# Defining the chromatin and transcriptional landscape of stem cell-derived islets

Punn Augsornworawat[1,2], Nathaniel J. Hogrebe[1], Matthew Ishahak[1], Erica Marquez[1], Marlie M. Maestas[1], Mason D. Schmidt[1], Daniel A. Veronese-Paniagua[1], Sarah E. Gale[1], Julia R. Miller[1,2], Leonardo Velazco-Cruz[1], and Jeffrey R. Millman[1,2]*

[1]Division of Endocrinology, Metabolism and Lipid Research

Washington University School of Medicine

MSC 8127-057-08

660 South Euclid Avenue

St. Louis, MO 63110

USA

[2]Department of Biomedical Engineering

Washington University in St. Louis

1 Brookings Drive

St. Louis, MO 63130

USA

*To whom correspondence should be addressed:

Jeffrey R. Millman, jmillman@wustl.edu

**Abstract**

Transplantation of insulin-secreting β-cells differentiated from human pluripotent stem cells holds great potential as a cell therapy for treating insulin-dependent diabetes. While these stem cell-derived islets (SC-islets) are able to reverse diabetes in animal models, they are not fully equivalent to their *in vivo* counterparts. To better define the state of the cell types generated within these SC-islets and provide a resource for identifying deficiencies in lineage specification, we used single-cell multiomic sequencing to simultaneously measure the chromatin accessibility and transcriptional profiles of SC-islets at multiple time points as well as primary human islets. The integrated analysis of both the transcriptional and chromatin landscape for each cell provided greater resolution for defining cell identity, allowing us to derive novel gene lists for identifying each islet cell type. Furthermore, this multiomic analysis revealed that the difference between SC-β cells and enterochromaffin-like cells, which are a major off-target from *in vitro* differentiation, is a gradient of progressive cell states rather than a stark difference in identity. The chromatin landscape of primary human islets was much more restricted, suggesting that stem cell-derived cells are not fully locked into their cell fate. While long term culture of SC-islets both *in vitro* and *in vivo* does close overall chromatin state, only *in vivo* transplantation directs cells toward their correct identities. Collectively, our multiomic analysis demonstrates that both the chromatin and transcriptional landscapes play significant roles in islet cell identity, and these data can be used as a resource to identify specific deficiencies in the chromatin and transcriptional state of SC-islet cell types.

## Introduction

The development of methods to differentiate human pluripotent stem cells (hPSCs) into islet-like clusters has the potential to generate an unlimited number of insulin-producing stem cell-derived β (SC-β) cells for the treatment of insulin-dependent diabetes [1-3]. This process utilizes temporal combinations of small molecules and growth factors [4-6], microenvironmental cues [7], and other sorting or aggregation [4, 8-11] approaches to drive cells through several intermediate progenitor cell types. The resulting SC-β cells possess many features of primary human β-cells, including the expression of β-cell specific markers, glucose responsive insulin secretion, and the ability to reverse severe diabetes in animal models [5, 6, 12-15]. As a result, these cells have garnered great interest and potential to provide a functional cure for human patients with type 1 diabetes. However, these stem cell-derived islets (SC-islets) are a heterogeneous tissue that also contains stem cell-derived α (SC-α) and δ (SC-δ) cells, as well as an endocrine cell type of intestinal lineage denoted here as stem cell-derived enterochromaffin-like (SC-EC) cells [16]. The presence of this off-target population suggests that there are inefficiencies in lineage specification during these directed differentiation protocols [9]. Correcting this aberrant signaling could enhance specification to a β-cell identity and further improve the function of SC-β cells.

The specific pattern of gene expression and chromatin state within a cell directs differentiation and maintains its final identity [17-20]. Thus, characterizing both the transcriptional and chromatin landscape of cells during this differentiation process can provide insight into the degree that a particular cell type resembles its *in vivo* analogue. Single-cell RNA sequencing has been used to characterize the transcriptional environment of SC-islets, demonstrating that SC-β cells express many, but not all, of the important genes found in human primary β-cells [9, 13, 14, 21, 22]. Studies have also begun to investigate the chromatin state within pancreatic cell

differentiations using bulk approaches [23-25], and recent work has demonstrated the advantages of studying primary human islets using single-cell Assay for Transposase-Accessible Chromatin (ATAC) sequencing [26-28]. ATAC sequencing can describe whether particular chromatin regions are in an open, accessible state that is ready to be transcribed or interacted with, providing valuable information about cell identity that is missing when only transcriptional data is investigated. This data can define, for instance, the chromatin accessibility of promoters for the specific genes themselves that are capable of being transcribed or the DNA sequence motifs where specific transcription factors bind to help activate the transcription of different genes. Importantly, modulating the chromatin states of particular genes and motifs that are mis-expressed in *in vitro*-derived cell types are likely to drive cells closer to the identity of their *in vivo* counterparts. Despite the potential power of using a single-cell multiomic approach, a comprehensive indexing of the combined chromatin accessibility and transcriptional signatures of each *in vitro*-differentiated SC-islet cell type is lacking.

Here, we provide a multiomic analysis of the transcriptional and chromatin landscapes of SC-islets at the single-cell level. The simultaneous characterization of both the transcriptional and chromatin states for each cell allowed us to better define each cell type found within SC-islets, resulting in novel lists of important genes and motifs for each cell type. Interestingly, SC-EC and SC-β cells formed a gradient of cell identities rather than distinct cell types, with sub-populations of each cell type exhibiting characteristics of the other. Primary human islets had a much more restricted and defined chromatin state, while SC-islet cell types had open chromatin regions that were associated with other lineages. Transplantation into mice for 6 months closed many of these accessible chromatin regions and improved the expression of signature genes associated with each lineage, while extended *in vitro* culture did not have the same effect on

identity. Furthermore, we identified and modulated chromatin regulators that were important for SC-β cell identity, highlighting the significance of the chromatin landscape. Collectively, these data provide a resource that improves the characterization of cell identities found within SC-islets, facilitating the discovery of mis-expressed genes and motifs that can be targeted to further improve SC-β cell differentiation strategies.

**Multiomic characterization of SC-islets enables greater resolution of cell identity**

To better understand the cell types produced with directed differentiation strategies toward pancreatic islets, we sought to use an integrated analysis of both the transcriptional and chromatin landscapes for each cell to define cell identity and ascertain deficiencies in lineage specification more accurately. We differentiated human pluripotent stem cells (hPSCs) through our 6-stage protocol using specific temporal combinations of small molecules and growth factors to produce islet-like clusters [7, 29] (Extended Data Fig. 1a and b, Supplementary Table 1). These SC-islets were processed and sequenced using single-cell multiomics, obtaining both gene expression (mRNA) and chromatin accessibility (ATAC) information for each cell. (Fig. 1a, Supplementary Table 2). Analysis of gene expression and chromatin accessibility both individually and in combination enabled us to identify specific islet cell types, including SC-β, SC-α, and SC-δ cells as well as the off-target populations SC-EC cells, exocrine, and mesenchymal-like cells (Fig. 1b and c, Extended Data Fig. 1c and d). While both gene expression and chromatin accessibility information on their own were sufficient to separate most cell types in SC-islets, we were able to identify two subpopulations of SC-EC cells (denoted as SC-EC1 and SC-EC2) using the integrated analysis of both mRNA and ATAC data (Fig. 1b). These EC populations clustered adjacent to SC-β cells, suggesting their resemblance. In contrast,

SC-α and SC-δ populations formed more distinct, separated clusters (Fig. 1b).

Each of the identified populations had high expression of genes traditionally associated with their cell type (Fig. 1c). Furthermore, the SC-β, SC-α, and SC-δ populations had strong chromatin peaks represented by their respective *INS*, *GCG*, and *SST* gene regions as expected (Fig. 1d). Unexpectedly, however, the *INS* gene region had open accessibility across all detected cell types, including non-endocrine cell types to some extent. In contrast, *GCG* and *SST* gene regions had relatively few chromatin peaks outside of SC-α cells and SC-δ cells, respectively. We also found distinct motif groups enriched in each specific cell type (Fig. 1e, Extended Data Fig. 1e, Supplementary Table 4). Within the endocrine population, SC-β cells had enrichment of motifs that correspond to known β-cell associated transcription factors [9, 22, 30, 31]. Interestingly, this list includes the DNA-binding motif for MAFA, which is a transcription factor associated with primary β-cells but is expressed at a very low level in SC-β cells [16, 32]. Motifs that were specifically enriched in SC-EC cells include GATA6, CDX2, CTCF, and LMX1A, which are often associated with intestinal lineages [33, 34]. Interestingly, SC-β cells and SC-EC cells possessed shared enrichment of certain motifs, such as PDX1, PAX4, and NKX6-1, which have been reported to be important in human primary β-cells [27] (Fig. 1e).

In an effort to better characterize these cell populations within SC-islets, we cross-referenced transcription factor DNA-binding motif chromatin accessibility and gene expression information to identify active transcription factors that have both high expression and that can access their target binding motifs to promote transcription of other genes (Supplementary Table 4). Here, we highlight the top 10 identified active transcription factors enriched in specific endocrine populations compared to the average of the other endocrine populations (Fig. 1f). Highly active transcription factors in SC-β cells include of RFX1, PDX1 and PAX6, which have

been previously reported to be important for β-cell identity [27, 35]. Other highly active transcription factors, such as MAFG, EBF1, ISX, and PLAGL2, have not been highlighted in previous studies, demonstrating the utility of using both mRNA and chromatin accessibility data for identifying cell types. In contrast, SC-EC cells have co-enrichment of LMX1B, LMX1A, MNX1, and GATA6. Notably, the SC-β and SC-EC populations both have high activities of NKX6-1 and PAX4, suggesting that they share common features essential for their identities. This similarity is further demonstrated with motif chromatin accessibility plots, where NKX6-1 and PDX1 were enriched in both SC-β and SC-EC populations (Fig. 1e and g). In contrast, binding motifs for PAX6 were enriched in SC-β cells compared to SC-EC cells while LMX1A was enriched in SC-EC cells compared to SC-β cells. Unexpectedly, we found a mismatch in RNA expression and motif accessibility across cell types for certain transcription factors (Fig. 1f and g). In one instance, a high amount of *LMX1B* transcript was detected in all endocrine cell types, but motif accessibility was predominately open only in SC-EC cells, indicating that this transcription factor is likely helping activate transcription of its target genes only in SC-EC cells. In contrast, *ISX* transcripts were detected predominately in SC-β and SC-δ cells, but the associated DNA-binding motif was enriched in SC-β and SC-EC cells. Thus, while the motif is open in SC-EC cells, only SC-β cells are actively transcribing the transcription factor needed to utilize this open motif. Collectively, this multiomic analysis has generated novel lists of genes to better classify the cell types found within SC-islets, providing better resolution of cell identity and insights into the relative importance of different transcription factor activity within each cell type.

**SC-EC and SC-β cells form a gradient of progressive cell states rather than well-defined cell populations**

SC-EC cells are an off-target cell population that arises during *in vitro* SC-islet differentiation protocols but that does not positively contribute to tissue function [9, 36, 37] (Fig. 1b). Although both enterochromaffin and β-cells arise from a definitive endoderm precursor during *in vivo* development, intestinal and pancreatic origins respectively, they appear to share a common progenitor lineage during *in vitro* differentiation of SC-islets [9]. However, the field does not fully understand the nature of these cell populations that emerge from differentiation. Interestingly, the combined mRNA/ATAC clustering analysis allowed us to resolve two distinct SC-EC populations that were adjacent to the SC-β cell population (Fig. 1a), suggesting that there may be multiple or a gradient of cell states between SC-EC and SC-β cells. Given the apparent similarity of SC-β and the two SC-EC cell populations in our multiomic clustering (Fig. 1b), we performed a trajectory analysis to detect differences in both gene expression and chromatin accessibility between SC-β and SC-EC cells (Fig. 2a, Extended Data Fig. 2a, Supplementary Table 5). Enterochromaffin cell identity genes, such as *SLC18A1*, *TPH1*, and *FEV*, were most highly expressed in the SC-EC side of the trajectory map as expected. Similarly, β-cell identity genes, including *INS*, *PAX6*, and *ISL1*, were most highly expressed toward the SC-β cell end (Fig. 2b and c). Furthermore, gene sets relating to SC-EC or SC-β cells were upregulated in their respective sides of the trajectory map (Extended Data Fig. 2b). Differential motif chromatin accessibility analysis identified LMX1A, GATA6, and CTCF motifs to be enriched in the SC-EC cell end, while NEUROD1, HAND2, MAFA, PAX6, ONECUT2, and RFX2 motifs were enriched for in the SC-β cell side (Fig. 2b and c).

While many of these identified genes and motifs were enriched on the side normally

associated with that particular cell type, their expression was often a gradient that spilled over toward the other end of the trajectory map (Fig. 1b). To further probe the transcriptional and chromatin landscape of cells along this trajectory map, we performed subclustering analyses separately on the SC-β cell population and the SC-EC cell populations (Fig. 2d). This subclustering identified 4 separate sub-populations of both SC-β and SC-EC cells (Fig. 2d). Differential expression analyses revealed that each of the SC-β cell sub-populations had distinguishing gene expression and motif accessibility features (Fig. 2d and e, Extended Data Fig. 2c and d, Supplementary Table 6), including SC-β cells with *TPH1* expression and more open chromatin accessibility of the SC-EC cell-associated DNA-binding motifs for LMX1A and CTCF. The high *INS*-expressing SC-β cell population displayed greater motif accessibility to β-cell-associated transcription factors, such as PAX6, and less motif accessibility to EC-cell transcription factors, such as LMX1A. Similarly, differential expression analyses demonstrated that the four SC-EC cell sub-populations also had distinguishing gene expression and motif accessibility features, including SC-EC cells with elevated *INS* expression and more open chromatin accessibility of the SC-β cell-associated DNA-binding motifs for PAX6 and MAFG. Collectively, these data suggest that the SC-EC and SC-β cell populations produced from *in vitro* differentiation form a continuum of cell states rather than exhibiting clear exclusivity of gene expression and chromatin accessibility. The identification of cells with both SC-β and SC-EC cell features here is in contrast with prior studies that only used transcriptional profiling of SC-islets to conclude that the SC-β and SC-EC cells populations were distinct [9, 12], highlighting how this multiomic approach facilitates greater resolution of cell identity.

We were interested in understanding how chromatin regulators could influence cell fate identities between islet and intestinal endocrine cells during *in vitro* differentiation. Our

integrated multiomic analysis we identified the chromatin remodeler CCCTC-binding factor (CTCF) as the transcription factor binding motif having the greatest increase in accessibility in SC-EC cells compared to SC-β cells (Fig. 2c). Because CTCF is a chromatin remodeler that is involved in other developmental and differentiation processes [38], we sought to further examine its impact on pancreatic differentiation. We utilized a doxycycline inducible VP64-p65-Rta CRISPR activation (CRISPRa) stem cell line [39] to increase transcription of *CTCF* during differentiation (Fig. 2f). Unfortunately, we were unsuccessful in also knocking down its expression despite trying several approaches, which we would hypothesize could improve differentiation to SC-β cells. The introduction of *CTCF* guide RNA (gRNA) into the CRISPRa stem cell line induced the upregulation of *CTCF* expression upon activation of *dCas9* with doxycycline (Extended Data Fig. 3a). Upregulation of *CTCF* during stage 5 endocrine cell induction [7, 29] resulted in drastic reductions in the expression of β-cell identity genes, such as *INS* and *ISL1*, and significant upregulation of intestinal lineage associated genes, such as *SLC18A1* and *FEV* (Fig. 2g, Extended Data Fig. 3a). At the end of differentiation, *CTCF*-overexpressed cells displayed upregulated EC-cell genes including *SLC18A1*, *FEV*, *DDC*, *TPH1*, and *LMX1A* (Extended Data Fig. 3b). Immunocytochemistry staining revealed a greatly diminished SC-β cell population but increased SC-EC cell generation (Fig. 2h, Extended Data Fig. 3c). This change in lineage was further demonstrated with reduced total insulin content and impaired glucose-stimulated insulin secretion in *CTCF*-overexpressed cells (Extended Data Fig. 3d and e). The effects of *CTCF* overexpression were temporally dependent and had the greatest effect during the endocrine induction stage of the protocol, suggesting *CTCF* to be important for SC-EC versus SC-β cell fate selection (Extended Data Fig. 3f). Furthermore, single-cell multiomic sequencing demonstrated that CTCF overexpression caused endocrine cells to have increased

CTCF motif chromatin accessibility and SC-EC cell signatures while decreased accessibility of several β-cell-associated transcription factor binding motifs, such as PAX6, MAFA, PDX1, and ISX (Fig. 2i, Extended Data Fig. 3g to l). Taken together, these findings show that *CTCF* modulation drastically disrupted acquisition of a SC-β cell identity from pancreatic progenitors and instead redirected them to an intestinal enterochromaffin-like cell fate, highlighting the utility of this multiomics approach for identifying important regulators of SC-β cell fate determination.

**Multiomic characterization of primary islets reveals well-defined cell types**

In order to compare the transcriptional and chromatin signatures identified in our SC-islets to their *in vivo* counterparts, we sequenced and characterized primary human islets using the same multiomics approach (Fig. 3a, Extended Data Fig. 4a and b, Supplementary Table 2). Using mRNA and chromatin information either in isolation or in a combined analysis, we identified ten distinct populations that included both pancreatic endocrine and exocrine cell types as well as other minor cell populations (Fig. 3b and c, Extended Data Fig. 4c and d, Supplementary Table 3). Unlike SC-islets (Fig. 1b and 1c, Extended Data Fig. 1c). Clusters representing all cell types appear very distinct and exhibit robust expression of their respective gene markers by. In addition, the chromatin accessibility profiles demonstrate that primary islet endocrine cells contain distinct peak signals around the *INS*, *GCG*, *SST*, and *PPY* genomic regions of β-cells, α-cells, δ-cells, and PP-cells, respectively (Fig. 3d). The closed chromatin accessibility profile of *INS* across non-β-cell types in particular is in stark contrast to that observed in SC-islets (Fig. 1d), where chromatin accessibility of the *INS* gene is relatively open across all cell types. We also performed chromatin motif accessibility analyses in primary islets

and identified distinguishing features across these endocrine types (Fig. 3e, Extended Data Fig. 4e, Supplementary Table 4). Notably, primary β-cells have accessible binding sites for PDX1, NKX6-1, MAFA and ISX, similar to SC-β cells. Surprisingly, however, motifs identified as enriched in SC-EC cells, such as LMX1A and MNX1, were enriched in β-cells when compared to other endocrine populations. Furthermore, re-clustering analysis of the primary β-cells identified 3 sub-populations that displayed unique gene expression and accessible motif signatures (Extended Data Fig. 4f - j, Supplementary Table 6), consistent with previous studies utilizing either single-cell RNA or ATAC sequencing [27, 40-42].

Similar to our multiomic analysis of SC-islets, we also examined transcription factor activity in primary islet endocrine cells as assessed by both relative increases in mRNA transcripts and chromatin accessibility of the corresponding DNA-binding motif (Fig. 3f and g). For β-cells, NKX6-1 is the only transcription factor that is shared in the top 10 active factor list with SC-β cells (Fig. 1f vs Fig 3f). Other notable active transcription factors include TFEB, MLX, NR1D1 and the primary β-cell gene *MAFA*. δ-cells share multiple accessible motifs with β-cells, including ATF4, SOX9, INSM1, and PDX1. NFYB, EGR3, and SMAD3, however, are exclusive to δ-cells by both motif enrichment and mRNA. Notable α-cell active factors include CREM, ATF7, ATF3, and VDR, while PP-cells were enriched for ELF2, EBF1, MEIS1, and CUX1. While some of the active transcription factors are shared with the list from SC-islets, this analysis illustrates that primary islet cells have a unique transcriptional and chromatin landscape when compared to their stem cell-derived counterparts, highlighting specific deficiencies in lineage specification during directed differentiation protocols.

**Primary human islets have a more restricted chromatin landscape than SC-islets**

Previous studies have demonstrated that β-cells derived from *in vitro* differentiation of hPSCs are functionally and transcriptionally different than their *in vivo* counterparts [5, 6, 9, 13, 14, 16]. To obtain greater resolution of the differences between stem cell-derived endocrine cells and those found *in vivo*, we used single-cell multiomics to directly compare both the transcriptional and chromatin landscapes of SC-islets and primary human islets (Fig. 4a). Both SC-islets and primary islets contained β, α, and δ-cells (Extended Data Fig. 5a and b). However, only SC-islets contained EC-cells, while only primary islets contained PP-cells, consistent with prior analysis [9]. Pearson correlation analysis of gene expression demonstrated that SC-islet cell types were generally most similar to their primary cell counterparts (Fig. 4b). Surprisingly, this trend was not observed with promoter chromatin accessibility. Greater similarity was observed based on cell origin (*in vitro*-derived vs. primary) rather than by cell type. We explored this dissimilarity in promoter accessibility further by examining commonly used endocrine identity genes (Fig. 4c). In general, promoter accessibility of SC-islet cells was broadly open across cell types even though mRNA expression patterns were consistent with prior RNA-only analysis [5, 9, 13]. SC-β cells expressing high levels of many identity genes but yet maintaining low expression of *MAFA*, *UCN3*, *IAPP*, and *SIX2* (Fig. 4c). Additionally, SC-islet cells have open chromatin at gene regions associated with non-endocrine identities including exocrine, hepatic, esophagus, stomach, intestinal, and progenitor cells (Fig. 4d). Notably, primary β-cells had greater chromatin accessibility of the *INS* gene compared to SC-β cells (Fig. 4e). SC-α and δ-cells also shared chromatin accessibility peaks around the *INS* gene, which was not observed in primary α and δ-cells. The *GCG* and *SST* gene regions were open in α and δ-cells from both origins, respectively. Unlike with *INS*, however, these gene regions were relatively inaccessible in the other endocrine cell types, regardless of origin. Further analysis of the chromatin accessibility of

*MAFA* and *UCN3* (Fig. 4f), which are known to be transcriptionally upregulated in β-cells [9, 13, 43-45], demonstrated that predicted cis-regulatory elements present around the *MAFA* and *UCN3* gene regions differed considerably between SC-β and primary β-cells.

To further understand the similarities and differences in endocrine cell identity between SC-islets and primary islets, we performed differential gene expression and motif chromatin accessibility analyses on the β, α, and δ-cell populations (Fig. 4g, Extended Data Fig. 5c, Supplementary Table 8). The mRNA data from the multiomic analysis yielded results consistent with previous RNA-only studies, demonstrating SC-β cells had lower expression of *IAPP*, *INS*, *UCN3*, *G6PC2*, and *CHGB* than primary β-cells [9, 13, 40, 43, 45]. We also noted higher expression of *ARID1B* in SC-β cells, which is a chromatin regulator reported to play a role in identity of non-pancreatic tissues [46]. Surprisingly, *ONECUT2* had increased mRNA expression and DNA-binding motif chromatin accessibility in SC-β cells compared to primary β-cells, despite this being a gene whose increased expression is associated with adult human β-cells compared to juvenile β-cells [30]. The chromatin accessibility data demonstrated that SC-β cells had more enriched motifs that are associated with off-target or progenitor cell states compared to primary β-cells, such as CDX2, LMX1A, SOX9, and SOX2. Primary β-cells had enriched motifs relating to genes in the FOS/JUN family, whose gene expression is associated with improved function of SC-β cells after transplantation [13]. MAFA surprisingly had similar chromatin accessibility of the associated DNA-binding motif between SC-β and primary β-cells, even though SC-β cells have relatively low MAFA mRNA expression and lower chromatin accessibility to the MAFA gene itself (Fig. 4f and 4g). This suggests that the role of MAFA in β-cell identity may be regulated by the chromatin state of the associated gene and not accessibility of the transcription factor to its motif. Taken together, these data illustrate that key differences exist in both mRNA expression

and motif chromatin accessibility in SC-β cells compared to primary β-cells. The motif differences in particular demonstrate that SC-β cells have increased hallmarks of off-target cell types as well as a lack of key adult primary β-cell identity signatures, while primary β-cells have a more restricted chromatin state.

Earlier in Figures 1 and 3, we defined a list of the top 10 most relatively active transcription factors for each major endocrine cell for SC-islets and primary islets based on mRNA expression and motif accessibility. To better understand the differences and similarities between β, α, and δ-cells from SC-islets and primary islets, we examined the relative activity of these transcription factors in each cell type from both origins (Fig. 4h and Extended Data Fig. 5d). For the SC-β cell and primary β-cell lists, many of the transcription factors had similar mRNA expression and motif accessibility. Some notable exceptions exist, including enriched RFX1 motif accessibility, reduced *MAFA* RNA expression, and reduced NFE2L1 and BACH2 motif accessibility in SC-β cells compared to primary β-cells. To delineate the largest differences in transcription factor activity, we defined the top 10 most differentially active transcription factors in SC-β cells compared to primary β-cells. Finally, we also applied the active transcription factor list from SC-EC cells in Figure 1 to these populations. Interestingly, many of these transcription factors are enriched in SC-β cells compared to primary β-cells, further supporting the notion that SC-β cells retain some ambiguity in their cell identity due to either incorrect or incomplete cell fate specification. Taken together, these analyses highlight that while the cell types found within SC-islets generally correspond to their *in vivo* counterparts, fundamental differences still exist in cell identity in terms of mRNA expression, chromatin accessibility of gene-encoding regions, and transcription factor DNA-binding motifs

**Extended time *in vivo* but not *in vitro* improves chromatin and transcriptional SC-β cell signatures**

We and others have previously demonstrated that SC-β cells can acquire improved phenotypes with extended time *in vitro* or after transplantation [4, 9, 14, 24, 47-49]. To better characterize the chromatin and transcriptional changes that occur to cells within SC-islets over time under these two conditions, we performed multiomic sequencing on SC-islets that underwent extended culture *in vitro* or were transplanted into mice. First, stage 6 *in vitro* cultured SC-islets were sequenced at 2 weeks, 3 weeks, 4 weeks, 6 months, and 12 months (Fig. 5a, Extended Data Fig. 6a). All key SC-islet endocrine cell types (SC-β, SC-α, SC-δ, and SC-EC cells) were identified in the integrated datasets (Extended Data Fig. 6b). Interestingly, while SC-β cells displayed increased *INS* transcript with time, insulin secretion improved only until week 4 (Extended Data Fig. 6c-d). SC-α, SC-δ, and SC-EC cells demonstrated incremental increases in GCG, SST, and TPH1 expression, respectively, until week 4 but declined by month 6 (Extended Data Fig. 6c). Similarly, we found drastic reductions in the SC-EC population at month 6 and 12 and a decrease in SC-α numbers by month 12 (Extended Data Fig. 6e). Pearson correlation analysis using RNA expression indicated that SC-EC cells began to lose their identity at month 6, while SC-α cells lost their defining signatures at month 12 (Extended Data Fig. 6f). Interestingly, Pearson correlation using the ATAC information revealed that there were drastic shifts in chromatin accessibility at month 6 and month 12 for all endocrine cell types (Extended Data Fig. 6g).

Specifically in SC-β cells, expression of many β-cell transcription factors and their associated motifs were upregulated at week 4 of *in vitro* culture, but most of these motifs and many transcripts decreased in the long term (month 6 and month 12) (Fig. 5b - d). Expression of other β-cell genes and the accessibility to their respective promoters, including *INS*, *IAPP*, and

*UCN3*, increased over time, while the expression and promoter accessibility of genes associated with non-endocrine cells, such as the SC-EC genes LMX1A and TPH1, diminished in long term culture (Fig. 1b, d, Supplementary Table 9). Similarly, ATAC peaks demonstrated accessibility around the *INS* genomic region increased in SC-β cells over time but diminished in SC-α and SC-δ cells (Fig. 5c). Collectively, multiomic analysis of *in vitro* cultured SC-islets revealed that while the transcription of some important β-cell genes increased over time, the chromatin state of these cells became much more restricted in long-term culture. While this restriction helped to decrease off-target populations, many crucial β-cell transcription factors were also downregulated, possibly leading to the observed decrease in function.

In parallel, we also performed multiomic sequencing of SC-islets transplanted for 6 months under the kidney capsule of immunocompromised NOD.Cg-*Prkdc*$^{scid}$ *Il*2*rg*$^{tm1wjl}$/SzJ (NSG) mice (Fig. 5e, Extended Data Fig. 7a and b). Gene expression and ATAC promoter accessibility information demonstrated that endocrine identities became even more distinct than before transplantation (Extended Data Fig. 7c, d and e). Motif accessibility was also well defined for each cell type, with transplanted SC-β and SC-EC cells no longer displaying shared open motifs (Extended Data Fig. 7f). In particular, PDX1 was enriched in transplanted SC-β cells rather than in both SC-β and SC-EC cells as it had been before transplantation (Extended Data Fig. 7g). Comparison of these 6-month transplanted SC-islets to those before transplantation revealed that transplanted SC-islets acquired gene expression and chromatin accessibility signatures that were more similar to their primary cell counterparts, including increased *INS*, *MAFA*, and *IAPP* expression in SC-β cells (Extended Data Fig. 8a, b and c, Supplementary Table 10). These improvements in cell identity were also reflected in the chromatin accessibility around the *INS* gene, which exhibited diminished peak signals in non-β cell populations (Fig. 5f).

Gene set enrichment analysis illustrated that gene sets associated with β-cell identity were upregulated in transplanted SC-β cells, whereas gene sets relating to immature and off-target populations were higher in pre-transplanted SC-β cells (Extended Data Fig. 8d). Furthermore, the motif accessibility for many β-cell associated transcription factors were enriched in transplanted SC-islets (Fig. 5g). Collectively, transplanted SC-β cells were much more similar to primary β-cells than before transplantation in terms of gene expression, promoter accessibility, and motif accessibility, highlighting that transplantation improves both the transcriptional and chromatin landscape of SC-β cells (Extended Data Fig. 8f).

We then integrated the datasets of 2-week *in vitro* SC-islets, 6-month *in vitro* SC-islets, and 6-month transplanted SC-islets and found that transplanted SC-β cells had the highest enrichment of β-cell associated genes by both gene expression and promoter accessibility, such as MAFA, G6PC2, UCN3, and IAPP (Fig. 5h, Extended Data Fig. 9a and b). SC-EC cells also retained expression of identity genes when transplanted, in contrast to during 6-month *in vitro* culture. Overall, transplanted SC-islet cells displayed greater enrichment of genes and motifs for transcription factors associated with their cell types (Extended Data Fig. 9c). In particular, transplanted SC-β cells exhibited many upregulated and downregulated genes and motifs compared to their 6-month *in vitro* counterparts, allowing us to compile a list of markers relevant for β-cell identity (Fig. 5i-j, Extended Data Fig. 9d, Supplementary Table 11). In summary, extended *in vitro* culture tends to generally close chromatin regions, even some associated with β-cells. In contrast, extended time *in vivo* tends to further open chromatin regions associated with each cell's identity while restricting regions for other lineages, allowing transplanted SC-β cells to develop a chromatin and transcriptional landscape more similar to primary β-cells.

During our SC-islet multiomic analysis, the chromatin regulator *ARID1B* emerged

several times as being differentially expressed in β-cells (Fig. 4g, 5d, Extended Data Fig. 8c, 10a) and has also been reported to play a role in the identity of non-pancreatic tissues [46]. To study its role in SC-β cell identity, we utilized short hairpin RNAs (shRNAs) targeting *ARID1B* transcript to reduce its expression in fully differentiated SC-islets (Fig. 5k and Extended Data Fig. 10b and c). *ARID1B* knockdown resulted in upregulation of β-cell identity markers, including *INS*, *IAPP*, *DLK1* (Fig 5l, Extended Data Fig. 10c). This improvement in gene expression was also associated with elevated insulin protein content, reduced pro-insulin/insulin ratio, increased amylin (IAPP) expression, and increased SC-β cell yields (Fig. 5l and m, and Extended Data Fig. 10d and e*). ARID1B* knockdown, however, did not increase either glucagon (GCG) or somatostatin (SST) protein content (Extended Data Fig. 10f). ARID1B also helped improve insulin secretion in the presence of insulin secretagogues, such as exendin 4 (Extended Data Fig. 10g). Single-cell multiomic sequencing of SC-islets with *ARID1B* knockdown (Extended Data Fig. 10h) revealed increased gene expression and motif accessibility for signatures enriched in mature human β-cells and transplanted SC-β cells, including increased chromatin accessibility around the *IAPP* genomic region (Fig 5n, Extended Data Fig. 10i and j, Supplementary Table 12). Taken together, these data demonstrate that the chromatin landscape plays a significant role in islet cell identity. Furthermore, this approach exemplifies how single-cell multiomics can be used to identify important chromatin and transcriptional deficiencies within particular cell populations that can be targeted to improve directed differentiation to specific cell types.

**Discussion**

SC-β cells generated from established differentiation protocols possess many features of

primary human β-cells, including the expression of β-cell specific markers, glucose responsive insulin secretion, and the ability to reverse severe diabetes in animal models [7, 9, 12, 14, 16]. Thus, these cells have the potential to provide a functional cure for human patients with type 1 diabetes [50]. These *in vitro*-generated cells are recognized to not have exactly the same transcriptional and functional features as their *in vivo* counterparts. An improved understanding of the deficiencies in lineage specification during differentiation to SC-islet cell types could help prevent unintended generation of non-endocrine cell types, improve endocrine cell identity, and ultimately increase SC-β cell function. By using a single-cell multiomic approach of integrating mRNA transcriptional data and ATAC chromatin accessibility information for each individual cell, we were able to generate more robust definitions of the cell types found within SC-islets than using either data set in isolation. In particular, we provide novel lists of genes and transcription factor binding motifs that emerged from this integrated analysis as being important to the specification of the different cell types found within SC-islets. Importantly, these data provide a valuable resource to identify particular deficiencies in the chromatin and transcriptional landscape of SC-islet cell types. Furthermore, we demonstrate how the differentially expressed genes and accessible DNA-binding motifs identified within this single-cell multiomic data set can be targeted to alter directed differentiation to SC-islet cell types.

Our multiomic analysis identified several interesting observations about the dynamics of gene expression and chromatin state of SC-islet cell types over time and in relation to primary human islets. In general, SC-islet cell types were less distinct from each other by chromatin accessibility compared to their primary cell counterparts. This difference was driven mainly by continued open chromatin accessibility of genes expressed by progenitor cell types, such as *NEUROG3* [51] and *GP2* [52], or alternative cell fates that were closed in primary cells. SC-islet cell

types developed more distinguished transcriptional and chromatin accessibility signatures that matched their respective cell identities after transplantation into mice [13, 14], while extended *in vitro* culture broadly restricted access to chromatin regions, including those associated with a β-cell identity. The signals that the cells are receiving from the *in vivo* environment that are missing in *in vitro* culture are unknown. Furthermore, the mechanism behind this difference in cellular identity remains unclear, but similar differences may also be present within *in vitro* differentiation systems for other cell types where immaturity is commonly observed [53, 54].

A prior study discovered that enterochromaffin-like cells are generated during SC-islet differentiations *in vitro* and suggested that SC-EC and SC-β cells are distinct populations that arise from the same pancreatic progenitor population [9]. By obtaining greater resolution of cell identity by simultaneously analyzing both the transcriptional and chromatin state of each cell rather than mRNA alone, our data suggests that SC-EC and SC-β cells are not completely exclusive cell populations. Instead, they comprise a continuum of cell types that contain varying degrees of both enterochromaffin and β-cell features. Interestingly, increased DNA binding motif accessibility for the chromatin regulator CTCF was highly correlative with increased SC-EC cell features. *CTCF*-overexpression blocked differentiation to pancreatic endocrine and further redirected cell fate selection toward the enteroendocrine lineage, supporting the notion that chromatin accessibility is a major regulator of the fate decision between SC-EC and SC-β cells along this gradient. As SC-EC cells may be detrimental to SC-β cell function [9], or at the least is unlikely to be providing a benefit for SC-islet in diabetes cell replacement therapy, understanding how to reduce or eliminate this off-target cell population is a priority for the field. In addition, the apparent plasticity of identity in SC-islets also opens up the possible of cell fate switching like has been done with primary cells [55, 56].

Study of SC-islets is largely motivated by success of isolated primary islet transplantation into patients with Type 1 diabetes [57-59]. Several areas of investigation are being pursued in the field for which multiomic assessment could be beneficial. Since the early reports of generating SC-islets and other similar pancreatic endocrine [5, 6, 60], many teams has focused on changing the composition of the media or how the cells are cultured to improve SC-islet function [4, 7, 8, 14, 61, 62]. Controlled assembly of the three-dimensional SC-islet aggregates and circadian entrapment have also further improved function of SC-islets [4, 8, 24]. Proliferation of early endoderm and pancreatic progenitors have also been achieved to expand the number of cells available for continued differentiation to SC-islets and potentially improve them [63-66]. Increased efficacy of generation or purification of progenitors have also been pursued to generate SC-islets [10, 52, 67, 68]. Furthermore, genetic engineering to lessen immune recognition [69] of SC-islets in order to transplant the cells without immunosuppression has been reported [62, 70]. SC-islets have been generated from patients with diabetes for study of diabetes pathology and consideration as an autologous cell source of diabetes regenerative medicine [12, 71-76]. Finally, many materials are being studied to provide immune protection or otherwise benefit islets or SC-islets after transplantation [77-84].

Our data demonstrating that chromatin accessibility differs drastically between cell types as well as within a given cell type from different origins, such as *in vitro* differentiation compared to isolation from donors, suggesting that regulators of chromatin may play a substantial role during the generation and identity of islet cell types. Our modulations of the chromatin regulators CTCF and ARID1B to alter SC-β cell identity, highlight how better control of the chromatin landscape could lead to the development of SC-islet differentiation protocols with improved function and defined cellular composition. Here, we provide a comprehensive indexing of single-cell transcriptional state and chromatin accessibility in SC-islets to help

facilitate further development of these differentiation protocols, as both aspects of islet cell identity will likely need to be targeted to enhance SC-islet differentiation strategies.

## Methods

### Cell culture and differentiation

The HUES8 hESC line was provided by Douglas Melton (Harvard University)[5]. The H1 hESC line was provided by Lindy Barrett (Broad Institute) with permission from WiCell containing doxycycline-inducible dCas9-VPR transgene in the AAVS1 locus (CRISPRa system)[39]. hESC work was approved by the Washington University Embryonic Stem Cell Research Oversight Committee (approval no. 15-002) with appropriate conditions and consent. mTeSR1 (StemCell Technologies; 05850) was used for the culture of undifferentiated cell propagation. All cell culture was maintained in a humidified incubator at 5% $CO_2$ and 37°C. Cells were passaged every 4 days by washing cell with PBS and incubating with 0.2 mL TrypLE/$cm^2$ (Gibco; 12-604-013) for 10 min or less at 37°C. Cells were collected into a 15-mL conical containing equal volume mTeSR1 supplemented with 10 µM Y-27632 (Pepro Tech; 129382310MG). Cells were counted on Vi-Cell XR (Beckman Coulter) and spun at 300g for 3 min at RT. The supernatant was aspirated, and cells were seeded at a density of $0.8x10^5/cm^2$ for propagation onto a Matrigel (Corning; 356230) coated plate in mTeSR1 supplemented with 10µM Y-27632 (Pepro Tech; 129382310MG). After 24 hr, media was replaced daily with mTeSR1 alone. SC-islet differentiation was performed as described previously[7, 29]. Briefly, 24 hr after seeding hESCs at a density of $6.3x10^5$ cells/$cm^2$, daily mediums with added growth factors were started to initiate the differentiation process. The base media formulation and added growth factors can be found in Supplementary Table 1. After 7 days in stage 6, cells were placed on an Orbishaker (Benchmark) at 100 RPM and maintained in ESFM media. All *in vitro* data in the paper is from 14 days in stage 6 unless otherwise noted. Human islets were acquired from Prodo Laboratories for comparison in some experiments.

**Single-nuclei sample preparation and sequencing**

Cells collected were processed and delivered to the McDonnell Genome Institute at Washington University for library preparation and sequencing. Single-cell suspension samples were processed into nuclei according to 10X Multiome ATAC + Gene Expression (GEX) protocol (CGOOO338). Briefly, cell samples were collected and washed with PBS (with 0.04% BSA), lysed with chilled Lysis Buffer for 4 min, washed 3 times with wash buffer, and resuspended with 10x nuclei buffer at 3000-5000 nuclei/µl. Nuclei samples were subsequently processed using the Chromium 10x genomics instrument, with a target cell number of 7000-10000. The 10x Single Cell Multiome ATAC + Gene Expression v1 kit was used according to the manufacturer's instructions for library preparations. Sequencing of the library was performed using the NovaSeq 6000 System (Illumina).

**Multiomics sequencing analysis workflow**

Multiomic sequenced files were processed for demultiplexing and analyzed using Cell Ranger ARC v2.0. Genes were mapped and referenced using GRCh38. RStudio with R version 4.0.3 was used to perform subsequent analyses. Datasets were further analyzed using Seurat 4.01[85] and Signac 1.3.0[86]. For ATAC data, genomic positions were mapped and annotated with EnsDb.Hsapeins.v86 and hg38. Low quality cells including doublets, dead cells, and poor sequencing depth cells, were filtered by removing cells with low RNA counts (nCount_RNA < 1000) and low ATAC counts (nCount_ATAC < 1000); high RNA counts (ranging > 40000 – 50000) and high ATAC counts (ranging > 40000 – 50000); nucleosome signal > 1.25, and TSS enrichment < 2. In addition, mouse cells were removed by removing cells showing high

expression of kidney marker *TTC36*[13] for transplanted SC-islet samples. Additional details for filtering cells can be found in Supplementary Table 2 and Supplementary Figure 1. Gene expression data and ATAC data were processed using SCT transform for the former and RunTFIDF and RunSVD for the latter. Dataset integration was performed by anchoring using SCT normalized data and integrative gene expression and ATAC dimension reductions were carried out using PCA and LSI. Clusters were identified using differential gene expression analysis or expression of known marker genes. Promoter accessibilities were determined from ATAC information using GeneActivity by considering 2000 base pairs upstream of the transcription start site. ATAC peaks were called using MACS2 and linked using LinkPeaks and Cicero to determine cis-regulatory elements[86, 87]. Motif activity was computed by using the chromVAR package[88] and JASPAR version 2020 database[89]. Trajectory analysis was conducted using Monocle3 package[90].

**Transduction of *CTCF* gRNA in CRISPRa line**

CRISPRa genetic engineering of the H1 dCas9-VPR line[39] was mediated using custom guide RNAs (gRNAs) ordered from MilliporeSigma. *CTCF* targeting gRNAs for CRISPRa sequences are provided in Supplementary Table 1. gRNAs were resuspended to a final concentration of 100uM in water. Primers were phosphorylated and ligated together by adding T4 ligation buffer and T4 Polynucleotide kinase (PNK) enzyme (NEB; B0202A and M0201S) and running on a thermocycler under the following conditions: 37 °C for 30 min; 95 °C for 5 min; and ramp down to 25 °C at 5 °C/min. Oligos were then diluted with 90 µl of ultrapure water. These oligos were then inserted into the sgRNA library backbone (Lenti sgRNA(MS2)_puro) using the golden gate reaction. This is achieved by adding a 25ng ul

plasmid backbone to a master mix of Rapid Ligase Buffer 2X (Enzymatics: B1010L), Fast Digest Esp31 (Thermo: FD0454), DTT (Promega: PRP1171), BSA (NEB: B9000S), T7 DNA ligase, (Enzymatics: L6020L) and the diluted gRNA oligos in a total reaction volume of 25 ul. The Golden Gate assembly reaction is then preformed in a thermocycler under the following conditions: 15 cycles of 37 °C for 5 min, 20 °C for 5 min with final hold at 4°C. Lenti sgRNA (MS2) puromycin optimized backbone was a gift from Feng Zhang (Addgene plasmid # 73797). This final plasmid was then transfected into STBL3, following the same methods as mentioned in the previous shRNA lentivirus production section. To transfect the CRISPRa H1 dCas9-VPR stem cell line, lentiviral particles containing gRNA was added at MOI 5 with polybrene (5 µg/mL) in culture for 24 hr. At confluency, transfected stem cells were passaged and cultured with media containing puromycin (1 µg/mL) for selection. To induce CRISPRa expression, doxycycline (MilliporeSigma) was added at 1 µg/mL for 7 d during stage 5 of the differentiation protocol.

**Real-Time PCR**

Cells were lysed directly with RLT buffer from the RNeasy Mini Kit (74016; Qiagen) followed by RNA extraction following the manufacturer's instructions. cDNA was synthesized from the RNA using the High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems; 129382310MG) on a T100 thermocycler (BioRad). PowerUp SYBR Green Master Mix (Applied Biosystems; A257411) was used to run samples on the Quant Studio 6 Pro (Applied Biosystems) and results were analyzed using ΔΔCt methodology. Primer sequences used in this paper are listed in Supplementary Table 1.

**Microscopy**

Fluorescence images were taken on a Zeiss Cell Discoverer confocal 7 microscope. For immunocytochemistry (ICC), cells were fixed in 4% paraformaldehyde (PFA) (Electron Microscopy Science; 15714) for 30 min at RT. For staining, fixed cells were blocked in ICC solution (PBS (Fisher; MT21040CV), 0.1% Triton X (Acros Organics; 327371000), 5% donkey serum (Jackson Immunoresearch; 01700-121)) for 30 min at RT. Samples were subsequently treated with primary and secondary antibodies in ICC solution overnight at 4°C and 2 hr at RT respectively. DAPI (Invitrogen; D1306) was used for nuclear staining. Samples were incubated in DAPI for 12 min at RT, washed with ICC solution and stored in PBS until imaging. Antibody details and dilutions can be found in Supplementary Table 1. Image J was used for analysis. Quantification was performed by manual counting of cells from analyzed fluorescent images and can be found in Supplementary Figure 2.

**Flow cytometry**

Cells were single-cell dispersed by washing with PBS and adding TrypLE/cm$^2$ for 10 min at 37°C. Samples were washed with PBS, fixed for 30 min in 4% PFA at 4°C and washed with PBS. For staining, samples were treated with ICC solution for 45 min at RT. Primary antibodies were prepared in ICC solution and incubated on cells overnight at 4°C. Samples were washed with PBS and incubated for 2 hr with secondary antibodies in ICC at 4°C. Cells were washed twice with PBS and filtered before running on the LSR Fortessa flow cytometer (BD Bioscience). FlowJo v10.8.0 (Becton, Dickinson, and Company) was used for analysis. Gating strategy for flow cytometry analysis can be found in Supplementary Figure 3.

**Hormone content**

Insulin, somatostatin, glucagon, and proinsulin were measured from differentiated SC-islets. Cells were collected, rinsed, and incubated in acid-ethanol solution for 48 hr at -20°C. Sample solutions were neutralized with 1M TRIS buffer (Millipore Sigma; T6066) for downstream processing. Measurements of insulin, somatostatin, glucagon and proinsulin were made using the following ELISA kits: Human insulin ELISA (ALPCO; 80-INSHU-E01.1), Somatostatin EIA (Phoenix Pharmaceuticals; EK-060-03), Glucagon ELISA (Crystal Chem; 81520), and human proinsulin ELISA (Mercodia; 10-1118-01). ELISAs were performed according to the manufacturer instructions. Results were normalized to cell counts performed on the Vi-Cell XR (Beckman Coulter).

**Mouse transplantations and SC-islet cell retrieval**

Mice that were 7-week old, male, and with the NOD.Cg-*Prkdc*$^{scid}$ *Il*2*rg*$^{tm1wjl}$/SzJ (NSG) background (Jackson Laboratories; 005557) were randomly assigned to experimental groups. Mice were housed in a facility with a 12-hr light/dark cycle and were fed on chow diet. Animal studies were done in accordance with Washington University International Care and Use Committee (IACUC) guidelines (Approval 21-0240). Animal studies were performed by unblinded individuals in accordance with Washington University International Care and Use Committee (IACUC) guidelines. Mice randomly assigned to experimental groups were anesthetized using isoflurane and injected with ~5x10$^{6}$ SC-islets under the kidney capsule. At 6 months post-transplantation, mice were euthanized, and the kidney transplanted with SC-islets

was removed. A razor blade was used to mince the kidney prior to placing it into a solution of 2 mg/mL collagenase D (Sigma; 11088858001) in RPMI (GIBCO; 1187-085). The tissue was incubated for 40 min at 37°C before diluting with PBS and mechanical disrupting with a pipette and filtering through a 70-µm strainer (Corning; 431751). The flow through was centrifuged and the remaining cell pellet was resuspended in MACS buffer (0.05% BSA in PBS). The Miltenyi mouse cell isolation kit (Miltenyi; 130-104-694; LS column, 130-042-401) was used to remove any excess mouse cells from the cell solution. The flow through of cells was collected, centrifuged, counted, and resuspended in PBS with 0.04% BSA for nuclei processing and sequencing.

**Lentiviral design, preparation, and transduction**

Gene KD of *ARID1B* was performed using pLKO.1 TRC plasmids containing shRNA sequences targeting *ARID1B* and *GFP* (Control) (Supplementary Table 1). Glycerol stocks were grown, and plasmid DNA was isolated using Qiagen Mini-prep kit (Qiagen; 27115). Plasmid DNA was transfected into One Shot™ Stbl3™ Chemically Competent E. coli (Invitrogen; C737303) and spread on an agar plate. After 18 hr a single colony was selected, cultured, and DNA was extracted using the Qiagen Maxi-prep- plus kit (Qiagen; 12981). Viral particles were generated using Lenti-X 293T cells (Takara; 632180) cultured in DMEM with 10% heat inactivated fetal bovine serum (MilliporeSigma; F4135), and 0.01mM Sodium Pyruvate (Corning; 25-000-CL) in 10cm tissue culture treated plates (Falcon; 353003). Confluent Lenti-X 293T cells were transfected with 6 µg of shRNA plasmid, 4.5 µg of psPAX2 (Addgene; 12260; gift from Didier Trono), and 1.5 µg pMD2.G (Addgene; 12259; gift from Didier Trono) packaging plasmids in 600 µL of Opti-MEM (Life Technologies; 31985-070) and 48 µL of

Polyethylenimine ‘Max’ MW 40,000 Da (Polysciences; 24765-2) per plate. Media was refreshed after 16 hr. Viral containing supernatant was collected at 96 hr post transfection and concentrated using Lenti-X concentrator (Takara; 631232). Collected lentivirus was tittered using Lenti-X™ GoStix™ Plus (Takara;631280). Lentiviral transduction was initiated at the first day of Stage 6 with an MOI of 5 for 24 hr.

**Statistics and reproducibility**

Cell selection criteria or exclusion methods for single-cell analysis can be found in Supplementary Figure 1. For *in vitro* experiments, we performed unpaired or paired parametric t-tests (two-sided) to determine significance. All *in vitro* experiment data points presented are biological replicates. Significant values are marked based on p-values using ns > 0.05, * < 0.05, ** < 0.01, *** < 0.001, and **** < 0.0001.

**Data availability**

Single-cell multiomics sequencing analysis and source data can be found in the supplementary tables. Fluorescent images for quantification can be found in Supplementary Figure 2. Individual flow cytometry data can be found in Supplementary 4. All sequencing datasets have been deposited in the Gene Expression Omnibus (GEO) – NCBI under accession number GSE199636.

**Code availability**

Codes used for analyzing single-nuclei multiomic sequencing are available on https://github.com/punnaug.

## References


1. Lanzoni, G. & Ricordi, C. Transplantation of stem cell-derived pancreatic islet cells. *Nature reviews. Endocrinology* **17**, 7-8 (2021).
2. Millman, J.R. & Pagliuca, F.W. Autologous Pluripotent Stem Cell-Derived beta-Like Cells for Diabetes Cellular Therapy. *Diabetes* **66**, 1111-1120 (2017).
3. Aguayo-Mazzucato, C. & Bonner-Weir, S. Stem cell therapy for type 1 diabetes mellitus. *Nat Rev Endocrinol* **6**, 139-148 (2010).
4. Velazco-Cruz, L. *et al.* Acquisition of Dynamic Function in Human Stem Cell-Derived beta Cells. *Stem cell reports* **12**, 351-365 (2019).
5. Pagliuca, F.W. *et al.* Generation of functional human pancreatic beta cells in vitro. *Cell* **159**, 428-439 (2014).
6. Rezania, A. *et al.* Reversal of diabetes with insulin-producing cells derived in vitro from human pluripotent stem cells. *Nature biotechnology* **32**, 1121-1133 (2014).
7. Hogrebe, N.J., Augsornworawat, P., Maxwell, K.G., Velazco-Cruz, L. & Millman, J.R. Targeting the cytoskeleton to direct pancreatic differentiation of human pluripotent stem cells. *Nature biotechnology* **38**, 460-470 (2020).
8. Nair, G.G. *et al.* Recapitulating endocrine cell clustering in culture promotes maturation of human stem-cell-derived beta cells. *Nature cell biology* **21**, 263-274 (2019).
9. Veres, A. *et al.* Charting cellular identity during human in vitro beta-cell differentiation. *Nature* **569**, 368-373 (2019).
10. Mahaddalkar, P.U. *et al.* Generation of pancreatic beta cells from CD177(+) anterior definitive endoderm. *Nature biotechnology* **38**, 1061-1072 (2020).
11. Song, J. & Millman, J.R. Economic 3D-printing approach for transplantation of human stem cell-derived beta-like cells. *Biofabrication* **9**, 015002 (2016).
12. Maxwell, K.G. *et al.* Gene-edited human stem cell-derived beta cells from a patient with monogenic diabetes reverse preexisting diabetes in mice. *Sci Transl Med* **12** (2020).

13. Augsornworawat, P., Maxwell, K.G., Velazco-Cruz, L. & Millman, J.R. Single-Cell Transcriptome Profiling Reveals beta Cell Maturation in Stem Cell-Derived Islets after Transplantation. *Cell Rep* **32**, 108067 (2020).

14. Balboa, D. *et al.* Functional, metabolic and transcriptional maturation of human pancreatic islets derived from stem cells. *Nature biotechnology* **40**, 1042-1055 (2022).

15. Maxwell, K.G., Kim, M.H., Gale, S.E. & Millman, J.R. Differential Function and Maturation of Human Stem Cell-Derived Islets After Transplantation. *Stem Cells Transl Med* **11**, 322-331 (2022).

16. Cuesta-Gomez, N. *et al.* Characterization of stem-cell-derived islets during differentiation and after implantation. *Cell Rep* **40**, 111238 (2022).

17. Chen, T. & Dent, S.Y. Chromatin modifiers and remodellers: regulators of cellular differentiation. *Nature Reviews Genetics* **15**, 93-106 (2014).

18. Dixon, J.R. *et al.* Chromatin architecture reorganization during stem cell differentiation. *Nature* **518**, 331-336 (2015).

19. Meshorer, E. & Misteli, T. Chromatin in pluripotent embryonic stem cells and differentiation. *Nature reviews Molecular cell biology* **7**, 540-546 (2006).

20. Lund, R.J., Närvä, E. & Lahesmaa, R. Genetic and epigenetic stability of human pluripotent stem cells. *Nature Reviews Genetics* **13**, 732-744 (2012).

21. Weng, C. *et al.* Single-cell lineage analysis reveals extensive multimodal transcriptional control during directed beta-cell differentiation. *Nat Metab* **2**, 1443-1458 (2020).

22. Augsornworawat, P. & Millman, J.R. Single-cell RNA sequencing for engineering and studying human islets. *Curr Opin Biomed Eng* **16**, 27-33 (2020).

23. Xie, R. *et al.* Dynamic chromatin remodeling mediated by polycomb proteins orchestrates pancreatic differentiation of human embryonic stem cells. *Cell stem cell* **12**, 224-237 (2013).

24. Alvarez-Dominguez, J.R. *et al.* Circadian Entrainment Triggers Maturation of Human In Vitro Islets. *Cell stem cell* **26**, 108-122 e110 (2020).

25. Hu, M. *et al.* Chromatin 3D interaction analysis of the STARD10 locus unveils FCHSD2 as a regulator of insulin secretion. *Cell Rep* **34**, 108881 (2021).

26. Chiou, J. *et al.* Interpreting type 1 diabetes risk with genetics and single-cell epigenomics. *Nature* **594**, 398-402 (2021).

27. 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).

28. Su, C. *et al.* 3D chromatin maps of the human pancreas reveal lineage-specific regulatory architecture of T2D risk. *Cell metabolism* **34**, 1394-1409.e1394 (2022).

29. Hogrebe, N.J., Maxwell, K.G., Augsornworawat, P. & Millman, J.R. Generation of insulin-producing pancreatic beta cells from multiple human stem cell lines. *Nature protocols* **16**, 4109-4143 (2021).

30. Arda, H.E. *et al.* Age-Dependent Pancreatic Gene Regulation Reveals Mechanisms Governing Human beta Cell Function. *Cell metabolism* **23**, 909-920 (2016).

31. Hrvatin, S. *et al.* Differentiated human stem cells resemble fetal, not adult, beta cells. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 3038-3043 (2014).

32. Velazco-Cruz, L., Goedegebuure, M.M. & Millman, J.R. Advances Toward Engineering Functionally Mature Human Pluripotent Stem Cell-Derived beta Cells. *Front Bioeng Biotechnol* **8**, 786 (2020).

33. Gross, S. *et al.* The novel enterochromaffin marker Lmx1a regulates serotonin biosynthesis in enteroendocrine cell lineages downstream of Nkx2.2. *Development (Cambridge, England)* **143**, 2616-2628 (2016).

34. Egozi, A. *et al.* Insulin is expressed by enteroendocrine cells during human fetal development. *Nat Med* **27**, 2104-2107 (2021).

35. Van Gurp, L. *et al.* Generation of human islet cell type-specific identity genesets. *Nature communications* **13**, 1-14 (2022).

36. Haber, A.L. *et al.* A single-cell survey of the small intestinal epithelium. *Nature* **551**, 333-339 (2017).

37. Grün, D. *et al.* Single-cell messenger RNA sequencing reveals rare intestinal cell types. *Nature* **525**, 251-255 (2015).

38. Phillips, J.E. & Corces, V.G. CTCF: master weaver of the genome. *Cell* **137**, 1194-1211 (2009).

39. Hazelbaker, D.Z. *et al.* A multiplexed gRNA piggyBac transposon system facilitates efficient induction of CRISPRi and CRISPRa in human pluripotent stem cells. *Sci Rep* **10**, 635 (2020).

40. Augsornworawat, P. & Millman, J.R. Single-cell RNA sequencing for engineering and studying human islets. *Current Opinion in Biomedical Engineering* (2020).

41. Segerstolpe, Å. *et al.* Single-cell transcriptome profiling of human pancreatic islets in health and type 2 diabetes. *Cell metabolism* **24**, 593-607 (2016).

42. Baron, M. *et al.* A single-cell transcriptomic map of the human and mouse pancreas reveals inter- and intra-cell population structure. *Cell systems* **3**, 346-360. e344 (2016).

43. Artner, I. *et al.* MafA and MafB regulate genes critical to beta-cells in a unique temporal manner. *Diabetes* **59**, 2530-2539 (2010).

44. Aguayo-Mazzucato, C. *et al.* MAFA and T3 Drive Maturation of Both Fetal Human Islets and Insulin-Producing Cells Differentiated From hESC. *J Clin Endocrinol Metab* **100**, 3651-3659 (2015).

45. Blum, B. *et al.* Functional beta-cell maturation is marked by an increased glucose threshold and by expression of urocortin 3. *Nature biotechnology* **30**, 261 (2012).

46. Santen, G.W. *et al.* Mutations in SWI/SNF chromatin remodeling complex gene ARID1B cause Coffin-Siris syndrome. *Nature genetics* **44**, 379-380 (2012).

47. Hogrebe, N.J., Augsornworawat, P., Maxwell, K.G., Velazco-Cruz, L. & Millman, J.R. Targeting the cytoskeleton to direct pancreatic differentiation of human pluripotent stem cells. *Nature Biotechnology* **38**, 460-470 (2020).

48. Du, Y. *et al*. Human pluripotent stem-cell-derived islets ameliorate diabetes in non-human primates. *Nature Medicine* **28**, 272-282 (2022).

49. Maxwell, K.G., Kim, M.H., Gale, S.E. & Millman, J.R. Differential Function and Maturation of Human Stem Cell-Derived Islets After Transplantation. *Stem cells translational medicine* **11**, 322-331 (2022).

50. Ramzy, A. *et al.* A century long journey from the discovery of insulin to the implantation of stem cell derived islets. *Endocr Rev* (2022).

51. Gu, G., Dubauskaite, J. & Melton, D.A. Direct evidence for the pancreatic lineage: NGN3+ cells are islet progenitors and are distinct from duct progenitors. *Development (Cambridge, England)* **129**, 2447-2457 (2002).

52. Cogger, K.F. *et al.* Glycoprotein 2 is a specific cell surface marker of human pancreatic progenitors. *Nat Commun* **8**, 331 (2017).

53. Patterson, M. *et al*. Defining the nature of human pluripotent stem cell progeny. *Cell Res* **22**, 178-193 (2012).

54. Murry, C.E. & Keller, G. Differentiation of embryonic stem cells to clinically relevant populations: lessons from embryonic development. *Cell* **132**, 661-680 (2008).

55. Bramswig, N.C. *et al.* Epigenomic plasticity enables human pancreatic α to β cell reprogramming. *The Journal of clinical investigation* **123**, 1275-1284 (2013).

56. Zhou, Q., Brown, J., Kanarek, A., Rajagopal, J. & Melton, D.A. In vivo reprogramming of adult pancreatic exocrine cells to beta-cells. *Nature* **455**, 627-632 (2008).

57. Markmann, J.F. *et al.* Insulin independence following isolated islet transplantation and single islet infusions. *Annals of surgery* **237**, 741 (2003).

58. Shapiro, A.M. *et al.* Islet transplantation in seven patients with type 1 diabetes mellitus using a glucocorticoid-free immunosuppressive regimen. *N Engl J Med* **343**, 230-238 (2000).

59. Barton, F.B. *et al*. Improvement in outcomes of clinical islet transplantation: 1999–2010. *Diabetes Care* **35**, 1436-1445 (2012).

60. Russ, H.A. *et al.* Controlled induction of human pancreatic progenitors produces functional beta-like cells in vitro. *The EMBO journal* **34**, 1759-1772 (2015).

61. Ghazizadeh, Z. *et al.* ROCKII inhibition promotes the maturation of human pancreatic beta-like cells. *Nat Commun* **8**, 298 (2017).

62. Yoshihara, E. *et al.* Immune-evasive human islet-like organoids ameliorate diabetes. *Nature* **586**, 606-611 (2020).

63. Cheng, X. *et al.* Self-renewing endodermal progenitor lines generated from human pluripotent stem cells. *Cell stem cell* **10**, 371-384 (2012).

64. Sneddon, J.B., Borowiak, M. & Melton, D.A. Self-renewal of embryonic-stem-cell-derived progenitors by organ-matched mesenchyme. *Nature* **491**, 765-768 (2012).

65. Nakamura, A. *et al.* Long-term feeder-free culture of human pancreatic progenitors on fibronectin or matrix-free polymer potentiates β cell differentiation. *Stem cell reports* **17**, 1215-1228 (2022).

66. Ma, X. *et al.* Human expandable pancreatic progenitor-derived β cells ameliorate diabetes. *Science advances* **8**, eabk1826 (2022).

67. Nostro, M.C. *et al.* Efficient generation of NKX6-1+ pancreatic progenitors from multiple human pluripotent stem cell lines. *Stem cell reports* **4**, 591-604 (2015).

68. Mukherjee, S., French, D.L. & Gadue, P. Loss of TBX3 enhances pancreatic progenitor generation from human pluripotent stem cells. *Stem cell reports* **16**, 2617-2627 (2021).

69. Deuse, T. *et al.* Hypoimmunogenic derivatives of induced pluripotent stem cells evade immune rejection in fully immunocompetent allogeneic recipients. *Nature biotechnology* **37**, 252-258 (2019).

70. Sintov, E. *et al.* Whole-genome CRISPR screening identifies genetic manipulations to reduce immune rejection of stem cell-derived islets. *Stem cell reports* **17**, 1976-1990 (2022).

71. Maehr, R. *et al.* Generation of pluripotent stem cells from patients with type 1 diabetes. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 15768-15773 (2009).

72. Sui, L. *et al.* beta-Cell Replacement in Mice Using Human Type 1 Diabetes Nuclear Transfer Embryonic Stem Cells. *Diabetes* **67**, 26-35 (2018).

73. Kishore, S. *et al.* A Non-Coding Disease Modifier of Pancreatic Agenesis Identified by Genetic Correction in a Patient-Derived iPSC Line. *Cell stem cell* **27**, 137-146.e136 (2020).

74. Ng, N.H.J. *et al.* HNF4A Haploinsufficiency in MODY1 Abrogates Liver and Pancreas Differentiation from Patient-Derived Induced Pluripotent Stem Cells. *iScience* **16**, 192-205 (2019).

75. Kahraman, S. *et al.* Abnormal exocrine-endocrine cell cross-talk promotes β-cell dysfunction and loss in MODY8. *Nat Metab* **4**, 76-89 (2022).

76. Leite, N.C. *et al.* Modeling Type 1 Diabetes In Vitro Using Human Pluripotent Stem Cells. *Cell Rep* **32**, 107894 (2020).

77. Vegas, A.J. *et al.* Long-term glycemic control using polymer-encapsulated human stem cell-derived beta cells in immune-competent mice. *Nat Med* **22**, 306-311 (2016).

78. Lei, J. *et al.* FasL microgels induce immune acceptance of islet allografts in nonhuman primates. *Science advances* **8**, eabm9881 (2022).

79. Bochenek, M.A. *et al.* Alginate encapsulation as long-term immune protection of allogeneic pancreatic islet cells transplanted into the omental bursa of macaques. *Nature biomedical engineering* **2**, 810-821 (2018).

80. Wang, X. *et al.* A nanofibrous encapsulation device for safe delivery of insulin-producing cells to treat type 1 diabetes. *Sci Transl Med* **13** (2021).

81. Stock, A.A. *et al.* Conformal Coating of Stem Cell-Derived Islets for β Cell Replacement in Type 1 Diabetes. *Stem cell reports* **14**, 91-104 (2020).

82. Jiang, K. *et al.* 3-D physiomimetic extracellular matrix hydrogels provide a supportive microenvironment for rodent and human islet culture. *Biomaterials* **198**, 37-48 (2019).

83. Brusko, T.M., Russ, H.A. & Stabler, C.L. Strategies for durable β cell replacement in type 1 diabetes. *Science (New York, N.Y* **373**, 516-522 (2021).

84. Yang, K. *et al.* A therapeutic convection-enhanced macroencapsulation device for enhancing β cell viability and insulin secretion. *Proceedings of the National Academy of Sciences of the United States of America* **118** (2021).

85. Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587 e3529 (2021).

86. Stuart, T., Srivastava, A., Madad, S., Lareau, C.A. & Satija, R. Single-cell chromatin state analysis with Signac. *Nat Methods* **18**, 1333-1341 (2021).

87. Pliner, H.A. *et al.* Cicero Predicts cis-Regulatory DNA Interactions from Single-Cell Chromatin Accessibility Data. *Mol Cell* **71**, 858-871 e858 (2018).

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

89. Fornes, O. *et al.* JASPAR 2020: update of the open-access database of transcription factor binding profiles. *Nucleic acids research* **48**, D87-D92 (2020).

90. Cao, J. *et al.* The single-cell transcriptional landscape of mammalian organogenesis. *Nature* **566**, 496-502 (2019).

**Acknowledgements**

This work was funded by NIH (R01DK114233, R01DK127497), JDRF (5-CDA-2017-391-A-N), and startup funds from the Washington University School of Medicine Department of Medicine. N.J.H. was supported by a JDRF Advanced Postdoctoral Fellowship (3-APF-2020-930-A-N). M.I. was supported by Rita Levi-Montalcini Postdoctoral Fellowship in Regenerative Medicine and the NIH (T32DK007120). M.M.M. was supported by the NIH (T32GM139774). J.R.Miller was supported by a Washington University BioSURF award. D.A.V.P. was supported by the NSF Graduate Research Fellowship Program (DGE-2139839 and DGE-1745038). L.V.C. was supported by the NIH (F31DK125068). Microscopy was performed through the Washington University Center for Cellular Imaging, which is supported by the Washington University School of Medicine, the Children's Discovery Institute (CDI-CORE-2015-505) and the Foundation for Barnes-Jewish Hospital (3770). Microscopy analysis was supported by the Washington University Diabetes Research Center (P30DK020579) and Center of Regenerative Medicine. We thank the Genome Technology Access Center at the McDonnell Genome Institute at Washington University School of Medicine for help with genomic analysis, which is supported by NIH (P30CA91842) support to the Siteman Cancer Center and by ICTS/CTSA (UL1TR002345) from the National Center for Research Resources (NCRR). We thank N. Udomkittivorakul for the graphics design, D. Melton (Harvard University) for the HUES8 cell line, and L. Barret (Broad Institute of MIT and Harvard University) for the CRISPRa VPR cell line.

**Contributions**

P.A. and J.R.Millman designed all experiments and performed all *in vivo* experiments. P.A. performed all computational analysis and associated cell culture. P.A., E.M., M.M.M., M.D.S.,

M.I., D.A.V.P., J.R.Miller, S.E.G., and L.V.C. performed *in vitro* experiments. P.A., N.J.H., M.I., and J.R.Millman wrote the manuscript. All authors revised and approved the manuscript.

**Competing interests**

P.A., N.J.H., L.V.C., and J.R.Millman are inventors on related patents and patent applications. P.A., L.V.C., and J.R.Millman are co-founders of Salentra Biosciences. J.R.Millman is a consultant for Sana Biotechnology. L.V.C. is currently employed by Sana Biotechnology.

**Additional information**

Supplementary Information is available for this paper. Correspondence and requests for materials should be addressed to J.R.Millman Reprints and permissions information is available at www.nature.com/reprints.

**Figure Legends**

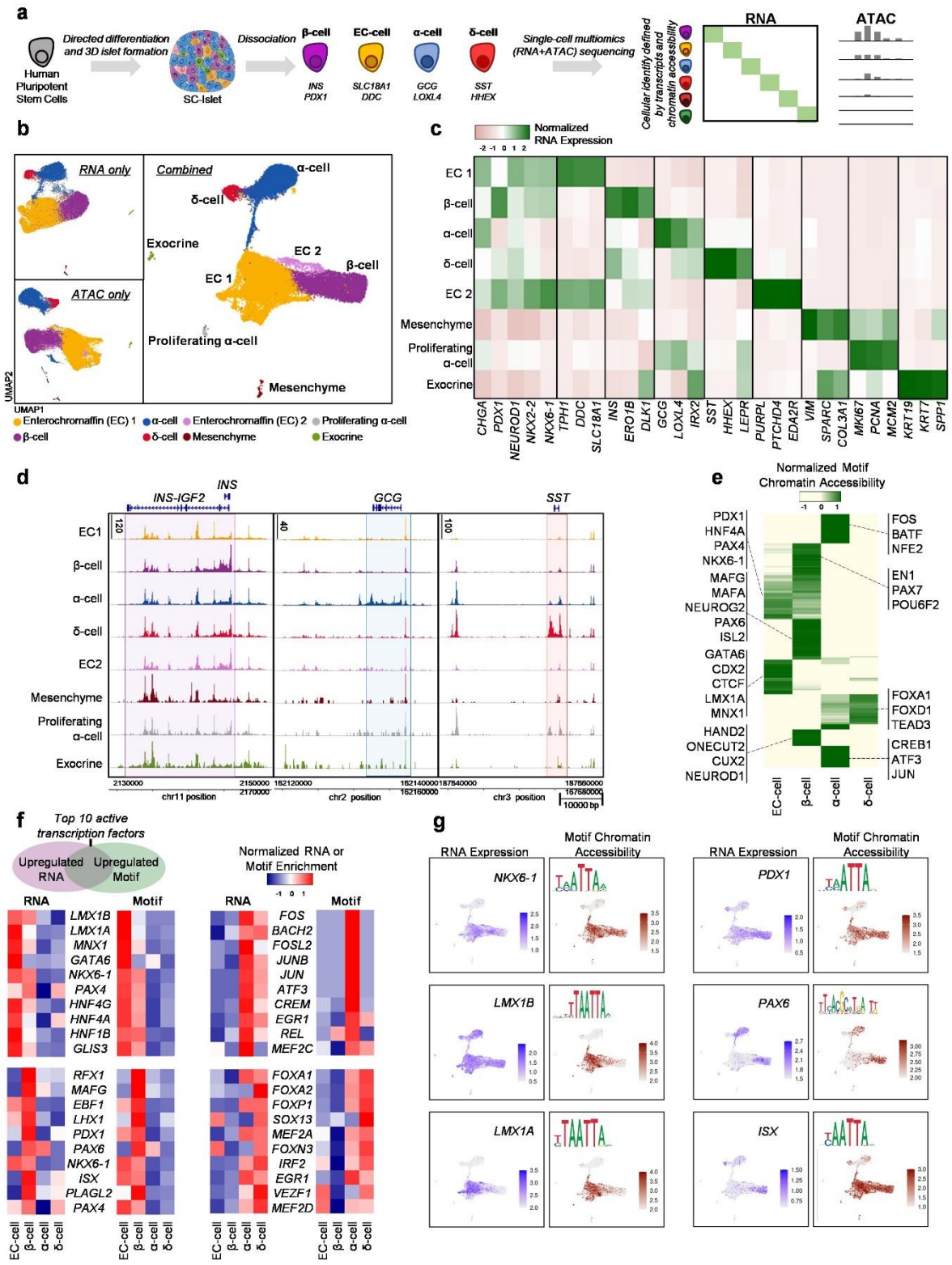



**Fig 1. | Multiomic profiling of stem-cell derived islets shows unique chromatin accessibility signatures in endocrine cell types.** a, Schematic of SC-islet differentiation and multiomic

sequencing. b, UMAPs showing identified cell types in SC-islets 2 weeks into stage 6 using both and either chromatin accessibility (ATAC) or gene (mRNA) information. c, Heatmap showing gene expression of markers associated with each cell type. d, ATAC plots showing chromatin accessibility of SC-islet cell types around the INS, GCG, and SST genomic regions. e, Heatmap showing the top 200 variable DNA-binding motif accessibility within endocrine cell populations and highlighting markers for each cell type. f, Heatmaps highlighting gene expression and ATAC motif accessibility of top 10 active transcription factors co-enriched with both features in SC-β, SC-α, SC-δ, and SC-EC cells. g, UMAP showing gene expression and motif accessibility of selected transcription factors associated with SC-β cells or SC-EC cells. SC, Stem cell derived; EC, enterochromaffin

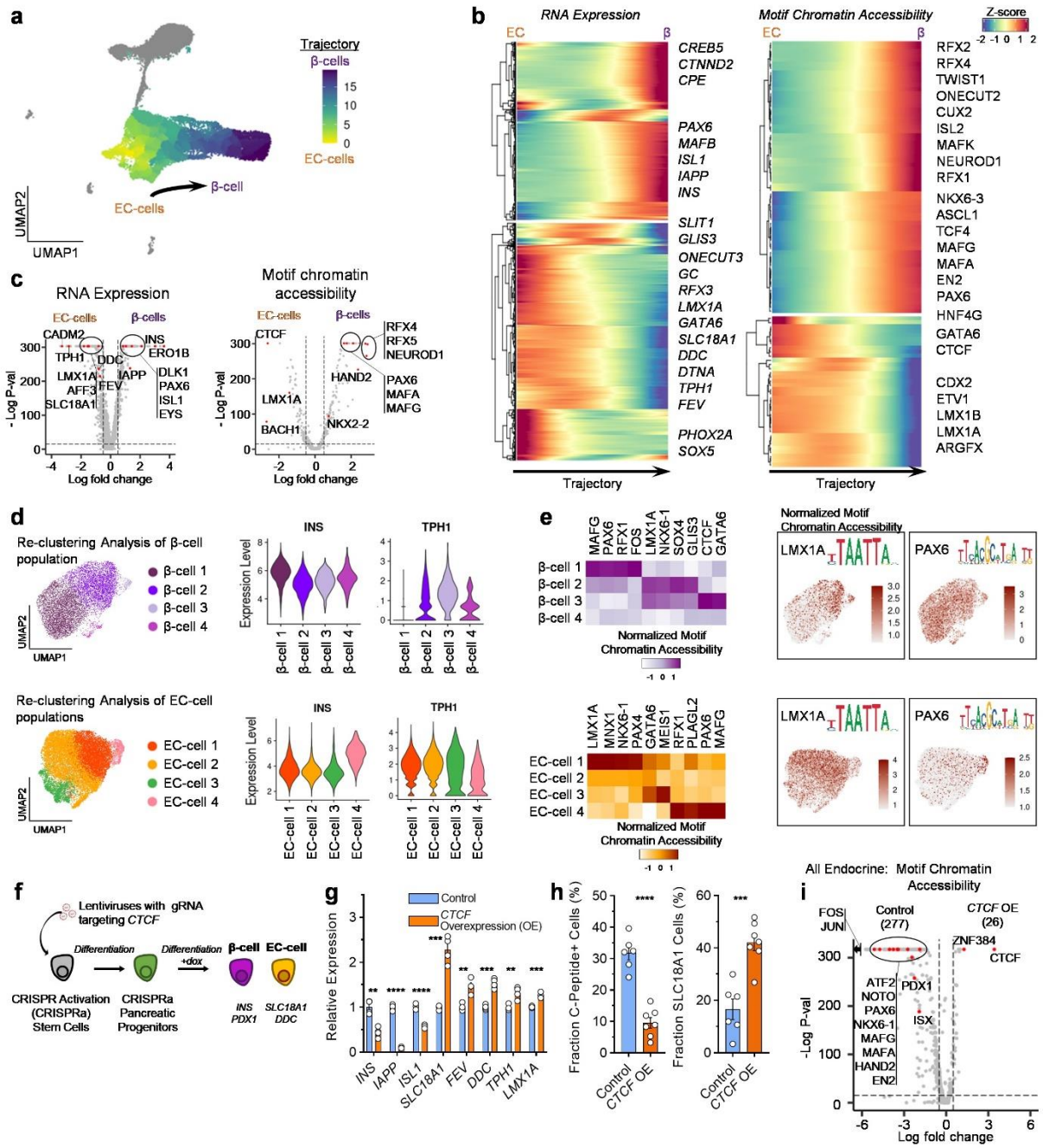



**Fig 2. | SC-EC and SC-β cells have unique and common transcriptional and chromatin accessibility signatures.** a, UMAP showing the trajectory of cells from the SC-β, SC-EC1 and SC-EC2 population. b, Trajectory heatmap showing dynamic changes of gene expression and motif accessibility enriched in β, and EC groups. Selected genes and motifs that are relevant to SC-β, and SC-EC identity and development are highlighted. c, Volcano plots showing

differential gene expression analysis (left) and differential motif accessibility analysis (right), highlighting relevant genes associated with SC-β and SC-EC cell populations. d, UMAP showing sub-populations by reclustering the SC-β population or SC-EC cell populations. Violin plots show gene marker expressions of *INS* and *TPH1*, highlighting the presence of off-target genes in sub-populations. e, Heatmap showing DNA-binding motif accessibility associated with SC-β and SC-EC in subpopulations. Selected transcription factors are plotted to show distribution of cells with target or off-target motif accessibility. f, Schematic of CRISPRa experiment for overexpression of *CTCF* in differentiating pancreatic progenitor cells. g, qPCR analysis of differentiated SC-islets with *CTCF* over expression during endocrine induction, plotting mean ± s.e.m. (n = 4), showing expression differences of genes associated with β-cells (*INS*, P = 0.0010; *IAPP*, P = $1.5 \times 10^{-7}$; *ISL1*, P = $3.0 \times 10^{-5}$) and EC cells (*SLC18A1*, P = $1.6 \times 10^{-4}$; *FEV*, P = 0.0019; *DDC*, P = $1.2 \times 10^{-4}$; *TPH1*, P = 0.0047; *LMX1A*, P = $4.6 \times 10^{-4}$). Statistical significance was assessed by unpaired two-sided t-test. h, Immunocytochemistry quantification of cells expressing C-peptide protein (P = $5.3 \times 10^{-6}$) and SLC18A1 protein (P = $3.1 \times 10^{-4}$) with or without *CTCF* over expression, plotting mean ± s.e.m. (control; n = 6, doxycycline; n = 7). Statistical significance was assessed by unpaired two-sided t-test. i, Volcano plots showing differential motif chromatin accessibility analysis of SC-endocrine population comparing control and *CTCF* over expression. SC, Stem cell derived; EC, enterochromaffin cells; CRISPRa, CRISPR activation.

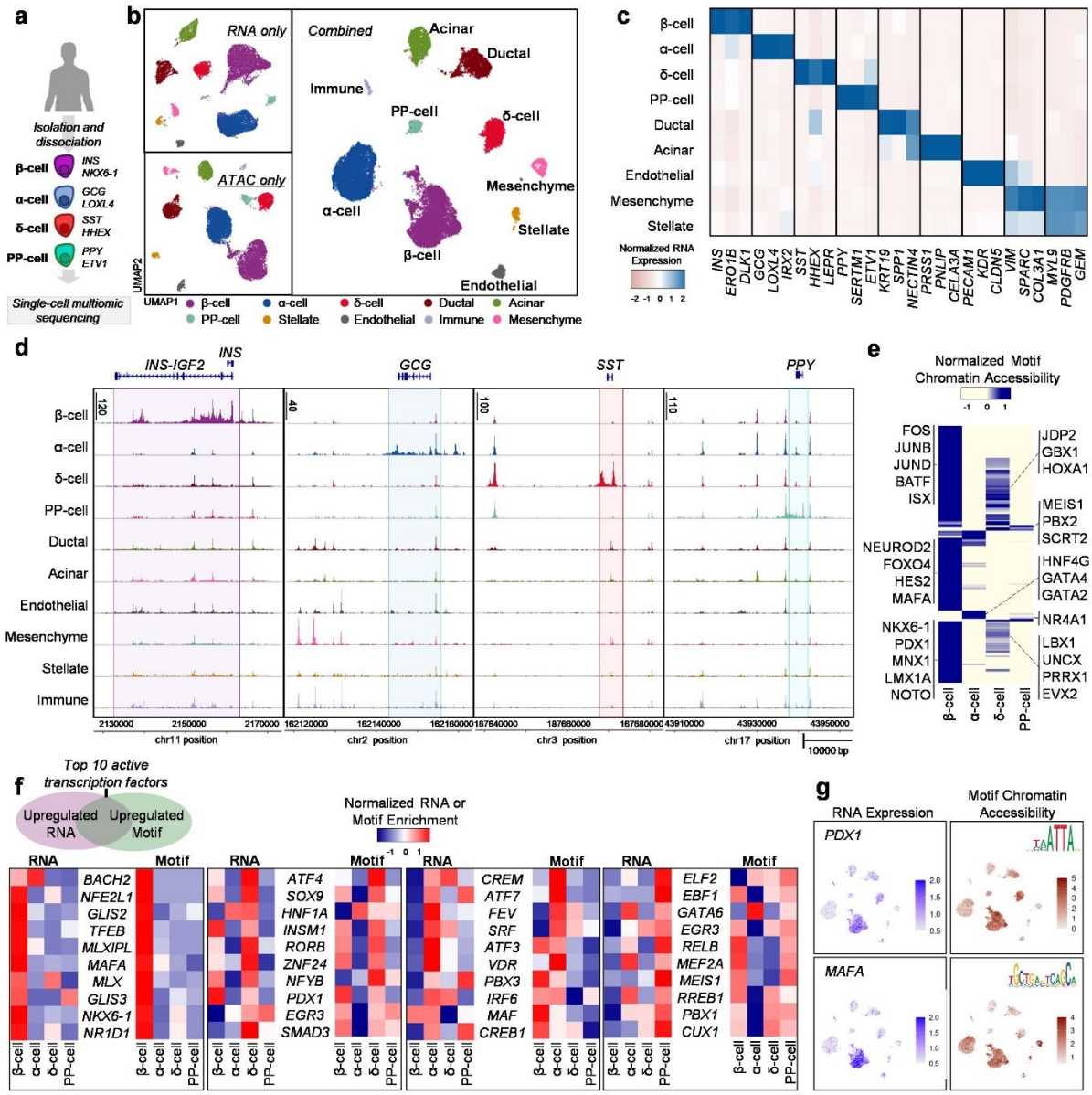



**Fig 3. | Multiomic profiling of human adult primary islets shows unique chromatin accessibility signatures in endocrine cell types.** a, Schematic of human adult primary islets for mutiomic sequencing. b, UMAPs showing identified cell types in primary human islets using both and either chromatin accessibility (ATAC) or gene (mRNA) information. c, Heatmap showing gene expression of markers associated with each cell type. d, ATAC plots showing chromatin accessibility primary islet cell types around the *INS*, *GCG*, *SST*, and *PPY* genomic

regions. e, Heatmap showing the top 200 variable DNA-binding motif accessibility within endocrine cell populations and highlighting motif markers for each cell type. f, Heatmaps highlighting gene expression and ATAC motif accessibility of top 10 active transcription factors co-enriched with both features in primary β, α, δ, and PP-cells. g, UMAP showing gene expression and motif accessibility of selected transcription factors associated with primary β-cells. PP, pancreatic polypeptide.

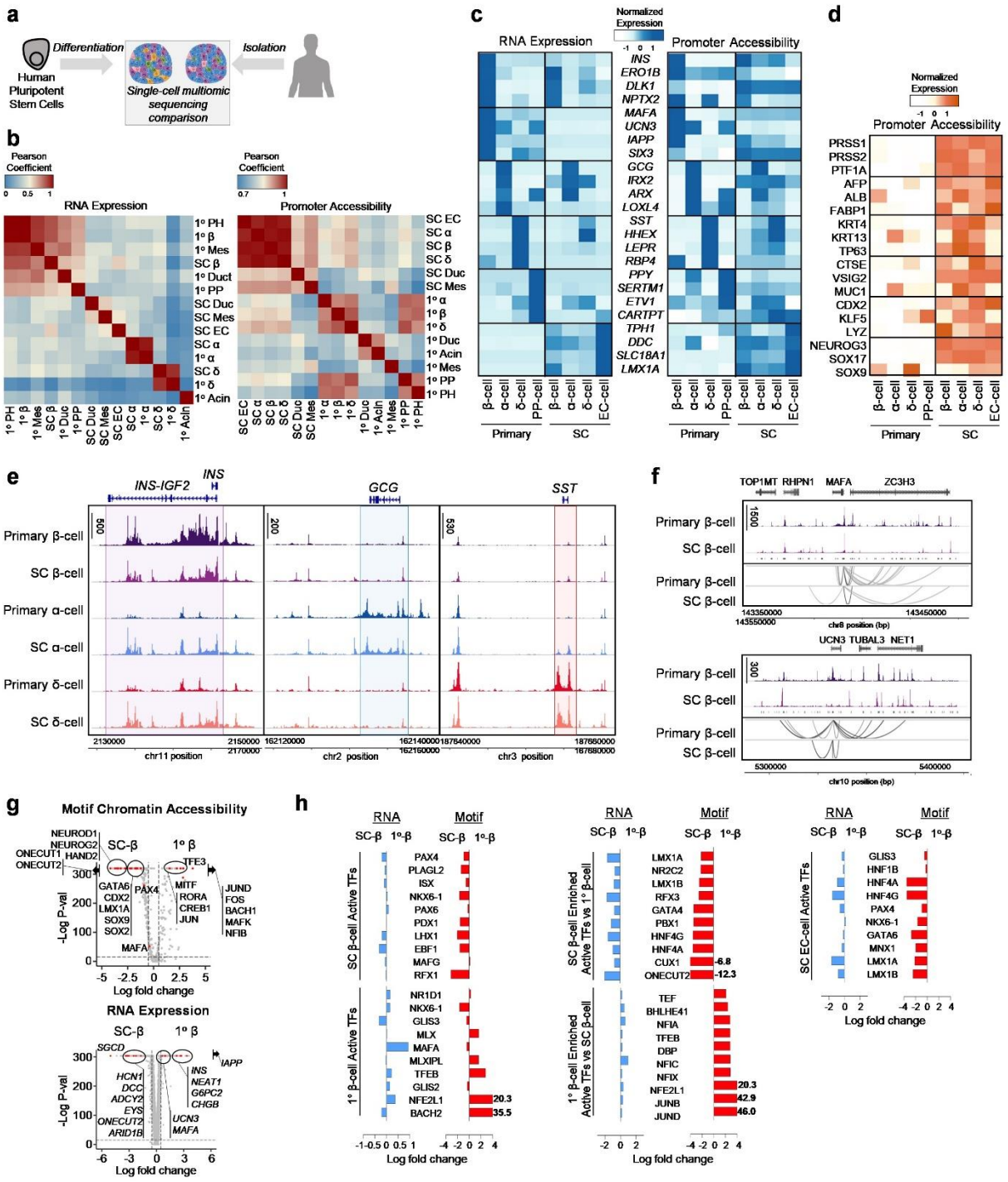



**Fig 4. | Comparative analysis of SC-islets and primary human islets shows differences in chromatin accessibility signatures associated with islet identity.** a, Schematic showing comparisons of SC-islet and primary human islets. b, Pearson correlation analysis comparing cell types in SC-islets and primary islets using gene expression and ATAC promoter accessibility. c,

Heatmap highlighting and comparing identity (β, adult β, α, δ, PP and EC) associated gene expression and ATAC promoter accessibility of endocrine cells in SC-islets and primary islets. d, Heatmap highlighting and comparing off-target identity (Exocrine, hepatic, esophagus, stomach, intestinal, pancreatic progenitor) associated gene expression and ATAC promoter accessibility in SC-islet cells and primary islet cells. e, ATAC plots comparing chromatin accessibility around the *INS*, *GCG*, and *SST* genomic regions in β, α, and δ-cells from SC-islets and primary islets. f, ATAC peaks from SC-β cells and 1° β-cells showing chromatin accessibility around β-cell identity marker, *MAFA* (top) and *UCN3* (bottom). Peaks were linked and analyzed to depict differences in the number of cis-regulatory elements. c, Volcano plots showing differential motif accessibility analysis (top) and differential gene expression analysis (bottom) comparing SC-β cells and primary β-cells. h, Bar graphs showing fold change differences between SC-β cells and primary β-cells, showing gene expression and motif accessibility of identified transcription factors associated with SC-β cells, primary β-cells, and SC-EC cells. SC, stem cell derived; EC, enterochromaffin cells.

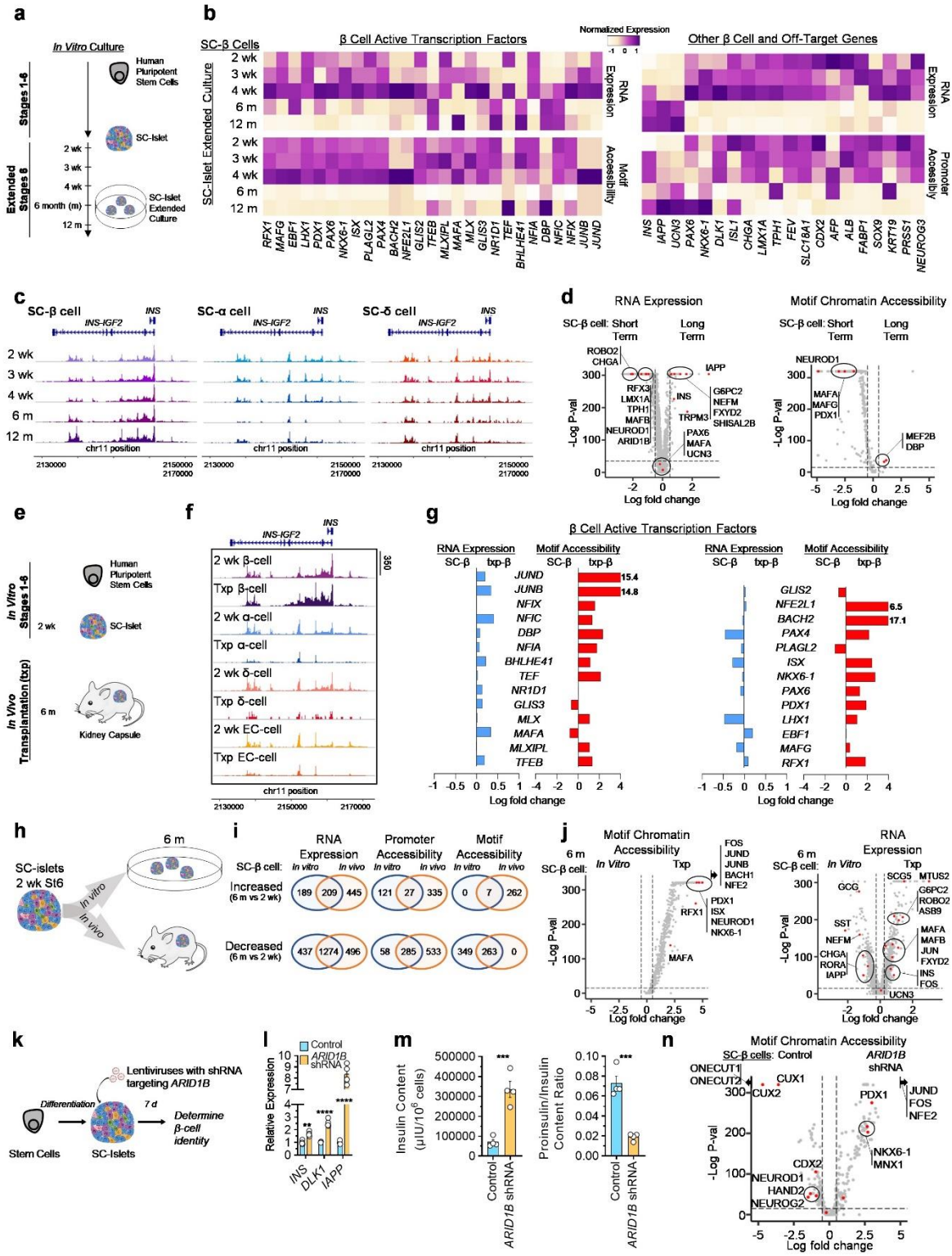
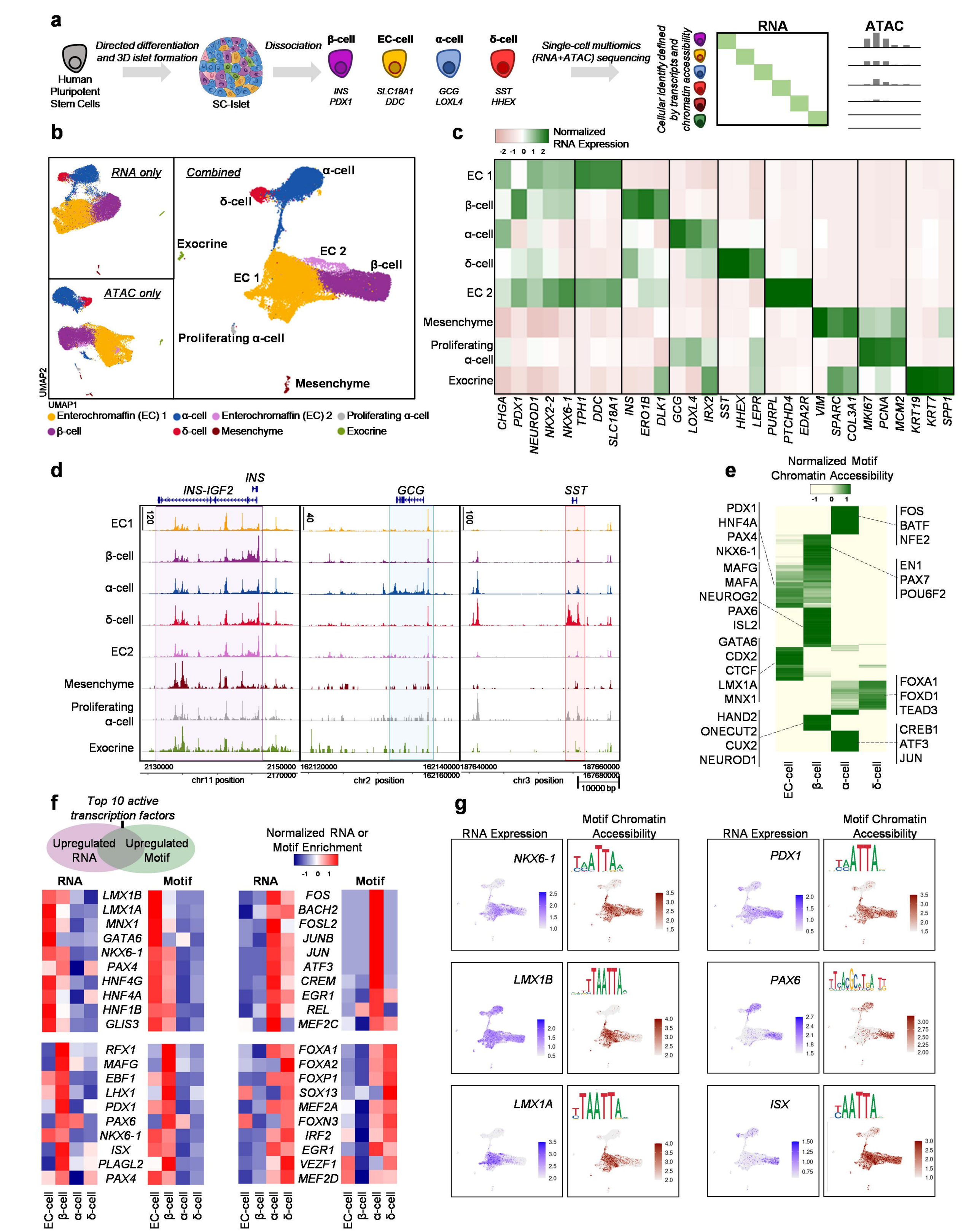
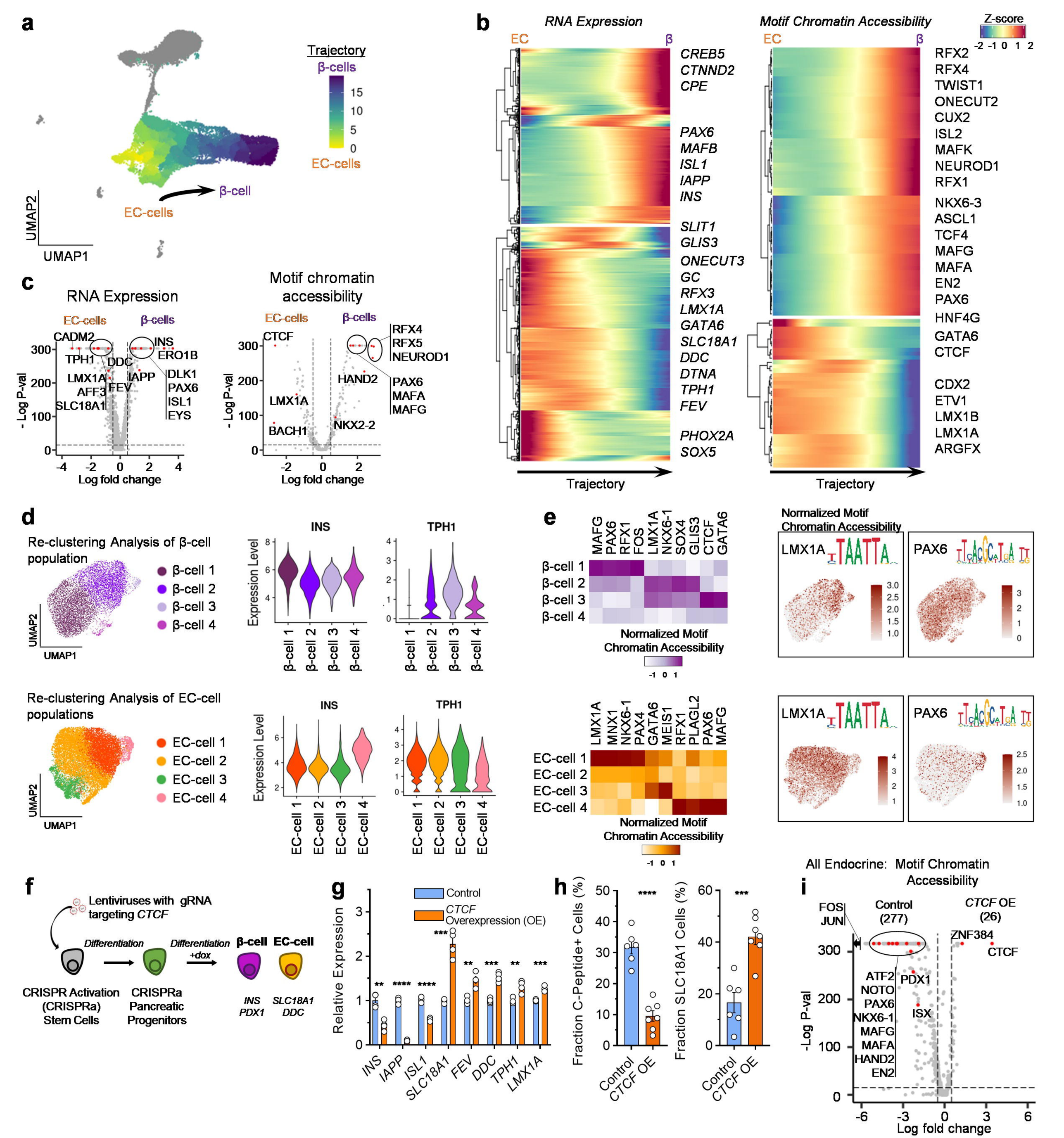


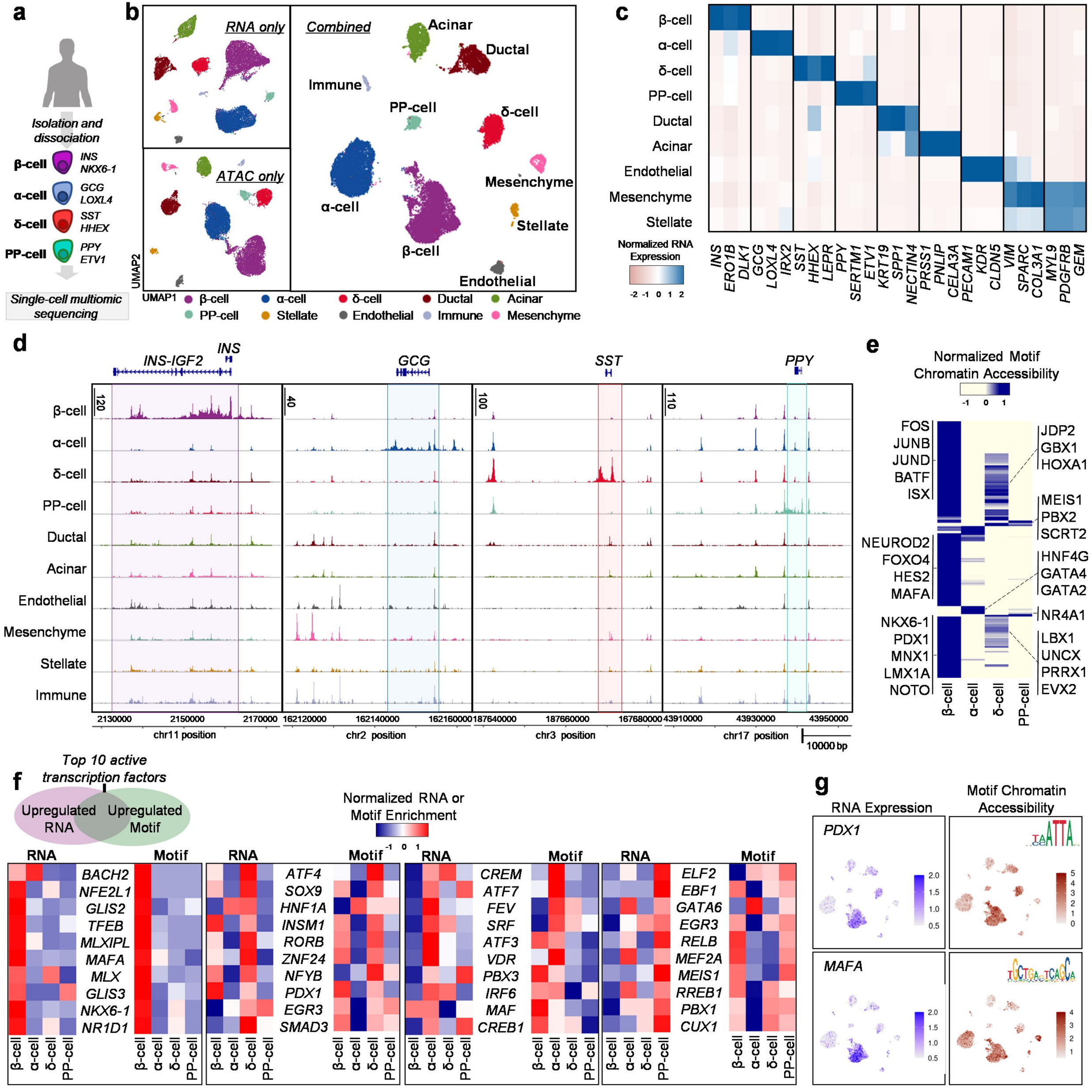

**Fig 5. | Interrogating SC-islet identity with time *in vitro* and after transplantation.** a, Schematic of SC-islet *in vitro* culture with extended time course. b, Heatmap showing gene

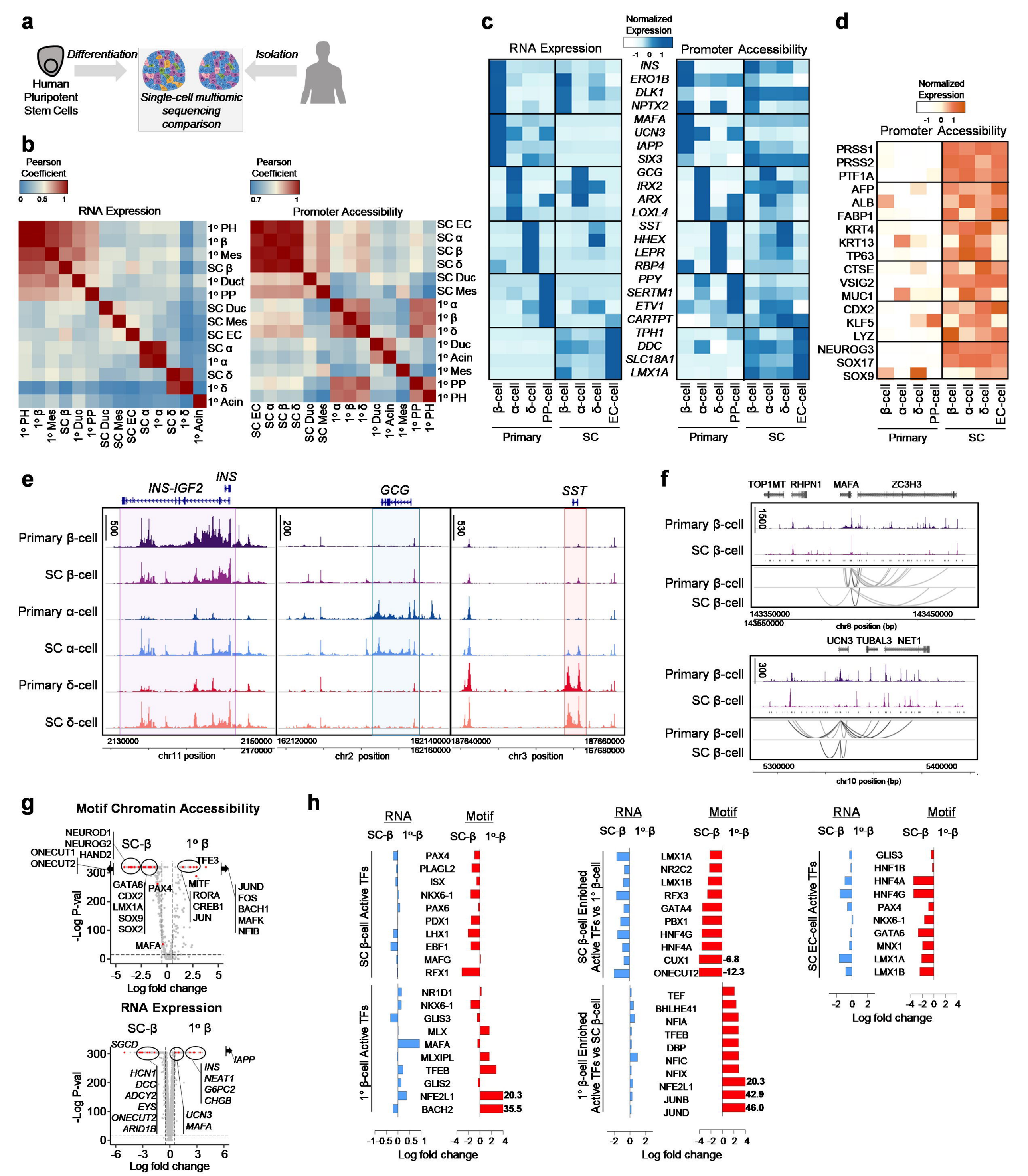
expression, motif chromatin accessibility, or ATAC promoter accessibility of gene markers and transcription factors of SC-β cells cultured *in vitro* with extended time (2, 3, 4 weeks and 6, 12 months). c, Graphs comparing time course of chromatin accessibility around the *INS* genomic region in SC-β, α, and δ cells. Peak signals around the *INS* gene in SC-β cell increases over time, but decreases in SC-α, and δ cells. d, Volcano plots showing differential gene expression analysis (left) and differential motif accessibility analysis (right) comparing SC-β cells cultured short term (week 2, 3 and 4) and SC-β cells cultured long term (month 6 and 12). e, Schematic of differentiated SC-islets maintained long-term by *in vivo* transplantation. f, ATAC chromatin accessibility around the *INS* genomic region showing increase of peak signals in SC-β cells after transplantation and decrease in SC-α and δ cells. g, Bar graphs showing fold change differences of β-cell associated active transcription factors in SC-β cells and transplanted SC-β cells. h, Schematic showing comparison of changes associated with 6 months *in vitro* culture and 6 months *in vivo* transplants. i, Number of genes, by gene expression or ATAC promoter accessibility, or motif chromatin accessibility, downregulated or upregulated in 6 months *in vitro* and *in vivo* SC-β cells. Number of features were determined by differential gene, promoter accessibility or motif accessibility analysis comparing 2 week SC-β cells with 6 months SC-β cells *in vitro* or 6 months SC-β cells *in vivo*. j, Volcano plots showing differential motif accessibility analysis (left) and differential gene expression analysis (right) comparing 6 moths SC-β cells from *in vitro* and *in vivo* SC-islets. k, Schematic of SC-islets transfected with shRNA lentivirus for *ARID1B* gene knockdown. l, qPCR plots of SC-islets with ARID1B shRNA showing mean ± s.e.m. (n = 4) of expression of β-cell associated genes, *INS* (P = 0.0014), *DLK1* (P = $4.3 \times 10^{-5}$), and *IAPP* (P = $3.0 \times 10^{-6}$). Statistical significance was assessed by unpaired two-sided t-test. m, Protein quantification plot showing mean ± s.e.m. (n = 4) of human insulin

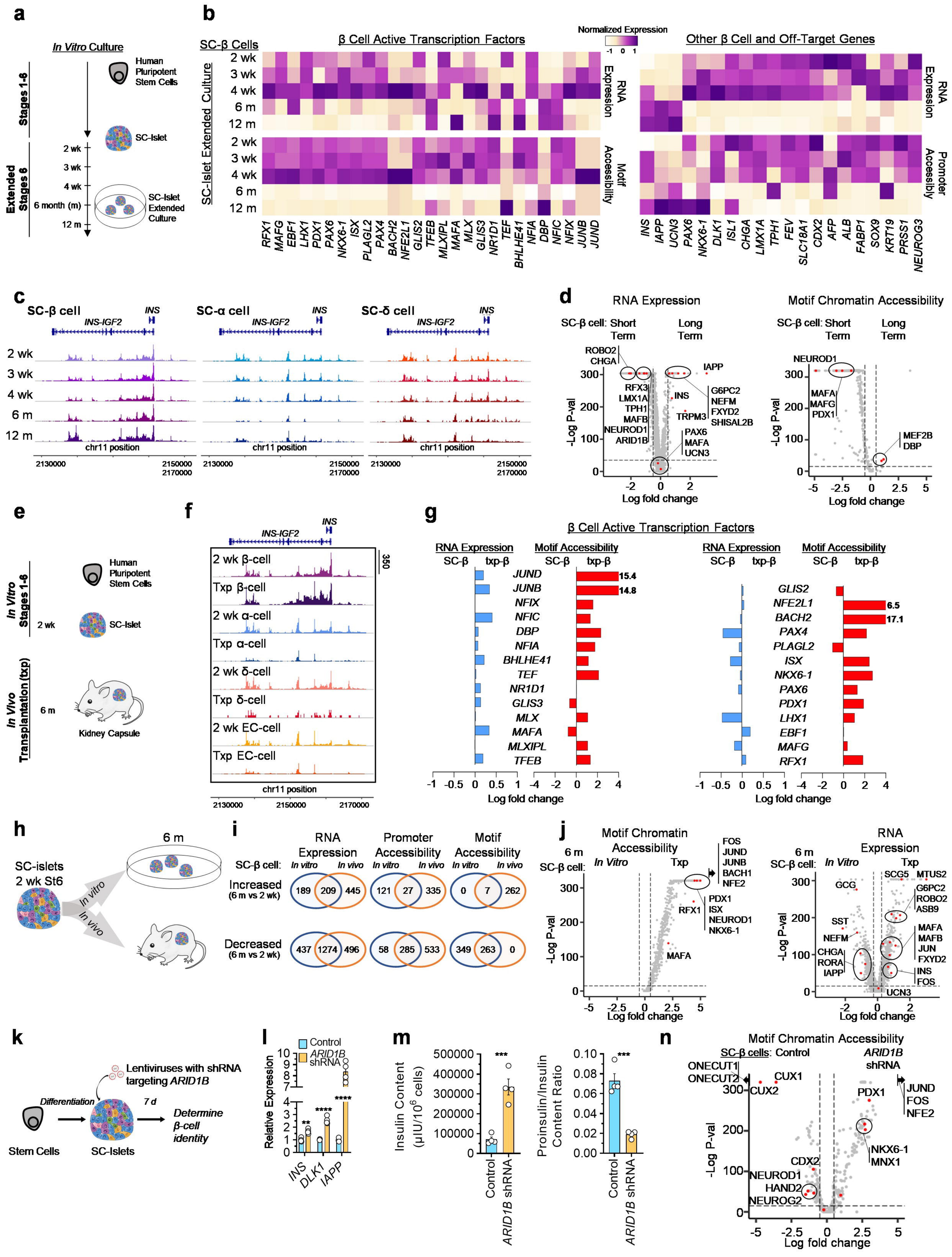
content ($P = 7.6 \times 10^{-4}$) by ELISA (left), and proinsulin/insulin ratio ($P = 3.6 \times 10^{-4}$) by ELISA (right). Statistical significance was assessed by unpaired two-sided t-test. n, Volcano plot from single-cell multiomics comparing motif chromatin accessibility of SC-β cells from control and *ARID1B* shRNA condition. SC, stem cell derived; EC, enterochromaffin cells; ELISA, enzyme-linked immunosorbent assay.


a
Human Pluripotent Stem Cells
Directed differentiation and 3D islet formation
SC-Islet
Dissociation
β-cell
EC-cell
α-cell
δ-cell
INS PDX1
SLC18A1 DDC
GCG LOXL4
SST HHEX
Single-cell multiomics (RNA+ATAC) sequencing
Cellular identify defined by transcripts and chromatin accessibility
RNA
ATAC
b
RNA only
ATAC only
Combined
α-cell
δ-cell
Exocrine
EC 2
β-cell
EC 1
Proliferating α-cell
Mesenchyme
UMAP1
UMAP2
Enterochromaffin (EC) 1
α-cell
Enterochromaffin (EC) 2
Proliferating α-cell
β-cell
δ-cell
Mesenchyme
Exocrine
c
Normalized RNA Expression
-2 -1 0 1 2
EC 1
β-cell
α-cell
δ-cell
EC 2
Mesenchyme
Proliferating α-cell
Exocrine
CHGA PDX1 NEUROD1 NKX2-2 NKX6-1 TPH1 DDC SLC18A1 INS ERO1B DLK1 GCG LOXL4 IRX2 SST HHEX LEPR PURPL PTCHD4 EDA2R VIM SPARC COL3A1 MKI67 PCNA MCM2 KRT19 KRT7 SPP1
d
INS-IGF2
INS
GCG
SST
EC1
β-cell
α-cell
δ-cell
EC2
Mesenchyme
Proliferating α-cell
Exocrine
chr11 position
chr2 position
chr3 position
10000 bp
e
Normalized Motif Chromatin Accessibility
-1 0 1
PDX1
HNF4A
PAX4
NKX6-1
MAFG
MAFA
NEUROG2
PAX6
ISL2
GATA6
CDX2
CTCF
LMX1A
MNX1
HAND2
ONECUT2
CUX2
NEUROD1
FOS
BATF
NFE2
EN1
PAX7
POU6F2
FOXA1
FOXD1
TEAD3
CREB1
ATF3
JUN
EC-cell
β-cell
α-cell
δ-cell
f
Top 10 active transcription factors
Upregulated RNA
Upregulated Motif
Normalized RNA or Motif Enrichment
-1 0 1
RNA
Motif
LMX1B
LMX1A
MNX1
GATA6
NKX6-1
PAX4
HNF4G
HNF4A
HNF1B
GLIS3
RFX1
MAFG
EBF1
LHX1
PDX1
PAX6
NKX6-1
ISX
PLAGL2
PAX4
FOS
BACH2
FOSL2
JUNB
JUN
ATF3
CREM
EGR1
REL
MEF2C
FOXA1
FOXA2
FOXP1
SOX13
MEF2A
FOXN3
IRF2
EGR1
VEZF1
MEF2D
EC-cell
β-cell
α-cell
δ-cell
g
RNA Expression
Motif Chromatin Accessibility
NKX6-1
PDX1
LMX1B
PAX6
LMX1A
ISX


a
Trajectory
β-cells
15
10
5
0
EC-cells
β-cell
EC-cells
UMAP2
UMAP1
b
RNA Expression
Motif Chromatin Accessibility
Z-score
-2 -1 0 1 2
EC
β
CREB5
CTNND2
CPE
PAX6
MAFB
ISL1
IAPP
INS
SLIT1
GLIS3
ONECUT3
GC
RFX3
LMX1A
GATA6
SLC18A1
DDC
DTNA
TPH1
FEV
PHOX2A
SOX5
Trajectory
RFX2
RFX4
TWIST1
ONECUT2
CUX2
ISL2
MAFK
NEUROD1
RFX1
NKX6-3
ASCL1
TCF4
MAFG
MAFA
EN2
PAX6
HNF4G
GATA6
CTCF
CDX2
ETV1
LMX1B
LMX1A
ARGFX
Trajectory
c
RNA Expression
EC-cells
β-cells
CADM2
INS
TPH1
DDC
ERO1B
LMX1A
IAPP
DLK1
AFF3
FEV
PAX6
SLC18A1
ISL1
EYS
- Log P-val
Log fold change
Motif chromatin accessibility
EC-cells
β-cells
CTCF
RFX4
RFX5
NEUROD1
HAND2
PAX6
MAFA
MAFG
LMX1A
BACH1
NKX2-2
- Log P-val
Log fold change
d
Re-clustering Analysis of β-cell population
β-cell 1
β-cell 2
β-cell 3
β-cell 4
UMAP2
UMAP1
INS
TPH1
Expression Level
β-cell 1
β-cell 2
β-cell 3
β-cell 4
Re-clustering Analysis of EC-cell populations
EC-cell 1
EC-cell 2
EC-cell 3
EC-cell 4
UMAP2
UMAP1
INS
TPH1
Expression Level
EC-cell 1
EC-cell 2
EC-cell 3
EC-cell 4
e
MAFG
PAX6
RFX1
FOS
LMX1A
NKX6-1
SOX4
GLIS3
CTCF
GATA6
β-cell 1
β-cell 2
β-cell 3
β-cell 4
Normalized Motif Chromatin Accessibility
-1 0 1
Normalized Motif Chromatin Accessibility
LMX1A
PAX6
3.0
2.5
2.0
1.5
1.0
LMX1A
MNX1
NKX6-1
PAX4
GATA6
MEIS1
RFX1
PLAGL2
PAX6
MAFG
EC-cell 1
EC-cell 2
EC-cell 3
EC-cell 4
Normalized Motif Chromatin Accessibility
-1 0 1
LMX1A
PAX6
f
Lentiviruses with gRNA targeting CTCF
Differentiation
Differentiation +dox
β-cell
EC-cell
CRISPR Activation (CRISPRa) Stem Cells
CRISPRa Pancreatic Progenitors
INS
PDX1
SLC18A1
DDC
g
Control
CTCF Overexpression (OE)
Relative Expression
INS
IAPP
ISL1
SLC18A1
FEV
DDC
TPH1
LMX1A
h
Fraction C-Peptide+ Cells (%)
Control
CTCF OE
Fraction SLC18A1 Cells (%)
Control
CTCF OE
i
All Endocrine: Motif Chromatin Accessibility
FOS
JUN
Control (277)
CTCF OE (26)
ZNF384
CTCF
ATF2
NOTO
PAX6
NKX6-1
MAFG
MAFA
HAND2
EN2
PDX1
ISX
-Log P-val
Log fold change


a
Isolation and dissociation
β-cell INS NKX6-1
α-cell GCG LOXL4
δ-cell SST HHEX
PP-cell PPY ETV1
Single-cell multiomic sequencing
b
RNA only
ATAC only
Combined
Acinar
Ductal
Immune
PP-cell
δ-cell
α-cell
Mesenchyme
Stellate
β-cell
Endothelial
UMAP1
UMAP2
β-cell α-cell δ-cell Ductal Acinar
PP-cell Stellate Endothelial Immune Mesenchyme
c
β-cell
α-cell
δ-cell
PP-cell
Ductal
Acinar
Endothelial
Mesenchyme
Stellate
Normalized RNA Expression
-2 -1 0 1 2
INS ERO1B DLK1 GCG LOXL4 IRX2 SST HHEX LEPR PPY SERTM1 ETV1 KRT19 SPP1 NECTIN4 PRSS1 PNLIP CELA3A PECAM1 KDR CLDN5 VIM SPARC COL3A1 MYL9 PDGFRB GEM
d
INS-IGF2
INS
GCG
SST
PPY
β-cell
α-cell
δ-cell
PP-cell
Ductal
Acinar
Endothelial
Mesenchyme
Stellate
Immune
120
40
100
110
2130000 2150000 2170000
162120000 162140000 162160000
187640000 187660000 167680000
43910000 43930000 43950000
chr11 position
chr2 position
chr3 position
chr17 position
10000 bp
e
Normalized Motif Chromatin Accessibility
-1 0 1
FOS
JUNB
JUND
BATF
ISX
NEUROD2
FOXO4
HES2
MAFA
NKX6-1
PDX1
MNX1
LMX1A
NOTO
JDP2
GBX1
HOXA1
MEIS1
PBX2
SCRT2
HNF4G
GATA4
GATA2
NR4A1
LBX1
UNCX
PRRX1
EVX2
β-cell α-cell δ-cell PP-cell
f
Top 10 active transcription factors
Upregulated RNA
Upregulated Motif
Normalized RNA or Motif Enrichment
-1 0 1
RNA
Motif
BACH2 NFE2L1 GLIS2 TFEB MLXIPL MAFA MLX GLIS3 NKX6-1 NR1D1
ATF4 SOX9 HNF1A INSM1 RORB ZNF24 NFYB PDX1 EGR3 SMAD3
CREM ATF7 FEV SRF ATF3 VDR PBX3 IRF6 MAF CREB1
ELF2 EBF1 GATA6 EGR3 RELB MEF2A MEIS1 RREB1 PBX1 CUX1
β-cell α-cell δ-cell PP-cell
g
RNA Expression
Motif Chromatin Accessibility
PDX1
MAFA
2.0 1.5 1.0 0.5
5 4 3 2 1 0
4 3 2 1


a
Human Pluripotent Stem Cells
Differentiation
Single-cell multiomic sequencing comparison
Isolation
b
Pearson Coefficient
RNA Expression
Promoter Accessibility
c
Normalized Expression
RNA Expression
Promoter Accessibility
INS
ERO1B
DLK1
NPTX2
MAFA
UCN3
IAPP
SIX3
GCG
IRX2
ARX
LOXL4
SST
HHEX
LEPR
RBP4
PPY
SERTM1
ETV1
CARTPT
TPH1
DDC
SLC18A1
LMX1A
Primary
SC
d
Normalized Expression
Promoter Accessibility
PRSS1
PRSS2
PTF1A
AFP
ALB
FABP1
KRT4
KRT13
TP63
CTSE
VSIG2
MUC1
CDX2
KLF5
LYZ
NEUROG3
SOX17
SOX9
Primary
SC
e
INS-IGF2
INS
GCG
SST
Primary β-cell
SC β-cell
Primary α-cell
SC α-cell
Primary δ-cell
SC δ-cell
chr11 position
chr2 position
chr3 position
f
TOP1MT RHPN1 MAFA ZC3H3
Primary β-cell
SC β-cell
chr8 position (bp)
UCN3 TUBAL3 NET1
chr10 position (bp)
g
Motif Chromatin Accessibility
SC-β
1° β
NEUROD1
NEUROG2
ONECUT1
HAND2
ONECUT2
GATA6
CDX2
LMX1A
SOX9
SOX2
PAX4
MAFA
MITF
RORA
CREB1
JUN
TFE3
JUND
FOS
BACH1
MAFK
NFIB
-Log P-val
Log fold change
RNA Expression
SGCD
HCN1
DCC
ADCY2
EYS
ONECUT2
ARID1B
INS
NEAT1
G6PC2
CHGB
UCN3
MAFA
IAPP
h
RNA
Motif
SC-β
1°-β
SC β-cell Active TFs
PAX4
PLAGL2
ISX
NKX6-1
PAX6
PDX1
LHX1
EBF1
MAFG
RFX1
1° β-cell Active TFs
NR1D1
NKX6-1
GLIS3
MLX
MAFA
MLXIPL
TFEB
GLIS2
NFE2L1
BACH2
20.3
35.5
SC β-cell Enriched Active TFs vs 1° β-cell
LMX1A
NR2C2
LMX1B
RFX3
GATA4
PBX1
HNF4G
HNF4A
CUX1
ONECUT2
-6.8
-12.3
1° β-cell Enriched Active TFs vs SC β-cell
TEF
BHLHE41
NFIA
TFEB
DBP
NFIC
NFIX
NFE2L1
JUNB
JUND
20.3
42.9
46.0
SC EC-cell Active TFs
GLIS3
HNF1B
HNF4A
HNF4G
PAX4
NKX6-1
GATA6
MNX1
LMX1A
LMX1B
Log fold change


a
In Vitro Culture
Stages 1-6
Human Pluripotent Stem Cells
SC-Islet
Extended Stages 6
2 wk
3 wk
4 wk
6 month (m)
12 m
SC-Islet Extended Culture
b
SC-β Cells
β Cell Active Transcription Factors
Normalized Expression
Other β Cell and Off-Target Genes
SC-Islet Extended Culture
RNA Expression
Motif Accessibility
Promoter Accessibly
RFX1 MAFG EBF1 LHX1 PDX1 PAX6 NKX6-1 ISX PLAGL2 PAX4 BACH2 NFE2L1 GLIS2 TFEB MLXIPL MAFA MLX GLIS3 NR1D1 TEF BHLHE41 NFIA DBP NFIC NFIX JUNB JUND
INS IAPP UCN3 PAX6 NKX6-1 DLK1 ISL1 CHGA LMX1A TPH1 FEV SLC18A1 CDX2 AFP ALB FABP1 SOX9 KRT19 PRSS1 NEUROG3
c
SC-β cell
SC-α cell
SC-δ cell
INS-IGF2
INS
chr11 position
2130000
2150000
2170000
d
RNA Expression
Motif Chromatin Accessibility
SC-β cell: Short Term
Long Term
-Log P-val
Log fold change
ROBO2
CHGA
RFX3
LMX1A
TPH1
MAFB
NEUROD1
ARID1B
IAPP
INS
TRPM3
G6PC2
NEFM
FXYD2
SHISAL2B
PAX6
MAFA
UCN3
NEUROD1
MAFA
MAFG
PDX1
MEF2B
DBP
e
In Vitro Stages 1-6
2 wk
Human Pluripotent Stem Cells
SC-Islet
In Vivo Transplantation (txp)
6 m
Kidney Capsule
f
INS-IGF2
INS
2 wk β-cell
Txp β-cell
2 wk α-cell
Txp α-cell
2 wk δ-cell
Txp δ-cell
2 wk EC-cell
Txp EC-cell
350
chr11 position
2130000
2150000
2170000
g
β Cell Active Transcription Factors
RNA Expression
Motif Accessibility
SC-β
txp-β
JUND
JUNB
NFIX
NFIC
DBP
NFIA
BHLHE41
TEF
NR1D1
GLIS3
MLX
MAFA
MLXIPL
TFEB
15.4
14.8
GLIS2
NFE2L1
BACH2
PAX4
PLAGL2
ISX
NKX6-1
PAX6
PDX1
LHX1
EBF1
MAFG
RFX1
6.5
17.1
Log fold change
h
6 m
SC-islets
2 wk St6
In vitro
In vivo
i
RNA Expression
Promoter Accessibility
Motif Accessibility
SC-β cell:
In vitro
In vivo
Increased (6 m vs 2 wk)
189 209 445
121 27 335
0 7 262
Decreased (6 m vs 2 wk)
437 1274 496
58 285 533
349 263 0
j
Motif Chromatin Accessibility
RNA Expression
6 m
SC-β cell:
In Vitro
Txp
-Log P-val
Log fold change
FOS
JUND
JUNB
BACH1
NFE2
PDX1
ISX
NEUROD1
NKX6-1
RFX1
MAFA
SCG5
MTUS2
GCG
G6PC2
ROBO2
ASB9
SST
NEFM
CHGA
RORA
IAPP
MAFA
MAFB
JUN
FXYD2
INS
FOS
UCN3
k
Lentiviruses with shRNA targeting ARID1B
Differentiation
7 d
Stem Cells
SC-Islets
Determine β-cell identity
l
Control
ARID1B shRNA
Relative Expression
INS
DLK1
IAPP
m
Insulin Content (μIU/10^6 cells)
Proinsulin/Insulin Content Ratio
Control
ARID1B shRNA
n
Motif Chromatin Accessibility
SC-β cells: Control
ARID1B shRNA
ONECUT1
ONECUT2
CUX1
CUX2
PDX1
JUND
FOS
NFE2
NKX6-1
MNX1
CDX2
NEUROD1
HAND2
NEUROG2
-Log P-val
Log fold change