www.ncbi.nlm.nih.gov/bioproject/>) under accession number PRJNA388786. Processed data generated and analyzed during this study are available in the Gene Expression Omnibus repository (GEO; <http://www.ncbi.nlm.nih.gov/geo/>) under accession number GSE99954.

## References

- Mohlke, K. L. & Boehnke, M. Recent advances in understanding the genetic architecture of type 2 diabetes. *Hum. Mol. Genet.* **24**, R85–R92 (2015).
- Ashcroft, F. M. & Rorsman, P. Diabetes mellitus and the  $\beta$  cell: the last ten years. *Cell* **148**, 1160–1171 (2012).
- Li, J. *et al.* Single-cell transcriptomes reveal characteristic features of human pancreatic islet cell types. *EMBO Rep.* **17**, 178–187 (2016).
- 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. gr.* **212720**, 116, <https://doi.org/10.1101/gr.212720.116> (2016).
- Segerstolpe, Å. *et al.* Single-Cell Transcriptome Profiling of Human Pancreatic Islets in Health and Type 2 Diabetes. *Cell Metab.* **24**, 593–607 (2016).
- Xin, Y. *et al.* RNA Sequencing of Single Human Islet Cells Reveals Type 2 Diabetes Genes. *Cell Metab.* **24**, 608–615 (2016).
- Muraro, M. J. *et al.* A Single-Cell Transcriptome Atlas of the Human Pancreas. *Cell Syst.* <https://doi.org/10.1016/j.cels.2016.09.002> (2016).
- Baron, M. *et al.* A Single-Cell Transcriptomic Map of the Human and Mouse Pancreas Reveals Inter- and Intra-cell Population Structure. *Cell Syst.* **3**, 346–360.e4 (2016).
- Wang, Y. J. *et al.* Single-Cell Transcriptomics of the Human Endocrine Pancreas. *Diabetes* **65**, 3028–3038 (2016).
- Nica, A. C. *et al.* Cell-type, allelic, and genetic signatures in the human pancreatic beta cell transcriptome. *Genome Res.* **23**, 1554–1562 (2013).

11. Bramswig, N. C. *et al.* Epigenomic plasticity enables human pancreatic  $\alpha$  to  $\beta$  cell reprogramming. *J. Clin. Invest.* **123**, 1275–1284 (2013).
12. Dorrell, C. *et al.* Transcriptomes of the major human pancreatic cell types. *Diabetologia* **54**, 2832–2844 (2011).
13. Blodgett, D. M. *et al.* Novel Observations From Next-Generation RNA Sequencing of Highly Purified Human Adult and Fetal Islet Cell Subsets. *Diabetes* **64**, 3172–3181 (2015).
14. Ackermann, A. M., Wang, Z., Schug, J., Naji, A. & Kaestner, K. H. Integration of ATAC-seq and RNA-seq identifies human alpha cell and beta cell signature genes. *Mol. Metab.* **5**, 233–244 (2016).
15. Lawlor, N., Khetan, S., Ucar, D. & Stitzel, M. L. Genomics of Islet (Dys)function and Type 2 Diabetes. *Trends Genet. TIG* **33**, 244–255 (2017).
16. Ishihara, H. *et al.* Pancreatic beta cell line MIN6 exhibits characteristics of glucose metabolism and glucose-stimulated insulin secretion similar to those of normal islets. *Diabetologia* **36**, 1139–1145 (1993).
17. Choi, S.-E. *et al.* IL-6 protects pancreatic islet beta cells from pro-inflammatory cytokines-induced cell death and functional impairment *in vitro* and *in vivo*. *Transpl. Immunol.* **13**, 43–53 (2004).
18. Liu, Y. *et al.* Characterization of Zinc Influx Transporters (ZIPs) in Pancreatic  $\beta$  Cells: ROLES IN REGULATING CYTOSOLIC ZINC HOMEOSTASIS AND INSULIN SECRETION. *J. Biol. Chem.* **290**, 18757–18769 (2015).
19. Asfari, M. *et al.* Establishment of 2-mercaptoethanol-dependent differentiated insulin-secreting cell lines. *Endocrinology* **130**, 167–178 (1992).
20. Hartley, T. *et al.* Endoplasmic reticulum stress response in an INS-1 pancreatic  $\beta$ -cell line with inducible expression of a folding-deficient proinsulin. *BMC Cell Biol.* **11**, 59 (2010).
21. Lorenz, M. A., Azzouny, M. A. E., Kennedy, R. T. & Burant, C. F. Metabolome Response to Glucose in the  $\beta$ -Cell Line INS-1 832/13. *J. Biol. Chem.* **288**, 10923–10935 (2013).
22. Kim, M. *et al.* Comparison of pancreatic beta cells and alpha cells under hyperglycemia: Inverse coupling in pAkt-FoxO1. *Diabetes Res. Clin. Pract.* <https://doi.org/10.1016/j.diabres.2017.05.017>
23. Perelis, M. *et al.* Pancreatic  $\beta$  cell enhancers regulate rhythmic transcription of genes controlling insulin secretion. *Science* **350**, aac4250 (2015).
24. Wakae-Takada, N., Xuan, S., Watanabe, K., Meda, P. & Leibel, R. L. Molecular basis for the regulation of islet beta cell mass in mice: the role of E-cadherin. *Diabetologia* **56**, 856–866 (2013).
25. Poutout, V. *et al.* Morphological and functional characterization of beta TC-6 cells—an insulin-secreting cell line derived from transgenic mice. *Diabetes* **44**, 306–313 (1995).
26. Efrat, S. *et al.* Beta-cell lines derived from transgenic mice expressing a hybrid insulin gene-oncogene. *Proc. Natl. Acad. Sci. USA* **85**, 9037–9041 (1988).
27. Diao, J., Asghar, Z., Chan, C. B. & Wheeler, M. B. Glucose-regulated Glucagon Secretion Requires Insulin Receptor Expression in Pancreatic  $\alpha$ -Cells. *J. Biol. Chem.* **280**, 33487–33496 (2005).
28. Vasu, S., Moffett, R. C., McClenaghan, N. H. & Flatt, P. R. Differential molecular and cellular responses of GLP-1 secreting L-cells and pancreatic alpha cells to glucotoxicity and lipotoxicity. *Exp. Cell Res.* **336**, 100–108 (2015).
29. Piro, S. *et al.* Chronic Exposure to GLP-1 Increases GLP-1 Synthesis and Release in a Pancreatic Alpha Cell Line ( $\alpha$ -TC1): Evidence of a Direct Effect of GLP-1 on Pancreatic Alpha Cells. *PLOS ONE* **9**, e90093 (2014).
30. Powers, A. C. *et al.* Proglucagon processing similar to normal islets in pancreatic alpha-like cell line derived from transgenic mouse tumor. *Diabetes* **39**, 406–414 (1990).
31. Hamaguchi, K. & Leiter, E. H. Comparison of cytokine effects on mouse pancreatic alpha-cell and beta-cell lines. Viability, secretory function, and MHC antigen expression. *Diabetes* **39**, 415–425 (1990).
32. Ravassard, P. *et al.* A genetically engineered human pancreatic  $\beta$  cell line exhibiting glucose-inducible insulin secretion. *J. Clin. Invest.* **121**, 3589–3597 (2011).
33. Benazra, M. *et al.* A human beta cell line with drug inducible excision of immortalizing transgenes. *Mol. Metab.* **4**, 916–925 (2015).
34. Scharfmann, R. *et al.* Development of a conditionally immortalized human pancreatic  $\beta$  cell line. *J. Clin. Invest.* **124**, 2087–2098 (2014).
35. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218 (2013).
36. 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).
37. Schaffer, A. E. *et al.* Nkx6.1 Controls a Gene Regulatory Network Required for Establishing and Maintaining Pancreatic Beta Cell Identity. *PLOS Genet.* **9**, e1003274 (2013).
38. Taylor, B. L., Liu, F.-F. & Sander, M. Nkx6.1 Is Essential for Maintaining the Functional State of Pancreatic Beta Cells. *Cell Rep.* **4**, 1262–1275 (2013).
39. Tessem, J. S. *et al.* Nkx6.1 regulates islet  $\beta$ -cell proliferation via Nr4a1 and Nr4a3 nuclear receptors. *Proc. Natl. Acad. Sci.* **111**, 5242–5247 (2014).
40. Guo, T. *et al.* ISL1 Promotes Pancreatic Islet Cell Proliferation. *PLoS ONE* **6** (2011).
41. Ediger, B. N. *et al.* Islet-1 Is essential for pancreatic  $\beta$ -cell function. *Diabetes* **63**, 4206–4217 (2014).
42. Gao, T. *et al.* Pdx1 Maintains  $\beta$  Cell Identity and Function by Repressing an  $\alpha$  Cell Program. *Cell Metab.* **19**, 259–271 (2014).
43. Fujimoto, K. & Polonsky, K. S. Pdx1 and other factors that regulate pancreatic beta-cell survival. *Diabetes Obes. Metab.* **11**(Suppl 4), 30–37 (2009).
44. Piccand, J. *et al.* Rfx6 Maintains the Functional Identity of Adult Pancreatic  $\beta$  Cells. *Cell Rep.* **9**, 2219–2232 (2014).
45. Subramanian, L. *et al.* Transcription factor Lhx2 is necessary and sufficient to suppress astrogliogenesis and promote neurogenesis in the developing hippocampus. *Proc. Natl. Acad. Sci. USA* **108**, E265–274 (2011).
46. Kitajima, K., Kawaguchi, M., Iacovino, M., Kyba, M. & Hara, T. Molecular functions of the LIM-homeobox transcription factor Lhx2 in hematopoietic progenitor cells derived from mouse embryonic stem cells. *Stem Cells Dayt. Ohio* **31**, 2680–2689 (2013).
47. Spaeth, J. M. *et al.* The FOXP1, FOXP2 and FOXP4 transcription factors are required for islet alpha cell proliferation and function in mice. *Diabetologia* **58**, 1836–1844 (2015).
48. Xu, Z., Huang, S., Chang, L.-S., Agulnick, A. D. & Brandt, S. J. Identification of a TAL1 Target Gene Reveals a Positive Role for the LIM Domain-Binding Protein Ldb1 in Erythroid Gene Expression and Differentiation. *Mol. Cell Biol.* **23**, 7585–7599 (2003).
49. Hunter, C. S. *et al.* Islet  $\alpha$ -,  $\beta$ -, and  $\delta$ -cell development is controlled by the Ldb1 coregulator, acting primarily with the islet-1 transcription factor. *Diabetes* **62**, 875–886 (2013).
50. Galloway, J. R. *et al.* SSBP3 Interacts With Islet-1 and Ldb1 to Impact Pancreatic  $\beta$ -Cell Target Genes. *Mol. Endocrinol.* **29**, 1774–1786 (2015).
51. Uittenbogaard, M. & Chiaramello, A. Expression of the bHLH transcription factor Tcf12 (ME1) gene is linked to the expansion of precursor cell populations during neurogenesis. *Brain Res. Gene Expr. Patterns* **1**, 115–121 (2002).
52. Badinga, L., Song, S., Simmen, R. C. & Simmen, F. A. A distal regulatory region of the insulin-like growth factor binding protein-2 (IGFBP-2) gene interacts with the basic helix-loop-helix transcription factor, AP-4. *Endocrine* **8**, 281–289 (1998).
53. Yoneyama, T. *et al.* Identification of IGFBP2 and IGFBP3 As Compensatory Biomarkers for CA19-9 in Early-Stage Pancreatic Cancer Using a Combination of Antibody-Based and LC-MS/MS-Based Proteomics. *PLOS ONE* **11**, e0161009 (2016).
54. Newman, A. M. *et al.* Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods* **12**, 453–457 (2015).

55. Maaten, L. van der. Accelerating t-SNE using Tree-Based Algorithms. *J. Mach. Learn. Res.* **15**, 3221–3245 (2014).
56. Corces, M. R. *et al.* Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution. *Nat. Genet.* **48**, 1193–1203 (2016).
57. Petri, A. *et al.* The effect of neurogenin3 deficiency on pancreatic gene expression in embryonic mice. *J. Mol. Endocrinol.* **37**, 301–316 (2006).
58. Artner, I. *et al.* MafB: an activator of the glucagon gene expressed in developing islet alpha- and beta-cells. *Diabetes* **55**, 297–304 (2006).
59. Brink, C., Chowdhury, K. & Gruss, P. Pax4 regulatory elements mediate beta cell specific expression in the pancreas. *Mech. Dev.* **100**, 37–43 (2001).
60. Lorenzo, P. I. *et al.* PAX4 Defines an Expandable  $\beta$ -Cell Subpopulation in the Adult Pancreatic Islet. *Sci. Rep.* **5**, srep15672 (2015).
61. Hang, Y. *et al.* The MafA Transcription Factor Becomes Essential to Islet  $\beta$ -Cells Soon After Birth. *Diabetes* **63**, 1994–2005 (2014).
62. Nishimura, W., Takahashi, S. & Yasuda, K. MafA is critical for maintenance of the mature beta cell phenotype in mice. *Diabetologia* **58**, 566–574 (2015).
63. Song, Z. *et al.* EGFR signaling regulates beta cell proliferation in adult mice. *J. Biol. Chem.* jbc.M116.747840 <https://doi.org/10.1074/jbc.M116.747840> (2016).
64. Roper, J., O'Carroll, A.-M., Young, W. & Lolait, S. The vasopressin Avpr1b receptor: Molecular and pharmacological studies. *Stress Amst. Neth.* **14**, 98–115 (2011).
65. Oshikawa, S., Tanoue, A., Koshimizu, T., Kitagawa, Y. & Tsujimoto, G. Vasopressin Stimulates Insulin Release from Islet Cells through V1b Receptors: a Combined Pharmacological/Knockout Approach. *Mol. Pharmacol.* **65**, 623–629 (2004).
66. Courtney, M. *et al.* The Inactivation of Arx in Pancreatic  $\alpha$ -Cells Triggers Their Neogenesis and Conversion into Functional  $\beta$ -Like Cells. *PLoS Genet.* **9**, e1003934 (2013).
67. Collombat, P. *et al.* Embryonic endocrine pancreas and mature beta cells acquire alpha and PP cell phenotypes upon Arx misexpression. *J. Clin. Invest.* **117**, 961–970 (2007).
68. Butler, A. E. *et al.* Beta-cell deficit and increased beta-cell apoptosis in humans with type 2 diabetes. *Diabetes* **52**, 102–110 (2003).
69. Talchai, C., Xuan, S., Lin, H. V., Sussel, L. & Accili, D. Pancreatic  $\beta$ -Cell Dedifferentiation As Mechanism Of Diabetic  $\beta$ -Cell Failure. *Cell* **150**, 1223–1234 (2012).
70. Cinti, F. *et al.* Evidence of  $\beta$ -Cell Dedifferentiation in Human Type 2 Diabetes. *J. Clin. Endocrinol. Metab.* **101**, 1044–1054 (2016).
71. Wang, Z., York, N. W., Nichols, C. G. & Remedi, M. S. Pancreatic  $\beta$ -cell Dedifferentiation in Diabetes and Re-differentiation following Insulin Therapy. *Cell Metab.* **19**, 872–882 (2014).
72. Gale, J. E. *et al.* Disruption of circadian rhythms accelerates development of diabetes through pancreatic beta-cell loss and dysfunction. *J. Biol. Rhythms* **26**, 423–433 (2011).
73. Lee, J. *et al.* Circadian control of  $\beta$ -cell function and stress responses. *Diabetes Obes. Metab.* **17**, 123–133 (2015).
74. Watanabe, S. *et al.* Oxytocin Protects against Stress-Induced Cell Death in Murine Pancreatic  $\beta$ -Cells. *Sci. Rep.* **6**, srep25185 (2016).
75. Benayoun, B. A. *et al.* H3K4me3 Breadth Is Linked to Cell Identity and Transcriptional Consistency. *Cell* **158**, 673–688 (2014).
76. van der Meulen, T. *et al.* Virgin Beta Cells Persist throughout Life at a Neogenic Niche within Pancreatic Islets. *Cell Metab.* **25**, 911–926.e6 (2017).
77. DiGrucio, M. R. *et al.* Comprehensive alpha, beta and delta cell transcriptomes reveal that ghrelin selectively activates delta cells and promotes somatostatin release from pancreatic islets. *Mol. Metab.* **5**, 449–458 (2016).
78. Arda, H. E. *et al.* Age-Dependent Pancreatic Gene Regulation Reveals Mechanisms Governing Human  $\beta$  Cell Function. *Cell Metab.* **23**, 909–920 (2016).
79. Dai, C. *et al.* Islet-enriched gene expression and glucose-induced insulin secretion in human and mouse islets. *Diabetologia* **55**, 707–718 (2012).
80. Shibasaki, T. *et al.* Essential role of Epac2/Rap1 signaling in regulation of insulin granule dynamics by cAMP. *Proc. Natl. Acad. Sci.* **104**, 19333–19338 (2007).
81. Dzhura, I. *et al.* Phospholipase C- $\epsilon$  links Epac2 activation to the potentiation of glucose-stimulated insulin secretion from mouse islets of Langerhans. *Islets* **3**, 121–128 (2011).
82. Benner, C. *et al.* The transcriptional landscape of mouse beta cells compared to human beta cells reveals notable species differences in long non-coding RNA and protein-coding gene expression. *BMC Genomics* **15**, 620 (2014).
83. Thurman, R. E. *et al.* The accessible chromatin landscape of the human genome. *Nature* **489**, 75–82 (2012).
84. Roadmap Epigenomics, C. *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).
85. Parker, S. C. J. *et al.* Chromatin stretch enhancer states drive cell-specific gene regulation and harbor human disease risk variants. *Proc. Natl. Acad. Sci. USA* **110**, 17921–17926 (2013).
86. Hnisz, D. *et al.* Super-Enhancers in the Control of Cell Identity and Disease. *Cell* **155**, 934–947 (2013).
87. Denas, O. *et al.* Genome-wide comparative analysis reveals human-mouse regulatory landscape and evolution. *BMC Genomics* **16**, 87 (2015).
88. Schmidt, E. M. *et al.* GREGOR: evaluating global enrichment of trait-associated variants in epigenomic features using a systematic, data-driven approach. *Bioinformatics* **31**, 2601–2606 (2015).
89. Dooley, J. *et al.* Genetic predisposition for beta cell fragility underlies type 1 and type 2 diabetes. *Nat. Genet.* **48**, 519–527 (2016).
90. van de Bunt, M. *et al.* Transcript Expression Data from Human Islets Links Regulatory Signals from Genome-Wide Association Studies for Type 2 Diabetes and Glycemic Traits to Their Downstream Effectors. *PLoS Genet.* **11** (2015).
91. Khoo, C. *et al.* Research Resource: The Pdx1 Cistrome of Pancreatic Islets. *Mol. Endocrinol.* **26**, 521–533 (2012).
92. Pasquali, L. *et al.* Pancreatic islet enhancer clusters enriched in type 2 diabetes risk-associated variants. *Nat. Genet.* **46**, 136–143 (2014).
93. Liu, J. *et al.* Islet-1 Regulates Arx Transcription during Pancreatic Islet  $\alpha$ -Cell Development. *J. Biol. Chem.* **286**, 15352–15360 (2011).
94. Ait-Lounis, A. *et al.* The Transcription Factor Rfx3 Regulates  $\beta$ -Cell Differentiation, Function, and Glucokinase Expression. *Diabetes* **59**, 1674–1685 (2010).
95. Cabrera, O. *et al.* The unique cytoarchitecture of human pancreatic islets has implications for islet cell function. *Proc. Natl. Acad. Sci. USA* **103**, 2334–2339 (2006).
96. Brissova, M. *et al.* Assessment of human pancreatic islet architecture and composition by laser scanning confocal microscopy. *J. Histochem. Cytochem. Off. J. Histochem. Soc.* **53**, 1087–1097 (2005).
97. Varshney, A. *et al.* Genetic regulatory signatures underlying islet gene expression and type 2 diabetes. *Proc. Natl. Acad. Sci.* **114**, 2301–2306 (2017).
98. Gaulton, K. J. *et al.* Genetic fine mapping and genomic annotation defines causal mechanisms at type 2 diabetes susceptibility loci. *Nat. Genet.* **47**, 1415–1425 (2015).
99. Fuchsberger, C. *et al.* The genetic architecture of type 2 diabetes. *Nature* advance online publication (2016).
100. DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium *et al.* Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. *Nat. Genet.* **46**, 234–244 (2014).
101. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
102. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *ArXiv13033997 Q-Bio* (2013).

103. The Broad Institute. Picard-Tools. Available at: <http://broadinstitute.github.io/picard/> (2013).
104. Zhang, Y. *et al.* Model-based Analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
105. Ross-Innes, C. S. *et al.* Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature* **481**, 389–393 (2012).
106. Ramírez, F., Dündar, F., Diehl, S., Grüning, B. A. & Manke, T. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res.* **42**, W187–W191 (2014).
107. Krueger, F. Taking appropriate QC measures for RRBS-type or other -Seq applications with Trim Galore! Available at: [https://www.bioinformatics.babraham.ac.uk/projects/trim\\_galore/](https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) (2015).
108. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
109. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011).
110. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinforma. Oxf. Engl.* **26**, 139–140 (2010).
111. Risso, D., Ngai, J., Speed, T. P. & Dudoit, S. Normalization of RNA-seq data using factor analysis of control genes or samples. *Nat Biotechnol* **32** (2014).
112. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinforma. Oxf. Engl.* **25**, 2078–2079 (2009).
113. Swinton, J. Vennable: Venn and Euler area-proportional diagrams. R package version 3.1.0.9000. Available at: <https://github.com/js229/Vennable>. (2016).
114. Purcell, S. *et al.* PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).

## Acknowledgements

We thank Jane Cha for aid in graphic design. We thank all members of the Stitzel and Ucar labs for helpful discussion and feedback on this study and manuscript, and to the anonymous reviewers whose comments and suggestions improved the quality and clarity of this manuscript. This study was made possible by generous financial support of the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) under award number DK092251 (to MLS) and by The Jackson Laboratory startup funds (to DU and MLS). Opinions, interpretations, conclusions, and recommendations are solely the responsibility of the authors and do not necessarily represent the official views of the National Institutes of Health (NIH).

## Author Contributions

M.L.S. and D.U. conceived the experimental design and data analysis strategies. R.K. performed all cell culture, library preparation, and sequencing experiments. N.L. and A.Y. analyzed all corresponding data and wrote the manuscript. All authors reviewed the manuscript prior to submission.

## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-017-12335-1>.

**Competing Interests:** The authors declare that they have no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017