Title

Single-cell chromatin accessibility of developing murine pancreas identifies cell state-specific gene regulatory programs

Permalink

https://escholarship.org/uc/item/5dh37542

Authors

de la O, Sean
Yao, Xinkai
Chang, Sean
et al.

Publication Date

2023-07-01

DOI

10.1016/j.molmet.2023.101735

Peer reviewed

MOLECULAR METABOLISM

Single-cell chromatin accessibility of developing murine pancreas identifies cell state-specific gene regulatory programs

Sean de la O [1,2,3,4], Xinkai Yao [1,2,3,4], Sean Chang [1,2,3,4], Zhe Liu [1,2,3,4], Julie B. Sneddon [1,2,3,4,]*

ABSTRACT

Numerous studies have characterized the existence of cell subtypes, along with their corresponding transcriptional profiles, within the developing mouse pancreas. The upstream mechanisms that initiate and maintain gene expression programs across cell states, however, remain largely unknown. Here, we generate single-nucleus ATAC-Sequencing data of developing murine pancreas and perform an integrated, multi-omic analysis of both chromatin accessibility and RNA expression to describe the chromatin landscape of the developing pancreas at both E14.5 and E17.5 at single-cell resolution. We identify candidate transcription factors regulating cell fate and construct gene regulatory networks of active transcription factor binding to regulatory regions of downstream target genes. This work serves as a valuable resource for the field of pancreatic biology in general and contributes to our understanding of lineage plasticity among endocrine cell types. In addition, these data identify which epigenetic states should be represented in the differentiation of stem cells to the pancreatic beta cell fate to best recapitulate *in vitro* the gene regulatory networks that are critical for progression along the beta cell lineage *in vivo*.

1. INTRODUCTION

Development of the mammalian pancreas requires the coordination of multiple cell lineages over time, culminating in the generation of a highly branched, mature organ consisting of both an exocrine and endocrine compartment. Specification of the murine pancreas begins at embryonic day (E) 8.5 with the expression of the transcription factor (TF) pancreatic and duodenal homeobox 1 (*Pdx1*) in a focal region of the endoderm-derived primitive foregut [1,2]. These Pdx1(+) cells give rise to all of the epithelial lineages of the pancreas (duct, endocrine, and acinar) [3] and by E9 evaginate into the surrounding mesenchyme and begin to form a stratified epithelium. As branching morphogenesis progresses, regionalization of the epithelium results in the formation of both trunk and tip domains by E12.5. Cells located at the tip, marked by the expression of *Cpa1*, serve as multipotent progenitors that give rise to all three epithelial cell types until E13.5, at which point they undergo fate restriction to only give rise to acinar cells [3—5]. Epithelial cells located in the trunk give rise to either ductal or endocrine lineages, a fate choice dependent on levels of Notch signaling [5].
Endocrine progenitor (EP) cells derive from a subset of ductal epithelial cells that experience lower levels of Notch and then activate expression of the TF neurogenin3 (*Neurog3*) [5]. *Neurog3* expression marks early EP cells, which give rise to the main hormone-producing endocrine cells in the pancreas: alpha, beta, delta, and gamma [3,6]. Gene knockout studies in mice revealed that the expression of a number of TFs that are critical for differentiation and maintenance of pancreatic endocrine lineages, such as paired box gene 4 (*Pax4*) and 6 (*Pax6*), neurogenic differentiation 1 (*Neurod1*), and LIM-homeodomain protein Islet 1 (*Isl1*), is dependent on *Neurog3* [6].
Endocrine cell identity is specified and maintained by a complex network of TFs, many of which play dynamic roles throughout developmental time [7]. For instance, early in development *Pdx1* is required for specification of pancreatic progenitors, but later in development it is also important for the generation of beta cells and for the maintenance of beta cell identity [8,9]. Along with *Pdx1* and *Pax4*, the TFs NK2 homeobox 2 (*Nkx2-2*) and NK6 homeobox 1 (*Nkx6-1*) are critical factors for beta cell differentiation, while aristaless related homeobox (*Arx*) is essential for alpha cell differentiation. *Arx* and *Pax4* play mutually opposing roles in the differentiation of alpha and beta cells, with *Arx* promoting the generation of alpha at the expense of beta and delta cells [10] and *Pax4* regulating the decision towards beta and delta at the expense of alpha and epsilon cell fate [10—12]. When both

[1]Department of Cell and Tissue Biology, University of California, San Francisco, San Francisco, CA, 94143, USA [2]Department of Anatomy, University of California, San Francisco, San Francisco, CA, 94143, USA [3]Diabetes Center, University of California, San Francisco, San Francisco, CA, 94143, USA [4]Eli and Edythe Broad Center of Regeneration Medicine and Stem Cell Research, University of California, San Francisco, San Francisco, CA, 94143, USA

*Corresponding author. Department of Cell and Tissue Biology, University of California, San Francisco, San Francisco, CA, 94143, USA. E-mail: julie.sneddon@ucsf.edu (J.B. Sneddon).

Arx and *Pax4* are lost, delta cells persist but both alpha and beta cells are lost [11]. Expression of *Nkx2-2* and *Nkx6-1* follows that of *Pdx1* in early pancreatic progenitors, then becomes progressively restricted to endocrine cells [13,14]. Deletion of *Nkx2-2* results in a significant reduction of the four major endocrine cell types and an increase in ghrelin-producing epsilon cells [12,14]. *Nkx6-1* functions downstream of *Nkx2-2* and is necessary for beta cell neogenesis through the maintenance and/or expansion of beta cell precursors following *Neurog3* expression but prior to the production of insulin, while later it is lost from developing alpha cells [13,15].

Extrinsic signals derived from non-epithelial cells are also important in guiding pancreatic organogenesis. Early pioneering work using pancreatic explants *ex vivo* showed that when E11 epithelial buds were cultured without their surrounding mesenchymal tissue, epithelial growth and differentiation were arrested [16]. More recently, genetic ablation studies have demonstrated the requirement for pancreatic mesenchyme for expansion of the pool of early pancreatic progenitor cells early in development and for proliferation of differentiated cells later in development [17,18]. Although the pancreatic mesenchyme is broadly appreciated as playing an important role in pancreatic organogenesis, however, it is still not well understood whether there exist biologically relevant sub-populations of mesenchyme with distinct lineages and/or functional roles.

Recent single-cell RNA-Sequencing (scRNA-Seq) studies have highlighted previously unappreciated levels of cellular heterogeneity among the epithelial cells of the developing murine pancreas, particularly within the endocrine compartment [19—23]. Although relatively less attention has been given to elucidating potential cellular heterogeneity within the mesenchymal compartment, evidence from scRNA-Seq and classical genetic lineage tracing experiments suggests that transcriptionally distinct mesenchymal cell types also exist during development [19,20,24]. As a result of this body of work, we now have a greater understanding of the transcriptomic cues governing cell states across pancreatic development, but we still lack an understanding of the upstream epigenetic features that regulate cell fate decisions. In particular, integration of gene expression data and chromatin accessibility data would permit identification of active transcription factor binding to accessible chromatin within a given cell type.

In recent years, Assay for Transposase-Accessible Chromatin followed by Sequencing (ATAC-Seq) has been developed to profile genome-wide chromatin accessibility for epigenetic analysis in a given cell type or tissue [25]. This technique has been applied to sorted populations of endocrine cells from the murine pancreas to investigate the chromatin landscape of developing EP cells [20,26]. These studies, however, lacked single-cell resolution to capture the chromatin states of the various subpopulations of developing endocrine cells that have been described [19—23]. More recently, single-nucleus ATAC-Seq (snATAC-Seq) has emerged as a technology to provide insights into chromatin accessibility at single-cell resolution [27,28]. snATAC-Seq has been used to profile the chromatin landscape of many developing tissue types and has revealed cell-type specific *cis*- and *trans*-regulatory elements governing gene expression and cell fate decisions [29—33]. Furthermore, integration of scRNA- and snATAC-Seq data for multi-omic analysis permits refinement of expressed TFs to a further parsed subset that are not only expressed but are also likely binding TF motifs in open regions of chromatin and actively regulating expression of downstream target genes.

Here, we generate snATAC-Seq data of developing murine pancreas and perform an integrated multi-omic analysis of both chromatin accessibility and RNA expression. We describe at single-cell resolution the chromatin landscape of the developing epithelium at E14.5 and E17.5, stages at which the dynamic processes of expansion, differentiation, and morphogenesis are actively underway. We identify candidate TFs regulating transitions across the endocrine lineages and construct gene regulatory networks (GRNs) of active TFs binding to regulatory regions of downstream target genes. Additionally, we generate a snATAC-Seq dataset of developing pancreatic mesenchyme at E14.5, which to our knowledge represents the first ATAC-Seq dataset (bulk or single-nucleus) of this cell type. We believe that these datasets and analyses will serve as a valuable resource for the field of pancreatic biology in general, and will contribute to our understanding of lineage plasticity among endocrine cell types. In addition, these data will serve as a reference as to which epigenetic states should be represented in the differentiation of stem cells to the pancreatic beta cell fate to best recapitulate *in vitro* the gene regulatory networks that are critical for progression along the beta cell lineage *in vivo*.

2. METHODS

2.1. Animal studies

All mouse procedures were approved by the University of California, San Francisco (UCSF) Institutional Animal Care and Use Committee (IACUC). Mice were housed in a 12-hour light—dark cycle in a controlled temperature climate. Noon of the day of a vaginal plug was considered embryonic day (E) 0.5.

eFev-EYFP (ePet1-EYFP) mice were kindly donated by Dr. Evan Deneris, and have been previously described [34,35]. Mice were maintained on a C57BL/6J background. Wildtype C57BL/6J mice used for breeding and for the whole pancreas samples were obtained from the Jackson Laboratory. Genotyping of eFev-EYFP mice was conducted on tail DNA, with forward primer TGCGATGGGAAGATAAGAGGGG and reverse primer GAAGTTCACCTTGATGCCGTTC.

2.2. Histology, immunofluorescence, and imaging

E14.5 pancreata were dissected in ice cold 1x PBS, then fixed in 4% paraformaldehyde (PFA) overnight at 4 C. After washing three times in 1x PBS, tissues were preserved in 30% sucrose in PBS at 4 C overnight and then embedded in Optimal Cutting Temperature (O.C.T.) compound (Tissue-Tek) and flash frozen prior to sectioning at 10 μm thickness.

For immunofluorescence staining, cryosections were washed 3 times in 1x PBS, permeabilized in 0.5% triton X-100 in PBS (PBT) for 10 min at room temperature (RT), and then blocked with 5% normal donkey serum (NDS) in 0.1% PBT for 1 h. Sections were stained overnight at 4 C using primary antibodies against GFP (1:500, Abcam Cat. ab13970), Chga (1:250, Abcam Cat. ab15160), or E-cadherin (1:100, BD Transduction Laboratories Cat. 610182). The next day, sections were washed three times in 1x PBS and then incubated with species-specific Alexa 488-, 555-, or 647-conjugated secondary antibodies and DAPI in 5% NDS in 0.1% PBT for 1 h at RT. Sections were washed three times in 1x PBS and covered in Fluoromount-G mounting solution (SouthernBiotech, Cat. 0100—01).

Images were captured with an SP8 Leica confocal laser scanning microscope. Maximum intensity Z-projections were then prepared using Image J software [36].

2.3. *In situ* hybridization

Multiplexed *in situ* hybridization/immunofluorescence was performed with RNAscope technology using probes purchased from Advanced Cell Diagnostics, Inc. Probes against mouse *Fev* (Cat. 413241), *Neurog3* (Cat. 422401), *Mlxipl* (Cat. 558141), *Etv1* (Cat. 557891), *Nhlh1*

(Cat. 585751), *Spp1* (Cat. 435191), and *Rest* (Cat. 316251) were used according to the manufacturer's instructions for the RNAscope multiplex fluorescent detection V2 kit (Advanced Cell Diagnostics, Inc., Cat. 323110). 10 μm thick cryosections were brought to RT, washed with PBS to remove O.C.T., and treated with hydrogen peroxide and proteinase III. Tissue was hybridized with the probe mixture for 2 h at 40 C. Hybridization signals were amplified via sequential hybridization of amplifier AMP1, AMP2, and AMP3 and label probes Opal 570 (1:1500, PerkinElmer, Cat. FP1488001KT), Opal 650 (1:1500, PerkinElmer, Cat. FP1496001KT), and Opal 690 (1:1500, PerkinElmer, Cat. FP1497001KT).

Following signal amplification of the target probes, sections were incubated in 1x blocking buffer for 1 h at RT, followed by staining with primary antibodies against GFP (1:500, Abcam Cat. ab13970), Epcam (1:100, BD Biosciences, Cat. 552370), Ins (1:100, Abcam Cat. ab7842), or Gcg (1:250, Cell Signaling Technology Cat. 2760S). The next day, sections were washed three times with 1x PBS and then incubated with species-specific Alexa 488- or Alexa 555-secondary antibodies and DAPI in 5% NDS in 0.1% PBT for 1 h at RT. Sections were then washed three times in 1x PBS, mounted with ProLong Gold Antifade Mountant (Invitrogen, Cat. P36930), and stored at 4 C prior to imaging. Optical sectioning images were taken with a Leica confocal laser scanning SP8 microscope equipped with white light sources. 10 steps X 1 mm thickness Z-sections were captured for each imaging area.

2.4. Dissociation and sorting of murine pancreas tissue for quantitative RT-PCR

E14.5 pancreata were dissected from embryos of pregnant eFev-EYFP dams and kept in separate wells of a 96-well plate. EYFP fluorescence was assessed under a microscope to confirm the genotype of each pancreas. Pancreata with EYFP fluorescence (EYFP($+$)) were then pooled together and transferred to a 1.5 ml microcentrifuge tube, then dissociated into single cells by incubating with 250 ul of TrypLE Express dissociation reagent (Gibco, Cat. 12604013)) at 37 C for 20 min, with pipet trituration at 5 min intervals. Dissociation was neutralized with FACS buffer (10% FBS $+$ 2 mM EDTA in phenol-red free HBSS), and the single-cell suspensions were passed through 30 μm cell strainers.

Cells were stained with SYTOX Blue dead cell stain (Invitrogen, Cat. S34857) to remove dead cells, then with a PE-conjugated antibody against mesenchymal marker CD140a (1:50; eBioscience Cat. 12-1401-81) and an APC-conjugated antibody against epithelial marker CD326/Epcam (1:50; eBioscience Cat. 17-5791-82) at 4 C for 30 min. Stained cells were washed twice in FACS buffer and sorted using a BD FACSAria II cell sorter (BD Biosciences). After size selection to remove debris and doublets and sorting on SYTOX Blue negative (live) events, cells were further subgated on CD140a($-$)/CD326($+$) (epithelial) cells and then on EYFP fluorescence.

RNA was extracted from EYFP($-$), EYFP-low, and EYFP-high sorted cells with the RNeasy Mini Kit (Qiagen, Cat. 74106). Reverse transcription was performed with the PrimeScript High Fidelity RT-PCR Kit (Takara, Cat. R022A). RT-PCR was run on a 7900HT Fast RT-PCR instrument (Applied Biosystems) with Taqman probes for *Fev* (assay ID: Mm00462220_m1, Cat. 4331182) and *GAPDH* (assay ID: Mm99999915_g1, Cat. 4331182) in triplicate. Data were normalized to *GAPDH*. Error bars represent standard error of the mean (SEM).

2.5. Dissociation and sorting of murine pancreas tissue for snATAC-Seq

For the whole pancreas samples, C57BL/6J embryonic pancreata (n $=$ 10 from 3 litters for E14.5 replicate 1 and n $=$ 16 from 3 litters for

E14.5 replicate 2; n $=$ 7 from 1 litter for E17.5 replicate 1 and n $=$ 3 from 1 litter for E17.5 replicate 2) were pooled into a single 1.5 ml microcentrifuge tube per replicate, per timepoint. Biological replicate experiments for each timepoint were performed on different days. For the eFev-EYFP sample, E14.5 pancreata (n $=$ 15) were dissected from embryos of two pregnant eFev-EYFP dams, and each pancreas was kept in a separate well of a 96-well plate. EYFP fluorescence was assessed under a microscope to confirm the genotype of each pancreas. E14.5 pancreata with EYFP fluorescence (EYFP($+$)) (n $=$ 5) were then pooled together, pancreata without EYFP fluorescence (EYFP($-$)) (n $=$ 10) were pooled together as a negative control, and each sample was transferred to a separate 1.5 ml microcentrifuge tube.

The E14.5 whole pancreas, E17.5 whole pancreas, E14.5 EYFP($+$), and E14.5 EYFP($-$) samples were dissociated into single cells by incubating with 250$-$350 ul per sample of TrypLE Express dissociation reagent (Gibco, Cat. 12604013)) at 37 C for 20 min, with pipet trituration at 5 min intervals. Dissociation was neutralized with FACS buffer (10% FBS $+$ 2 mM EDTA in phenol-red free HBSS) and the single-cell suspensions were passed through 37 μm cell strainers.

All samples were stained with SYTOX Blue dead cell stain (Invitrogen, Cat. S34857) to remove dead cells. Cells were washed twice with FACS buffer and sorted using a BD FACSAria II cell sorter (BD Biosciences). After size selection to remove debris and doublets, all cells were sorted on SYTOX Blue negative (live) events, and the E14.5 EYFP($+$) and EYFP($-$) samples were further subgated on EYFP fluorescence. Live cells from whole pancreas and live EYFP($+$) cells from the EYFP($+$) sample were collected into separate tubes containing 1x FACS buffer and immediately subjected to extraction of nuclei as described below.

2.6. Extraction of nuclei

All buffers (e.g., 0.1x lysis buffer, lysis dilution buffer, and wash buffer) were freshly prepared according to the 10x Genomics Demonstrated protocol (CG000212 RevC) and maintained at 4 C. Nuclei were isolated from whole pancreas (25,000 cells for E14.5 replicate 1 and 50,000 cells for E14.5 replicate 2; 1,000,000 cells for E17.5 replicate 1 and 500,000 cells for E17.5 replicate 2) or from Fev-high (EYFP($+$)) cells (25,000 cells for E14.5) using the demonstrated protocol. Sorted cells were added to a 2 ml microcentrifuge tube and centrifuged at 500 rcf for 5 min at 4 C. All supernatant was removed without disrupting the cell pellet. 100 ul chilled 0.1x lysis buffer was then added and pipetted 5 times to fully mix the buffer with the cells, then incubated for 3 min on ice to achieve full cell lysis. 1 ml chilled wash buffer was added to the lysed cells to terminate the lysis. Lysed cells were centrifuged at 500 rcf for 5 min at 4 C, and supernatant was gently removed. Nuclei were resuspended in 50 ul wash buffer, transferred to a 200 ul tube, and spun down and resuspended in 10 ul 1x Nuclei buffer (10x Genomics, Part Number 2000153). 2 ul of the suspension was loaded onto a hemacytometer to determine the concentration of nuclei and simultaneously assess nucleus quality. High-quality nuclei from the whole pancreas and eFev-EYFP samples were then used for downstream library construction and sequencing.

2.7. snATAC-seq capture, library construction, and sequencing

Input nuclei were subjected to transposition, partitioning, and library construction using 10x Genomics Chromium Next GEM Single Cell ATAC Reagent Kit v1.1 Chemistry, according to the manufacturer's instructions. An Agilent Fragment Analyzer was used for assessing the fragment distribution of both the whole pancreas and eFev-EYFP libraries, which were run on the Illumina NovaSeq 6000 platform.

2.8. Clustering of murine scRNA-Seq data

For clustering of murine scRNA-Seq data for integration with our snATAC-Seq data, we applied the clustering algorithm CellFindR [37] to our previously published scRNA-Seq dataset of developing murine pancreas tissue [19]. 10x Genomics outputs of E14.5 and E17.5 pancreata were downloaded from the Gene Expression Omnibus (GEO) (GSE101099; samples GSM3140916, GSM3140919, GSM3140920, GSM2699157, GSM3140917, and GSM3140918), and analyzed with Seurat v3.2.3. Seurat objects were created from each 10x output with Read10x() and CreateSeuratObject() and filtered to retain high quality cells (nFeature_RNA >1250 and percent.mt < 7 for GSM3140916; nFeature_RNA >1500 and percent.mt < 5 for GSM3140919 and GSM3140920; nFeature_RNA >500 and <3000 and percent.mt < 5 for GSM2699157; nFeature_RNA >500 and <5000 and percent.mt < 5 for GSM3140917 and GSM3140918). The datasets were then normalized and variable features calculated with NormalizeData() and FindVariableFeatures(), respectively. The samples were then integrated using Seurat's standard batch correction method [38] with SelectIntegrationFeatures(), FindIntegrationAnchors(), and IntegrateData(). The integrated object was then scaled with ScaleData() and principal component analysis (PCA) performed with RunPCA(). UMAP dimensional reduction was calculated with RunUMAP() with dims = 1:30. Neighbors were found in the dataset with FindNeighbors with dims = 1:30 and clustering performed with FindClusters(), resolution = 0.2 for E14.5 and 0.4 for E17.5. Next, broad cell types were manually annotated based on expression of known marker genes (i.e. *Col3a1* for mesenchyme) and used for subsequent iterative subclustering.

For iterative subclustering with CellFindR, each broad cell type (Mesenchyme, Mesothelium, Exocrine, and Endocrine) was subsetted individually. PCA, Neighbors, and UMAP were recalculated as described above (Mesenchyme: dims 1:15; Mesothelium: dims 1:15 for E14.5 and 1:10 for E17.5; Exocrine: dims 1:10 for E14.5 and 1:15 for E17.5; and Endocrine: dims 1:10) and the first clustering resolution calculated with find_res() from CellFindR. Iterative subclustering was then performed with sub_clustering(). Subclusters that displayed characteristics of doublets (expressing markers of more than one broad group e.g., Col3a1(+)/Cpa1(+) acinar cells) or low quality (e.g. clustering based on high mitochondrial gene content) were manually removed.

2.9. snATAC-Seq analysis

FASTQ files were generated from raw sequencing reads using the bcl2fastq function from Illumina. BAM files and single-cell accessibility counts were generated using the cellranger-atac count function from Cell Ranger software (version 1.0.1 for the E14.5 datasets, version 2.1.0 for the E17.5 datasets). Reference genome used was *Mus musculus* assembly mm10, annotation gencode.vM17.basic. Files processed with Cell Ranger ATAC were then analyzed using ArchR (version 1.0.1) [39].

Unless otherwise noted, parameters and function calls were kept the same between the E14.5 and E17.5 datasets. First, ArchR Arrow files were created with the ArrowFiles() function with default settings. For 14.5, an ArchR project was then created using both whole pancreas and EYFP(+) sorted cells with ArchRProject(), resulting in an initial dataset consisting of 46,453 nuclei. For E17.5, an ArchR project was created using the whole pancreas sample, resulting in an initial dataset of 113,071 nuclei. Each project was then filtered for high quality nuclei (TSS enrichment ≥ 10 and number of fragments ≥ 3,000) (39,845 nuclei retained in E14.5; 102,644 retained in E17.5), and doublets removed with addDoubletScores() and filterDoublets() (35,823 nuclei

retained in E14.5; 94,175 retained in E17.5). Next, iterative LSI was performed on both E14.5 and E17.5 datasets with the addIterativeLSI() function, with clustering parameters of resolution = 0.2, sampleCells = 10,000, n. start = 10, varFeatures = 25,000. Clustering was performed with addClusters() with resolution = 0.1 and method = "Seurat". UMAP dimensional reduction was performed with addUMAP(), with minDist = 0.5. Clusters were manually annotated based on the Gene Score of known marker genes with addImputeWeights() and then visualized by UMAP. For E14.5, cluster 9 was removed as no discernible markers for the known pancreatic cell types were detected, resulting in a final dataset consisting of 33,206 total cells with a median number of 14,756 fragments per cell and a median TSS enrichment score of 14.914. For E17.5, each Broad Group was subsetted and re-clustered at higher resolution to separate out potential doublet clusters or empty droplets that were not removed with the doublet filtering detailed above. Doublet/empty droplet clusters were removed from the Mesenchymal, Epithelial, and Endothelial Broad Groups, resulting in a final dataset consisting of 78,669 nuclei with a median number of 15,965 fragments per cell and a median TSS enrichment score of 21.037.

For epithelial analysis, epithelial (exocrine and endocrine) nuclei were subsetted based on accessibility of known marker genes (*Cpa1, Spp1, Chga*). Iterative LSI was recalculated with iterations = 4 for E14.5 and = 2 for E17.5, resolutions = 0.1, 0.2, 0.4, and 0.5 for E14.5 and = 0.5 for E17.5, sampleCells = 5,000 for E14.5 and 10,000 for E17.5, n. start = 10, and varFeatures = 15,000 for E14.5 and 25,000 for E17.5. Clustering was performed with resolution = 0.9 for E14.5 and = 1.2 for E17.5, and UMAP recalculated with minDist = 0.5. Clustered epithelial cells from the scRNA-Seq data described above were used for unconstrained integration with addGeneIntegrationMatrix(). Chromatin accessibility peaks were then called with Macs2 via ArchR with addGroupCoverage(), addReproduciblePeakSet() and addPeakMatrix(). Marker peaks within the epithelial compartment were calculated with getMarkerFeatures using the "PeakMatrix". For motif analysis within marker peaks, motif annotations were added with addMotifAnnotations() with the "cisbp" motif set and then calculated with peakAnnoEnrichment(). ChromVAR [40] analysis was performed with addBgdPeaks() and addDeviationsMatrix(). Correlated transcription factors were correlated between the "GeneIntegrationMatrix" (RNA expression from the unconstrained integration) and "MotifMatrix" (ChromVAR motif deviations) with correlateMatrices(), keeping TFs with a correlation >0.5, padj <0.01 and max delta greater than 0.5 of the upper quartile.

For pseudotime lineage calculations, we manually imputed the cell states for each cell lineage (Alpha and Beta for E14.5; Alpha, Beta, and Delta for E17.5) and computed the pseudotime values with addTrajectory().

To compare peaks across the E14.5 whole pancreas and EYFP(+) sorted samples, we first subsetted the epithelial dataset to retain all cell types that had sufficient cell numbers to call peaks after being split by sample. This included the following clusters: Neurog3(+), Fev(+)/Chgb(+), Pdx1(+)/Mafb(+), Alpha, and Beta cells. We then split this subsetted dataset according to the sample from which they were derived and then re-called peaks with addGroupCoverages(), addReproduciblePeakSet() and addPeakMatrix(). We extracted these peaks with getPeakset() for each sample and assessed overlap with findOverlaps() from the R package IRanges (https://github.com/Bioconductor/IRanges).

For mesenchymal and mesothelial analysis in the E14.5 datasets, nuclei were subsetted based on accessibility of known marker genes (*Col3a1, Wt1*). Iterative LSI was recalculated with iterations = 2,

resolution = 0.5, sampleCells = 10,000, n.start = 10, and varFeatures = 25,000. Clustering was performed with resolution = 0.3, and UMAP recalculated with minDist = 1.5. Clustered mesenchymal and mesothelial cells from the scRNA-Seq data described above were used for unconstrained integration. Peak calling, marker peak identification, motif analysis, ChromVAR analysis, and transcription factor correlation were performed as described above.

2.10. Gene regulatory network analysis

Gene regulatory networks (GRNs) were constructed as described by Lyu et al., 2021 [31] (https://github.com/Pinlyu3/IReNA-v2). Candidate cluster-enriched genes were calculated with the scRNA-Seq dataset of epithelial or mesenchymal cells with Seurat's FindAllMarkers() with min.pct = 0.1 and logfc.threshold = 0.25. DEGs were retained with an average logFC >0 and padj <0.01. DEGs were then mapped to specific cell types with the IReNA v2 function Process_DEGs_to_Celltypes().

Peak-to-gene linkage was performed with ArchR's addPeak2Gene-Links() function for both epithelium and mesenchyme using the "GeneIntegrationMatrix" (integrated RNA-Seq counts). Peak-to-gene links were then extracted with the IReNA v2 function Get_p2g_fun(). To identify potential *cis*-regulatory elements for each candidate gene, called correlated accessible regions (CARs), we separated the peak-to-gene links into three categories: TSS (when the peak lies within the transcription start site (TSS) for the gene), gene body (when the distance between peak and TSS is less than 100 kb, and the peak-to-gene score calculated above is significant), or distal (when the peak is 100 kb upstream or downstream of the TSS of correlated gene, and the peak-to-gene score is significant). These peak-to-gene links were then filtered to only include genes in the DEG list calculated above with Selection_peaks_to_one().

Next, we predicted the cell-type specific transcription factors binding in these CARs. We first took the snATAC fragments for each dataset (whole pancreas and Fev-high) and then extracted the fragments for each cell type. We converted these fragment lists to .BAM files and corrected the Tn5 insertion bias with TOBIAS [41] ATACorrect with default parameters except –read_shift 0 0. We then converted the TOBIAS output bigwig files to GRanges with the IReNA v2 function Check_normalized_Signal(). Next, we calculated TF binding motifs in our peaks with motifmatcher (https://github.com/GreenleafLab/motifmatchr), filtering calculated TFs out from the motif analysis if they were not enriched in each cell type by the DEG analysis. We calculated the NC (average bias-corrected Tn5 signal in the center of the motif), NL and NR (average bias-corrected Tn5 signal in the left and right flanking regions of the motif) scores with Calculate_footprint_celltypes() and filtered TFs with a score of $NC < -0.1$ and $NL > 0.1$ and $NR > 0.1$.

Next, we used MAGIC (Mining Algorithm for Genetic Controllers) [42] to compute correlation between TF and target gene gene expression. We retained the top and bottom 2.5% of correlations for our downstream analysis.

Lastly, we constructed the cell-type specific GRNs. We combined the peak-target links from our third step with the cell-type specific TF-peak links from our fourth step with Reg_one_cells_RPC_MG(). We then classified these interactions as either activating or repressing with our TF-target gene interactions calculated above with Add_Cor_-to_GRN_network_and_Filter(). We then identified feedback TF—TF pairs in our constructed GRN with FoundFeedBackPairs_new() and Process_the_Feedback_res().

2.11. Confirmation of *Fev* GRN peaks and targets

Peaks from Fev ChIP-Seq data detected in E12.5-E15.5 hindbrain were obtained from GEO (accession number GSE74315), from the supplementary file labeled "GSE74315_mycPet-1_ChIP_-peaks.mm10. bed.gz." Fev-occupied CARs from the Fev(+)/Chgb(+) GRN were taken from Supplementary Table 3 and written in .bed format. These peaks were then intersected using the bedtools software (https://github.com/arq5x/bedtools2) with default settings. Differentially expressed genes between Fev wildtype and knockout E15.5 hindbrains were downloaded from GEO (accession number GSE74315) under the supplementary file labeled "GSE74315_WT_vs_Pet1KO_5-HT_neuron_E15.5C_Clontech_cuffdiff.txt.gz."

3. RESULTS

3.1. Single-nucleus ATAC-Sequencing of the developing murine pancreas

To investigate chromatin accessibility in the developing pancreas, we aimed to capture a broad range of cell types, including both epithelial and non-epithelial populations. In addition, we were specifically interested in profiling EP cells, but given their rare numbers we searched for a method to achieve enrichment of this population. We utilized ePet1-EYFP mice (referred to hereafter as eFev-EYFP, as the gene *Pet1* is also known as *Fev*), where EYFP expression is driven by a *Fev* enhancer [34,35]. In previous work, we had identified *Fev* as a marker of an intermediate murine EP population downstream of the better-characterized Neurog3(+) population and upstream of differentiated, hormone-expressing endocrine cells [19].

As lineage reconstruction of scRNA-Seq data had revealed that this *Fev*-expressing EP population is likely the state at which endocrine lineage allocation occurs, we chose to focus on enrichment of Fev(+) cells. Although previous work with this eFev-EYFP mouse line had validated that EYFP expression faithfully reflected *Fev* expression in brain tissue, similar confirmation had not yet been performed in the pancreas [34]. We performed dual *in situ* hybridization (ISH)/immunofluorescence (IF) staining of E14.5 eFev-EYFP pancreas tissue to evaluate the architecture of EYFP expression with respect to the expression of *Neurog3* and *Fev* transcripts, as well as Chromogranin A (Chga) protein, a marker of differentiated hormone-producing endocrine cells. Expression of *Neurog3* and *Fev* transcripts was mutually exclusive (Figure 1A), as expected from our previous work demonstrating by genetic lineage tracing and scRNA-Seq that Fev-expressing cells are downstream of a Neurog3(+) state [19]. Also as expected, EYFP expression was only found in epithelial (E-cadherin(+)) cells, and mostly localized to ductal-like structures (Fig. S1A). A significant fraction, but not all, of EYFP(+) cells were actively expressing *Fev* transcript (Figure 1A). In addition, we observed EYFP(+) cells also expressing Chga (Fig. S1A). These data are consistent with a model of *Fev* expression in pancreatic EP cells in eFev-EYFP mice in which *Fev* transcript first begins to be expressed as *Neurog3* expression wanes, then expression of EYFP (under the control of the *Fev* enhancer) follows (Figure 1B). Persistence of EYFP in cells that no longer express *Fev* transcript likely reflects longer perdurance of EYFP fluorescent protein compared to *Fev* mRNA in these cells, similar to what has been observed for Neurog3-tdTomato [43], Neurog3-YFP [44], and Neurog3-EGFP transgenic mice [45].

We further validated the eFev-EYFP mouse line using fluorescence-activated cell sorting (FACS) and quantitative real-time polymerase chain reaction (qRT-PCR). Consistent with our IF staining, we observed little to no EYFP signal in cells that were negative for the epithelial marker Epcam (Fig. S1B). Within the population of cells positive for Epcam and negative for the mesenchymal marker CD140a, a bimodal distribution of EYFP signal was detected (Fig. S1B). TaqMan qRT-PCR analysis revealed that the EYFP-low population had higher expression
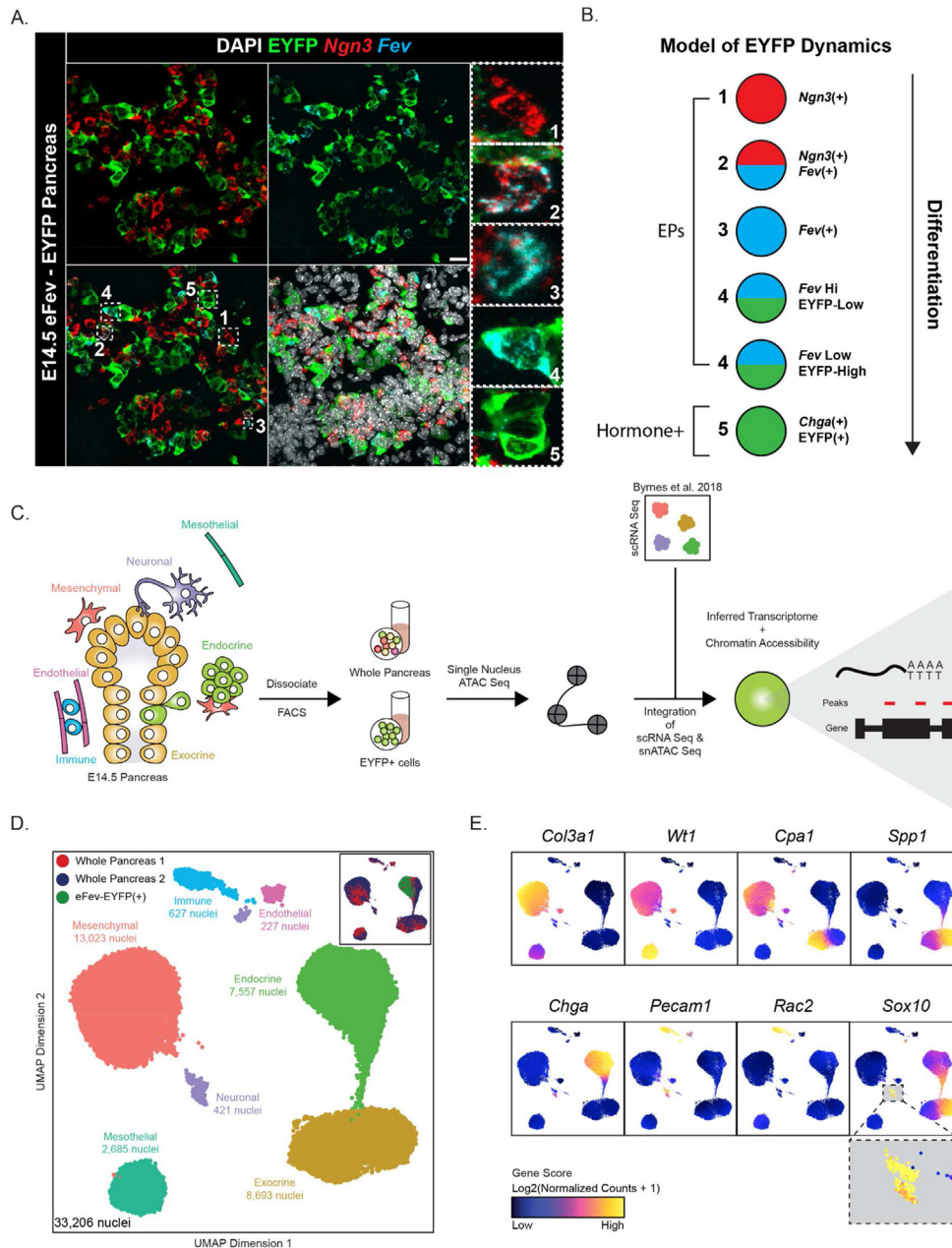
Figure 1: Single-nucleus ATAC-Seq of developing murine pancreas. (A) Multiplexed immunofluorescence and *in situ* hybridization staining of E14.5 pancreas from eFev-EYFP transgenic mouse embryos. *Neurog3* (*Ngn3*) transcript is shown in red and marks early endocrine progenitors (EPs), *Fev* transcript is shown in cyan and marks intermediate EPs, and EYFP protein is shown in green. DAPI marks nuclei in white. Five selected regions of interest (ROIs) are outlined by dashed white rectangles on the merged image and shown at higher magnification to the right. These ROIs highlight examples of cells that are 1) Ngn3(+)/Fev(−)/EYFP(−); 2) Ngn3(+)/Fev(+)/EYFP(−); 3) Ngn3(−)/Fev(+)/EYFP(−); 4) Ngn3(−)/Fev(+)/EYFP(+); and 5) Ngn3(−)/Fev(−)/EYFP(+). Scale bar is 20 μm. **(B)** Model of *Fev* and EYFP expression in EPs undergoing differentiation to a hormone-producing, Chromogranin A (Chga)-expressing state in eFev-EYFP reporter mice. Each circle represents a cell state across endocrine differentiation. **(C)** Overview of experimental approach for generating single-nucleus ATAC-Seq (snATAC-Seq) data. To enrich for Fev-high EPs, E14.5 eFev-EYFP murine pancreas was dissociated, and the resulting single cell suspension was subjected to FACS to enrich for EYFP(+) epithelial cells ("EYFP+ Cells"). In parallel, E14.5 pancreata from control (C57BL/6J) embryos were dissociated and subjected to FACS to isolate all live cells ("Whole Pancreas") to profile a broad spectrum of cell types, including non-epithelial cells. After subjecting samples to snATAC-Seq, data were then integrated with previously-published single-cell RNA-Sequencing (scRNA-Seq) datasets of E14.5 murine pancreas previously published by our laboratory [19]. **(D)** Uniform Manifold Approximation and Projection (UMAP) visualization of merged snATAC-Seq datasets from both Whole Pancreas (biological replicates 1 and 2) and EYFP(+) samples, comprising a total of 33,206 nuclei. Each dot represents a single cell, and each cell is colored according to cell type. Contribution of each sample (Whole Pancreas and eFev-EYFP(+)) to the total dataset is depicted in the inset, with the eFev-EYFP(+) sample contributing only to the endocrine cluster, as expected. **(E)** Feature plots depicting the Gene Scores (accessibility of the gene promoter plus the gene body) for some of the marker genes used to annotate the cell types in panel *D*. Inset for *Sox10* shows *Sox10*-expressing neuronal cells at higher magnification.

of *Fev* mRNA compared to EYFP-high cells (Fig. S1C). The EYFP-low population thus likely corresponds to a stage in which EYFP expression is on the rise and *Fev* expression is still present, whereas the EYFP-high population likely represents a stage where EYFP has reached higher expression but *Fev* itself has begun to wane (Figure 1B). Thus, we selected this EYFP-low population, enriched for Fev(+) cells, for snATAC-Seq using the 10x Genomics platform (Figure 1C, Fig. S1D). We included an additional two samples, biological replicates of whole pancreas tissue, to capture a broad range of cell types (Figure 1C).

Single cells were lysed to isolate nuclei, and chromatin was then subjected to the 10x Genomics pipeline and sequenced. The resulting dataset was analyzed with the computational package ArchR [39]. First, the datasets were filtered to retain high-quality nuclei by thresholding on the number of unique nuclear fragments, as well as the transcription start site (TSS) enrichment score (see Methods). This step provides enrichment of cells displaying a high fraction of fragments that map to the TSS versus other locations in the genome. Next, the datasets were subjected to doublet discrimination, resulting in a final dataset consisting of a combined total of 33,206 high-quality nuclei across the three samples. The data were then dimensionally reduced, clustered, and visualized in a 2D Uniform Manifold Approximation and Projection (UMAP) embedding (Figure 1D). The UMAP projection revealed that cells from both replicates of the whole pancreas contributed to each of the clusters within the dataset. As expected, cells from the eFev-EYFP(+) sample clustered only with endocrine cells from the two Whole Pancreas samples (Figure 1D, inset), reflecting successful enrichment of endocrine cells from the eFev-EYFP mouse line and effective integration of the three datasets. Each cluster was annotated as corresponding to a specific cell type found within the developing pancreas based on the gene score (accessibility of the gene promoter plus the gene body) of the following marker genes: *Col3a1* to mark mesenchymal cells, *Wt1* for mesothelial cells, *Cpa1* and *Spp1 for* exocrine cells, *Chga* for endocrine cells, *Pecam1* for endothelial cells, *Rac2* for immune cells, and *Sox10* for neuronal cells (Figure 1E, Supplementary Table 1).

3.2. Integration of single-cell transcriptional and chromatin accessibility data identifies epithelial heterogeneity in the developing murine pancreas

To reliably identify the heterogeneity of chromatin states within the epithelial cell types of the developing pancreas, we performed unconstrained integration of our snATAC-Seq data from all epithelial cells with E14.5 scRNA-Seq data previously published by our laboratory [19]. First, we computationally isolated the epithelial cells from the scRNA-Seq dataset (13,093 epithelial cells total) and performed iterative sub-clustering with the computational package CellFindR [37] to identify biologically relevant cell types. Next, we correlated the gene expression profiles of each of the cells within this scRNA-Seq dataset with the gene scores of each of the cells within our snATAC-Seq dataset. After identifying correlated cell pairs between the two datasets, cells in the snATAC-Seq dataset were assigned the cell type label, as well as the gene expression profile, of the cognate cell from the scRNA-Seq data.

This integration resulted in a final epithelial snATAC-Seq dataset comprised of 16,250 nuclei representing 10 distinct cell types, including Acinar, Ductal, Spp1(+)/Neurog3(+) double positive EPs, Neurog3(+) single positive EPs, Fev(+)/Chgb(+) intermediate progenitors, and Pdx1(+)/Mafb(+) beta cell precursors, as well as Alpha, Beta, Delta, and Epsilon cells (Figure 2A). As expected, the sorted EYFP(+) cells contributed highly to the endocrine but not the acinar or

ductal compartments of the overall dataset (Figure 2A, inset). The relative proportions of these annotated cell types in the snATAC-Seq dataset roughly matched the proportions of the epithelial cells in the scRNA-Seq dataset (Figure 2B). Integration scores, a reflection of confidence in the assignment of cell identity, were highest among terminally differentiated cell types (e.g., exocrine and hormone-expressing endocrine cells) (Fig. S2A), indicating less ambiguity in chromatin accessibility once cell fate is determined. Even in the absence of integration with scRNA-Seq data, all cell types were identified when clustering on chromatin accessibility alone (Fig. S2B). Next, we confirmed the cell type annotations by assessing chromatin accessibility (gene score), as well as the transferred RNA expression from the integration (gene expression). We observed high concordance between chromatin accessibility and RNA expression for the marker genes defining our cell types (Figure 2C). Additionally, we observed cell-type specific chromatin accessibility of each marker gene locus (Fig. S2C). To assess the reproducibility of called peaks across both biological samples, we subsetted the dataset by sample and re-called peaks on a subset of endocrine cells (see Methods). We detected 83,055 total peaks in the final endocrine peakset for the whole pancreas sample replicate 1, 119,809 total peaks for whole pancreas replicate 2, and 106,301 total peaks in the eFev-EYFP(+) sorted sample (Fig. S2D). We observed an overlap of 71,169 peaks between the two whole pancreas peaksets (representing an 86% overlap for whole pancreas sample 1 and a 59% overlap for whole pancreas sample 2). The number of peaks common to all three samples was 64,082, which represents an overlap of 77% for whole pancreas 1, 53% for whole pancreas 2, and 60% for the the EYFP(+) sample (Fig. S2D). When assaying differentially-accessible or -expressed genes, we observed far fewer significantly differentially accessible genes (n = 1,066) compared to differentially expressed (n = 4,567) (Figure 2D, Supplementary Table 1). Among these differentially-accessible genes were top markers of each cluster identified by differential gene expression analysis of our scRNA-Seq dataset (Supplementary Table 1). Taken together, these data confirm the existence of heterogeneous epithelial populations initially identified by scRNA-Seq, here by an orthogonal method.
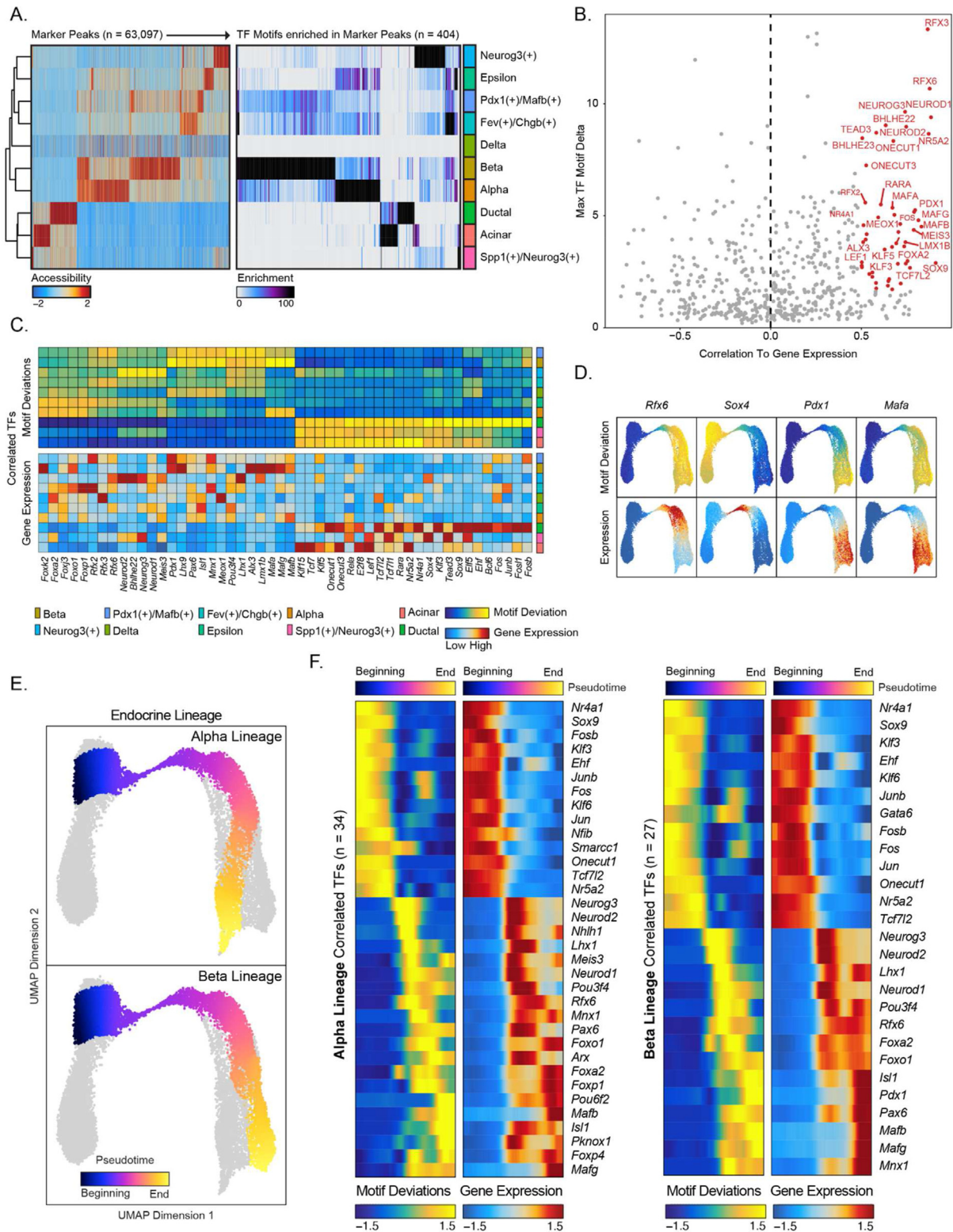
3.3. Identification of candidate regulators of epithelial cell fate

To identify regulators of cell fate decisions in the developing pancreatic epithelium, we applied the peak calling algorithm MACS2 [46] to our dataset. We identified 232,056 peaks across all epithelial clusters, with 63,097 peaks exhibiting differential accessibility across cell types (Figure 3A, Supplementary Table 2). Next, we assayed for TF motif enrichment in these differential peaks, identifying 404 enriched motifs. A number of TFs in the same family were deemed enriched due to the similarities in DNA binding motifs. For instance, TFs with enriched motifs included known regulators of pancreatic epithelial development, such as Sox family members (*Sox2, Sox4, Sox9*; Ductal), Hox family members (*Hoxb4, Hoxc4, Hoxa4*; Beta) and members of the Rfx family (*Rxf3* through *Rfx7*; Fev(+)/Chgb(+) and Pdx(+)/Mafb (+)) (Supplementary Table 2). To distinguish among TFs with similar DNA motifs identified in a given cell type, we next identified significant TF motif deviations (calculated as deviation of motif enrichment in accessible peaks from the expected distribution based on the average across all cells) of each cell type using ChromVAR [40]. The TFs from ChromVAR were then correlated with their gene expression profiles from the integrated RNA expression matrix, thereby identifying so-called "correlated TFs" that are both expressed and have significant motif deviation (Figure 3B, Supplementary Table 3). By breaking this down further on a per-cluster basis, we then were able to observe the
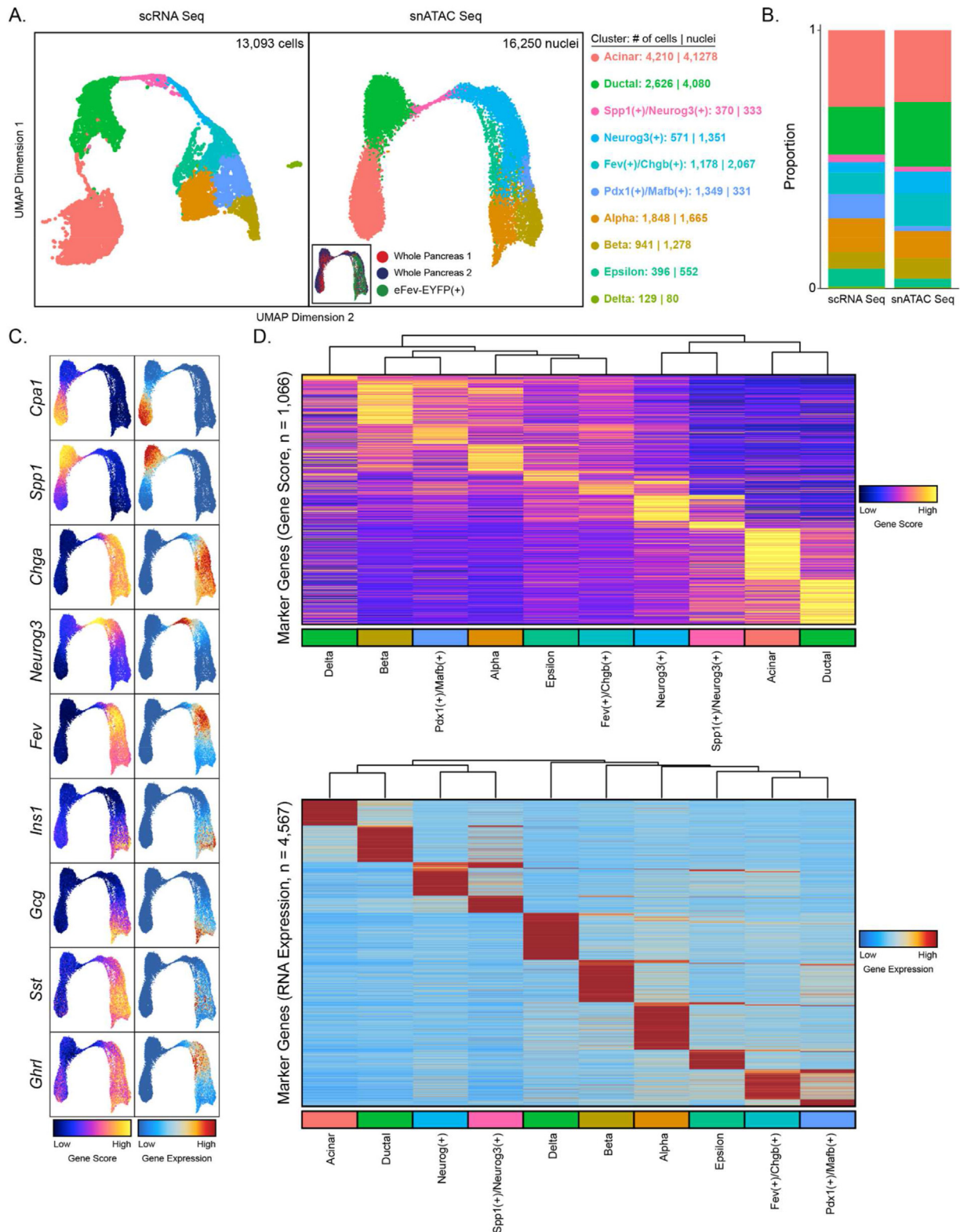
Figure 2: Integration of single-cell transcriptional and chromatin accessibility data identifies epithelial heterogeneity in the developing murine pancreas. (A) UMAP plots enabling visualization of scRNA-Seq (left) and snATAC-Seq (right) data for all epithelial cells in the E14.5 pancreas. Numbers of cells or nuclei are depicted on the right, along with cell type annotations. The scRNA-Seq dataset was previously published by our group [19]. (B) Bar graph depicting the proportion of all each cell type as a fraction of all epithelial cells in the scRNA-Seq and snATAC-Seq datasets. Colors match the cell types in (A). (C) Feature plots showing chromatin accessibility (Gene Score; left) or Gene Expression (right) of genes that mark each epithelial cell type. *Cpa1, Spp1, Chga, Neurog3, Fev, Ins1, Gcg, Sst,* and *Ghrl* mark acinar, ductal, pan-differentiated endocrine, early endocrine progenitor (EP), intermediate EP, beta, alpha, delta, and epsilon cells, respectively. (D) Heatmaps depicting genes that are differentially accessible (gene score; top heatmap; 1,066 genes) or differentially expressed (gene integration matrix; bottom heatmap; 4,567 genes) across the epithelial clusters. Genes listed were selected from the set of genes determined to be differentially expressed among scRNA-Seq clusters in panel (A).

Figure 3: Identification of candidate correlated transcription factors governing pancreatic epithelial cell fate. (**A**) Heatmaps depicting enriched marker peaks (left; n = 63,097) and transcription factor (TF) motifs enriched in marker peaks (right; n = 404) for each epithelial cell type. (**B**) Dot plot shows so-called "correlated" TFs (those with high correlation between motif deviation score and gene expression) in all epithelial cells. (**C**) Heatmaps revealing cell type-specific motif deviation scores (top) and gene expression values (bottom) of positive TFs identified in (**B**). (**D**) Feature plots displaying motif deviation (top) and gene expression (bottom) of selected positively correlated TFs at single-cell resolution. (**E**) Pseudotemporal ordering of epithelial cells along the Alpha and Beta lineages based on chromatin accessibility. (**F**) Heatmaps depicting positively correlated TFs across pseudotime (from left to right) for the Alpha lineage (left heatmaps; n = 34 correlated TFs) and Beta lineage (right heatmaps; n = 27 correlated TFs).

cell type-specific motif deviations and gene expression of the correlated TFs, narrowing the number of TFs with enriched motifs from 404 (Figure 3A) to 50 correlated TFs (Figure 3B,C). Correlated TFs included multiple members of the Fox family (*Foxk2* and *Foxo1*, Beta; *Foxp1*, Fev; *Foxa2*, Delta), as well as the Sox family (*Sox4*, Spp1(+)/Ngn3(+); *Sox9*, Ductal) (Figure 3C). The relationship between motif deviation and gene expression is further shown at single-cell resolution for several endocrine genes in Figure 3D. Interestingly, by observing not only motif deviation but also gene expression, we were able to determine that although both Ductal cells and Spp1(+)/Neurog3(+) EPs showed high motif deviation of *Sox4*, expression was significantly higher in the latter population (Figure 3D). This is in line with previously published work that shows that *Sox4* works with *Neurog3* to induce endocrine differentiation in the developing murine pancreas [47].

We next sought to understand the correlated TFs across the Alpha and Beta cell lineages. We first calculated the pseudotime values of cells along both trajectories (Figure 3E) and then applied the same motif deviation and gene expression correlation analysis for the genes and enriched motifs along these lineages (Figure 3F, Supplementary Table 3). Across the Alpha lineage (including Ductal, Spp1(+)/Neurog3(+), Neurog3(+), Fev(+)/Chgb(+), and Alpha cells) we identified 34 correlated TFs. This included TFs in Ductal cells (*Nr4a1, Sox9*), progenitor cells (*Neurog3, Neurod2*), and Alpha cells (*Foxp1, Isl1, Mafb, Mafg*). For the Beta lineage, we identified 27 correlated TFs, including *Mnx1, Mafb, Mafg, Pdx1*, and *Foxo1*. In summary, the multilayered approach taken here has further distilled the subset of TFs that likely play an important role in governing fate selection during endocrinogenesis.

3.4. Gene regulatory networks controlling epithelial cell fate

Our analyses thus far identified accessible chromatin and correlated TFs within the epithelial compartment of the developing endocrine pancreas. How and where these TFs bind and affect downstream target genes to govern cell fate decisions is not as well understood, however. To address this gap in knowledge, we next sought to construct a gene regulatory network (GRN) for Acinar, Ductal, and endocrine cells of the Alpha and Beta lineages (Figure 4A). We utilized the computational pipeline Integrated Regulatory Network Analysis (IReNA) v2 [48] (Figure 4B, Fig. S3A), which combines both scRNA-Seq and snATAC-Seq data to predict TF binding of downstream target genes in a cell type-specific manner. First, we performed differential gene expression analysis on our scRNA-Seq dataset to identify genes enriched in each cell type (Fig. S3B, Supplementary Table 4). We then performed peak-to-gene linkage analysis in our integrated scRNA- and snATAC-Seq datasets, identifying accessible regions of chromatin (peaks) that are either positively or negatively significantly correlated with gene expression (genes) (Fig. S3C). These peak-to-gene peaks were then further filtered and annotated as correlated accessible regions (CARs) belonging to one of three categories: TSS (when the peak lies in the transcription start site (TSS) for the gene), gene body (when the distance between peak and TSS is less than 100 kb, and the peak-to-gene score calculated above is significant), or distal (when the peak is 100 kb upstream or downstream of the TSS of a correlated gene, and the peak-to-gene score is significant). We observed varying proportions of CAR categories among the cell types, with TSS representing the highest proportion, followed by roughly equivalent proportions of positive and negative CARs (Fig. S3D). We next predicted the cell-type specific TF binding in these CARs by searching for TF DNA binding motifs in the CARs. Once identified, we then filtered the TFs by calculating their TF footprint score, retaining TFs with a score deemed significant by IReNA.

We observed the highest number of GRN TFs in the Fev(+)/Chgb(+) population (39 TFs), followed by Neurog3(+) (33 TFs) and Ductal (28 TFs) cells (Fig. S3E, Supplementary Table 5). Among our network of TFs and target gene-associated CARs, we binned these interactions as either activating or repressing by correlating the expression of each TF-target gene pair. Genes with a positive TF-target gene correlation were annotated as being activated by their given TF, while those with a negative TF-target gene correlation were annotated as being repressed. The Ductal and the Fev(+)/Chgb(+) populations had the highest number of regulations, followed by Pdx1(+)/Mafb(+) and Neurog3(+) (Fig. S3F). The regulations among all the populations examined were relatively evenly split between activating and repressing. Lastly, from the GRN constructed above, we identified pairs of TFs that regulated one another; for each TF, we identified target genes that are also TFs and mapped these pairs as either activating or repressing depending on the correlation of gene expression of the target TF in the given cell type. This analysis permitted us to identify TFs that regulate the expression of other TFs in given cell types (Fig. S3G, Supplementary Table 6).

To examine the TFs comprising this epithelial GRN in more depth, we first focused on the hormone(+) populations within our dataset. We found that genes enriched in the Beta cell population were largely repressed in the Acinar, Ductal, and Spp1(+)/Neurog3(+) GRNs, then gradually became activated as endocrine differentiation proceeded (Figure 4C). Within the Alpha cell population, Beta cell enriched genes were almost exclusively activated, consistent with previous studies investigating gene expression of individual TFs revealing that beta and alpha cells share common expression of genes needed for proper development and function [7]. We then inquired within all of the GRNs defined for epithelial cell types, which TFs either activate or repress genes enriched in the Beta cell population. Among the top activating TFs were known regulators of Beta cell development, such as *Mafb, Neurod1*, and *Pdx1* (Figure 4D). Targets of these activating TFs identified by our GRN analysis included numerous genes, both known and novel (Supplementary Table 5). Repressors of Beta cell enriched genes were largely contained within the Ductal GRN and included TFs such as *Tcf3, Tead2, Sox4*, and *Rest* (Figure 4E). As in Beta cell enriched genes, Alpha enriched genes followed a similar pattern of activation and repression across the epithelial GRN (Fig. S4A). TFs activating Alpha cell enriched genes included Beta cell activating TFs such as *Mafb* and *Neurod1* (Figure 4F). TFs repressing Alpha cell enriched genes also included *Tcf3, Tead2, Sox4*, and *Rest* (Figure 4G), suggesting that these TFs repress global hormone(+) cell gene signatures. The six TFs that overlapped between the Alpha (18 TFs) and Beta (13 TFs) GRNs comprised known endocrine regulators *Pax6, Mafb, Neurod1*, and *Isl1*, as well as TFs less well studied in pancreas, *Zfp516* and *Meis2* (Figure 4H). TFs unique to the Beta GRN included known regulators of beta cell fate, such as *Nkx6-1, Pdx1*, and *Foxo1*, while those less well characterized included *Mlxipl*. We performed multiplexed immunofluorescence and *in situ* hybridization on sections of wildtype E14.5 pancreas tissue and validated the enrichment of *Mlxipl* expression in Ins1(+) Beta cells (Fig. S4B). Examples of TFs unique to the Alpha GRN were *Irx2* and *Arx*, known regulators of Alpha cell fate. Less well characterized TFs included *Pbx1*, which is required for proper pancreas development [49], *Bbx, Peg3*, and *Etv1*. We also confirmed high enrichment of *Etv1* expression in Gcg(+) Alpha cells, as well as some expression in Fev(+) cells (Fig. S4C).

Among the top activating TFs of Alpha cell enriched genes was *Etv1* (Figure 4F), reported to be a direct or indirect target of *Nkx2-2* [50]. Within beta cells, failure to properly degrade *Etv1, Etv4*, and *Etv5* results in impaired insulin secretion [51]. We took a closer look at the
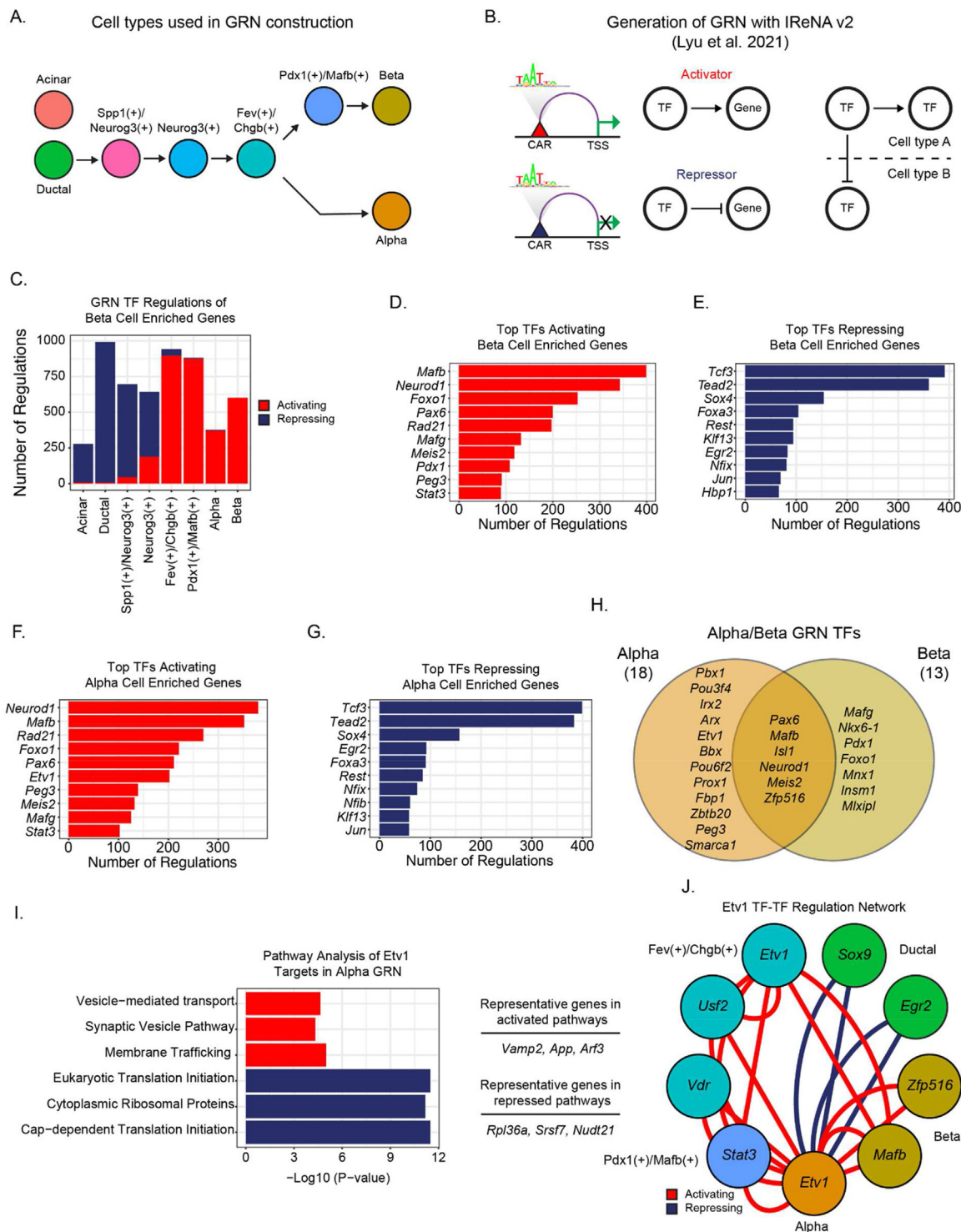
Figure 4: Mapping the gene regulatory networks active in the Alpha and Beta cell populations. (**A**) Cell populations used as input for IReNA. Arrows denote known lineage relationships. (**B**) Schematic depicting application of the Integrated Regulatory Network Analysis (IReNA) v2 pipeline to identify gene regulatory networks (GRNs) within specific cell types. The GRN refers to active, cell type-specific transcription factors (TFs) and their target genes. (**C**) Bar graph indicating the number of downstream target genes for TFs enriched in the Beta cell type. Activating (red bars) and repressing (blue bars) refers to positive or negative correlation between gene expression levels of the TF and the target gene. (**D–E**) Bar graphs showing the top 10 TFs with the highest number of activating (**D**) or repressing (**E**) regulations of target genes enriched in the Beta cell type. (**F–G**) Bar graphs showing the top 10 TFs with the highest number of activating (**F**) or repressing (**G**) regulations of target genes enriched in the Alpha cell type. (**H**) Venn diagram depicting the overlap between Alpha and Beta GRN TFs. (**I**) Bar graph depicting top most significantly (p-value <0.01) enriched pathways of genes activated (red bars) or repressed (blue bars) by the TF *Etv1* in the Alpha cell GRN. (**J**) Network diagram representing regulations between *Etv1* and interacting TFs. Each TF is represented by a circle (node) that is colored by the cell type in which that TF is active in the GRN. Activating regulations are depicted by red lines, while repressing are depicted by blue lines.

downstream targets of *Etv1* in the Alpha cell GRN and found that *Etv1* was predicted to activate 150 genes and repress 130 genes in the Alpha GRN (Supplementary Table 5). When performing pathway analysis on these genes, we observed that *Etv1* activated pathways related to vesicle mediated transport, synaptic vesicles, and membrane trafficking (Figure 4I). Other targets of *Etv1* included Alpha-cell related genes such as *Mafb* and *Irx2*, as well as endocrine hormones such as *Pyy* (Supplementary Table 5). Pathways repressed by *Etv1* included those related to translation initiation, cytoplasmic ribosomal proteins, and cap-dependent translation initiation. When examining TF—TF interactions, we found that *Etv1* is repressed by *Sox9* and *Egr2* in the Ductal population, and activated by *Stat3* in the Pdx1(+)/Mafb(+) population and by *Usf2* and *Vdr* in the Fev(+)/Chgb(+) population (Figure 4J).

In summary, the computational analyses described in this section have permitted the construction of a gene regulatory network of the acinar, ductal, and major endocrine lineages in the developing mouse pancreas. This workflow, which is dependent on integration of both chromatin accessibility and transcriptional data, has identified regulators of alpha and beta cell fate that can serve as the subjects of further experimental study.

3.5. Gene regulatory networks governing the initiation of endocrine differentiation

During mammalian pancreatic development, a subset of cells within the branching ductal epithelium activate the expression of the master regulator of endocrine differentiation, *Neurog3*. These rare Neurog3(+) cells represent the earliest known EP population, and considerable attention has been devoted to understanding this ductal to EP transition. In addition to *Neurog3*, which is required for mouse endocrine differentiation [6], numerous other TFs have been identified that are also important for endocrinogenesis. Investigation of *NEUROG3* binding across the genome in human pluripotent stem cell (hPSC)-derived EP cells revealed widespread regulation of 138 TFs, some with known roles in endocrine development and others with unknown function [52]. Further studies in human cells used inducible and knockout models in hPSC-derived endocrine cells to identify predicted targets of multiple endocrine TFs, including *NEUROG3*, *PDX1,* and *RFX6* [53]. Generation of an Neurog3-timer fluorescent reporter mouse line that permitted the specific isolation of early Neurog3-expressing cells identified numerous putative direct targets of *Neurog3* in mouse EPs [54]. These studies have highlighted the need for a broad, integrated analysis of all TFs and downstream targets that control the initiation of endocrine differentiation.

Towards this end, we began by investigating the GRN regulating the transition from a ductal to EP cell state. TFs in the Ductal GRN promoting the expression of Spp1(+)/Neurog3(+) EP genes included known regulators of endocrine cell fate, such as *Sox4* [47,55] (Figure 5A). Also identified was *Sox9*, which induces the expression of *Neurog3* and thus positively regulates the endocrine lineage before then being turned off during differentiation [5,55]. Interestingly, *Tead2* and *Tcf3* activated the most genes enriched in the Spp1(+)/Neurog3(+) and Neurog3(+) EP populations (Figure 5B), indicating that these TFs are important initiators of an endocrine cell fate. The Yap/Tead signaling complex has previously been shown to activate multipotent progenitor cell enhancers and regulate epithelial outgrowth during human pancreatic development [56]. *Tcf3*, also known as *E47*, has been shown in a human cell line to dimerize with *NEUROG3* to bind to the promoter region of the *INSM1* gene [57], which is required in mice to maintain mature beta cell function [58]. TFs involved in the transition from a Spp1(+)/Neurog3(+) to a Neurog3(+) EP cell state

include well known regulators of endocrine differentiation, such as *Nkx2-2*, *Pax4*, and *Neurod2* (Figure 5A). Other TFs with no known endocrine function included *Nhlh1*, whose expression was enriched in Neurog3(+) cells by E14.5 scRNA-Seq and confirmed *in vivo* (Fig. S5A). Major repressors of EP enriched genes in the Ductal GRN included TFs such as *Rest* and *Hes1* (Figure 5C). The classification of *Hes1* as repressing is consistent with what is known about the role of Notch signaling in the initiation of EP cell fate [5,59]. *Rest* is a master regulator of neurogenesis and has been previously described to inhibit direct reprogramming of pancreatic exocrine to endocrine cells by inhibiting the binding of *Pdx1* to key endocrine differentiation-related genes [60]. In addition, loss of *Rest* results in increased generation of pancreatic endocrine cells during development [61,62]. Consistent with previous studies, we confirmed expression of *Rest* in *Spp1*-expressing Ductal cells by *in vivo* staining of E14.5 pancreas and by scRNA-Seq (Fig. S5B). Among TF—TF pairs identified in Ductal, Spp1(+)/Neurog3(+), and Neurog3(+) cells, most were classified as activating, with the exception of *Nfib* (Ductal) (Fig. S5C). Taken together, our GRN analysis has identified novel candidate regulators, such as *Tcf3* and *Tead2*, of the ductal to EP cell state transition. These results expand upon our knowledge of this key developmental transition and serve as a resource for future studies.

We next focused on Fev(+)/Chgb(+) cells, as our previous work indicated that this cell state represents the bifurcation point at which the Alpha or Beta lineage is established (Figure 4A) [19]. As expected, we observed that the Acinar and Ductal cell types largely repress genes that are enriched in the Fev(+)/Chgb(+) population (Figure 5D). These genes begin to be activated as an endocrine cell fate is established (Spp1(+)/Neurog3(+) and Neurog3(+) cell types) and are fully activated by the Fev(+)/Chgb(+) cell stage. Curiously, the Fev(+)/Chgb(+) enriched genes are not repressed in the Alpha and Beta cell types, suggesting that Alpha/Beta cell fate is due more to activation of key Alpha/Beta genes as opposed to the repression of progenitor-associated genes. Among the top TF activators of Fev(+)/Chgb(+) enriched genes (Supplementary Table 5), we observed that known regulators of endocrine cell fate such as *Neurod1, Mafb*, and *Pax6* activated the most genes (Figure 5E). Activators also included less well described TFs, such as *Rad21* and *Peg3*, as well as *Foxo1* and *Etv1* (Figure 5E; Supplementary Table 5). Conversely, TF repressors of Fev(+)/Chgb(+) enriched genes included the TFs *Tead2*, *Tcf3*, *Rest, Sox4*, and *Nfib,* among others (Figure 5F). Next, we constructed a network diagram of TF pairs that either activate or repress TFs enriched in the Fev(+)/Chgb(+) population (Figure 5G). Consistent with our observations in Figure 5D, TF—TF regulations were entirely repressive in the Acinar and Ductal cell states, and a mix of activating and repressing in the Spp1(+)/Neurog3(+) and Neurog3(+) cell states (Figure 5G; Supplementary Table 6). In contrast, TF—TF regulations in the Alpha, Beta, and Pdx1(+)/Mafb(+) states were exclusively activating (Figure 5G; Supplementary Table 6).

When comparing the GRNs among all progenitors and precursors, we identified 18 TFs unique to the Fev(+)/Chgb(+) population (Fig. S5D, Supplementary Table 6). Among these TFs identified within the Fev(+)/Chgb(+) GRN, *Rad 21* and *Mafb* were among the top activators of the transition from Fev(+)/Chgb(+) to either Pdx1(+)/Mafb(+), Alpha, or Beta cell states (Figs. S5E—G). Among the top 10 TFs with the highest number of activating regulations across the transition from Fev(+)/Chgb(+) to Alpha cells was Vitamin D receptor (*Vdr*), whose expression has been linked to beta cell function and diabetes (Fig. S5F) [63—65]. As expected from Figure 4F, *Etv1* had a higher number of activating regulations for Alpha cell-enriched genes compared to Pdx1(+)/Mafb(+) or Beta populations (Figs. S5E—G).
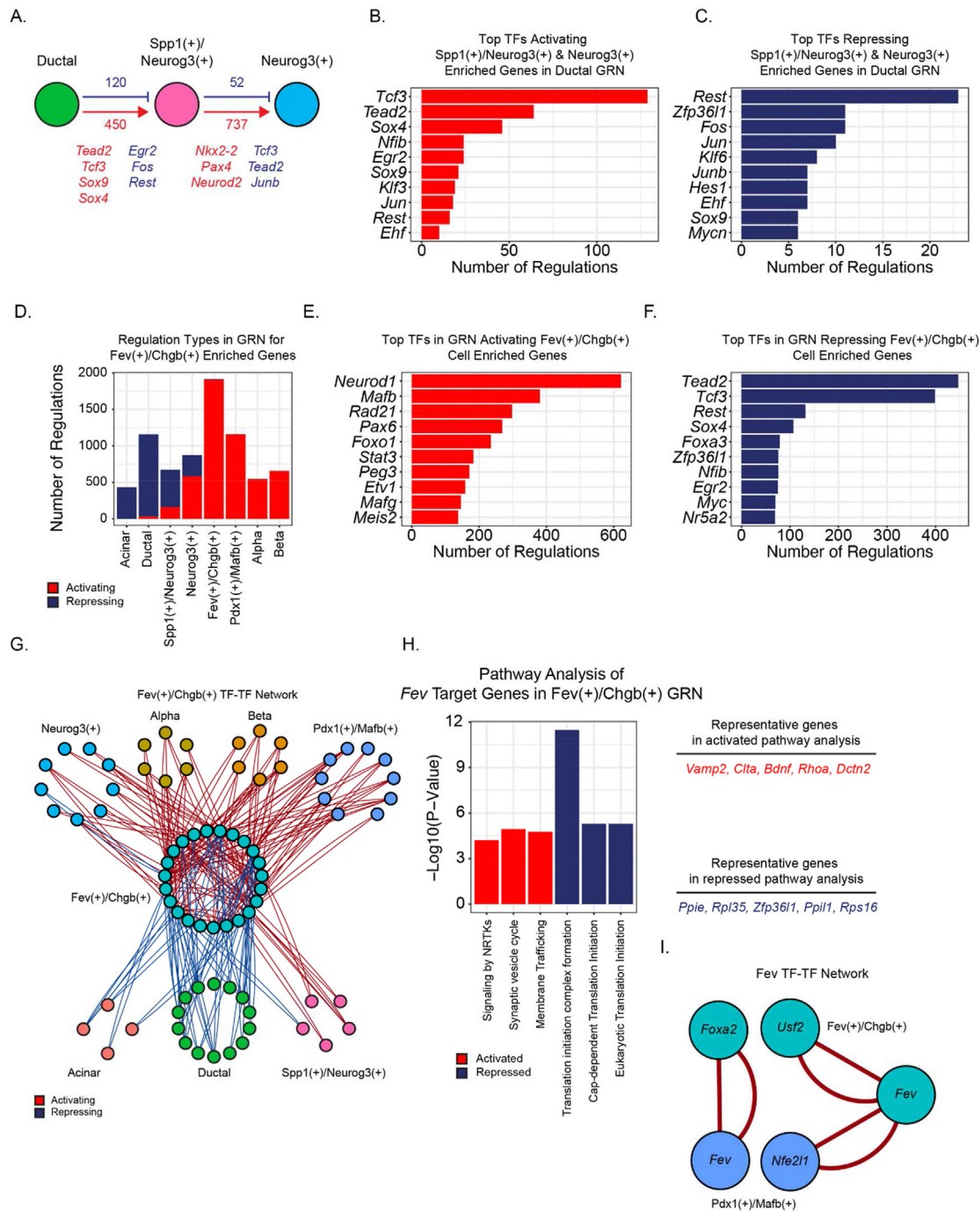
Figure 5: Mapping the gene regulatory networks active in the *Fev*-expressing pancreatic endocrine population. (A) Cartoon depicting the number of activating (red) and repressing (blue) regulations between transcription factors (TFs) in Ductal, Spp1(+)/Neurog3(+), and Neurog3(+) populations, with representative TFs indicated. (B–C) Bar graphs depicting the top 10 TFs in the Ductal GRN with the highest number of activating (B) or repressing (C) regulations of target genes that are enriched in the Spp1(+)/Neurog3(+) and Neurog3(+) cell types. (D) Bar graph depicting the number of TF regulations of target genes enriched in the Fev(+)/Chgb(+) cell type. Activating (red bars) and repressing (blue bars) refers to positive or negative correlation, respectively, of gene expression between the TF and target gene. (E-F) Bar graphs depicting the top 10 transcription factors with the highest number of activating (E) or repressing (F) interactions of target genes enriched in the Fev(+)/Chgb(+) cell type. (G) Network diagram depicting all TF−TF regulations between TFs enriched in the Fev(+)/Chgb(+) GRN and all other GRNs. TFs are denoted by each node, which is colored by the cell type in which the TF is found. Each activating regulation is depicted by a red line, while each repressing regulation is depicted by a blue line. (H) Bar graph depicting top significant (p-value <0.01) pathways of genes activated (red bars) or repressed (blue bars) by *Fev* in the GRN analysis. (I) Network diagram depicting the TFs activated (red lines) by *Fev* in the GRN analysis. Each node (TF) is colored according to the cell type in which it is expressed.

The gene *Fev* was initially described as a prototypical serotonergic transcription factor in the brain [66], then as a gene expressed in developing and adult mouse pancreatic islets [67]. More recently, we found that in the pancreas *Fev* marks an intermediate progenitor of the mouse endocrine lineage [19]. Our GRN analysis computed 83 genes activated and 57 genes repressed by *Fev* (Supplementary Table 5). Pathway analysis of activated genes included pathways related to synaptic vesicle cycle, signaling by NRTKs, and membrane trafficking (Figure 5H). These data corroborate previous findings in which full body knockout of *Fev* resulted in decreased pancreatic insulin content, an impairment of insulin secretion, and consequently defects in glucose clearance [67]. Pathway analysis of repressed genes included many pathways related to translation, such as translation initiation complex formation, cap-dependent translation initiation, and eukaryotic translation initiation (Figure 5H). Downstream TF interactions of *Fev* were all activating and included the TFs *Foxa2*, *Usf2*, and *Nfe2l1* (Figure 5I).

Next, we sought to confirm the occupied peaks and downstream targets of *Fev* in our GRN analysis. Although Chromatin Immunoprecipitation followed by Sequencing (ChIP-Seq) data of Fev binding in the endocrine pancreas is not available, a previous study published ChIP-Seq data from *Fev*-expressing 5-HT (+) neurons in the E12.5-E15.5 developing hindbrain [35]. Given the documented similarities in transcriptional networks between serotonergic neurons and pancreatic endocrine cells [67], we asked how much shared overlap there is between our gene-associated CARs in the pancreatic *Fev* GRN and serotonergic neuron ChIP-Seq peaks. When intersecting these peak sets, we observed that 90 of the 270 (33%) *Fev* GRN CARs overlapped with peaks from the ChIP-Seq dataset (Fig. S6A). Target genes of these gene-associated CARs included *Myo6*, *Mef2a*, and *Tmem30a* (deemed activated by *Fev* in our GRN analysis), as well as *Rrs1, Rpl41, and Cxadr* (deemed repressed by *Fev* in our GRN analysis). Non-overlapping activated gene-associated CARs included *Vamp3*, *Chga*, *Foxa2*, and *Klf7*. Repressed gene-associated CARs included *Notch2*, *Nr5ac,* and *Apex1*. The overlap of *Fev* GRN CAR and ChIP-Seq peak for *Mef2a* was visualized at the *Mef2a* gene locus in the Fev(+)/Chgb(+) and Pdx1(+)/Mafb(+) clusters (Fig. S6B). Next, to confirm the potential gene targets of *Fev* in our GRN analysis, we interrogated published RNA-Seq data from wildtype and *Fev* knockout 5-HT (+) serotonergic neurons from E15.5 mouse hindbrains. We intersected the target genes of *Fev* in our GRN analysis with differentially expressed genes between *Fev* wildtype and knockout mice (Fig. S6C) [35]. We found that 38 of the targets of *Fev* in the embryonic pancreas were differentially expressed between the *Fev* wildtype and knockout embryonic serotonergic neurons, including genes such as *Celf1, Pcsk1, Mcm6, Chga, Dctn3,* and *Fabp5*.

Taken together, our data identify both known and novel regulators of pro-Alpha and pro-Beta cell fates that are active in the Fev(+)/Chgb(+) stage, the cell state that represents the bifurcation point in the endocrine differentiation trajectory. Our analyses also yield a comprehensive view of potential targets of *Fev,* as well as insights regarding its function in regulating the machinery required for the production of endocrine hormone-containing vesicles. Lastly, our analysis has identified conserved *Fev*-occupied peaks and target genes between serotonergic neurons and Fev(+) EPs, as well as potential tissue-specific regions of *Fev*-mediated regulation.

3.6. Characterization of chromatin accessibility and identification of GRNs within pancreatic mesenchymal cell types

Although proper development of the pancreatic epithelium depends on signals from the surrounding mesenchyme, the lineage and function of pancreatic mesenchymal cells remains vastly understudied. In previous work, we and others have used scRNA-seq to identify multiple transcriptionally distinct mesenchymal populations, including mesothelium, within the developing murine pancreas [19,20]. Still, the upstream genetic regulators that maintain these distinct cell states are not defined. Data from snATAC-Seq of pancreatic mesenchymal cells would shed light on whether distinct states of chromatin accessibility correspond to transcriptionally distinct cell subpopulations and would reveal which TFs and binding sites are actively involved in controlling mesenchymal cell state.

We integrated the snATAC-Seq data from the mesenchymal populations within our merged E14.5 Whole Pancreas dataset (replicates 1 and 2) with the age-matched (E14.5) scRNA-Seq data, using methods as described above. Clustering of the scRNA-Seq dataset identified six populations of mesenchymal cells, including one cluster enriched in the expression of *Gap43* (Gap43(+)), another cluster enriched in expression of *Sfrp2* (Sfrp2(+)), two clusters expressing chemokines (Cxcl12(+) and Cxcl13(+)), Vascular Smooth Muscle cells (VSM; Acta2(+)), and finally Mesothelium (Wt1(+)) (Figure 6A). Integration and cell label transfer classified all populations in our snATAC-Seq dataset (Figure 6A,B). Clustering on chromatin accessibility alone, without integration with scRNA-Seq data, still resulted in the identification of similar clusters as in the integrated dataset (Fig. S7A).

Once the mesenchymal clusters were identified, we next asked whether there were secreted factors uniquely produced by any of these mesenchymal subpopulations that may be signaling to any of the epithelial subpopulations identified above. We performed cell—cell communication analysis using CellChat [68] on the mesenchymal and epithelial compartments in our E14.5 scRNA-Seq dataset (Fig. S7B). We found 33 cell—cell communication signaling pathways that were significantly active in the mesenchymal and epithelial datasets (Supplementary Table 7), including communication based on ECM signaling (Mk, Ptn, Mpz, Laminin) and secreted factors (Bmp, Ephb, Ngf, Notch). When focusing on mesenchymal/mesothelial to epithelial signaling, we found that the ECM pathways Mk, Ptn, and Collagen showed high relative strength for incoming signaling to Acinar and Ductal cells (Fig. S7B), with high outgoing strength across multiple mesenchymal populations. Signaling to Beta cells included Grn, likely being received by the Cxcl12(+) mesenchymal population, which showed the strongest outgoing strength. Jam signaling, which is important in cell migration, was highest in outgoing signaling from Mesothelial cells, while highest in incoming signaling in the Acinar, Ductal, Spp1(+)/Neurog3(+), and Neurog3(+) populations. Taken together, these data elucidate the potential cell—cell communication between the developing mesenchymal, mesothelial, and epithelial cells of the embryonic pancreas.

Next, we identified differentially accessible peaks across all of the mesenchymal populations and found a total of 182,020 peaks, with 59,707 peaks displaying differential accessibility (Supplementary Table 2). The majority of these differentially accessible peaks were enriched in the Gap43(+) and Mesothelial populations, with the remaining populations showing more modest numbers. Motif enrichment in these differentially accessible peaks identified 217 enriched motifs (Supplementary Table 2). Correlation of gene expression and motif deviation scores identified 34 correlated TFs (Figure 6C,D), including known regulators of mesenchymal cell fate, such as *Wt1*, *Twist2*, and *Hand2*.

We next assembled the mesenchymal GRN with IReNA v2 as described above. Among the mesenchymal cell types, the Gap43(+) and Mesothelial populations had the highest number of total regulations (Fig. S7C) as well as the highest numbers of TFs in each of the GRNs (Fig. S7D). In the TF—TF network, TFs active in the GRN of
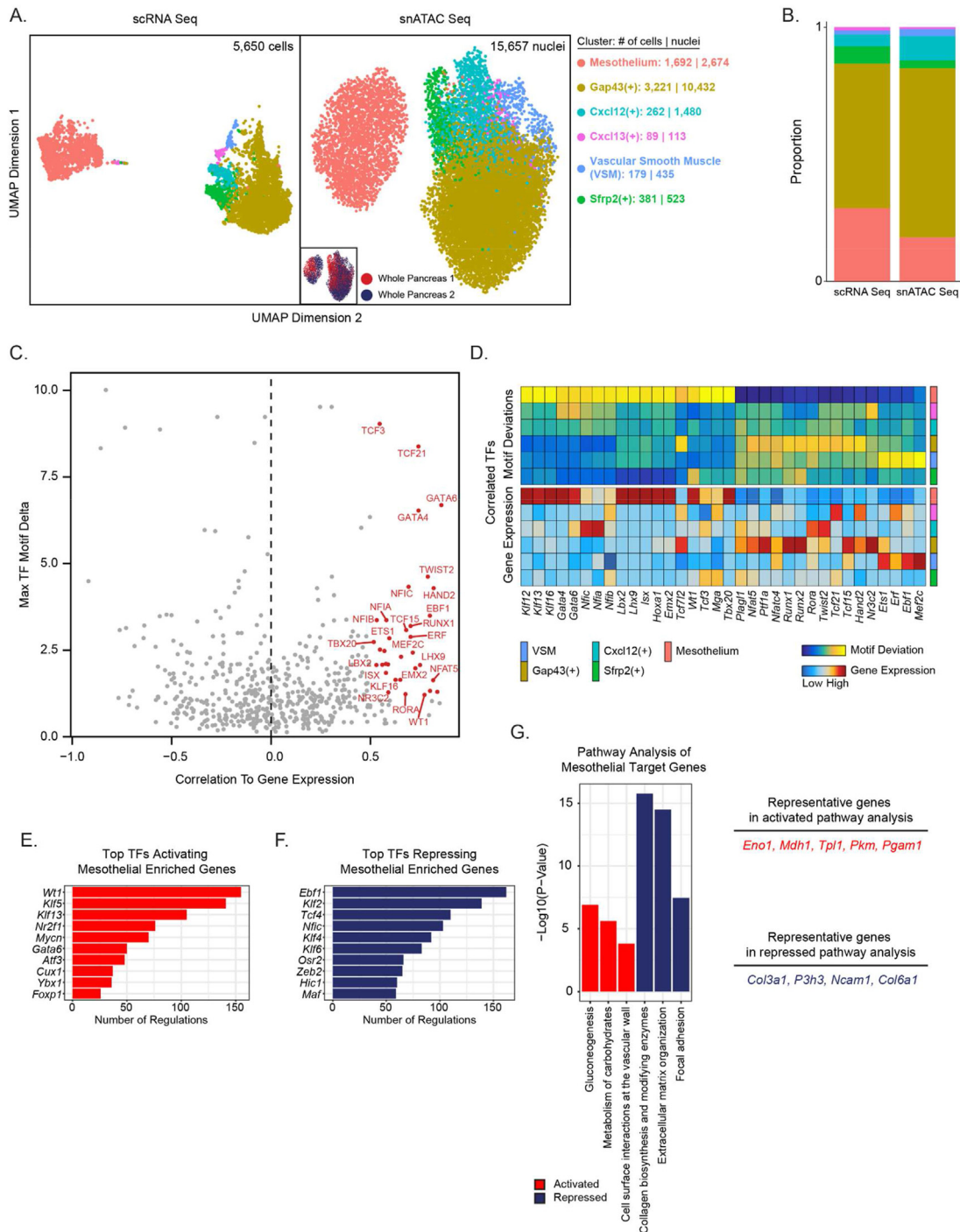
Figure 6: Chromatin accessibility and gene regulatory networks in the developing pancreatic mesenchyme. (**A**) UMAP plots enable visualization of scRNA-Seq (left) and snATAC-Seq (right) data for all mesenchymal cells in the E14.5 pancreas. Numbers of cells/nuclei are depicted on the right, along with cell type annotations. The scRNA-Seq dataset is from our previously published work [19]. (**B**) Bar graph depicts the proportion of each cell type in the scRNA-Seq and snATAC-Seq datasets. Colors match the cell types in (**A**). (**C**) Dot plot showing correlated transcription factors (as determined by correlation between motif deviation score and gene expression) in mesenchymal and mesothelial cells. (**D**) Heatmaps reveal cell type-specific motif deviation scores (top) and gene expression values (bottom) of positive transcription factors identified in (**C**). (**E, F**) Bar graphs depicting the top TFs activating (**E**) and repressing (**F**) genes enriched in the mesothelial population. (**G**) Bar graph depicting top significant (p-value <0.01) pathways of genes activated (red bars) or repressed (blue bars) by mesothelial cells in the GRN analysis. VSM, vascular smooth muscle.

Mesothelium largely repressed the expression of Gap43(+)-associated TFs, whereas the Gap43(+) population activated Cxcl12(+)-associated TFs (Fig. S7E). We next assessed which GRN TFs were either exclusive or shared among the mesenchymal populations (Fig. S7F). The GRN in the Gap43(+) population contained 18 specific TFs, including genes involved in epithelial—mesenchymal transition (EMT) such as *Snai2, Twist2,* and *Zeb1*, suggesting that cells within the Gap43(+) population may be actively undergoing EMT processes during pancreatic development. Other Gap43(+) cell type-specific TFs included *Hoxb4* and *Prrx1*. TFs exclusive to the Cxcl12(+) population included members of the nuclear factor 1 (*Nfia* and *Nfic*), Ap-1 (*Junb* and *Jund*), and Klf families (*Klf6; Klf2* shared with the Gap43+ population). TFs specific to the Sfrp2(+) GRN included *Hoxb5, Meis2,* and *Bcl11a*, while *Pbx1* was shared with the Cxcl12(+) population, and *Barx1* was shared with the Mesothelial population.

We next focused on mesothelial cells, which consist of a monolayer of specialized cells that line the pleura and internal organs of adult tissues and serve numerous functions in the adult, including lubrication of tissue and immune surveillance [69,70]. In the developing lung, lineage tracing studies have demonstrated that mesothelium also acts as a progenitor for certain specialized mesenchymal cell subtypes [71]. Furthermore, previous work in our lab predicted the downstream lineages of mesothelial cells in the developing pancreas based on pseudotemporal ordering of scRNA-Seq data [19]. Despite this, relatively little is known about how mesothelial cells are formed and maintained during pancreatic development. Within the snATAC-Seq dataset, we identified 20 TFs uniquely active within the GRN of the Mesothelial population (Fig. S7F). Top TFs activating Mesothelial-enriched genes included *Wt1,* which has previously been shown to be a master regulator of mesothelial formation [72], along with *Klf13,* whose role in mesothelial cell development and homeostasis is not well understood and thus warrants further study (Figure 6E). Top repressing TFs of Mesothelial-enriched genes included *Klf2, Ebf1,* and *Klf6* (Figure 6F). Pathway analysis of activated and repressed genes in the Mesothelium GRN identified pathways related to metabolism as activated, and ECM formation and deposition as repressed (Figure 6G). Taken together, profiling of the chromatin accessibility within the cells of the developing mouse pancreatic mesenchyme has determined differentially accessible peaks among these populations and identified TFs potentially important in mesenchymal development through the use of motif enrichment and gene expression correlation analyses. Lastly, we have constructed a set of mesenchymal GRNs, identifying active TFs and their downstream target genes. These data will provide a resource for future work geared towards studying mesenchymal biology and understanding how this important but understudied non-epithelial population is maintained.

3.7. Mapping the gene regulatory networks of pancreatic epithelium across developmental time

Previous work from our laboratory [19] used scRNA-Seq to profile the murine pancreas across three developmental timepoints. Although this timecourse analysis revealed shifts in endocrine cell differentiation, proliferation, and gene expression programs across developmental time, the regulatory elements governing these changes remain poorly understood. To better characterize how gene regulatory networks within the pancreatic epithelium shift across developmental time, we performed snATAC-Seq on two independent samples of whole pancreas tissue at E17.5, each sample comprising at least n = 3 pancreata. After performing quality control and filtering steps (see Methods), we generated a merged dataset consisting of 78,669 high-quality nuclei comprised of the expected broad pancreatic cell types

(Fig. S8A). Of these broad groups, 8,859 were labeled as endocrine based on the Gene Score for *Chga* (Figure 7A, Figs. S8A—B). We then performed unconstrained integration of our E17.5 snATAC-Seq epithelial cells with E17.5 scRNA-Seq [19] as described above for E14.5. This integration resulted in a final epithelial snATAC-Seq dataset comprised of 8 distinct cell types, including Acinar, Ductal, Spp1(+)/Neurog3(+) double positive EPs, Neurog3(+) single positive EPs, Fev(+)/Chgb(+) intermediate progenitors, Alpha, Beta, and Delta cells. (Figure 7A). All epithelial cell populations annotated in the scRNA-Seq dataset were successfully captured in the snATAC-Seq dataset, although the relative proportions of some cell types varied, presumably due to cell isolation and nuclei capture (Figure 7B). We then identified correlated TFs across the hormone(+) lineages as detailed above. Across the Alpha lineage, we identified 39 correlated TFs, including those in Ductal (*Nr5a2, Sox9*), progenitor (*Elf5, Neurog3, Foxa2*), and Alpha cells (*Rfx3, Rfx6, Arx, Isl1, Mafb, Mafg*) (Fig. S8C). For the Beta lineage, we identified 32 correlated TFs, including *Arid3a, Pax6,* and *Pdx1* (Fig. S8D). For the Delta lineage, we identified 34 correlated TFs, including *Rfx3* and *6, Lhx1, Pou3f4, Pax6, Arx,* and *Pdx1* (Figure 7C). Next, we sought to construct GRNs for each of the epithelial cell types in the E17.5 pancreas utilizing the IReNA v2 pipeline as detailed above for the E14.5 sample. After constructing the GRN, we observed the highest number of GRN TFs in the Neurog3(+) population [57], followed by the Fev(+)/Chgb(+) [50] and Ductal [44] populations (Fig. S8E, Supplementary Table 4). As in the E14.5 GRN, the Ductal, Fev(+)/Chgb(+), and Neurog3(+) populations had the highest number of regulations (Fig. S8F). We observed that the Acinar population displayed a marked reduction in the number of regulations at E17.5 compared to E14.5 (Fig. S8F, Fig. S3F). Next, we focused on the hormone(+) cell types in our GRN analysis. The E17.5 Alpha cell population contained 23 total GRN TFs, 19 of which were unique to the Alpha cell population (Figure 7D). These Alpha GRN TFs included known regulators such as *Irx2, Arx,* and *Neurod1*, as well as unknown regulators such as *Bptf, Thap11,* and *Nfat5*. When comparing Alpha GRN TFs at E17.5 [23] to those at E14.5 [18], we found that 7 TFs were unique to E17.5, including *Foxa2, Foxp1,* and *Bptf*, while two were unique to E14.5 (*Pbx1* and *Fbp1*) (Fig. S8G). Alpha GRN TFs common between E14.5 and E17.5 [16] included key Alpha cell regulators such as *Irx2, Arx, Isl1,* and *Neurod1*, as well as *Etv1* and *Peg3*. The E17.5 Beta GRN contained six total TFs, two of which (*Insm1* and *Mlxipl*) were not found in the Alpha or Delta GRNs (Figure 7D). While a number of known regulators of beta cell fate such as *Nkx6-1, Mnx1, Mafg,* and *Isl1* were unique to the E14.5 Beta GRN, only *Banp* was unique to the E17.5 Beta GRN. Shared GRN TFs were *Pax6, Pdx1, Mafb, Insm1,* and *Mlxipl* (Fig. S8H). Lastly, the E17.5 Delta GRN contained six TFs: known Delta regulator *Hhex*, along with *Pdx1, Zbtb20, Junb, Fos,* and *Klf7*. (Figure 7D). Top TFs regulating Delta cell-enriched genes included *Mafb, Neurod1, Klf7, Rad21,* and *Pax6* (Figure 7E), while TFs repressing Delta cell-enriched genes included *Tcf3, Tead2,* and *Jun* (Figure 7F).

Lastly, we sought to understand how the targets of GRN TFs change across developmental time. We focused on activated target genes of *Fev* in the Fev(+)/Chgb(+) GRN and identified 91 targets in the E17.5 GRN compared to 83 in the E14.5 GRN (Figure 7G, Supplementary Table 3). Targets unique to E14.5 included genes such as *Vamp3, Dctn3, Mef2a, Klf10,* and *Ptpm2,* while shared genes included *Fam174a, Serinc1, Tmem30a, Chga,* and *Tmed8*. Targets unique to the E17.5 GRN included genes involved in pancreatic secretion, beta-catenin independent Wnt signaling, and signaling by NTRK1 (TRKA) (Figure 7H), such as *Pdx1, Itpr3, Rab6a, Gng12,* and *Pak3*. The predicted TF targets of *Fev* at E17.5 were *Nfix* (Spp1 (+)/Neurog3(+)),
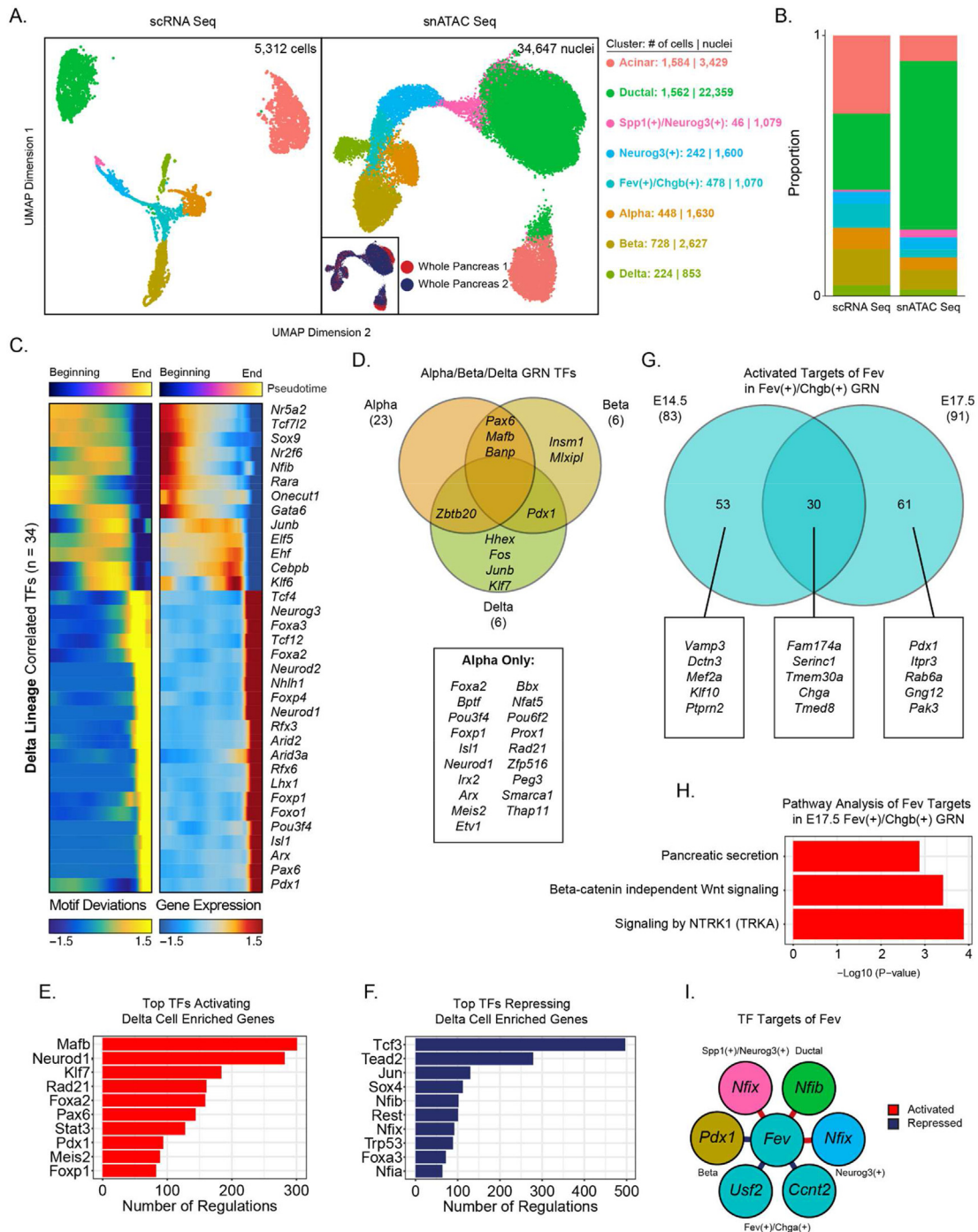
Figure 7: Mapping the gene regulatory networks of pancreatic epithelium across developmental time. (A) UMAP plots enable visualization of scRNA-Seq (left) and snATAC-Seq (right) data for all epithelial cells in the E17.5 mouse pancreas. Numbers of cells or nuclei are depicted to the right, along with cell type annotations. The scATAC-Seq dataset is comprised of two independent biological replicates (n = 7 pancreata from 1 litter for replicate 1; n = 3 pancreata from 1 litter for replicate 2) of whole pancreas. Contribution of each sample (replicate 1 and replicate 2) to the merged dataset is depicted in the inset. The scRNA-Seq dataset is from our previously published work [19]. **(B)** Bar graph depicts the proportion of each cell type within the scRNA-Seq and snATAC-Seq datasets. Colors match the cell types in (A). **(C)** Heatmaps depicting positively correlated TFs across pseudotime (from left to right) for the Delta lineage (Ductal to Spp1(+)/Neurog3(+), to Neurog3(+), to Fev(+)/Chgb(+), to Delta). **(D)** Venn diagram depicting the overlap in TFs among Alpha, Beta, and Delta GRNs. **(E–F)** Bar graphs showing the top 10 TFs with the highest number of activating (E) or repressing (F) regulations of target genes enriched in the Delta cell type. **(G)** Venn diagram depicting both the common and the timepoint-specific activated target genes of *Fev* in the E14.5 and E17.5 Fev(+)/Chgb(+) GRNs. Selected TFs within each group are listed in the boxes at the bottom. **(H)** Bar graph depicting pathways significantly (p-value <0.01) enriched among target genes activated by *Fev* that are unique to the E17.5 Fev(+)/Chgb(+) GRN. **(I)** Network diagram representing regulations between Fev and interacting TFs. Each TF is represented by a circle (node) that is colored by the cell type in which that TF is active in the GRN. Activating regulations are depicted by red lines, while repressing are depicted by blue lines.

Nfib (Ductal), *Pdx1* (Beta), *Usf2* (Fev(+)/Chga(+)), and *Ccnt2* (Fev(+)/Chga(+)) (Figure 7I, Supplementary Table 4). Taken together, these data have permitted the construction of a gene regulatory network of E17.5 exocrine and endocrine cells in the developing mouse pancreas and permitted a direct comparison of the important TFs and target genes across these two timepoints to yield a more refined understanding of the regulatory dynamics of endocrine maturation across developmental time.

4. DISCUSSION

Numerous studies have used scRNA-Seq to characterize developing mouse pancreas tissue, providing important insights into cellular heterogeneity and key transcriptional programs expressed in developing cell types [19—23]. Still, these datasets lack information about which of the expressed TFs are active and binding, and about how the TFs are organized into regulatory networks. Profiling of the chromatin accessibility landscape at single-cell resolution has emerged as a powerful approach for generating new insights about regulatory programs governing development and cell fate decisions across multiple tissue types [31,32,73,74], and we have now extended this approach to developing mouse pancreas tissue. Given that we were particularly interested in interrogating mechanisms underlying endocrine lineage allocation, we utilized a genetic tool to achieve significant enrichment of EP cells. Previous work from our laboratory had identified the transcription factor *Fev* as a marker of a novel EP state, and lineage reconstruction analysis indicated that it is at this *Fev*(+) state that lineage allocation is executed [19]. Here, we validated the use of an eFev-EYFP transgenic mouse line for enriching for *Fev*-expressing endocrine cells in the developing pancreas.

We have generated a comprehensive map of chromatin accessibility in the developing E14.5 and E17.5 murine pancreas, including enriched endocrine populations as well as non-endocrine cell types. Although previous studies have investigated chromatin accessibility in the developing pancreatic epithelium through bulk ATAC-Seq of sorted populations [20,26], to our knowledge our study represents the first to examine chromatin accessibility at single-cell resolution. By integrating both scRNA-Seq and snATAC-Seq data, we successfully generated a refined list of correlated TFs that are not only expressed, but also likely binding to open regions of chromatin to control cell fate decisions across developmental time. The analyses performed in this study utilized separate, computationally integrated scRNA and snATAC Seq datasets; in the future, performing multi-omic profiling of the transcriptional and chromatin states within the same cell may lend even greater confidence in the GRN reconstruction and identification of TFs governing cell fate.

Here, we have constructed cell-type specific GRNs describing active TFs and their putative target genes through the binding of *cis*-regulatory regions. Future work could leverage the datasets presented here to catalog in greater depth the gene-associated CARs to identify specific *cis*-regulatory elements operant in specific subtypes of cells on a gene-by-gene basis in a manner reported in other tissues, such as the human retina [75]. An advantage of our single cell level data is that it permits identification of not only cell type-specific, but even cell subtype-specific differences in active regulatory elements. Our analysis identified a number of known regulators of endocrine cell fate, such as TFs *Pdx1* and *Nkx6-1* in beta cells and *Arx* in alpha cells, as well as identified novel candidate TFs, such as *Mlxipl* in beta cells and *Pbx1* and *Peg3* in alpha cells. Furthermore, we identified a number of potential developmental timepoint-specific GRN TFs that may drive changes in proliferation and maturation that occur in endocrine cells as

development progresses [19—21]. Although here our focus within the epithelial compartment was on the endocrine lineages, our dataset also provides a rich resource for future interrogation of gene regulatory networks controlling acinar and ductal cell fates. Identification of these networks will inform efforts underway at generating stem cell-derived exocrine cells *in vitro* for studies aimed at understanding exocrine cell physiology and modeling of diseases such as cystic fibrosis, pancreatitis, and pancreatic cancer [76,77].

Traditional single-gene studies, along with genomic studies, have led to the identification of numerous TFs that play a functional role in regulating pancreatic endocrine differentiation. Although some individual TF—TF interaction pairs have been identified through these methods, the field still lacks an understanding of how these TFs are broadly arranged in regulatory networks across cell types and developmental stages. Our analysis permitted the creation of a TF—TF regulatory network, identifying TFs that control cell fate decisions through the binding and regulation of other important TFs. The assembly of these networks has identified both known and novel TF—TF interacting pairs whose associations can be experimentally validated in future studies using tools such as ChIP-Seq for confirmation of binding to specific DNA regions. Furthermore, CRISPR-mediated gene editing can be used to assess the downstream consequences of loss of individual candidate TFs on cell fate outcomes.

One TF that we focused on in this work is the gene *Fev*. In a beta cell line, *Fev* has been demonstrated to bind not only to serotonergic genes, reflective of common transcriptional cascades that drive the differentiation of both serotonergic neurons and of beta cells [78], but also to a conserved insulin gene regulatory element [67]. To investigate ubiquitous versus tissue-specific regions bound by Fev, we compared target peaks of *Fev* from our GRN analysis to Fev ChIP-Seq data obtained from developing serotonergic neurons and found some overlap of Fev-occupied chromatin between the two datasets. To probe downstream targets of Fev, we assessed the intersection of target genes of Fev in our embryonic pancreatic endocrine GRN with genes differentially expressed between *Fev* wildtype and knockout embryonic serotonergic neurons. This analysis identified candidate genes directly regulated by Fev in both tissue types. Generation of ChIP-Seq data from embryonic pancreas will permit further validation of our GRN analysis in a tissue-specific manner.

In contrast to the pancreatic epithelium, the cellular composition and transcriptional features of the pancreatic mesenchyme have been less well described. We and others have applied scRNA-Seq to mesenchymal tissue to identify transcriptionally distinct sub-populations [19,20] and infer lineage relationships among some of these cell subtypes [19]. In addition, functional heterogeneity among pancreatic mesenchymal cells has begun to be explored. For instance, one study reported that expression of *Pbx1* in a subset of Nkx2-5+ mesenchymal cells defines an anatomically specialized, pro-endocrine niche [24]. Which genes, including TFs, govern the acquisition of mesenchymal subpopulation identity, however, is poorly understood. Our work begins to investigate novel TFs regulating mesenchymal cell fate and will serve as an important resource for understanding mesenchymal development and function. Our dataset identifies which gene networks should be activated in order to generate not only organ-specific mesenchyme, but even mesenchymal subtypes from pluripotent stem cell sources [79]. Generation of stem cell-derived mesenchymal subtypes would then permit their co-culture with stem cell-derived endocrine cells to recapitulate *in vitro* the endogenous signaling events between mesenchyme and endocrine cells during development. In addition, the CellChat analysis presented here provides a complementary set of data by identifying candidate secreted mediators

of cell—cell signaling between mesenchymal and endocrine sub-populations.

The map of chromatin accessibility generated here not only provides deeper understanding of fundamental mechanisms underlying genetic control of developmental programs, but also holds relevance to the translational goals of beta cell regeneration and cell replacement therapy. For instance, our comprehensive characterization of chromatin state across endocrine development provides insights into the lineage plasticity observed among endocrine cells [7], and future work can leverage information about active endocrine cell type-specific GRNs to improve strategies for trans-differentiation of non-beta endocrine cells to the beta cell fate. Furthermore, the generation of functionally mature beta cells from stem cells remains a strong focus of cell replacement therapeutic strategies for patients with diabetes, and such *in vitro* protocols would benefit from an improved understanding of the dynamics in chromatin accessibility across endocrine development *in vivo*. Our dataset identifies which GRNs should be modulated *in vitro* to better approximate *in vivo* development. For instance, it will be interesting to benchmark recently published multi-omic datasets of human pluripotent stem cells undergoing differentiation to a beta cell fate [80,81] against our multi-omic dataset generated here to evaluate the fidelity of cells generated *in vitro* to their *in vivo* counterparts.

AUTHOR CONTRIBUTIONS

Sean de la O: Conceptualization, Methodology, Software, Formal Analysis, Investigation, Data Curation, Writing - original draft, Writing - review and editing, Visualization. Xinkai Yao: Methodology, Validation, Investigation, Visualization. Sean Chang: Validation, Investigation, Visualization. Zhe Liu: Conceptualization, Methodology, Validation, Formal Analysis, Investigation, Visualization. Julie B. Sneddon: Conceptualization, Methodology, Formal Analysis, Resources, Writing - original draft, Writing - review and editing, Visualization, Supervision, Project Administration, Funding Acquisition.

DECLARATION OF COMPETING INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

DATA AVAILABILITY

Raw and processed data for the single-nucleus ATAC-Sequencing analyses reported in this manuscript have been deposited to the GEO database under accession number GSE210569.

ACKNOWLEDGMENTS

APPENDIX A. SUPPLEMENTARY DATA

Supplementary data to this article can be found online at https://doi.org/10.1016/j.molmet.2023.101735.

REFERENCES

[1] Guz Y, Montminy MR, Stein R, Leonard J, Gamer LW, Wright CV, et al. Expression of murine STF-1, a putative insulin gene transcription factor, in beta cells of pancreas, duodenal epithelium and pancreatic exocrine and endocrine progenitors during ontogeny. Development 1995 Jan;121(1):11—8.

[2] Ohlsson H, Karlsson K, Edlund T. IPF1, a homeodomain-containing trans-activator of the insulin gene. EMBO J 1993 Nov;12(11):4251—9.

[3] Gu G, Dubauskaite J, Melton DA. Direct evidence for the pancreatic lineage: NGN3+ cells are islet progenitors and are distinct from duct progenitors. Development 2002 May;129(10):2447—57.

[4] Zhou Q, Law AC, Rajagopal J, Anderson WJ, Gray PA, Melton DA. A multipotent progenitor domain guides pancreatic organogenesis. Dev Cell 2007 Jul;13(1):103—14.

[5] Shih HP, Kopp JL, Sandhu M, Dubois CL, Seymour PA, Grapin-Botton A, et al. A Notch-dependent molecular circuitry initiates pancreatic endocrine and ductal cell differentiation. Development 2012 Jul;139(14):2488—99.

[6] Gradwohl G, Dierich A, LeMeur M, Guillemot F. neurogenin3 is required for the development of the four endocrine cell lineages of the pancreas. Proc Natl Acad Sci USA 2000 Feb 15;97(4):1607—11.

[7] van der Meulen T, Huising MO. Role of transcription factors in the trans-differentiation of pancreatic islet cells. J Mol Endocrinol 2015 Apr;54(2):R103—17.

[8] Ahlgren U, Jonsson J, Jonsson L, Simu K, Edlund H. beta-cell-specific inactivation of the mouse Ipf1/Pdx1 gene results in loss of the beta-cell phenotype and maturity onset diabetes. Genes Dev 1998 Jun 15;12(12):1763—8.

[9] Yang Y-P, Thorel F, Boyer DF, Herrera PL, Wright CVE. Context-specific α- to-β-cell reprogramming by forced Pdx1 expression. Genes Dev 2011 Aug 15;25(16):1680—5.

[10] Collombat P, Mansouri A, Hecksher-Sorensen J, Serup P, Krull J, Gradwohl G, et al. Opposing actions of Arx and Pax4 in endocrine pancreas development. Genes Dev 2003 Oct 15;17(20):2591—603.

[11] Collombat P, Hecksher-Sørensen J, Broccoli V, Krull J, Ponte I, Mundiger T, et al. The simultaneous loss of Arx and Pax4 genes promotes a somatostatin-producing cell fate specification at the expense of the alpha- and beta-cell lineages in the mouse endocrine pancreas. Development 2005 Jul;132(13):2969—80.

[12] Prado CL, Pugh-Bernard AE, Elghazi L, Sosa-Pineda B, Sussel L. Ghrelin cells replace insulin-producing beta cells in two mouse models of pancreas development. Proc Natl Acad Sci USA 2004 Mar 2;101(9):2924—9.

[13] Sander M, Sussel L, Conners J, Scheel D, Kalamaras J, Dela Cruz F, et al. Homeobox gene Nkx6.1 lies downstream of Nkx2.2 in the major pathway of beta-cell formation in the pancreas. Development 2000 Dec;127(24):5533—40.

[14] Sussel L, Kalamaras J, Hartigan-O'Connor DJ, Meneses JJ, Pedersen RA, Rubenstein JL, et al. Mice lacking the homeodomain transcription factor Nkx2.2 have diabetes due to arrested differentiation of pancreatic beta cells. Development 1998 Jun;125(12):2213—21.

[15] Schaffer AE, Taylor BL, Benthuysen JR, Liu J, Thorel F, Yuan W, et al. Nkx6.1 controls a gene regulatory network required for establishing and maintaining pancreatic Beta cell identity. PLoS Genet 2013 Jan 31;9(1):e1003274.

[16] Golosow N, Grobstein C. Epitheliomesenchymal interaction in pancreatic morphogenesis. Dev Biol 1962 Apr;4:242—55.

[17] Bhushan A, Itoh N, Kato S, Thiery JP, Czernichow P, Bellusci S, et al. Fgf10 is essential for maintaining the proliferative capacity of epithelial progenitor cells

during early pancreatic organogenesis. Development 2001 Dec;128(24): 5109—17.

[18] Landsman L, Nijagal A, Whitchurch TJ, Vanderlaan RL, Zimmer WE, Mackenzie TC, et al. Pancreatic mesenchyme regulates epithelial organogenesis throughout development. PLoS Biol 2011 Sep 6;9(9):e1001143.

[19] Byrnes LE, Wong DM, Subramaniam M, Meyer NP, Gilchrist CL, Knox SM, et al. Lineage dynamics of murine pancreatic development at single-cell resolution. Nat Commun 2018 Sep 25;9(1):3922.

[20] Scavuzzo MA, Hill MC, Chmielowiec J, Yang D, Teaw J, Sheng K, et al. Endocrine lineage biases arise in temporally distinct endocrine progenitors during pancreatic morphogenesis. Nat Commun 2018 Aug 22;9(1):3356.

[21] Bastidas-Ponce A, Tritschler S, Dony L, Scheibner K, Tarquis-Medina M, Salinno C, et al. Comprehensive single cell mRNA profiling reveals a detailed roadmap for pancreatic endocrinogenesis. Development 2019 Jun 17;(12): 146.

[22] Krentz NAJ, Lee MYY, Xu EE, Sproul SLJ, Maslova A, Sasaki S, et al. Single-cell transcriptome profiling of mouse and hESC-derived pancreatic progenitors. Stem Cell Rep 2018 Dec 11;11(6):1551—64.

[23] Yu X-X, Qiu W-L, Yang L, Zhang Y, He M-Y, Li L-C, et al. Defining multistep cell fate decision pathways during pancreatic development at single-cell resolution. EMBO J 2019 Apr 15;38(8).

[24] Cozzitorto C, Mueller L, Ruzittu S, Mah N, Willnow D, Darrigrand J-F, et al. A specialized niche in the pancreatic microenvironment promotes endocrine differentiation. Dev Cell 2020 Oct 26;55(2):150—62. e6.

[25] Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat Methods 2013 Dec;10(12):1213—8.

[26] Duvall E, Benitez CM, Tellez K, Enge M, Pauerstein PT, Li L, et al. Single-cell transcriptome and accessible chromatin dynamics during endocrine pancreas development. Proc Natl Acad Sci USA 2022 Jun 28;119(26):e2201267119.

[27] Buenrostro JD, Wu B, Litzenburger UM, Ruff D, Gonzales ML, Snyder MP, et al. Single-cell chromatin accessibility reveals principles of regulatory variation. Nature 2015 Jul 23;523(7561):486—90.

[28] Cusanovich DA, Daza R, Adey A, Pliner HA, Christiansen L, Gunderson KL, et al. Multiplex single cell profiling of chromatin accessibility by combinatorial cellular indexing. Science 2015 May 22;348(6237):910—4.

[29] Ranzoni AM, Tangherloni A, Berest I, Riva SG, Myers B, Strzelecka PM, et al. Integrative single-cell RNA-seq and ATAC-seq analysis of human developmental hematopoiesis. Cell Stem Cell 2021 Mar 4;28(3):472—87. e7.

[30] Duong TE, Wu Y, Sos BC, Dong W, Limaye S, Rivier LH, et al. A single-cell regulatory map of postnatal lung alveologenesis in humans and mice. Cell Genomics 2022 Mar 9;2(3).

[31] Lyu P, Hoang T, Santiago CP, Thomas ED, Timms AE, Appel H, et al. Gene regulatory networks controlling temporal patterning, neurogenesis, and cell-fate specification in mammalian retina. Cell Rep 2021 Nov 16;37(7):109994.

[32] Ziffra RS, Kim CN, Ross JM, Wilfert A, Turner TN, Haeussler M, et al. Single-cell epigenomics reveals mechanisms of human cortical development. Nature 2021 Oct 6;598(7879):205—13.

[33] Trevino AE, Müller F, Andersen J, Sundaram L, Kathiria A, Shcherbina A, et al. Chromatin and gene-regulatory dynamics of the developing human cerebral cortex at single-cell resolution. Cell 2021 Sep 16;184(19):5053—69. e23.

[34] Scott MM, Wylie CJ, Lerch JK, Murphy R, Lobur K, Herlitze S, et al. A genetic approach to access serotonin neurons for in vivo and in vitro studies. Proc Natl Acad Sci USA 2005 Nov 8;102(45):16472—7.

[35] Wyler SC, Spencer WC, Green NH, Rood BD, Crawford L, Craige C, et al. Pet-1 switches transcriptional targets postnatally to regulate maturation of serotonin neuron excitability. J Neurosci 2016 Feb 3;36(5):1758—74.

[36] Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. Nat Methods 2012 Jul;9(7):671—5.

[37] Yu KS, Frumm SM, Park JS, Lee K, Wong DM, Byrnes L, et al. Development of the mouse and human cochlea at single cell resolution. bioRxiv 2019. https://doi.org/10.1101/739680.

[38] Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM, et al. Comprehensive integration of single-cell data. Cell 2019 Jun 13;177(7):1888—902. e21.

[39] Granja JM, Corces MR, Pierce SE, Bagdatli ST, Choudhry H, Chang H, et al. ArchR: an integrative and scalable software package for single-cell chromatin accessibility analysis. Nat Gen 2021 Feb;53:403—11.

[40] Schep AN, Wu B, Buenrostro JD, Greenleaf WJ. chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. Nat Methods 2017 Oct;14(10):975—8.

[41] Bentsen M, Goymann P, Schultheis H, Klee K, Petrova A, Wiegandt R, et al. ATAC-seq footprinting unravels kinetics of transcription factor binding during zygotic genome activation. Nat Commun 2020 Aug 26;11(1):4267.

[42] Roopra A. MAGIC: a tool for predicting transcription factors and cofactors driving gene sets using ENCODE data. PLoS Comput Biol 2020 Apr 6;16(4): e1007800.

[43] Sugiyama T, Benitez CM, Ghodasara A, Liu L, McLean GW, Lee J, et al. Reconstituting pancreas development from purified progenitor cells reveals genes essential for islet differentiation. Proc Natl Acad Sci USA 2013 Jul 30;110(31):12691—6.

[44] Mellitzer G, Martín M, Sidhoum-Jenny M, Orvain C, Barths J, Seymour PA, et al. Pancreatic islet progenitor cells in neurogenin 3-yellow fluorescent protein knock-add-on mice. Mol Endocrinol 2004 Nov;18(11):2765—76.

[45] White P, May CL, Lamounier RN, Brestelli JE, Kaestner KH. Defining pancreatic endocrine precursors and their descendants. Diabetes 2008 Mar;57(3):654—68.

[46] Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, et al. Model-based analysis of ChIP-seq (MACS). Genome Biol 2008 Sep 17;9(9): R137.

[47] Xu EE, Krentz NAJ, Tan S, Chow SZ, Tang M, Nian C, et al. SOX4 cooperates with neurogenin 3 to regulate endocrine pancreas formation in mouse models. Diabetologia 2015 May;58(5):1013—23.

[48] jiang J, Lyu P, Li J, Huang S, Tao J, Blackshaw S, Qian J, Wang J. IReNA: Integrated regulatory network analysis of single-cell transcriptomes and chromatin accessibility profiles. iScience 2022 Nov 18;25(11):105359.

[49] Kim SK, Selleri L, Lee JS, Zhang AY, Gu X, Jacobs Y, et al. Pbx1 inactivation disrupts pancreas development and in Ipf1-deficient mice promotes diabetes mellitus. Nat Genet 2002 Apr;30(4):430—5.

[50] Anderson KR, White P, Kaestner KH, Sussel L. Identification of known and novel pancreas genes expressed downstream of Nkx2.2 during development. BMC Dev Biol 2009 Dec 10;9:65.

[51] Suriben R, Kaihara KA, Paolino M, Reichelt M, Kummerfeld SK, Modrusan Z, et al. β-Cell insulin secretion requires the ubiquitin ligase COP1. Cell 2015 Dec 3;163(6):1457—67.

[52] Schreiber V, Mercier R, Jiménez S, Ye T, García-Sánchez E, Klein A, et al. Extensive NEUROG3 occupancy in the human pancreatic endocrine gene regulatory network. Mol Metabol 2021 Nov;53:101313.

[53] Zhu Z, Li QV, Lee K, Rosen BP, González F, Soh C-L, et al. Genome editing of lineage determinants in human pluripotent stem cells reveals mechanisms of pancreatic development and diabetes. Cell Stem Cell 2016 Jun 2;18(6):755—68.

[54] Miyatsuka T, Li Z, German MS. Chronology of islet differentiation revealed by temporal cell labeling. Diabetes 2009 Aug;58(8):1863—8.

[55] Lynn FC, Smith SB, Wilson ME, Yang KY, Nekrep N, German MS. Sox9 coordinates a transcriptional network in pancreatic progenitor cells. Proc Natl Acad Sci USA 2007 Jun 19;104(25):10500—5.

[56] Cebola I, Rodríguez-Seguí SA, Cho CH-H, Bessa J, Rovira M, Luengo M, et al. TEAD and YAP regulate the enhancer network of human embryonic pancreatic progenitors. Nat Cell Biol 2015 May;17(5):615—26.

[57] Breslin MB, Wang H-W, Pierce A, Aucoin R, Lan MS. Neurogenin 3 recruits CBP co-activator to facilitate histone H3/H4 acetylation in the target gene INSM1. FEBS Lett 2007 Mar 6;581(5):949—54.

[58] Jia S, Ivanov A, Blasevic D, Müller T, Purfürst B, Sun W, et al. Insm1 co-operates with Neurod1 and Foxa2 to maintain mature pancreatic β-cell function. EMBO J 2015 May 12;34(10):1417—33.

[59] Murtaugh LC, Stanger BZ, Kwan KM, Melton DA. Notch signaling controls multiple steps of pancreatic differentiation. Proc Natl Acad Sci USA 2003 Dec 9;100(25):14920—5.

[60] Elhanani O, Salame TM, Sobel J, Leshkowitz D, Povodovski L, Vaknin I, et al. REST inhibits direct reprogramming of pancreatic exocrine to endocrine cells by preventing PDX1-mediated activation of endocrine genes. Cell Rep 2020 May 5;31(5):107591.

[61] Rovira M, Atla G, Maestro MA, Grau V, García-Hurtado J, Maqueda M, et al. REST is a major negative regulator of endocrine differentiation during pancreas organogenesis. Genes Dev 2021 Sep 1;35(17—18):1229—42.

[62] Martin D, Kim Y-H, Sever D, Mao C-A, Haefliger J-A, Grapin-Botton A. REST represses a subset of the pancreatic endocrine differentiation program. Dev Biol 2015 Sep 15;405(2):316—27.

[63] Takiishi T, Gysemans C, Bouillon R, Mathieu C. Vitamin D and diabetes. Endocrinol Metab Clin N Am 2010 Jun;39(2):419—46 [table of contents].

[64] Chen C, Luo Y, Su Y, Teng L. The vitamin D receptor (VDR) protects pancreatic beta cells against Forkhead box class O1 (FOXO1)-induced mitochondrial dysfunction and cell apoptosis. Biomed Pharmacother 2019 Sep;117:109170.

[65] Morró M, Vilà L, Franckhauser S, Mallol C, Elias G, Ferré T, et al. Vitamin D receptor overexpression in β-cells ameliorates diabetes in mice. Diabetes 2020 May;69(5):927—39.

[66] Fyodorov D, Nelson T, Deneris E. Pet-1, a novel ETS domain factor that can activate neuronal nAchR gene transcription. J Neurobiol 1998 Feb 5;34(2):151—63.

[67] Ohta Y, Kosaka Y, Kishimoto N, Wang J, Smith SB, Honig G, et al. Convergence of the insulin and serotonin programs in the pancreatic β-cell. Diabetes 2011 Dec;60(12):3208—16.

[68] Jin S, Guerrero-Juarez CF, Zhang L, Chang I, Ramos R, Kuan C-H, et al. Inference and analysis of cell-cell communication using CellChat. Nat Commun 2021 Feb 17;12(1):1088.

[69] Mutsaers SE. Mesothelial cells: their structure, function and role in serosal repair. Respirology 2002 Sep;7(3):171—91.

[70] Mutsaers SE. The mesothelial cell. Int J Biochem Cell Biol 2004 Jan;36(1):9—16.

[71] Que J, Wilm B, Hasegawa H, Wang F, Bader D, Hogan BLM. Mesothelium contributes to vascular smooth muscle and mesenchyme during lung development. Proc Natl Acad Sci USA 2008 Oct 28;105(43):16626—30.

[72] Ariza L, Cañete A, Rojas A, Muñoz-Chápuli R, Carmona R. Role of the Wilms' tumor suppressor gene Wt1 in pancreatic development. Dev Dynam 2018 Jul;247(7):924—33.

[73] Domcke S, Hill AJ, Daza RM, Cao J, O'Day DR, Pliner HA, et al. A human cell atlas of fetal chromatin accessibility. Science 2022;370:eaba7612.

[74] Chung C-Y, Ma Z, Dravis C, Preissl S, Poirion O, Luna G, et al. Single-cell chromatin analysis of mammary gland development reveals cell-state transcriptional regulators and lineage relationships. Cell Rep 2019 Oct 8;29(2):495—510. e6.

[75] Thomas ED, Timms AE, Giles S, Harkins-Perry S, Lyu P, Hoang T, et al. Cell-specific cis-regulatory elements and mechanisms of non-coding genetic disease in human retina and retinal organoids. Dev Cell 2022 Mar 28;57(6):820—36. e6.

[76] Simsek S, Zhou T, Robinson CL, Tsai S-Y, Crespo M, Amin S, et al. Modeling cystic fibrosis using pluripotent stem cell-derived human pancreatic ductal epithelial cells. Stem Cells Transl Med 2016 May;5(5):572—9.

[77] Delaspre F, Massumi M, Salido M, Soria B, Ravassard P, Savatier P, et al. Directed pancreatic acinar differentiation of mouse embryonic stem cells via embryonic signalling molecules and exocrine transcription factors. PLoS One 2013 Jan 17;8(1):e54243.

[78] Wilson ME, Scheel D, German MS. Gene expression cascades in pancreatic development. Mech Dev 2003 Jan;120(1):65—80.

[79] Kishimoto K, Iwasawa K, Sorel A, Ferran-Heredia C, Han L, Morimoto M, et al. Directed differentiation of human pluripotent stem cells into diverse organ-specific mesenchyme of the digestive and respiratory systems. Nat Protoc 2022 Nov;17(11):2699—719.

[80] Zhu H, Wang G, Nguyen-Ngoc K-V, Kim D, Miller M, Goss G, et al. Understanding cell fate acquisition in stem-cell-derived pancreatic islets using single-cell multiome-inferred regulomes. Developmental Cell 2023 April 10;58(9):727—743.E11.

[81] Augsornworawat P, Marquez E, Maestas MM, Ishahak M, Gale SE, Schmidt MD, et al. Multiomic profiling defines cell fate plasticity of in vitro-derived islets. bioRxiv 2022. https://doi.org/10.1101/2022.02.26.482126.