bioRxiv preprint doi: https://doi.org/10.1101/2022.02.26.482126; this version posted February 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

# Multiomic profiling defines cell fate plasticity of in vitro-derived islets

Punn Augsornworawat[1,2], Erica Marquez[1], Marlie M. Maestas[1], Matthew Ishahak[1], Sarah E. Gale[1], Mason D. Schmidt[1], Daniel A. Veronese-Paniagua[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

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.26.482126; this version posted February 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.


## Abstract

Insulin-secreting β-cells within the pancreatic islets of Langerhans are lost in patients with type 1 diabetes[1]. Islets can be generated in vitro by differentiation of human pluripotent stem cells[2-5], but understanding of the molecular events governing this process is limited. Here, we use single-cell multiomics to measure chromatin accessibility and transcriptional profiles of 120,064 cells to characterize and delineate maturation of islets from in vitro differentiation. We find distinguishing chromatin accessibility and gene expression signatures as well as dynamic profiles for each major islet cell type produced from in vitro differentiation. Furthermore, based on chromatin accessibility, β, α, and δ cells from in vitro differentiation are more ambiguous in their cellular identity than from isolated primary islets. However, extended time in vitro or with transplantation into mice drives more distinct states of chromatin accessibility for each cell type. Modulation of *CTCF* expression redirects cell fate from pancreatic islet endocrine to an intestinal enteroendocrine-like cell type. Additionally, knockdown of *ARID1B* enhances β-cell maturation transcriptionally and by chromatin accessibility. These results provide a comprehensive atlas and insights into cell fate identity and maturation of stem cell-derived islets, which will inform on their utility for therapy and disease modeling[6-8].

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.26.482126; this version posted February 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Human pluripotent stem cell-derived islets (SC-islets) containing stem cell-derived β (SC-β), α (SC-α), and δ (SC-δ) cells among other cell types have been differentiated in vitro[2-5]. This lab-grown tissue is of current immense interest due to their utility in diabetes cell replacement therapy and disease modeling[6-9]. These tissues can be generated using specific combinations of small molecules and growth factors[3,5,10], microenvironmental cues[2], and other sorting and aggregation[4,11-13] approaches. SC-islets are capable of secreting insulin and improving glucose homeostasis in mice[14,15]. However, a challenge with in vitro differentiation is that the resulting cells are immature and uncontrollably heterogeneous in composition, including the presence of off-target cell types such as enterochromaffin-like cells (SC-EC cells)[11].

Gene expression and chromatin accessibility are important in cellular differentiation[16-19]. Single-cell RNA sequencing has been used to characterize SC-islets and found that SC-β cells co-express insulin along with some, but not all, other transcripts that are expected of mature β-cells[11,15,20,21]. Bulk chromatin approaches have shown differences in a large number of genomic regions in various pancreatic cell differentiation contexts[22,23], but studies have not comprehensively indexed the chromatin accessibility and transcriptional signatures of each in vitro-differentiated SC-islet cell type. Particularly in terms of chromatin accessibility, the field lacks information on how similar each SC-islet cell type is to those from primary human islets, how plastic this in vitro-derived cell fate is, and what the role of chromatin regulators is in cell fate dynamics.

Given the immediate scientific and medical importance of SC-islets for diabetes, we utilize single-cell multiomic sequencing measuring gene expression and chromatin accessibility of in vitro-differentiated SC-islets to deeply index these molecular features for each cell type (Fig. 1a). We leverage our protocol[24] that differentiates human embryonic stem cells (hESCs)

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.26.482126; this version posted February 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

through 6 defined stages containing specific temporal combinations of small molecules and growth factors to produce SC-islets containing SC-β cells, SC-α cells, SC-δ cells, and other cell types (Extended Data Fig. 1a, Supplementary Table 1). When we apply these approaches, distinguished chromatin accessibility and gene expression signatures emerge delineating differential cell fates (Fig. 1b, Extended Data Fig. 1b).

**Multiomics identifies SC-islet cell fates**

Utilizing both gene expression and chromatin accessibility information, we performed clustering and identified specific islet cell identities, including SC-β, SC-α, and SC-δ cells along with the off-target population SC-EC cells[11,14,15] (Fig. 1b). Using this methodology, we also identified a new cell population that possess features of both SC-β and SC-EC cells but is discrete as its own identity, which we have designated as enterochromaffin-β like cells (EC-β cells) (Fig. 1b, Extended Data Fig. 1b-d). Differential gene expression analysis showed that the upregulation of EC-β associated genes is not represented in other endocrine population (Extended Data Fig. 1c,d). Gene set enrichment analysis of the other major gene markers validated the assignment of these identities (Extended Data Fig. 1d). The contiguous clustering of SC-β, SC-EC, and EC-β populations suggested their resemblance, whereas the SC-α and SC-δ populations formed separate clusters (Fig. 1b). We assessed the chromatin information by comparing the chromatin accessibility of the *INS*, *GCG*, and *SST* gene regions in each cell types. While the SC-β cells, SC-α cells, and SC-δ cells had strong chromatin peaks represented by their respective *INS*, *GCG*, and *SST* gene regions, there was to some extent shared detection of these peak across all endocrine cell types, including SC-EC and EC-β cells (Extended Data Fig. 1e,f).

We performed chromatin accessibility motif enrichment analysis and found distinct motif

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.26.482126; this version posted February 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

groups enriched in each specific cell type (Fig. 1c, Extended Data Fig. 1g,h). Within the endocrine population, SC-β cells had enrichment of motifs that correspond to β-cell associated transcription factors[11,21,25,26], such as MAFA, HNFIB, NEUROG2, PAX6, HAND2, ONECUT2, CUX2, and NEUROD1. SC-α cells had motif enrichment of FOS, BATF, CREB1, ATF3, and JUN, while SC-δ cells have motif enrichment of FOXA1, FOXD1, and TEAD3. Motifs that were specifically enriched in SC-EC cells include GATA6, CDX2, CTCF, and LMX1A, which were of notable interest as some of these transcription factors are associated with intestinal lineages[27,28]. Interestingly, SC-β cells and SC-EC cells possessed shared motifs enrichment, such as PDX1, PAX4, and NKX6-1, which were reported to be important in human primary β-cells[29]. This suggests the possibility that these in vitro-derived SC-EC cells differ substantially from their in vivo cell counterparts. Collectively, the ATAC profiles supports that SC-islet endocrine cells contain less distinctive chromatin states through detection of peaks, such as in the *INS*, *GCG*, and *SST* regions, but are committed in their lineages through motif enrichment.

**Islet endocrine cell fate is plastic**

Multiomics sequencing of SC-islets with extended time during the terminal stage (Stage 6) of the protocol revealed that SC-islet endocrine cell fate is dynamic with time (Fig. 1d). All key SC-islet endocrine cell types (SC-β, SC-α, SC-δ, and SC-EC cells) were identified in all timepoints during 4 weeks of Stage 6 (Extended Data Fig. 2a,b). Integrated clustering analysis of all timepoints showed that the assay for transposase-accessible chromatin (ATAC) dependent analysis of endocrine populations were more homogenous than the RNA dependent analysis (Extended Data Fig. 2c,d). Focusing on the 2000 most variable promoter accessibility regions, changes were observed in the chromatin accessibility profile of the endocrine populations over

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.26.482126; this version posted February 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

time (Extended Data Fig. 2e). Employing the top ATAC promoter accessibility regions for each cell types, we found that the SC-β, SC-α, SC-δ, and SC-EC cells have increased enrichment in on-target and decreased enrichment in off-target motifs (Fig. 1d,e, Extended Data Fig. 2f). This data illustrates that the chromatin accessibility signatures for each SC-islet endocrine cell type changes with time in culture, showing they have increased commitment to their respective cell fate while reducing off-target characteristics.

**EC-cell and β cell delineation**

While in vitro differentiation-derived SC-β and SC-EC cell populations form clusters in UMAPs that are adjacent, the differences in their identities are of interest since they normally form from distant pancreatic and intestinal origins in normal development in vivo[11,30,31]. Here, we performed pseudotime ordering analysis to delineate differences between SC-β cells and SC-EC cells in both gene expression and chromatin accessibility contexts (Fig. 2a). SC-EC cell identity genes, such as *SLC18A1*, *TPH1*, and *FEV*, are explicitly expressed in the early pseudotime segment of the SC-EC population (Fig. 2b, Extended Data Fig. 3a,b). Expression was dynamically shifted towards β-cell identity genes through pseudotime, which included *INS*, *PAX6*, and *ISL1* (Fig. 2b, Extended Data Fig. 3a,b). Differential chromatin accessibility motif enrichment analysis depicted LMX1A, GATA6, and CTCF motifs to be prominent for SC-EC cell identity (Fig. 2c,d, Extended Data Fig. 3d). Conversely, NEUROD1, HAND2, MAFA, PAX6, ONECUT2, and RFX2 motifs were prominent for β-cell identity (Fig. 2c, Extended Data Fig. 3d). These analyses catalogue and highlight the key gene expression and motif signatures that pertain to SC-EC and SC-β cell types, illustrating that many cells exist with chromatin accessibility profiles that are common with both cell types.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.26.482126; this version posted February 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

**Tissue lineage remodeled by *CTCF***

CCCTC-binding factor (CTCF) is a zinc finger transcription factor that has been reported to play numerous roles in various developmental and differentiation processes[32]. In our multiomic analysis of SC-EC and SC-β cells, we found CTCF to be one of the most differentially enriched chromatin accessibility motifs in the SC-EC population (Fig. 2c). Furthermore, the SC-β cell population had low enrichment of CTCF motifs when compared to other cell types (Fig. 2d). As the role of CTCF has not been explored in this context, we sought to further examine the impact of CTCF on human pancreatic and intestinal endocrine differentiation. We utilized a doxycycline inducible VP64-p65-Rta CRISPR activation (CRISPRa) stem cell line[33] to enhance the transcription and expression of *CTCF* during endocrine induction (Stage 5) of our differentiation protocol from pancreatic progenitors (Extended Data Fig. 1a, Extended Data Fig. 4a). 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. 4b,c). Upregulation of *CTCF* during the induction of endocrine cells resulted in drastic reductions in expression of β-cell identity genes, such as *INS* and *ISL1* (Extended Data Fig. 4d-f). We also observed significant upregulation in intestinal/enteroendocrine cell lineage associated genes, such as *SLC18A1*, *FEV*, and *CHGA*.

Single-cell multiomic sequencing further elucidated the effects of *CTCF* modulation on cell fate plasticity, showing major shifts in pancreatic islet endocrine identity and intestinal enteroendocrine-like cell identity (Fig. 2e, Extended Data Fig. 5a). Specifically, SC-β cell identity was greatly diminished, and surprisingly there was an increase in the appearance of a stem cell-derived intestinal enteroendocrine-like cell (SC-EE cell) population. Our analyses

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.26.482126; this version posted February 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

showed that these SC-EE cells closely resemble SC-EC cells, as SC-EC lineage genes are expressed, such as *SLC18A1*, *FEV*, and *DDC* (Fig. 2f, Extended Data Fig. 5b,c). In addition, SC-EE cells sustained endocrine identity, expressing *CHGA*, *NEUROD1*, and *NKX2-2* (Extended Data Fig. 5c). Nevertheless, SC-EE cells were still distinct from SC-EC cells in some important markers, notably their low expression of *TPH1*, a defining feature of SC-EC cells, and other genes and chromatin accessibility motifs (Extended Data Fig. 5d). Similar results were seen by protein expression, with substantial reduction in SC-β cell marker and increased SC-EE marker expression (Fig. 2g,h, Extended Data Fig. 6). Taken together, these findings show that *CTCF* modulation drastically disrupts acquisition of pancreatic endocrine identity from pancreatic progenitors and instead redirects it to an intestinal enteroendocrine-like cell fate, highlighting the importance of chromatin regulation in fate determination.

**Distinct chromatin states define maturation**

To further characterize SC-islet cell identities, we used single-cell multiomics to sequence primary cadaveric human islets and compared gene expression and chromatin accessibility profiles. Clustering of the integrated datasets, which included SC-islets and human islets both ATAC and RNA sequencing information, generated two β-cell populations (β-cell 1 and β-cell 2), α cells, δ cells, EC cells, and other cell types (Fig. 3a, Extended Data Fig. 7a-e). β-cell 1 and β-cell 2 populations are comprised mainly from SC-β islets and primary islets respectively (Fig. 3b, Extended Data Fig. 7b). Consistent with prior transcriptional analysis[11], there was no indication of EC cell populations in primary islets by chromatin accessibility analysis. Pearson correlation analysis using promoter accessibility showed that key islet endocrine cell types (SC-β cells, SC-α cells, and SC-δ cells) are relatively ambiguous in their

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.26.482126; this version posted February 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

chromatin accessibility, while their primary human islet endocrine cell counterparts are highly heterogenous and distinct (Fig. 3c).

To further interrogate these differences, we performed differential gene and motif enrichment analyses on the β, α, and δ-cell populations (Fig. 3d, Extended Data Fig. 7g-i). Chromatin accessibility motif enrichment analysis identified a multitude of motifs in SC-β cells that are expected in β-cells. However, SC-β cells were also enriched for many motifs of off-target/non-pancreatic markers and progenitor states, such as CDX2, LMX1A, SOX9, and SOX2. Furthermore, SC-β cells had substantially lower accessibility of many motifs associated with maturation in primary β-cells, including MTF, CREB1, MAFK, NFIB, NFE2, and the FOS/JUN family. We also performed gene regulatory network (GRN) analysis to determine if these motifs correspond to downstream expression of target genes (Fig. 3e). Activity of GRNs corresponding to motifs associated with β-cell immaturity (ONECUT2, GATA6, and PAX4) was higher in SC-β cells than in primary β-cells, while those corresponding with a mature β-cell state (MAFA, MAFB and JUND) are more active in primary β-cells.

Looking specifically at accessibility of *INS* and *MAFA*, known to be upregulated transcriptionally in mature β-cells[11,15,34,35], we found that the chromatin accessibility around these two gene regions differ considerably (Fig. 3f). Around the *INS* gene, we observed numerous peaks in the primary β-cell population that are predicted to be cis-regulatory enhancers, while the predicted cis-regulatory repressor peaks were present in SC-β cells. Similarly, numerous cis-regulatory elements enhancers were present around the *MAFA* gene region of primary β-cells. This was also reflected with other mature β-cell genes, where the promoter regions were more accessible in primary β-cells compared to SC-β cells (Extended Data Fig. 7j). Interestingly, while chromatin accessibility for the *MAFA* gene was lower in SC-β cells compared to primary

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.26.482126; this version posted February 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

β-cells (Fig. 3f), motif accessibility of MAFA was not differentially enriched (Fig. 3d). This suggests that the role of MAFA in maturation is regulated by the expression of the gene and not by accessibility of the loci it targets. Cross-referencing chromatin regulator genes[36] with our data yielded candidates for regulator proteins that modify chromatin accessibility in maturation, including *ARID1B* (Fig. 3g). Taken together, this analysis highlights the fundamental differences between SC-islet and primary islet cell types and that the ambiguity in cell identity is specific to SC-islets and not shared by primary islets.

**Ambiguity is overcome by transplantation**

We postulated that transplantation of SC-islets, particularly the SC-β cells, renders the chromatin accessibility to become mature-like and distinct[15]. We harvested transplanted SC-islets from the kidney capsule of immunocompromised NOD.Cg-*Prkdc*$^{scid}$ *Il*2*rg*$^{tm1wjl}$/SzJ (NSG) mice after 6 months, isolated human cells, and performed multiomic sequencing (Fig. 4a, Extended Data Fig. 8a). Using integrative single-cell analysis, comprising of transplanted (in vivo) SC-islets and in vitro-derived SC-islets, we identified all key SC-islet endocrine cell populations (Fig. 4b, Extended Data Fig. 8b,c). Transplanted SC-β, SC-α, SC-δ, and SC-EC cells acquired more mature gene expression signatures, including increased *INS* expression in SC-β cells (Extended Data Fig. 8c-e). In addition, Pearson correlation analysis using ATAC-based promoter accessibility showed that transplanted endocrine cells are more distinct in their respective cell types (Fig. 4c). Clustering using only transplanted datasets also highlighted distal clusters in ATAC-dependent analysis (Extended Data Fig. 8f) that paralleled our observations in primary islets (Extended Data Fig. 7e), including clear separation of SC-β cell and SC-EC cell populations that were merged in vitro (Fig. 1b). Importantly, chromatin accessibility motif

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.26.482126; this version posted February 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

enrichment analysis showed that transplantation caused enrichment in chromatin accessibility of mature motifs and reduced off-target motif activities (Fig. 4d, Extended Data Fig. 8g). Chromatin accessibility increases were also observed with *IAPP*, another mature β-cell gene, where several cis-regulatory enhancer peaks were present in transplanted β-cells (Fig. 4e). We also cross-referenced chromatin regulator genes[36] with our dataset, which yielded candidates for regulator proteins that modify chromatin accessibility in maturation, including *ARID1B* (Fig. 4f). Overall, these results support that SC-islet cell type identities are plastic based on chromatin accessibility, with transplantation improving the distinctiveness of each cell type and increasing their resemblance to mature islet cell types.

### *ARID1B* regulates SC-β cell maturation

Our comparisons of SC-β cells to primary β-cells and study of transplanted SC-β cells identified several common differentially expressed regulators of chromatin (Fig. 3g and Fig. 4f), suggesting the involvement of chromatin regulation on SC-β cell maturation. One such regulator that was present on both lists was *ARID1B*[37]. To study its role in SC-β cell maturation, we utilized short hairpin RNAs (shRNAs) targeting *ARID1B* to reduce its expression in differentiated SC-islets (Extended Data Fig. 9a,b). *ARID1B* knockdown showed upregulation of β-cell maturation markers, including *INS*, *IAPP*, *ISL1*, and *DLK1* (Fig 4g, Extended Data Fig. 9c). This improvement in gene expression was also associated with elevated insulin protein content, reduced pro-insulin/insulin ratio, increased SC-β-cell yields, and increased amylin (IAPP) expression (Fig 4h, Extended Data Fig. 9d,e). *ARID1B* knockdown, however, did not increase either glucagon or somatostatin protein content, suggesting that *ARID1B* has cell type-specific maturation effects (Extended Data Fig. 9f). Single-cell multiomic sequencing of SC-

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.26.482126; this version posted February 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

islets with *ARID1B* knockdown revealed increased gene expression and motif accessibility for signatures that were enriched in mature human β-cells and transplanted SC-β cells (Fig 4i, Extended Data Fig. 9g,h,i). Specifically, ONECUT1, ONECUT2, NEUROD1, HAND2, NEUROG2, and some off-target non-pancreatic motifs, such as CDX2, were decreased with *ARID1B* knockdown (Fig 4i). JUND, FOS, NFE2 and other mature motifs associated with mature β-cells were enriched with *ARID1B* knockdown (Fig 4i, Extended Data Fig. 9j). Taken together, this data shows that targeting chromatin regulators can affect maturation state of SC-β cells both transcriptionally and by chromatin accessibility.

## Discussion

SC-islets and other in vitro-derived tissues have incredible potential for the development of next-generation therapies. However, the exact process by which individual cell types acquire their cell identity and mature from in vitro differentiation is an outstanding question in the field. High-resolution indexing of transcriptional and chromatin accessibility changes of individual pancreatic islet cell types has allowed us to gain new insights into these cell fates. These dynamics include not only acquisition of cell type-defining features but also decreases in chromatin accessibility of off-target and progenitor genes and motifs. Our results show that transcriptional and chromatin accessibility signatures of differentiated SC-β cells and other cell types are plastic, changing with time in vitro and after transplantation, and can be disrupted to cause changes in cell fate selection and maturation. One unexpected finding of this study is that individual in vitro-differentiated islet cell types are much less distinguished from each other by chromatin accessibility compared to their primary cell counterparts. This is driven mainly by continued open chromatin accessibility of genes expressed by progenitors cell types (e.g.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.26.482126; this version posted February 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

*NEUROG3*[38] and *GP2*[39]) or alternative cell fates (e.g. *FEV* and *SOX2*) that are closed in primary cells. The reason in vitro-differentiated islet cells have less distinguished transcriptional and chromatin accessibility signatures than their primary cell counterparts remains unclear. These differences could also be present within in vitro differentiation systems for other cell types, as immaturity is commonly observed[40,41].

A prior study identified that enterochromaffin-like cells are generated during SC-islet differentiation in vitro and come from the same progenitor population as the islet endocrine cells[11]. Our results show that enterochromaffin-like cells and SC-β cells are not completely distinct. Instead, a continuum of cell types containing features unique to both cell types exist. DNA binding motif accessibility for the chromatin regulator CTCF was highly correlative, perturbation of which blocks differentiation to islet endocrine and further redirects cell fate selection to enteroendocrine-like cells, supporting the notion that chromatin accessibility is a major regulator of this fate decision. We therefore hypothesize that regulators of chromatin play a substantial role in generating and maturing islet cells and that better control of these processes will lead to development of SC-islet differentiation protocols with improved maturation and defined cellular composition. Given that these cell types are not seen in the in vivo pancreas from normal development in vivo[42,43], this fragility of in vitro differentiation should be understood, and microenvironmental cues strictly controlled to advance the field. This is further emphasized by the fact that extended culture in vitro, while understood to increase functional and transcriptional maturation in prior work[10,11,44], has relatively small changes on gene expression and chromatin accessibility signatures of SC-islets compared to mature cadaveric islets. Our observation that *ARID1B* knockdown increases SC-β cell maturation in vitro supports the notion that these maturation events are at least partially controlled by chromatin accessibility.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.26.482126; this version posted February 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

We provide a comprehensive indexing of transcriptional and chromatin accessibility in SC-islets that will facilitate further development of differentiation protocols. The multiomics approach leveraged here allows for more robust hypothesis generation than using transcriptional data alone. Not only are SC-β cells immature transcriptionally but also by chromatin accessibility, and consideration for both should be made in future studies pursuing next-generation differentiation protocols.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.26.482126; this version posted February 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

## Methods

### Cell culture and differentiation

The HUES8 hESC line was provided by Douglas Melton (Harvard University)[3]. 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)[33]. 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% $CO_2$ at 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[2,24]. 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 and Supplementary Table 2. After 7 days in stage 6, cells were placed on an Orbishaker (Benchmark) at 100 RPM and maintained in ESFM media.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.26.482126; this version posted February 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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[45] and Signac 1.3.0[46]. For ATAC data, genomic positions were mapped and annotated with EnsDb.Hsapeins.v86 and hg38. Low quality 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);

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.26.482126; this version posted February 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

nucleosome signal > 1.25, and TSS enrichment < 2. In addition, mouse cells were removed by removing cells showing high expression of kidney marker *TTC36*[15] for transplanted SC-islet samples. 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[46,47]. Motif activity was computed by using the chromVAR package[48] and JASPAR version 2020 database[49]. Pseudo-time ordering analysis was conducted using Monocle3 package[50]. Gene regulatory network (GRN) analysis was performed using the SCENIC tool set[51].

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

CRISPRa genetic engineering of the H1 dCas9-VPR line[33] was mediated using custom guide RNAs (gRNAs) ordered from MilliporeSigma. 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

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.26.482126; this version posted February 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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 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 3.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.26.482126; this version posted February 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

**Microscopy**

Fluorescence images were taken on a Zeiss Cell Discoverer confocal 7 microscope. For immunocytochemistry (ICC), differentiated SC-islet cells, and cells undergoing differentiation 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. Image J was used for analysis[52].

**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.

**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.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.26.482126; this version posted February 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.26.482126; this version posted February 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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). 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.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.26.482126; this version posted February 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

**Statistics and reproducibility**

For in vitro experiments, we performed unpaired 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. Gene expression values and promoter accessibility values from Fig. 1b were analyzed by using Z-score normalization to perform spearman correlation.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.26.482126; this version posted February 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

## References

1 Latres, E., Finan, D. A., Greenstein, J. L., Kowalski, A. & Kieffer, T. J. Navigating Two Roads to Glucose Normalization in Diabetes: Automated Insulin Delivery Devices and Cell Therapy. *Cell metabolism* **29**, 545-563, doi:10.1016/j.cmet.2019.02.007 (2019).

2 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, doi:10.1038/s41587-020-0430-6 (2020).

3 Pagliuca, F. W. *et al.* Generation of functional human pancreatic beta cells in vitro. *Cell* **159**, 428-439, doi:10.1016/j.cell.2014.09.040 (2014).

4 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, doi:10.1038/s41556-018-0271-4 (2019).

5 Rezania, A. *et al.* Reversal of diabetes with insulin-producing cells derived in vitro from human pluripotent stem cells. *Nature biotechnology* **32**, 1121-1133, doi:10.1038/nbt.3033 (2014).

6 Lanzoni, G. & Ricordi, C. Transplantation of stem cell-derived pancreatic islet cells. *Nat Rev Endocrinol* **17**, 7-8, doi:10.1038/s41574-020-00430-9 (2021).

7 Migliorini, A., Nostro, M. C. & Sneddon, J. B. Human pluripotent stem cell-derived insulin-producing cells: A regenerative medicine perspective. *Cell metabolism* **33**, 721-731, doi:10.1016/j.cmet.2021.03.021 (2021).

8 Maxwell, K. G. & Millman, J. R. Applications of iPSC-derived beta cells from patients with diabetes. *Cell reports. Medicine* **2**, 100238, doi:10.1016/j.xcrm.2021.100238 (2021).

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.26.482126; this version posted February 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

9 Yoshihara, E. *et al.* Immune-evasive human islet-like organoids ameliorate diabetes. *Nature* **586**, 606-611, doi:10.1038/s41586-020-2631-z (2020).

10 Velazco-Cruz, L. *et al.* Acquisition of Dynamic Function in Human Stem Cell-Derived beta Cells. *Stem cell reports* **12**, 351-365, doi:10.1016/j.stemcr.2018.12.012 (2019).

11 Veres, A. *et al.* Charting cellular identity during human in vitro beta-cell differentiation. *Nature* **569**, 368-373, doi:10.1038/s41586-019-1168-5 (2019).

12 Mahaddalkar, P. U. *et al.* Generation of pancreatic beta cells from CD177(+) anterior definitive endoderm. *Nature biotechnology* **38**, 1061-1072, doi:10.1038/s41587-020-0492-5 (2020).

13 Song, J. & Millman, J. R. Economic 3D-printing approach for transplantation of human stem cell-derived beta-like cells. *Biofabrication* **9**, 015002, doi:10.1088/1758-5090/9/1/015002 (2016).

14 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**, doi:10.1126/scitranslmed.aax9106 (2020).

15 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, doi:10.1016/j.celrep.2020.108067 (2020).

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

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

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.26.482126; this version posted February 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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

19 Lund, R. J., Närvä, E. & Lahesmaa, R. Genetic and epigenetic stability of human pluripotent stem cells. *Nature Reviews Genetics* **13**, 732-744, doi:10.1038/nrg3271 (2012).

20 Weng, C. *et al.* Single-cell lineage analysis reveals extensive multimodal transcriptional control during directed beta-cell differentiation. *Nat Metab* **2**, 1443-1458, doi:10.1038/s42255-020-00314-2 (2020).

21 Augsornworawat, P. & Millman, J. R. Single-cell RNA sequencing for engineering and studying human islets. *Curr Opin Biomed Eng* **16**, 27-33, doi:10.1016/j.cobme.2020.06.003 (2020).

22 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, doi:10.1016/j.stem.2012.11.023 (2013).

23 Alvarez-Dominguez, J. R. *et al.* Circadian Entrainment Triggers Maturation of Human In Vitro Islets. *Cell stem cell* **26**, 108-122 e110, doi:10.1016/j.stem.2019.11.011 (2020).

24 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, doi:10.1038/s41596-021-00560-y (2021).

25 Arda, H. E. *et al.* Age-Dependent Pancreatic Gene Regulation Reveals Mechanisms Governing Human beta Cell Function. *Cell metabolism* **23**, 909-920, doi:10.1016/j.cmet.2016.04.002 (2016).

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.26.482126; this version posted February 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

26 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, doi:10.1073/pnas.1400709111 (2014).

27 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, doi:10.1242/dev.130682 (2016).

28 Egozi, A. *et al.* Insulin is expressed by enteroendocrine cells during human fetal development. *Nat Med* **27**, 2104-2107, doi:10.1038/s41591-021-01586-1 (2021).

29 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, doi:10.1038/s41588-021-00823-0 (2021).

30 Haber, A. L. *et al.* A single-cell survey of the small intestinal epithelium. *Nature* **551**, 333-339, doi:10.1038/nature24489 (2017).

31 Grün, D. *et al.* Single-cell messenger RNA sequencing reveals rare intestinal cell types. *Nature* **525**, 251-255, doi:10.1038/nature14966 (2015).

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

33 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, doi:10.1038/s41598-020-57500-1 (2020).

34 Artner, I. *et al.* MafA and MafB regulate genes critical to beta-cells in a unique temporal manner. *Diabetes* **59**, 2530-2539, doi:10.2337/db10-0190 (2010).

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.26.482126; this version posted February 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

35 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, doi:10.1210/jc.2015-2632 (2015).

36 Medvedeva, Y. A. *et al.* EpiFactors: a comprehensive database of human epigenetic factors and complexes. *Database : the journal of biological databases and curation* **2015**, bav067, doi:10.1093/database/bav067 (2015).

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

38 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).

39 Cogger, K. F. *et al.* Glycoprotein 2 is a specific cell surface marker of human pancreatic progenitors. *Nat Commun* **8**, 331, doi:10.1038/s41467-017-00561-0 (2017).

40 Patterson, M. *et al.* Defining the nature of human pluripotent stem cell progeny. *Cell Res* **22**, 178-193, doi:10.1038/cr.2011.133 (2012).

41 Murry, C. E. & Keller, G. Differentiation of embryonic stem cells to clinically relevant populations: lessons from embryonic development. *Cell* **132**, 661-680, doi:10.1016/j.cell.2008.02.008 (2008).

42 Baron, M. *et al.* A Single-Cell Transcriptomic Map of the Human and Mouse Pancreas Reveals Inter- and Intra-cell Population Structure. *Cell Syst* **3**, 346-360 e344, doi:10.1016/j.cels.2016.08.011 (2016).

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.26.482126; this version posted February 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

43 Enge, M. *et al.* Single-Cell Analysis of Human Pancreas Reveals Transcriptional Signatures of Aging and Somatic Mutation Patterns. *Cell* **171**, 321-330 e314, doi:10.1016/j.cell.2017.09.004 (2017).

44 Velazco-Cruz, L. *et al.* SIX2 Regulates Human beta Cell Differentiation from Stem Cells and Functional Maturation In Vitro. *Cell Rep* **31**, 107687, doi:10.1016/j.celrep.2020.107687 (2020).

45 Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587 e3529, doi:10.1016/j.cell.2021.04.048 (2021).

46 Stuart, T., Srivastava, A., Madad, S., Lareau, C. A. & Satija, R. Single-cell chromatin state analysis with Signac. *Nat Methods* **18**, 1333-1341, doi:10.1038/s41592-021-01282-5 (2021).

47 Pliner, H. A. *et al.* Cicero Predicts cis-Regulatory DNA Interactions from Single-Cell Chromatin Accessibility Data. *Mol Cell* **71**, 858-871 e858, doi:10.1016/j.molcel.2018.06.044 (2018).

48 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, doi:10.1038/nmeth.4401 (2017).

49 Fornes, O. *et al.* JASPAR 2020: update of the open-access database of transcription factor binding profiles. *Nucleic acids research* **48**, D87-d92, doi:10.1093/nar/gkz1001 (2020).

50 Cao, J. *et al.* The single-cell transcriptional landscape of mammalian organogenesis. *Nature* **566**, 496-502, doi:10.1038/s41586-019-0969-x (2019).

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.26.482126; this version posted February 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

51 Aibar, S. *et al.* SCENIC: single-cell regulatory network inference and clustering. *Nat Methods* **14**, 1083-1086, doi:10.1038/nmeth.4463 (2017).

52 Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nature Methods* **9**, 671-675, doi:10.1038/nmeth.2089 (2012).

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.26.482126; this version posted February 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

**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. M.I. was supported by Rita Levi-Montalcini Postdoctoral Fellowship in Regenerative Medicine. 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 Dr. Udomkittivorakul for the graphics design, Dr. Melton (Harvard University) for the HUES8 cell line, and Dr. 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. and J.R.Millman wrote the manuscript. All authors revised and approved the manuscript.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.26.482126; this version posted February 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

## Competing interests

P.A., 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.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.26.482126; this version posted February 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

## Figures

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.26.482126; this version posted February 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

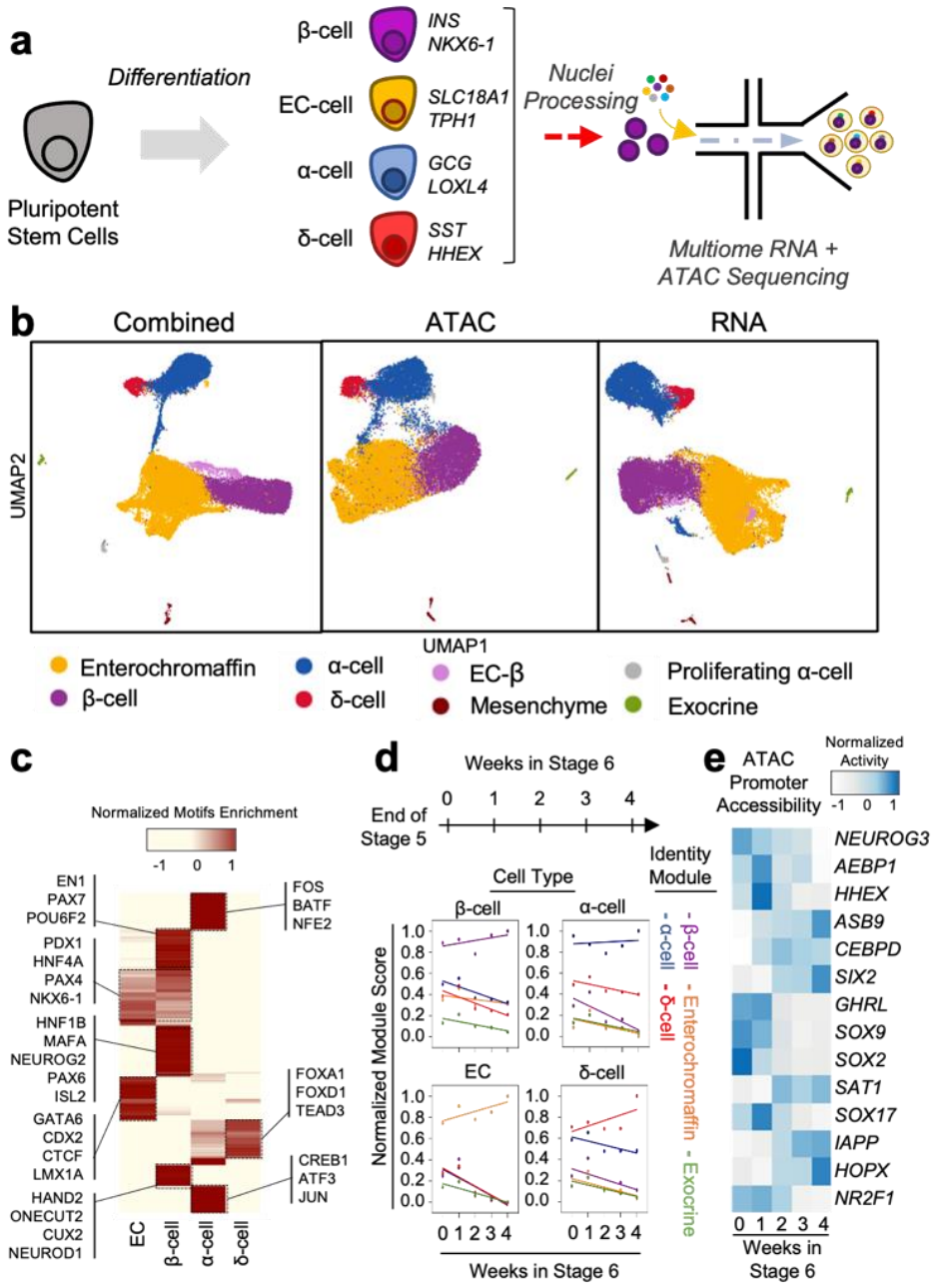



Fig. 1 | **Multiomic profiling of stem-cell derived islets shows unique chromatin accessibility signatures in endocrine cell types**. a, Schematic of SC-islet differentiation and multiomic sequencing. b, UMAPs showing identified cell types in SC-islets 14 days in stage 6 using both or either ATAC and gene (RNA) information. c, Heatmap showing the top 200 variable motifs within endocrine cell populations and highlighting motif markers for each cell type. d, Module

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.26.482126; this version posted February 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

scores of β, α, δ, EC, and exocrine cell identity based on top 50 promoter accessibilities in endocrine cell populations from time course multiomic sequencing of SC-islets cultured up to 4 weeks in stage 6 of the differentiation protocol. e, Heatmap showing time course promoter accessibilities during stage 6 for genes associated with β-cell, progenitor, and off-target identities in SC-β cells. EC, enterochromaffin cells; EC-β, enterochromaffin-β cells.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.26.482126; this version posted February 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

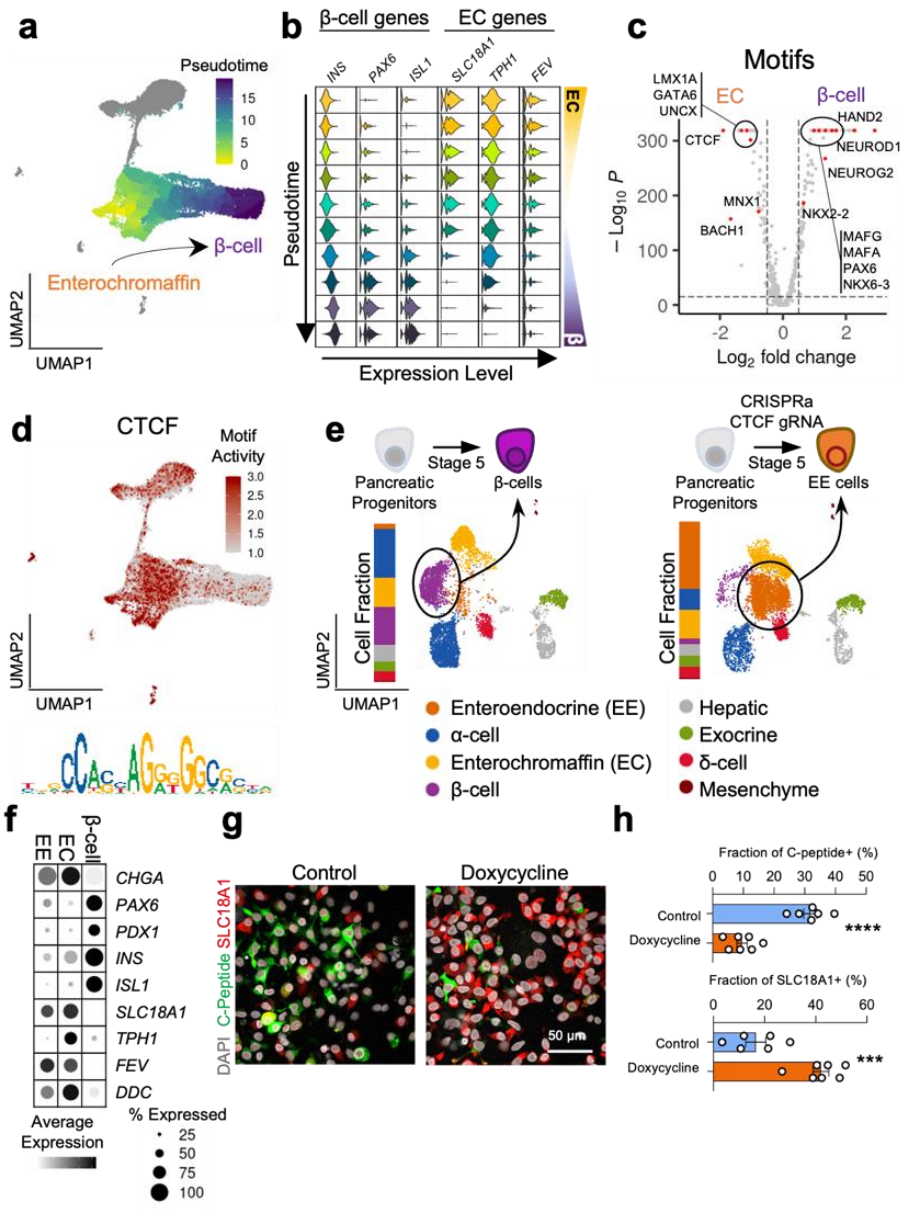



Fig. 2 | **SC-EC and SC-β cells have unique and common transcriptional and chromatin accessibility signatures.** a, UMAP showing the pseudotime ordering of cells from the SC-β, SC-EC and EC-β population at 14 days of stage 6. b, Gene expression (RNA) of key genes associated with SC-β cell and SC-EC cell identity along pseudotime. c, Differential motif enrichment analysis highlighting the top enriched motifs in the SC-β and SC-EC cell populations. d, Motif representing CTCF is highly active in the SC-EC cell population but not in

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.26.482126; this version posted February 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

the SC-β cell population. e, UMAPs of cells at the end of stage 5 with or without doxycycline induction of CTCF CRISPRa showing the identification of SC-EE cells. f, Dot plot showing the gene expression of genes associated with pancreatic endocrine (*CHGA, PAX6, PDX1*), β-cell (*INS, ISL1*), and EC-cell (*SLC18A1, TPH1, FEV, DDC*) in the SC-β cell, SC-EC, and SC-EE populations in the *CTCF* CRISPRa experiment. g, Immunocytochemistry staining showing expression of EE marker protein SLC18A1 (red) and β-cell marker protein C-peptide (green) with or without *CTCF* expression induced by doxycycline. h, 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* expression induced by doxycycline, plotting mean ± s.e.m. (control; n = 6, doxycycline; n = 7 biological replicates) EC, enterochromaffin cells; EC-β, EC-β, enterochromaffin-β cells; EE, enteroendocrine like cells; CRISPRa, CRISPR activation.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.26.482126; this version posted February 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

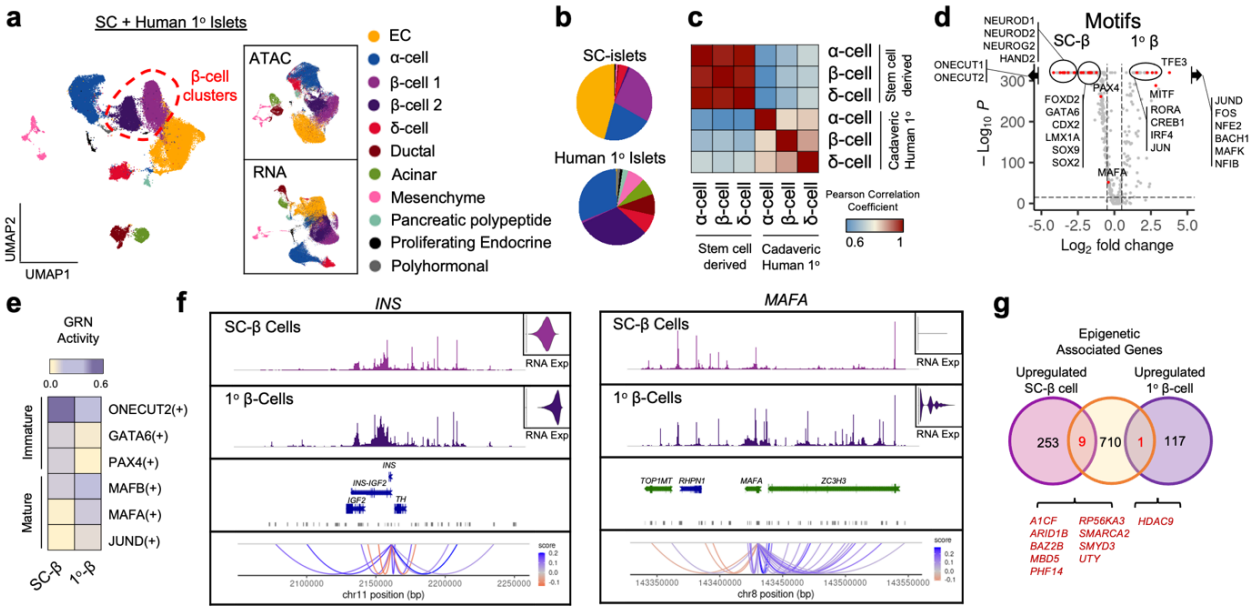



Fig. 3 | **Comparative analysis of SC-islets and cadaveric primary (1º) human islets shows differences in chromatin accessibility signatures associated with islet maturation.** a, UMAPs show clustering of multiomic sequencing datasets containing cells from SC-islets at 14 days of stage 6 and 1º islets using both or either ATAC and gene (RNA) expression information. b, Fractions of identified cell types in SC-islet and 1º islets. c, Pearson correlation analysis comparing SC β, α, and δ and 1º β, α, and δ cells. d, Differential motifs enrichment analysis highlighting the top enriched motifs in SC-β cells and 1º β-cells. e, Gene regulatory network analysis showing networks representing immature and mature states in β-cells. f, ATAC peaks from SC-β cells and 1º β-cells showing chromatin accessibility around the *INS* and *MAFA* genes. Peaks were linked and analyzed to depict cis-regulatory elements. Gene expression of *INS* and *MAFA* are presented as violin plots. g, Cross reference map of upregulated SC-β cell and 1º β-cell genes with epigenetic associated genes to highlight epigenetic regulators associated with each cell states. GRN, gene regulatory networks; 1º, primary.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.26.482126; this version posted February 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

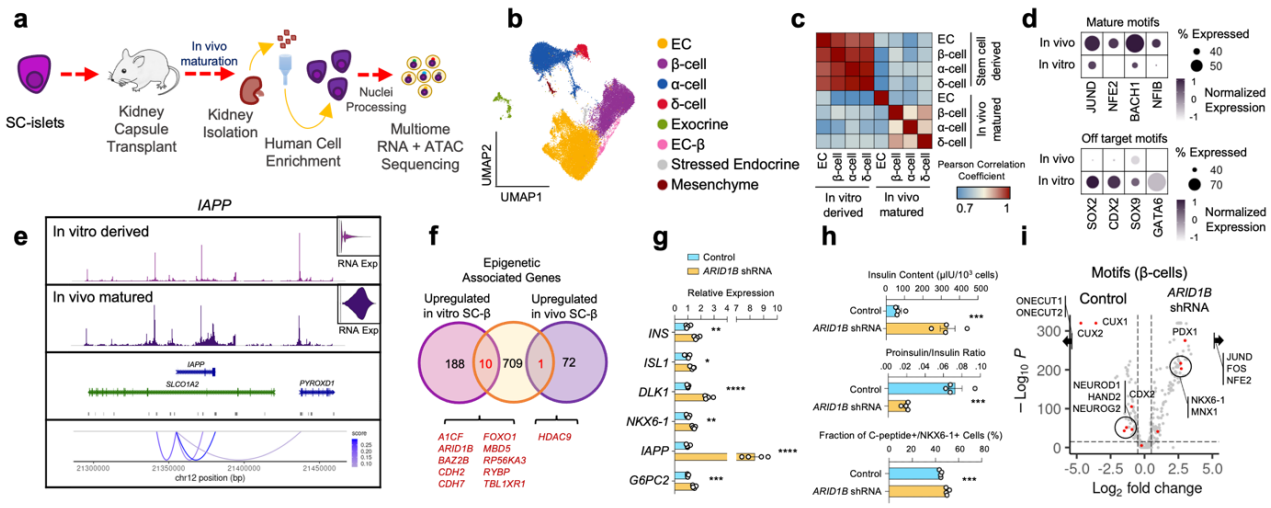



Fig. 4 | **Enhancing SC-islet chromatin accessibility and transcriptional profiles.** a, Schematic of transplantation of SC-islets and harvesting of transplanted cells for multiomic sequencing. b, UMAP showing clustering of multiomic sequencing datasets containing in-vitro derived SC-islets after 14 days of stage 6 and transplanted SC-islets after 6 months in vivo using both ATAC and gene (RNA) expression information. c, Pearson correlation analysis comparing endocrine cell types from in vitro and transplanted SC-islets. d, Dot plots of the β-cell populations, showing enrichment of motifs associated with β-cell maturity (top) and motifs associated with off target lineages. e, ATAC peaks of SC-β cells from in vitro and transplanted islets, around the *IAPP* gene region. Peaks were linked and analyzed to depict cis-regulatory elements. Gene expression of *IAPP* are presented as violin plots. f, Cross reference map of genes upregulated from in vitro SC-β cells and transplanted SC-β cells with chromatin associated genes. The list highlights chromatin regulators expressed in each condition. g-h, analysis of SC-islets after 14 days of stage 6 treated with shRNA for *ARID1B* knockdown on the first day of stage 6. Control = *GFP* knockdown. (g) Real-time PCR plots showing mean ± s.e.m. (n = 4, biological replicates) of expression of β-cell associated genes, *INS* (P = 0.0014), ISL1 (P = 0.022), *DLK1* (P = 4.3 x $10^{-5}$), *NKX6-1* (P = 0.0015), *IAPP* (P = 3.0 x $10^{-6}$), *G6PC2* (P = 2.1 x $10^{-4}$). (h) Protein quantification

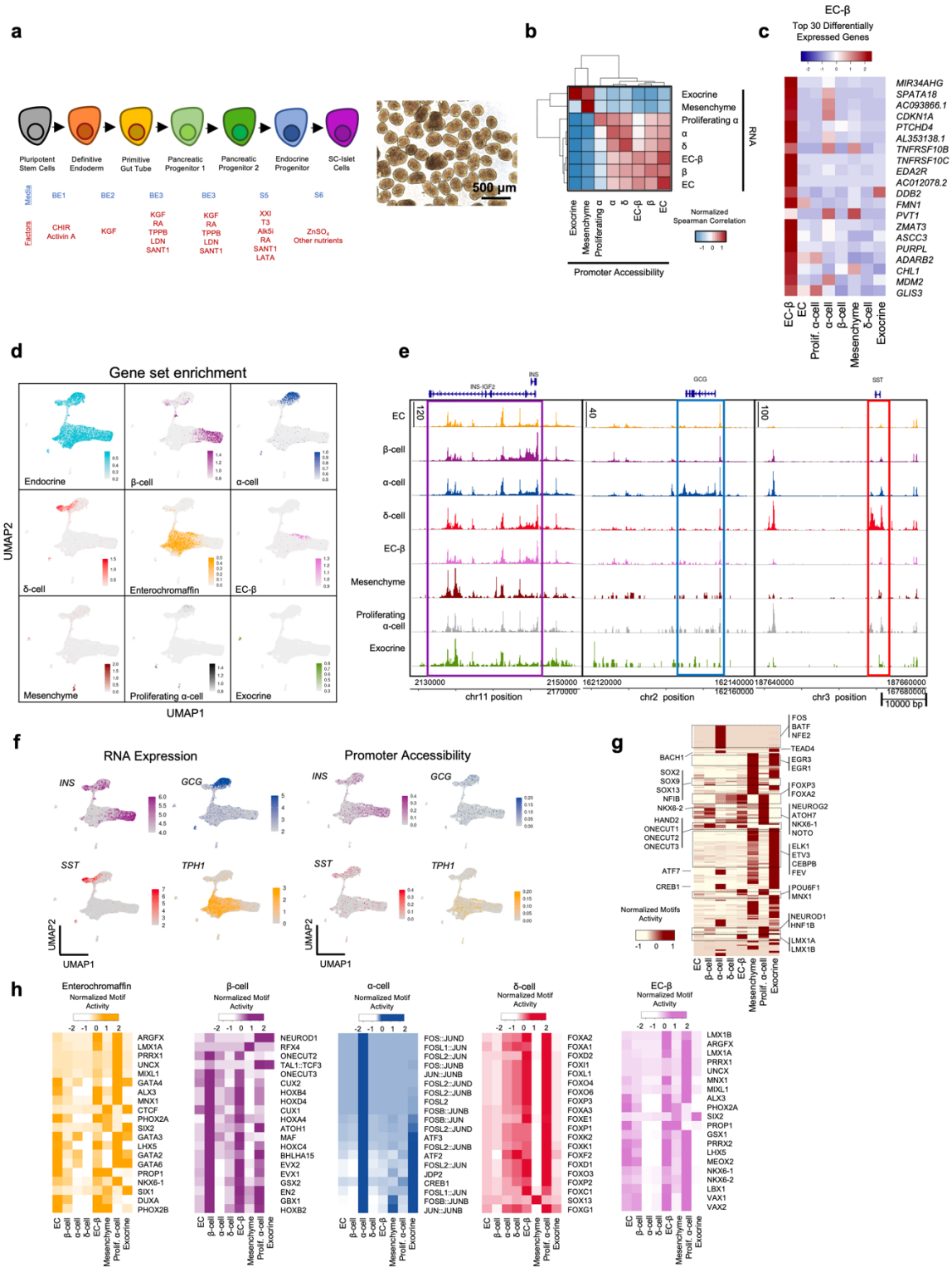
bioRxiv preprint doi: https://doi.org/10.1101/2022.02.26.482126; this version posted February 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

plot showing mean ± s.e.m. (n = 4, biological replicates) of human insulin content ($P = 7.6 \times 10^{-4}$) by ELISA (top), proinsulin/insulin ratio ($P = 3.6 \times 10^{-4}$) by ELISA (middle), and fraction of human C-peptide and NKX6-1 co-expression ($P = 4.3 \times 10^{-4}$) by flow cytometry (bottom). i, Multiomic sequencing differential motif chromatin accessibility enrichment analysis of in vitro SC-β cells with *ARID1B* knockdown. EC, enterochromaffin cells; EC-β, enterochromaffin-β cells.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.26.482126; this version posted February 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

## Additional Materials

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.26.482126; this version posted February 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

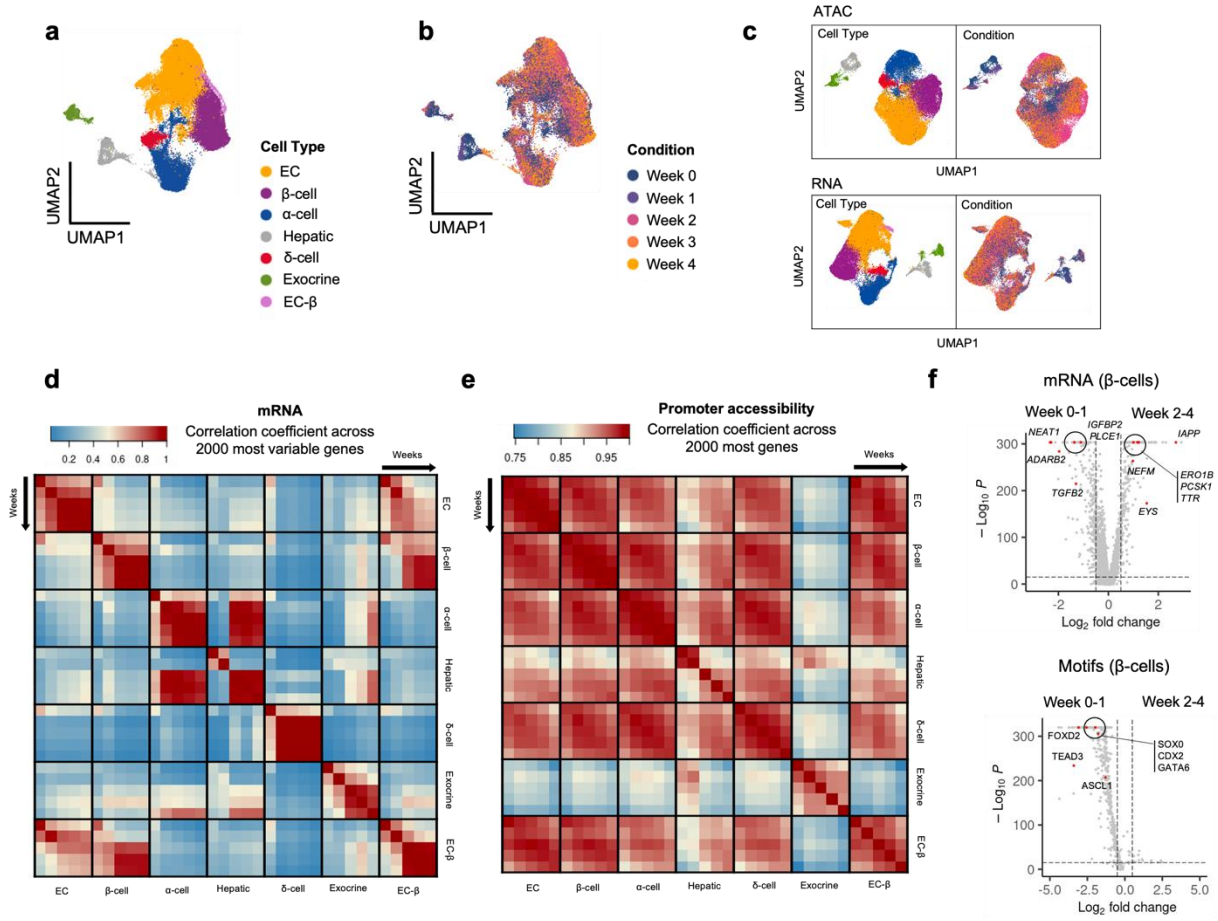



Extended Data Fig. 1 | **Single-cell Multiomic ATAC and gene expression characterization of SC-islets**. a, Schematic diagram outlining the differentiation protocol for generating SC-islets

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.26.482126; this version posted February 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

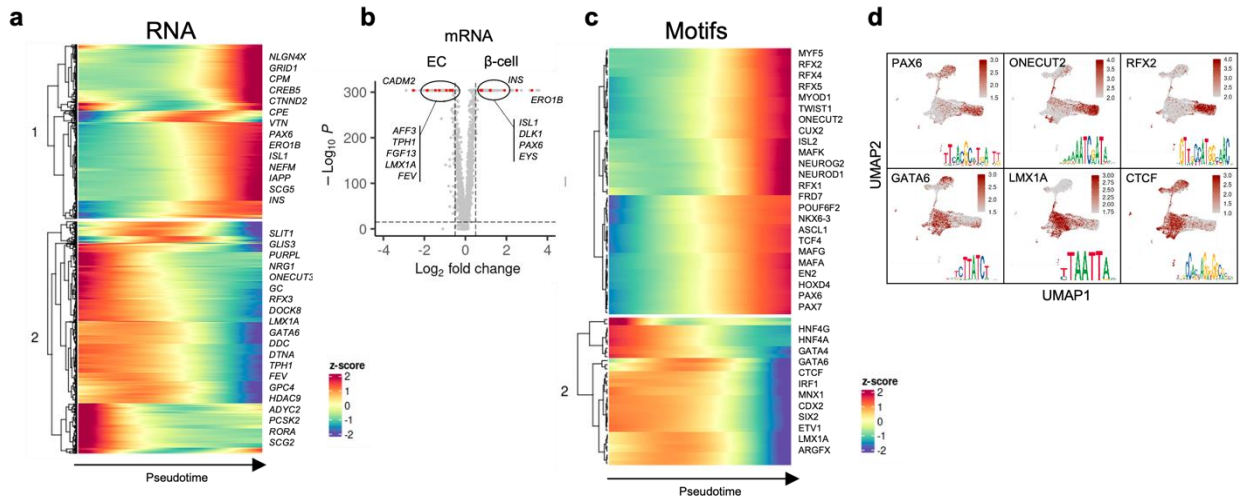
and brightfield image of SC-islets. b, Spearman correlation analysis comparing gene expression and ATAC promoter accessibility profiles using 15,984 commonly shared features. c, Differential gene expression analysis showing the top 20 upregulated genes in EC-β cell. d, Module scores showing gene set enrichment of genes associated with each identified cell type. e, ATAC peaks of major identified cell types in SC-islets around the *INS*, *GCG*, and *SST* gene regions. f, Feature plots displaying the RNA expression (left) and Promoter accessibility (right) of the *INS*, *GCG*, *SST*, and *TPH1* genes. g, Heatmap showing the top 200 variable motifs including all identified cell types and highlighting motifs associated with each cell type. h. Heatmaps showing the top 20 enriched motifs in SC-β, SC-α, SC-δ, SC-EC, and SC-EC-β cells. EC, enterochromaffin cells; EC-β, enterochromaffin-β cells; Prolif., Proliferating.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.26.482126; this version posted February 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



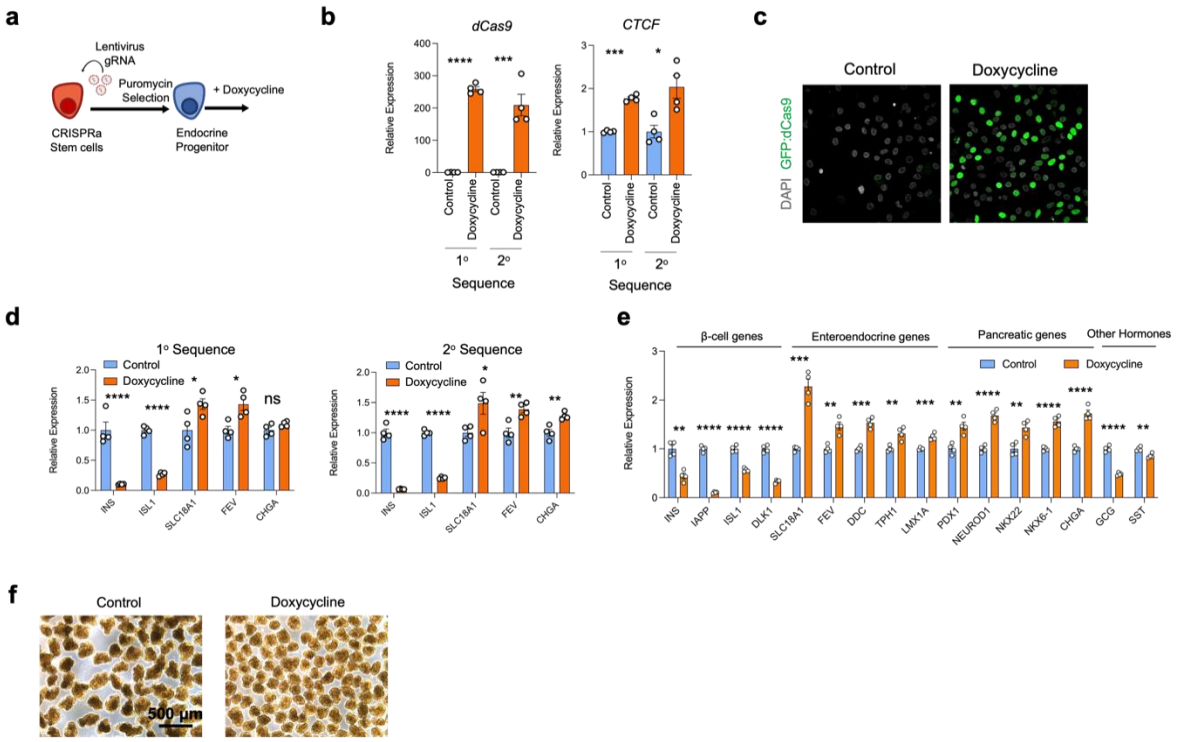

Extended Data Fig. 2 | **Time course single-cell multiomic sequencing analysis of SC-islets during stage 6 of the differentiation protocol**. a-c, Clustering of the integrative time course single-cell multiomic sequencing. (a) Clustering showing identified cell types obtained from both ATAC and gene expression information. (b) Clustering showing timepoint conditions obtained from both ATAC and gene expression information showing timepoint conditions. (c) Clustering using either ATAC or gene expression information. d-e, Pearson correction to compare all timepoints, week 0 – 4, of all identified cell types. (d) Using top 2000 most variable gene expression (mRNA). (e) Using top 2000 most variable promoter accessibility from ATAC. f, Differential gene (mRNA) expression analysis (top) and motif enrichment (bottom) analysis in SC-β from week 0-1 vs. week 2-4. EC, enterochromaffin cells; EC-β, enterochromaffin-β cells.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.26.482126; this version posted February 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.




Extended Data Fig. 3 | **Pseudotime ordering analysis of the SC-β cell, SC-EC cell, and EC-β cell population.** a, Pseudo-temporal expression pattern plot using gene expression. b, Differential gene expression (mRNA) analysis comparing SC-β cells with SC-EC cells. c, Pseudo-temporal expression patter plot using motif enrichment. d, Motif enrichment feature plot showing SC-β cell associated motifs, PAX6, ONECUT2, and RFX2, and SC-EC associated motifs, GATA6, LMX1A, and CTCF. EC, enterochromaffin cells; EC-β, enterochromaffin-β cells.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.26.482126; this version posted February 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

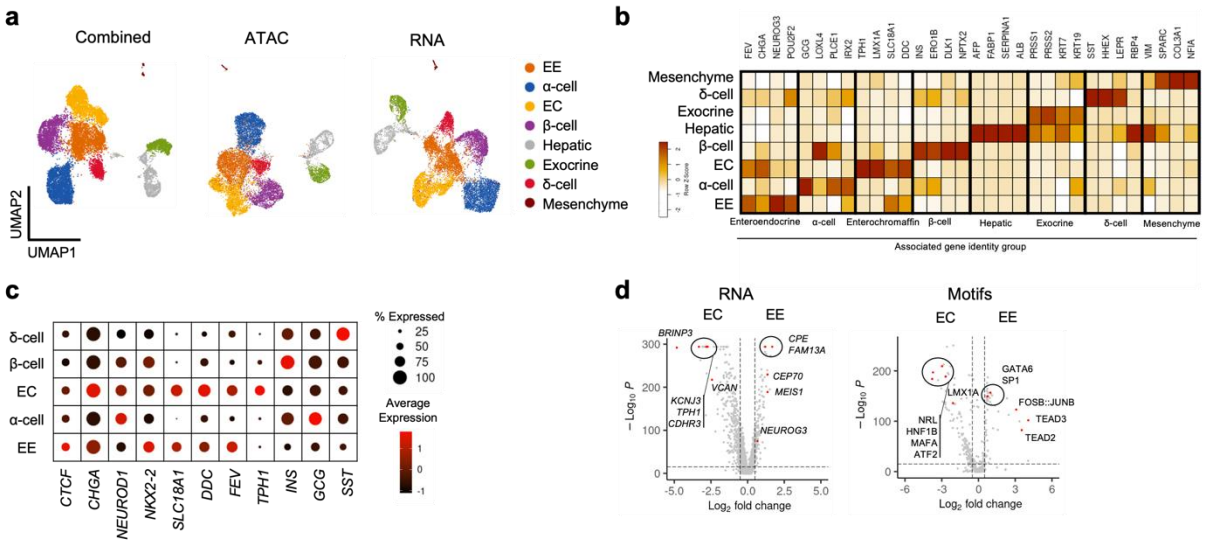



Extended Data Fig. 4 | ***CTCF* induction during differentiation using doxycycline inducible CRISPRa stem cells.** a, Schematic of lentivirus transduction of gRNA and induction of *CTCF* expression in CRISPRa stem cell differentiation system. b, qPCR analysis, plotting mean ± s.e.m. (n = 4, biological replicates), showing upregulation of dCas9 (1°, P = 4.2 x $10^{-8}$; 2°, P = 7.8 x $10^{-4}$) and *CTCF* (1°, P = 1.4 x $10^{-6}$; 2°, P = 0.014) upon activation using doxycycline. c, Confocal fluorescent image of dCas9 protein. d, qPCR analysis, plotting mean ± s.e.m. (n = 4, biological replicates), of *CTCF* induced SC-islets differentiated from CRISPRa stem cells with gRNAs of 2 sequences targeting *CTCF*. 1° sequence; *INS* (P = 5.6 x $10^{-4}$) *ISL1* (P = 1.1 x $10^{-6}$), *SLC18A1* (P = 0.034), *FEV* (P = 0.012), *CHGA* (ns, P = 0.12) and 2° sequence; *INS* (P = 4.4 x $10^{-6}$), *ISL1* (P = 7.8 x $10^{-8}$), *SLC18A1* (P = 0.042), *FEV* (P = 0.0069), *CHGA* (P = 0.0050). e, qPCR analysis, plotting mean ± s.e.m. (n = 4, biological replicates), of *CTCF* induced SC-islets later in differentiation showing expression differences of genes associated with β cells (*INS*, P = 0.0010; *IAPP*, P = 1.5 x $10^{-7}$; *ISL1*, P = 3.0 x $10^{-5}$; *DLK1*, P = 3.7 x $10^{-6}$), enteroendocrine cells

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.26.482126; this version posted February 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

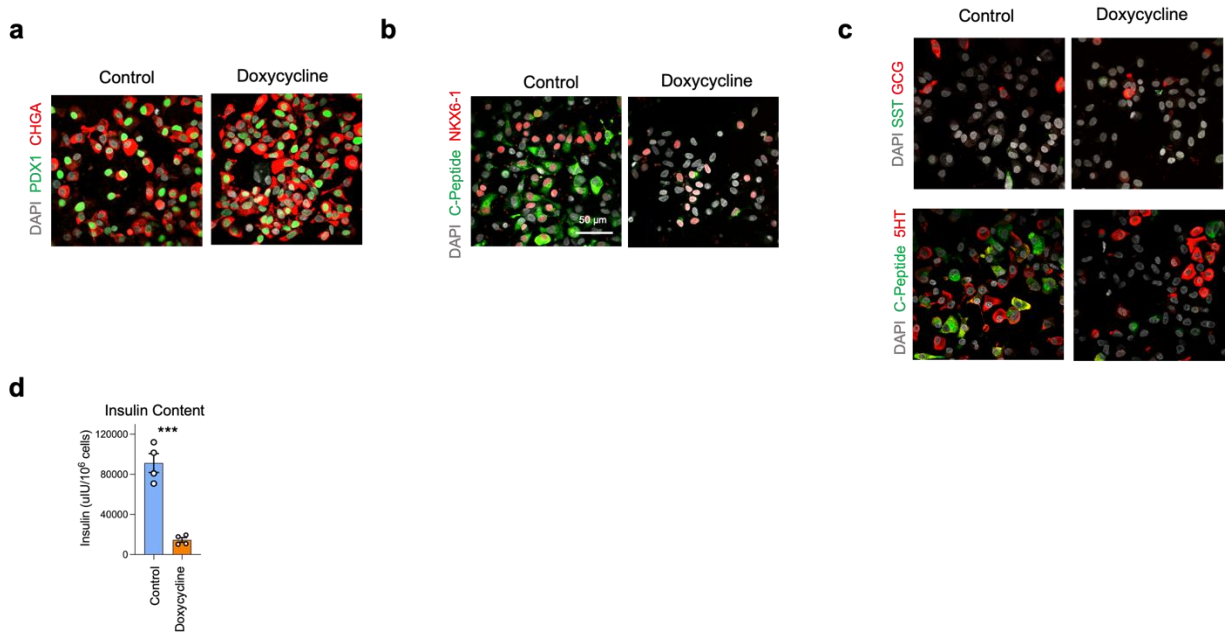
(SLC18A1, P = 1.6 x $10^{-4}$; *FEV*, P = 0.0019; *DDC*, P = 1.2 x $10^{-4}$; *TPH1*, P = 0.0047; *LMX1A*, P = 4.6 x $10^{-4}$), pancreatic identity (*PDX1*, P = 0.0031; *NEUROD1*, P = 6.2 x $10^{-5}$; *NKX2-2*, P = 0.0047; *CHGA*, P = 6.8 x $10^{-5}$), and other hormones (*GCG*, P = 5.9 x $10^{-6}$; *SST*, P = 0.0046). f, Brightfield images of differentiated SC-islets with *CTCF* induction. gRNA, guide RNA. ns, not significant. CRISPRa, CRISPR activation; gRNA, guide RNA.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.26.482126; this version posted February 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



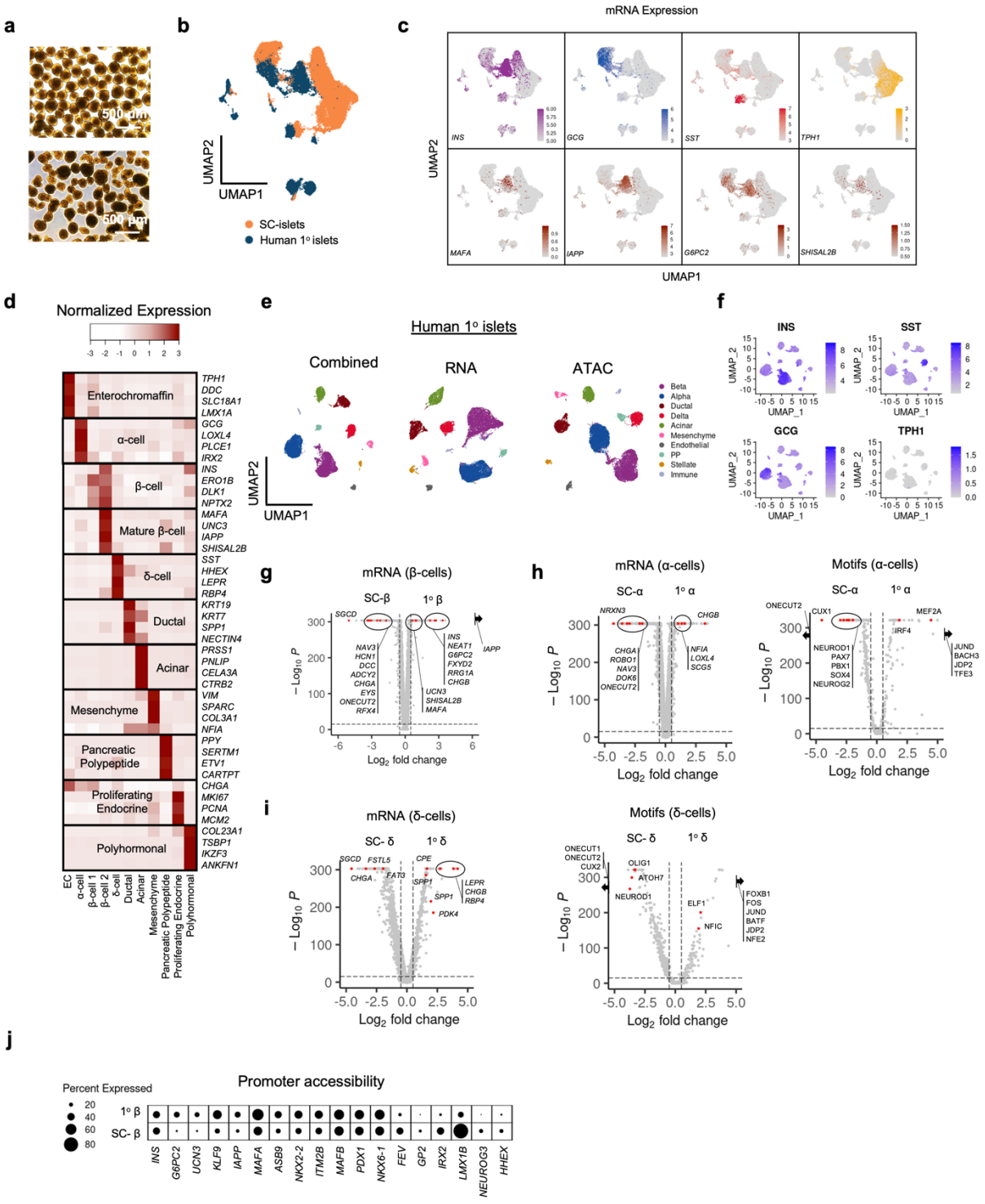

Extended Data Fig. 5 | **Multiomic single-cell sequencing of *CTCF* induced SC-islets reveals entero-endocrine cells.** a, UMAPs showing identified cell types in SC-islets induced with *CTCF* expression using both or either ATAC or gene (RNA) information. b, Heatmap depicting gene expression of major cell types identified in SC-islets and *CTCF* induced SC-islets. c, Dot plots highlighting the gene expression of *CTCF* and endocrine associated genes in the SC-β, SC-α, SC-δ, SC-EC, and SC-EE cell populations. d, Differential gene expression and motifs enrichment analysis comparing SC-EC and SC-EE cells. EC, enterochromaffin; EE, enteroendocrine-like.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.26.482126; this version posted February 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.




Extended Data Fig. 6 | **Validation of *CTCF* dependent fate shift to enteroendocrine-like cells.** a-c, Immunocytochemistry of SC-islets induced with *CTCF* expression during stage 5 of the differentiation protocol. (a) staining of endocrine markers, PDX1 (green) and CHGA (red). (b) staining of β-cell markers, C-peptide (green) and NKX6-1 (red) (c) staining of other pancreatic hormones (top), GCG (red) and SST (green), and (bottom), C-peptide (green) and serotonin;5HT (red). d, Quantification of insulin content ($P = 2.2 \times 10^{-4}$) using ELISA of SC-islets induced with *CTCF*, plotting mean ± s.e.m. (n = 4, biological replicates).

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.26.482126; this version posted February 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

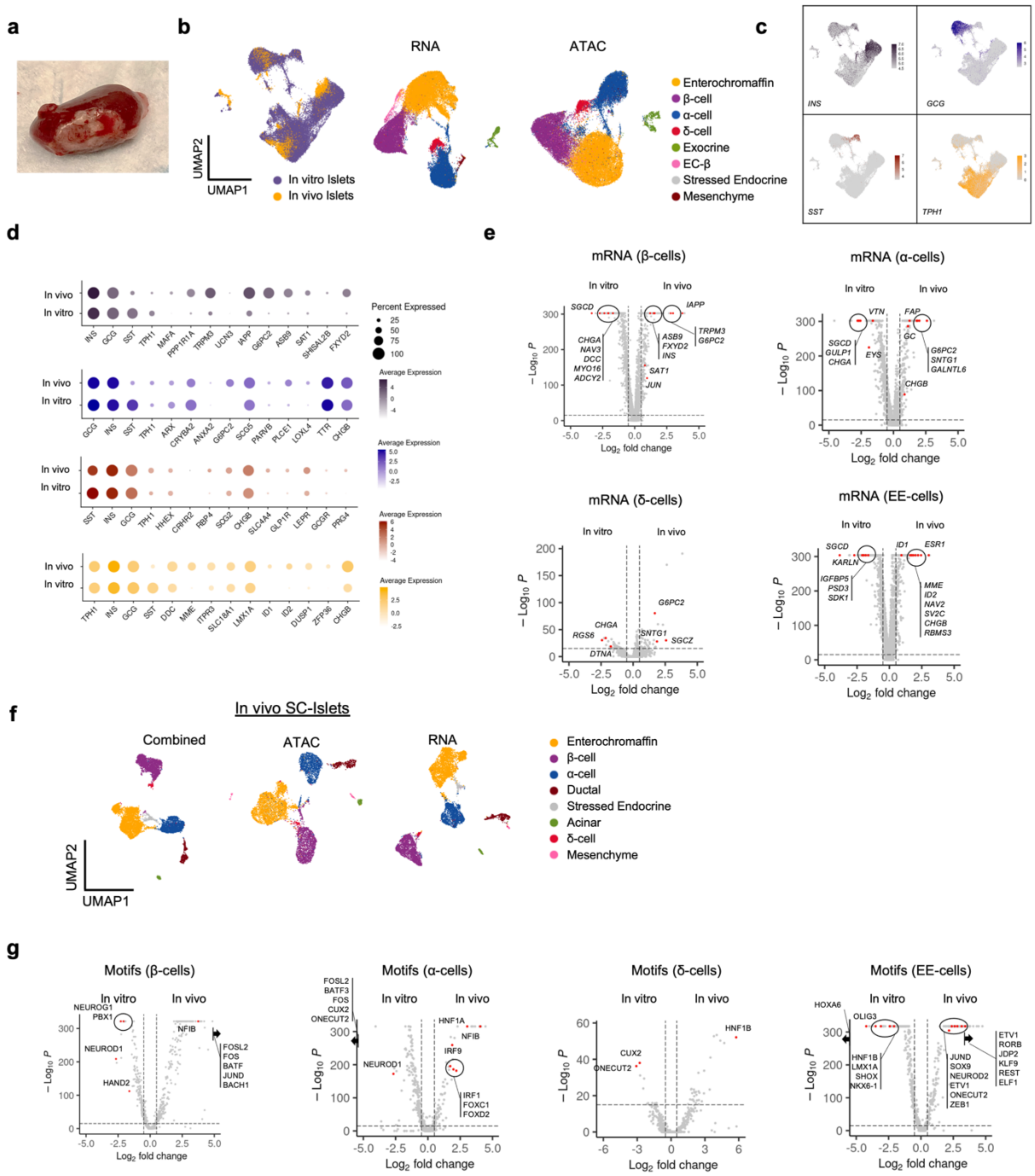



Extended Data Fig. 7 | **Multiomic sequencing comparison of SC-islets and primary human islets.** a, Brightfield images of SC-islets (top) and cadaveric 1° human islets (bottom). b,

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.26.482126; this version posted February 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Integrative UMAP showing cells from SC-islets and 1$^{o}$ islets. c, Feature plots showing gene expression of hormone genes, *INS*, *GCG*, and *SST*, EC marker gene, *TPH1*, and β-cell maturation marker genes, *MAFA*, *IAPP*, *G6PC2*, and *SHISAL2B*. d, Heatmap showing gene expression of identify markers of cell types identified in the analysis. e, UMAP clustering of only 1$^{o}$ human islet samples using both or either ATAC and gene (mRNA) information. f, Feature plot showing expression of hormones and absence of TPH1 in 1$^{o}$ human islet. g, Differential gene expression analysis, highlighting genes in SC-β and 1$^{o}$ β cells. h-i, Differential gene expression and motif enrichment analysis (h) α-cells, (i) δ-cells. j, Dot plots showing promoter accessibility of maturation associated genes in SC-β and 1$^{o}$ β cells. EC, enterochromaffin; 1$^{o}$, primary.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.26.482126; this version posted February 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

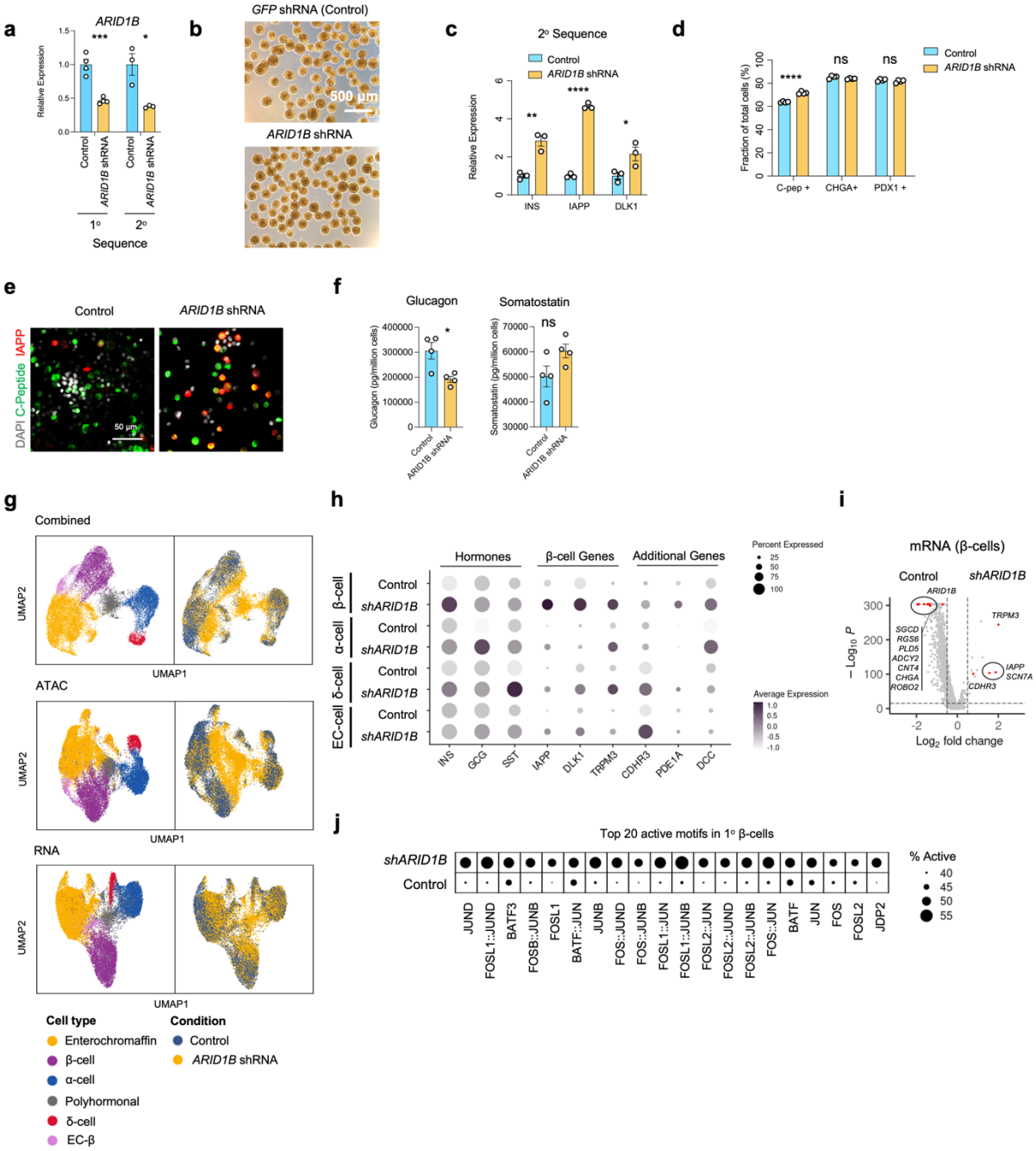



Extended Data Fig. 8 | **Multiomic sequencing comparison of in vitro SC-islets and transplanted in vivo SC-islets.** a, Image of SC-islets transplanted kidney after 6 months. b, Integrative UMAP clustering showing cells from in vitro SC-islets and in vivo SC-islets using both (left), and either gene (mRNA) expression (middle) or ATAC (right) information. c, Feature

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.26.482126; this version posted February 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

plots showing expression of hormone genes and *TPH1*. d, Dot plots comparing gene expression of β, α, δ and EC cell maturation associated genes in SC-β, SC-α, SC-δ and SC-EC cells from in vitro and in vivo SC-islets. e, Differential gene expression analysis of SC-β, SC-α, SC-δ and SC-EC cell populations from in vitro SC-islets and in vivo SC-islets. f, UMAP clustering of only transplanted SC-islet samples using both and either ATAC or gene (mRNA) information. g, Differential motif enrichment analysis of SC-β, SC-α, SC-δ and SC-EC cell populations from in vitro SC-islets and in vivo SC-islets. EC, enterochromaffin; EC-β, enterochromaffin-β cells.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.26.482126; this version posted February 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.




Extended Data Fig. 9 | ***ARID1B* knockdown increases expression of maturation genes and chromatin features in SC-β cells.** a, qPCR analysis, plotting mean ± s.e.m. (n = 4, biological replicates), showing reduced expression of *ARID1B* using lentiviruses with 2 shRNA sequences targeting *ARID1B* mRNA. b, Brightfield images of SC-islets transduced with lentivirus

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.26.482126; this version posted February 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

containing shRNA against *ARID1B* ($1^o$, P = 5.3 x $10^{-4}$; $2^o$, P = 0.018). c, qPCR, plotting mean ± s.e.m. (n = 4, biological replicates), depicting increased expression of β-cell associated genes with the $2^o$ shRNA sequence against *ARID1B* (*INS*, P = 0.0035; *IAPP*, P = 5.9 x $10^{-6}$; *DLK1*, P = 0.045). d, Flow cytometry analysis, plotting mean ± s.e.m. (n = 4, biological replicates), of SC-islets with *ARID1B* knockdown showing increased fraction of cells with C-peptide expression (P = 2.2 x $10^{-5}$), but unaltered fraction of cells expressing PDX1 (ns, P = 0.081), and CHGA (ns, P = 0.24). e, Confocal fluorescent images showing increased expression of amylin in SC-islets with *ARID1B* knockdown. f, ELISA quantification, plotting mean ± s.e.m. (n = 4, biological replicates) of glucagon (P = 0.017), and somatostatin (ns, P = 0.087) content. g, Single-cell multiomic sequencing UMAPs of *ARID1B* knockdown SC-Islets, showing identified cell types, and cell knockdown condition using both and either gene expression or ATAC information. h, Dot plots showing gene expression of β-cell associated genes and other hormones and genes. i, Differential gene expression analysis of SC-β cells comparing control and *ARID1B* knockdown. j, Comparing the top 20 $1^o$ β-cell enriched motifs in SC-islets with *ARID1B* knockdown. *shARID1B*, *ARID1B* knockdown; EC, enterochromaffin; EC-β, enterochromaffin-β cells; $1^o$, primary; $2^o$; secondary; ns, not significant.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.26.482126; this version posted February 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

| | Day | Media | Factor | Final Concentration |
|---|---|---|---|---|
| **Stage 0** | 1 | mTeSR1 | Y27632 | 10 μM |
| | | | | |
| | **Day** | **Media** | **Factor** | **Final Concentration** |
| **Stage 1** | 2 | BE1 | Activin A | 100 ng/mL |
| | | | CHIR9902 | 3 μM |
| **Stage 1** | 3 | BE1 | Activin A | 100 ng/mL |
| **Stage 1** | 4 | BE1 | Activin A | 100 ng/mL |
| **Stage 1** | 5 | BE1 | Activin A | 100 ng/mL |
| | | | | |
| | **Day** | **Media** | **Factor** | **Final Concentration** |
| **Stage 2** | 6 | BE2 | KGF | 50 ng/mL |
| | 7 | BE2 | KGF | 50 ng/mL |
| | | | | |
| | **Day** | **Media** | **Factor** | **Final Concentration** |
| **Stage 3** | 8 | BE3 | LDN19318 | 200 nM |
| | | | KGF | 50 ng/mL |
| | | | SANT1 | 0.25 μM |
| | | | RA | 2 μM |
| | | | TPPB | 500 nM |
| **Stage 3** | 9 | BE3 | LDN19318 | 200 nM |
| | | | KGF | 50 ng/mL |
| | | | SANT1 | 0.25 μM |
| | | | RA | 2 μM |
| | | | TPPB | 500 nM |
| | | | | |
| | **Day** | **Media** | **Factor** | **Final Concentration** |
| **Stage 4** | 10 | BE3 | KGF | 50 ng/mL |
| | | | SANT1 | 0.25 μM |
| | | | RA | 0.1 μM |
| | | | TPPB | 500 nM |
| | | | LDN19318 | 200 nM |
| | 11 | BE3 | KGF | 50 ng/mL |
| | | | SANT1 | 0.25 μM |
| | | | RA | 0.1 μM |
| | | | TPPB | 500 nM |
| | | | LDN19318 | 200 nM |
| | 12 | BE3 | KGF | 50 ng/mL |
| | | | SANT1 | 0.25 μM |
| | | | RA | 0.1 μM |
| | | | TPPB | 500 nM |
| | | | LDN19318 | 200 nM |
| | 13 | BE3 | KGF | 50 ng/mL |
| | | | SANT1 | 0.25 μM |
| | | | RA | 0.1 μM |
| | | | TPPB | 500 nM |
| | | | LDN19318 | 200 nM |

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.26.482126; this version posted February 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

| | Day | Media | Factor | Final Concentration |
|---|---|---|---|---|
| **Stage 5** | 14 | S5 | SANT1 | 0.25 μM |
| | | | RA | 0.1 μM |
| | | | XXI | 1 μM |
| | | | Alk5i | 10 μM |
| | | | T3 | 1 μM |
| | | | Latrunculin | 1 μM |
| | | | Betacellulin | 20 ng/mL |
| | 15 | S5 | SANT1 | 0.25 μM |
| | | | RA | 0.1 μM |
| | | | XXI | 1 μM |
| | | | Alk5i | 10 μM |
| | | | T3 | 1 μM |
| | | | Betacellulin | 20 ng/mL |
| | 16 | S5 | SANT1 | 0.25 μM |
| | | | RA | 0.1 μM |
| | | | XXI | 1 μM |
| | | | Alk5i | 10 μM |
| | | | T3 | 1 μM |
| | | | Betacellulin | 20 ng/mL |
| | 17 | S5 | SANT1 | 0.25 μM |
| | | | RA | 0.1 μM |
| | | | XXI | 1 μM |
| | | | Alk5i | 10 μM |
| | | | T3 | 1 μM |
| | | | Betacellulin | 20 ng/mL |
| | 18 | S5 | SANT1 | 0.25 μM |
| | | | RA | 0.1 μM |
| | | | XXI | 1 μM |
| | | | Alk5i | 10 μM |
| | | | T3 | 1 μM |
| | | | Betacellulin | 20 ng/mL |
| | 19 | S5 | SANT1 | 0.25 μM |
| | | | RA | 0.1 μM |
| | | | XXI | 1 μM |
| | | | Alk5i | 10 μM |
| | | | T3 | 1 μM |
| | | | Betacellulin | 20 ng/mL |
| | 20 | S5 | SANT1 | 0.25 μM |
| | | | RA | 0.1 μM |
| | | | XXI | 1 μM |
| | | | Alk5i | 10 μM |
| | | | T3 | 1 μM |
| | | | Betacellulin | 20 ng/mL |
| | | | | |
| **Stage 6** | Feed S6 every other day | | | |

Supplementary Table 1 | **Differentiation protocol for generating SC-islets from stem cells.**

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.26.482126; this version posted February 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

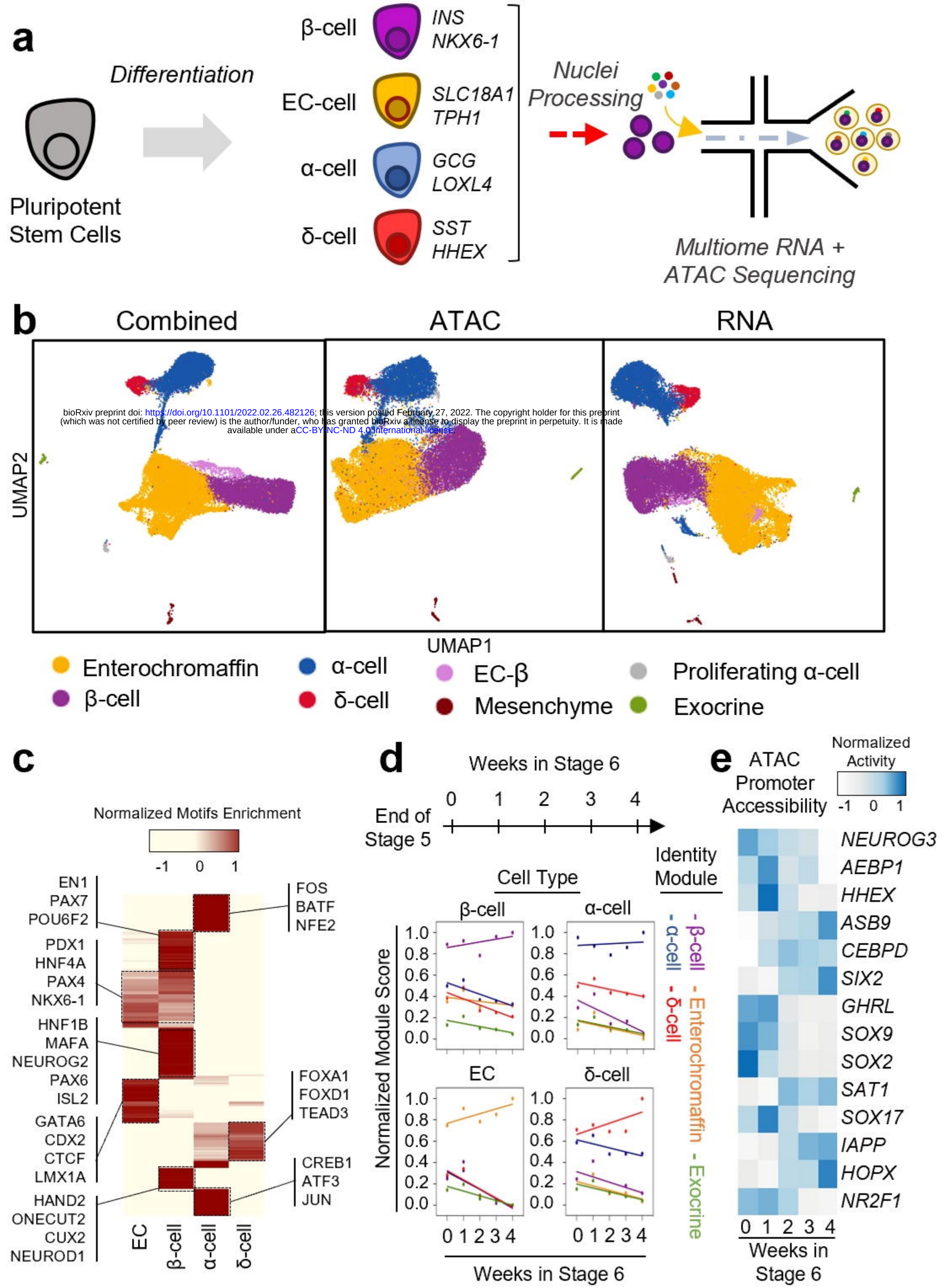
| Reagent | BE1 | BE2 | BE3 | S5 | S6 | Company | Catalog # |
|---|---|---|---|---|---|---|---|
| MCDB131 | 500 mL | 500 mL | 500 mL | 500 mL | 500 mL | Cellgro | 15-100-CV |
| Glucose | 0.8 g | 0.4 g | 0.22 g | 1.8 g | 0.23 g | MilliporeSigma | G7528 |
| NaHCO3 | 0.587 g | 0.587 g | 0.877 g | 0.877 g | N/A | MilliporeSigma | S3817 |
| FAF-BSA | 0.5 g | 0.5 g | 10 g | 10 g | 10.5 g | Proliant | 68700 |
| ITS-X | N/A | N/A | 2.5 mL | 2.5 mL | N/A | Invitrogen | 51500056 |
| Glutamax | 5 mL | 5 mL | 5 mL | 5 mL | 5.2 mL | Invitrogen | 35050079 |
| Vitamin C | N/A | 22 mg | 22 mg | 22 mg | N/A | MilliporeSigma | A4544 |
| Heparin | N/A | N/A | N/A | 5 mg | 10 µg/mL | MilliporeSigma | H3149 |
| Pen/Strep | N/A | N/A | N/A | 5 mL | 5.2 mL | Cellgro | 30-002-CI |
| Non-Essential Amino Acids | N/A | N/A | N/A | N/A | 5.2 mL | Corning | 25-025-CI |
| $ZnSO_4$ | N/A | N/A | N/A | N/A | 1 µM | MilliporeSigma | 10883 |
| Trace Elements A | N/A | N/A | N/A | N/A | 523 µL | Corning | 25-021-CI |
| Trace Elements B | N/A | N/A | N/A | N/A | 523 µL | Corning | 25-022-CI |

Supplementary Table 2 | **Base media formulations.**

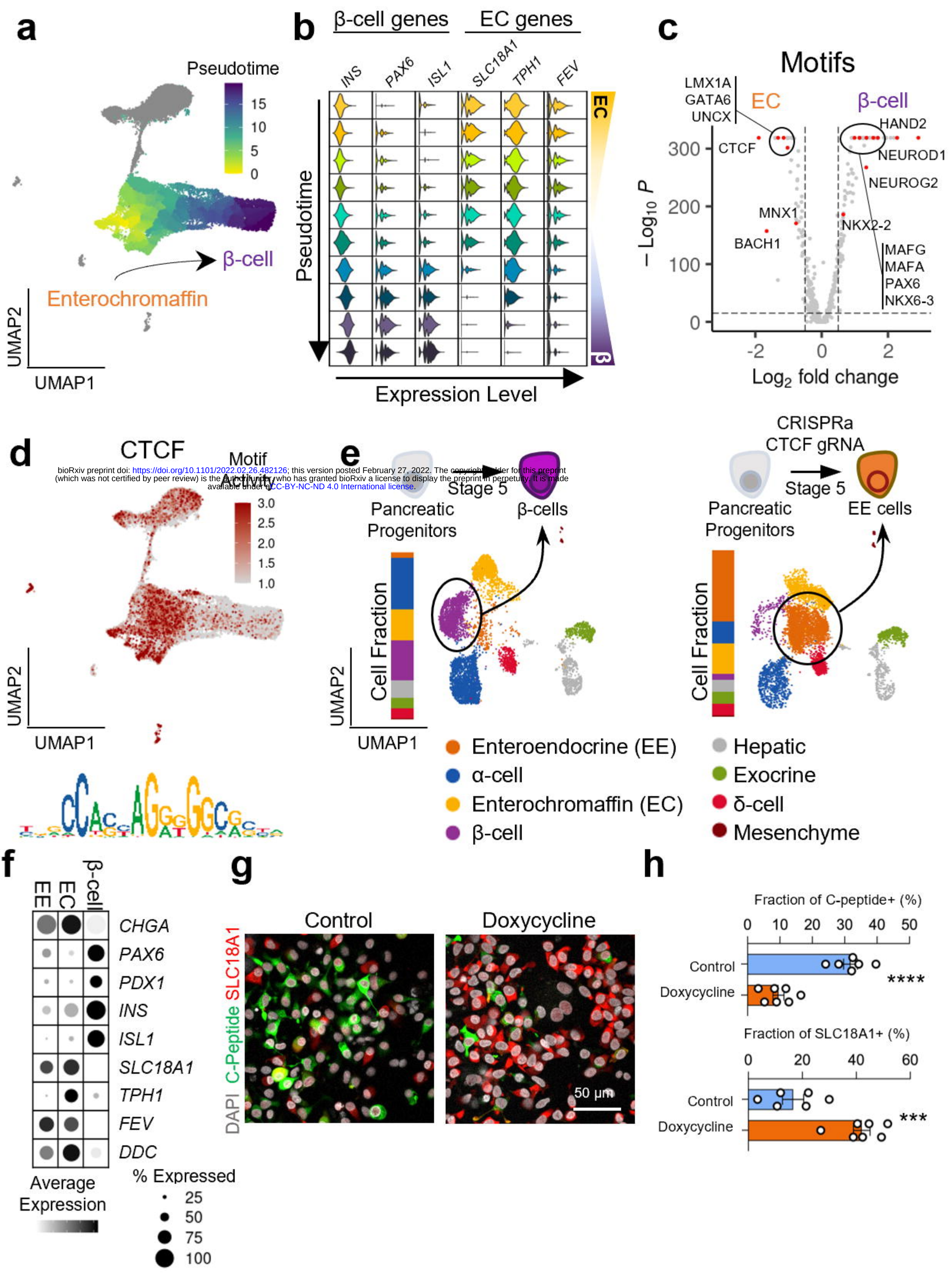
bioRxiv preprint doi: https://doi.org/10.1101/2022.02.26.482126; this version posted February 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

| Gene | Forward primer | Reverse primer |
|---|---|---|
| *ARID1B* | CAAGGGGATCAGAGCAACCC | CTACCTGGGATACTTGCAGGA |
| *CHGA* | TGACCTCAACGATGCATTTC | CTGTCCTGGCTCTTCTGCTC |
| *CTCF* | CAGTGGAGAATTGGTTCGGCA | CTGGCGTAATCGCACATGGA |
| *dCas9* | GTGACCGAGGGAATGAGAAA | AGCTGCTTCACGGTCACTTT |
| *DDC* | TGGGGACCACAACATGCTG | TCAGGGCAGATGAATGCACTG |
| *DLK1* | CTTTCGGCCACAGCACCTAT | TGTCATCCTCGCAGAATCCAT |
| *FEV* | CACGGCGAGTTCAAGCTCA | CTGGAAGTCGAAGCGGTAGG |
| *G6PC2* | TGGTATGTCATGGTAACCGC | CACTCCAAAGAAATGACCAGG |
| *GCG* | AGCTGCCTTGTACCAGCATT | TGCTCTCTCTTCACCTGCTCT |
| *IAPP* | ACATGTGGCAGTGTTGCATT | TCATTGTGCTCTCTGTTGCAT |
| *INS* | CAATGCCACGCTTCTGC | TTCTACACACCCAAGACCCG |
| *ISL1* | TCACGAAGTCGTTCTTGCTG | CATGCTTTGTTAGGGATGGG |
| *LMX1A* | GCAAAGGGGACTATGAGAAGGA | CGTTTGGGGCGCTTATGGT |
| *NEUROD1* | ATCAGCCCACTCTCGCTGTA | GCCCCAGGGTTATGAGACTAT |
| *NKX22* | GGAGCTTGAGTCCTGAGGG | TCTACGACAGCAGCGACAAC |
| *NKX6-1* | CCGAGTCCTGCTTCTTCTTG | ATTCGTTGGGGATGACAGAG |
| *PDX1* | CGTCCGCTTGTTCTCCTC | CCTTTCCCATGGATGAAGTC |
| *SLC18A1* | GTGGTGGTATTCGTCGCTTTG | CCGAGGTGCAGAGAAGAGT |
| *SST* | TGGGTTCAGACAGCAGCTC | CCCAGACTCCGTCAGTTTCT |
| *TPH1* | ACGTCGAAAGTATTTTGCGGA | ACGGTTCCCCAGGTCTTAATC |
| *TBP* | GCCATAAGGCATCATTGGAC | AACAACAGCCTGCCACCTTA |
| *GUSB* | CGTCCCACCTAGAATCTGCT | TTGCTCACAAAGGTCACAGG |

Supplementary Table 3 | **Primer sequences for qPCR.**

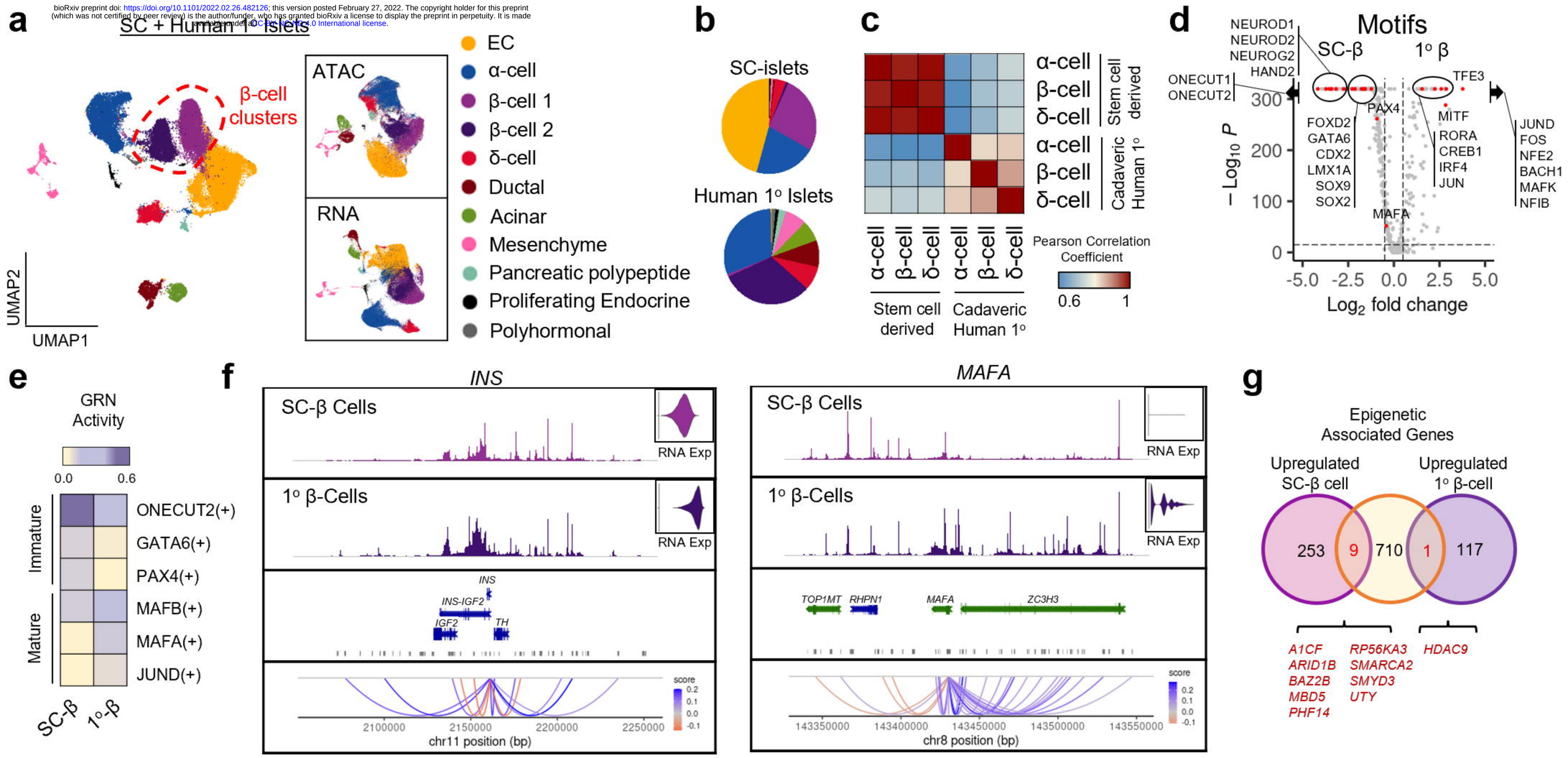


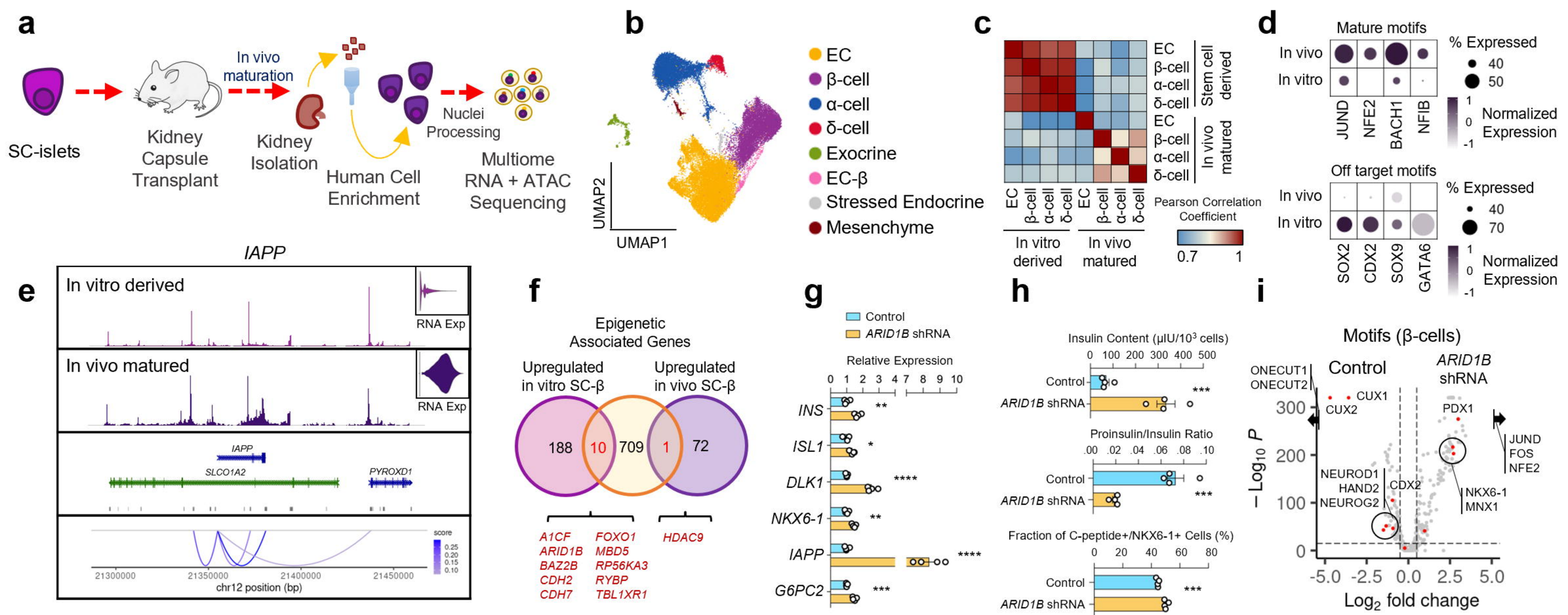

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.26.482126; this version posted February 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.




bioRxiv preprint doi: https://doi.org/10.1101/2022.02.26.482126; this version posted February 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.26.482126; this version posted February 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.




a
SC-islets
Kidney Capsule Transplant
In vivo maturation
Kidney Isolation
Human Cell Enrichment
Nuclei Processing
Multiome RNA + ATAC Sequencing
b
UMAP2
UMAP1
EC
β-cell
α-cell
δ-cell
Exocrine
EC-β
Stressed Endocrine
Mesenchyme
c
EC
β-cell
α-cell
δ-cell
Stem cell derived
In vivo matured
In vitro derived
Pearson Correlation Coefficient
0.7
1
d
Mature motifs
In vivo
In vitro
JUND
NFE2
BACH1
NFIB
% Expressed
40
50
Normalized Expression
Off target motifs
SOX2
CDX2
SOX9
GATA6
70
e
IAPP
In vitro derived
In vivo matured
RNA Exp
SLCO1A2
PYROXD1
score
chr12 position (bp)
f
Epigenetic Associated Genes
Upregulated in vitro SC-β
Upregulated in vivo SC-β
188
10
709
1
72
A1CF
ARID1B
BAZ2B
CDH2
CDH7
FOXO1
MBD5
RP56KA3
RYBP
TBL1XR1
HDAC9
g
Control
ARID1B shRNA
Relative Expression
INS
ISL1
DLK1
NKX6-1
IAPP
G6PC2
h
Insulin Content (μIU/10³ cells)
Proinsulin/Insulin Ratio
Fraction of C-peptide+/NKX6-1+ Cells (%)
i
Motifs (β-cells)
Control
ARID1B shRNA
ONECUT1
ONECUT2
CUX1
CUX2
PDX1
JUND
FOS
NFE2
NEUROD1
HAND2
NEUROG2
CDX2
NKX6-1
MNX1
−Log10 P
Log2 fold change