DEVELOPMENTAL BIOLOGY

The expression order determines the pioneer functions of NGN3 and NEUROD1 in pancreatic endocrine differentiation

Liu Yang[1,2]†, Xin-Xin Yu[1,2]†, Xin Wang[2,3], Chen-Tao Jin[1], Cheng-Ran Xu[1,2]*

Pioneer transcription factors (TFs) initiate chromatin remodeling, which is crucial for gene regulation and cell differentiation. In this study, we investigated how the sequential expression of neurogenin 3 (NGN3) and NEUROD1 affects their pioneering functions during pancreatic endocrine differentiation. Using a genetically engineered mouse model, we mapped NGN3-binding sites, confirming the pivotal role of this molecule in regulating chromatin accessibility. The pioneering function of NGN3 involves dose tolerance, and low doses are sufficient. Although NEUROD1 generally acts as a conventional TF, it can assume a pioneering role in the absence of NGN3. The sequential expression of *NeuroD1* and *Ngn3* predominantly drives α cell generation, which may explain the inefficient β cell induction observed in vitro. Our findings demonstrate that pioneer activity is dynamically shaped by temporal TF expression and inter-TF interactions, providing insights into transcriptional regulation and its implications for disease mechanisms and therapeutic targeting and enhancing in vitro differentiation strategies.

INTRODUCTION

Pioneer and master transcription factors (TFs) collaborate to regulate gene expression, driving cellular differentiation and ensuring precise developmental outcomes. Pioneer TFs recognize specific DNA sequences within compacted chromatin, initiating chromatin remodeling to make these regions accessible for transcription. This process enables master TFs to activate extensive gene networks, maintain cell identity, and refine gene expression patterns. This sequential interplay is essential for orchestrating the temporal and spatial regulation of genes during development and differentiation. Disruptions in the expression or function of these TFs can lead to aberrant cellular processes and disease.

The temporal order of TF expression plays a critical role in coordinating developmental programs. In various systems, pioneer TFs lay the foundation for subsequent transcriptional events by master TFs. In liver development, FOXA1 acts as a pioneer TF, remodeling compacted chromatin to enable HNF4α binding and activate liver-specific genes (*1*, *2*). In neurogenesis, ASCL1 primes chromatin for BRN2, driving the expression of neuronal differentiation genes (*3*). Similarly, PAX7 primes loci for TPIT binding during melanotrope development, initiating lineage-specific gene expression (*4*). In cardiac development, ISL1 opens critical chromatin regions, facilitating GATA4 activation of cardiac-specific genes (*5*). During hematopoiesis, CCAAT enhancer binding protein α creates accessible chromatin regions for PU.1, promoting macrophage differentiation (*6*). These examples highlight how the temporal sequence of TF expression ensures precise chromatin accessibility and transcriptional activation, enabling developmental programs to proceed in the correct cellular context.

The pancreas, an essential organ for endocrine and exocrine functions, offers an excellent model for studying TF dynamics during differentiation (*7*–*11*). In mice, pancreatic development begins with the emergence of multipotent progenitor (MP) cells at embryonic day 8.5 (E8.5), which differentiate into tip and trunk domains. Tip cells give rise to acinar cells, while trunk cells differentiate into ductal and endocrine progenitors (EPs). EPs undergo four developmental stages (EP1 to EP4), during which endocrine lineage specification occurs, leading to the formation of β, δ, α, and pancreatic polypeptide–producing cells.

Neurogenin 3 (NGN3), a basic helix-loop-helix (bHLH) TF, is a master regulator of trunk-to-EP differentiation (*12*, *13*). Its transient expression peaks at the EP2 stage and declines in EP4 (*8*, *14*–*17*). At the protein level, $NGN3^{+}$ cells display low expression ($Ngn3^{low}$) to high expression ($Ngn3^{high}$) as they transition from the trunk cells to EPs (*18*, *19*), orchestrating a transcriptional cascade critical for endocrine lineage commitment (*8*, *14*, *15*, *20*–*24*). In the absence of *Ngn3*, endocrine cells fail to form, resulting in diabetes and postnatal lethality (*16*). NGN3's role extends beyond gene activation to chromatin remodeling, suggesting its potential as a pioneer TF (*24*). However, direct in vivo evidence of NGN3's pioneering activity and chromatin-specific functions remains limited.

NEUROD1, another bHLH TF, acts downstream of NGN3 and is expressed from the EP2 stage onward, persisting through endocrine differentiation (*14*, *17*, *25*, *26*). NEUROD1 is essential for endocrine cell maturation and functional maintenance, and its absence disrupts cell identity, leading to neonatal diabetes. Previous studies suggest that NEUROD1 may also exhibit pioneer-like functions under specific conditions (*27*, *28*), although its role has primarily been described as a conventional TF.

Despite extensive research on these TFs, the mechanisms underlying their sequential expression and their impact on chromatin remodeling and transcriptional activation remain poorly understood. Specifically, how NGN3 primes chromatin for NEUROD1 and how NEUROD1's role changes in the absence of NGN3 are open questions.

In this study, we leverage the pancreatic EP differentiation model to investigate the distinct and complementary roles of NGN3 and NEUROD1, emphasizing how their sequential interplay orchestrates chromatin remodeling and gene activation during endocrine differentiation. Through comparative analyses of in vivo and in vitro systems, we uncover critical insights into the temporal regulation of these TFs and their impact on cell fate decisions. These findings not

[1]State Key Laboratory of Female Fertility Promotion, Department of Medical Genetics, School of Basic Medical Sciences, Peking University, Beijing 100191, China. [2]Peking-Tsinghua Center for Life Sciences, Academy for Advanced Interdisciplinary Studies, Peking University, Beijing 100871, China. [3]School of Life Sciences, Peking University, Beijing 100871, China.
*Corresponding author. Email: cxu@pku.edu.cn
†These authors contributed equally to this work.

only advance our understanding of transcriptional hierarchies during pancreatic development but also address key challenges in generating functional β cells in vitro. By elucidating these mechanisms, our study provides a framework for improving differentiation protocols and offers potential therapeutic strategies for diabetes and other endocrine-related disorders.

RESULTS

A genetically engineered *Ngn3-Flag-P2A-EGFP* mouse line

Because of the lack of a chromatin immunoprecipitation (ChIP)–grade NGN3 antibody for studying the regulatory dynamics of NGN3 during the transition from $Ngn3^{low}$ to $Ngn3^{high}$ expression in endocrine differentiation (*18*, *19*), we engineered an *Ngn3-Flag-P2A-EGFP* mouse line. This line features a FLAG tag at the C terminus of the NGN3 protein, followed by a self-cleaving P2A peptide sequence and an enhanced green fluorescent protein (EGFP) reporter (Fig. 1A). Immunofluorescence (IF) analysis confirmed the colocalization of the NGN3-FLAG fusion protein with endogenous NGN3 in the E14.5 pancreas (Fig. 1B). Flow cytometric analysis was used to categorize *Ngn3*-expressing cells into $Ngn3^{low}$ and $Ngn3^{high}$ populations on the basis of EGFP fluorescence intensity (Fig. 1C). Single-cell RNA sequencing (scRNA-seq) via the Smart-seq2 method (*29*) confirmed that E14.5 $Ngn3^{low}$ cells were predominantly trunk and EP1 cells, whereas $Ngn3^{high}$ cells were primarily EP2 to EP4 and endocrine cells, which is consistent with previous Smart-seq2 scRNA-seq data (fig. S1, A to C) (*14*, *15*). This engineered mouse line thus enables precise mapping of NGN3-binding sites without affecting EP development, providing a robust tool for studying the regulatory role of NGN3 in endocrine differentiation.

NGN3 occupancy during EP development

To investigate NGN3 occupancy during endocrine development, we used Cleavage Under Targets and Release Using Nuclease (CUT&RUN) method (*30*) with an anti-FLAG antibody on E13.5 $Ngn3^{low}$ and $Ngn3^{high}$ cells from our engineered mouse line (Fig. 1A). This approach identified 4682 NGN3-binding sites in $Ngn3^{high}$ cells, with lower binding levels in $Ngn3^{low}$ cells (Fig. 1, D and E, and table S1). While only a small fraction of binding peaks was found in $Ngn3^{low}$ cells, most sites presented higher signal intensities than did random genomic regions (fig. S1D). The sequence motifs from these sites generally match known NGN3 motifs from human EPs in vitro (*31*) and high-throughput systematic evolution of ligands by exponential enrichment (HT-SELEX) studies (*32*), validating our CUT&RUN analysis (Fig. 1F).

On the basis of previously published ULI-NChIP (ultralow-input micrococcal nuclease-based native ChIP)–sequencing (NChIP-seq) data of histone modifications associated with enhancers (H3K4me1 and H3K27ac) and promoters (H3K4me3 and H3K27me3) (*28*), we observed that most NGN3-binding sites were within enhancer regions marked by H3K4me1 but not H3K4me3 or other distal regions more than 2 kb from the transcription start site (TSS), accounting for 81.0% of the binding locations (Fig. 1G and fig. S1E). In contrast, 19.0% of the loci were associated with promoter regions located less than 2 kb from the TSS (Fig. 1G and fig. S1E).

Genes associated with NGN3 peaks were significantly up-regulated during the transition from the $Ngn3^{-}$ trunk stage to the $Ngn3^{low}$ (trunk-EP and EP1) stage and then to the $Ngn3^{high}$ (EP2 to EP4) stage, as shown by the scRNA-seq data (Fig. 1H and fig. S1B) (*14*). Gene Ontology (GO) analysis revealed that these genes are involved in critical biological processes such as cell differentiation, morphogenesis, and endocrine functions, including ion transport and glucose homeostasis (Fig. 1I and table S2). Single or multiple NGN3-binding sites were observed at the promoter region of *Pax4* and *NeuroD1* and the enhancer regions of *Nkx6-1*, *Pax6*, *Sox4*, *Pou3f4*, and *Nkx2-2*, which are crucial for endocrine lineage development (Fig. 1J and fig. S1, F and G). NGN3 also bound its own enhancer region (Fig. 1J), suggesting a positive autofeedback mechanism for its regulation (*33*). In summary, this genome-wide analysis of NGN3 occupancy underscores its central role in orchestrating the molecular landscape of endocrine differentiation and development.

Dual roles for NGN3 in EP specification

To elucidate the role of NGN3 in directing endocrine specification, we examined the chromatin accessibility dynamics within NGN3-bound regions during the differentiation of endocrine cell. Using the 10x Genomics single-nucleus assay for transposase-accessible chromatin sequencing (snATAC-seq), we analyzed pancreatic epithelial cells labeled with *Pdx1-Cre;Rosa26-RFP* at E13.5 and E15.5 and pancreatic endocrine cells labeled with *Ngn3-Cre^{ER};Rosa26-RFP* at E14.5, E15.5, and E16.5 (Fig. 2A). After quality control, the dataset comprised 36,296 cells, revealing distinct cellular clusters corresponding to the tip, trunk, EP1 to EP4, endocrine, and exocrine lineages (fig. S2A and table S3). The identification of these cell types was confirmed by the chromatin accessibility of lineage-specific markers (fig. S2B).

Chromatin accessibility dynamics in tip, trunk, and EP1 to EP4 cells revealed two groups of NGN3-bound regions, classified using the shared nearest neighbor (SNN) clustering algorithm based on the fold changes of aggregated peak signals between the EP stages and earlier tip and trunk stages (Fig. 2B and fig. S2C). To ensure robust classification, we tested multiple resolution settings in the clustering analysis. While subtle heterogeneity emerged at varying resolutions, two biologically meaningful groups remained consistent across settings (fig. S2D). The group 1 regions presented substantially greater peak counts and increased chromatin accessibility starting at the EP1 stage, which coincided with elevated *Ngn3* expression (Fig. 2, B and C, and fig. S2E) (*14*), indicating the pioneering role of NGN3. These regions were associated with TFs essential for endocrine development, including *NeuroD1*, *Pax6*, *Sox4*, *Pax4*, *Foxo1*, and *Ngn3* (Fig. 2D and fig. S2F). Conversely, group 2 regions presented greater chromatin accessibility from the tip stage through the EP stage, suggesting a collaborative role of NGN3 in these areas (Fig. 2, B and C). Combined with the results of scRNA-seq analysis (*14*), these results showed that genes in group 1 presented lower expression levels at the trunk stage than did those in group 2 but increased expression from EP1 onward, whereas group 2 genes maintained stable expression throughout this process (Fig. 2E).

We found that group 1 contained a greater proportion of enhancer-associated regions compared to group 2 (Fig. 2F). To characterize the chromatin landscapes of these NGN3-bound groups, we examined histone modifications associated with enhancers and promoters during the transition from $Ngn3^{low}$ to $Ngn3^{high}$ using previous NChIP-seq data (*28*) and recently generated CUT&RUN data for p300 binding (fig. S2G). Both groups exhibited a shift to a more active chromatin state at enhancers in $Ngn3^{high}$ cells, marked by increased H3K4me1, H3K27ac, and p300 levels (Fig. 2G). In contrast, histone modifications at promoter regions remained largely unchanged in

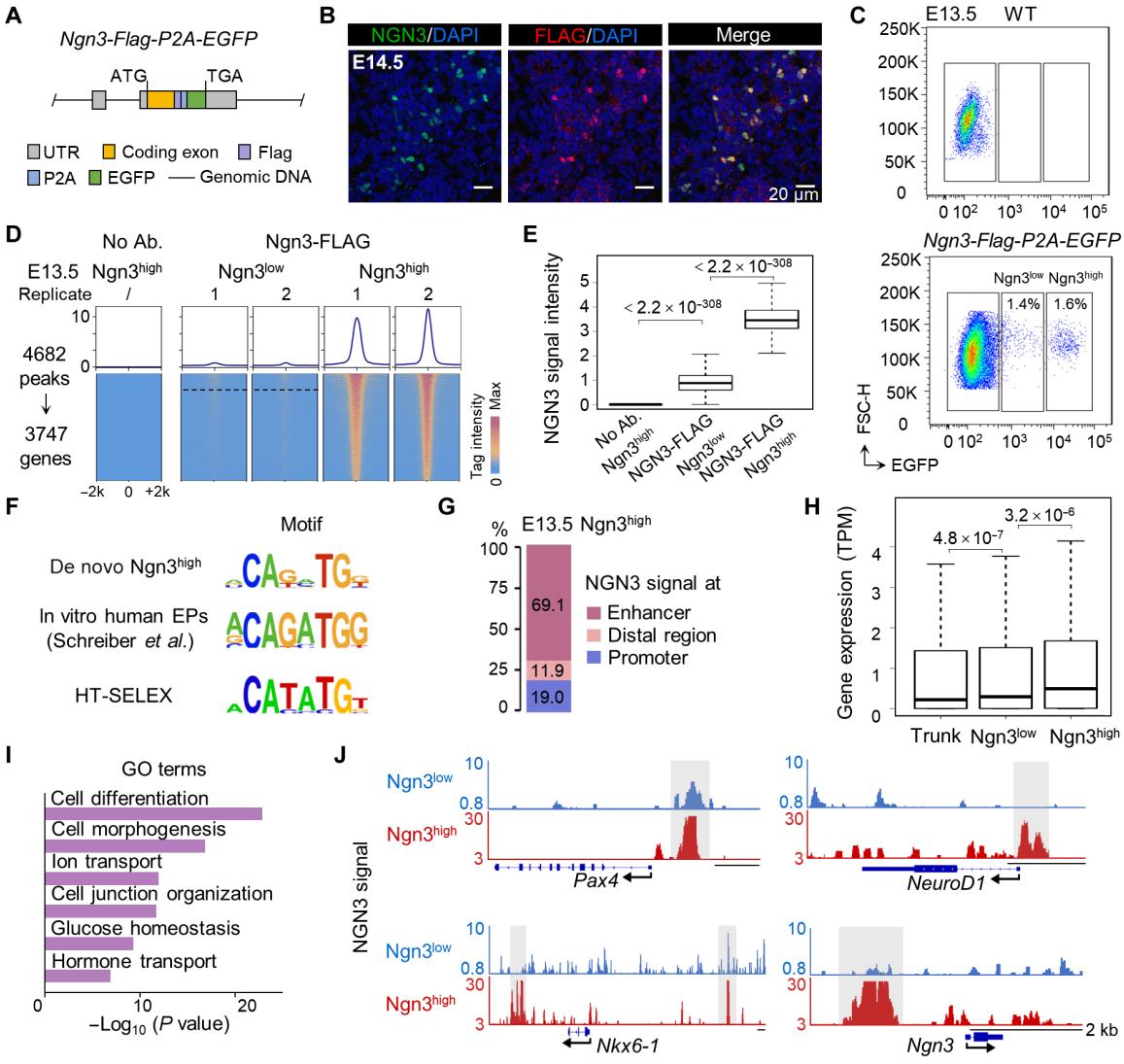

Fig. 1. Identification of NGN3 targets via the *Ngn3-Flag-P2A-EGFP* mouse line. (**A**) Generation of the *Ngn3-Flag-P2A-EGFP* strain by inserting the Flag-P2A-EGFP sequence before the stop codon of *Ngn3*. The FLAG tag was fused to NGN3, and P2A enabled EGFP expression as a marker for *Ngn3*. (**B**) IF staining of NGN3 and NGN3-FLAG fusion proteins in the E14.5 *Ngn3-Flag-P2A-EGFP* pancreas. Scale bars, 20 μm. (**C**) FACS gating strategies for purifying E13.5 Ngn3-Flag-EGFP^{low} and Ngn3-Flag-EGFP^{high} cells. WT (wild-type) pancreata served as negative controls. (**D**) Metaplots and heatmaps illustrating NGN3 binding in no-antibody control (No Ab.), E13.5 Ngn3^{low}, and Ngn3^{high} cells. Each line represents a peak, with color intensity indicating NGN3 binding at ±2 kb from peak center. The peaks above the dashed line are binding peaks identified in Ngn3^{low} cells. Replicate 1 and replicate 2 denote two independent experiments. (**E**) Intensity levels of NGN3 binding signals in the No Ab., Ngn3^{low}, and Ngn3^{high} cells. All *P* values were below 2.2×10^{-308} (Wilcoxon rank-sum test). (**F**) De novo motif enriched by NGN3 CUT&RUN compared to motif identified in human EPs in vitro (*31*) and through HT-SELEX studies (*32*). (**G**) Genomic distribution of NGN3-binding sites in E13.5 Ngn3^{high} cells. (**H**) Boxplots showing mRNA expression levels of genes with NGN3 binding in E13.5 Ngn3^{high} cells during development from trunk (Ngn3^{-}) to Ngn3^{low} (trunk-EP and EP1) and Ngn3^{high} (EP2 to EP4) cells. *P* values were calculated via the Wilcoxon rank-sum test. (**I**) Selected Gene Ontology (GO) terms for genes with NGN3 binding in E13.5 Ngn3^{high} cells. (**J**) Bedgraph plots showing normalized NGN3 binding signals of representative genes in E13.5 Ngn3^{low} and Ngn3^{high} cells. The peaks in Ngn3^{high} cells are highlighted in gray. Scale bars, 2 kb.

both groups. Notably, group 2 regions displayed a more active chromatin state compared to group 1, with higher H3K4me3 levels and lower H3K27me3 levels (Fig. 2G). To further evaluate the chromatin state of group 1 regions, we performed NChIP-seq (*34*) for H3K4me1, a marker of primed chromatin, in GFP^{high} cells sorted from E11.5 *Pdx1-GFP* mice (fig. S2H) (*35*). These GFP^{high} cells, referred to as Pdx1^{high} cells, represent MP-late, tip, and early trunk stages before NGN3 expression. Group 1 regions exhibited very low levels of H3K4me1 in Pdx1^{high} cells (fig. S2I), indicating that they were in a closed chromatin state rather than a primed state before NGN3 expression.

To confirm the multifaceted roles of NGN3, we knocked out (KO) *Ngn3* in *Ngn3-GFP/GFP* homozygous mice (*36*), where GFP replaces the *Ngn3* coding sequence (CDS), resulting in a complete loss of NGN3 protein function from the outset. scRNA-seq analysis of E13.5 Ngn3-GFP^{+} cells from the *Ngn3*-KO mice confirmed complete arrest in the trunk state, with failure to progress to the EP1 stage (Fig. 3A and fig. S2J). A recent ATAC-seq analysis of E15.5

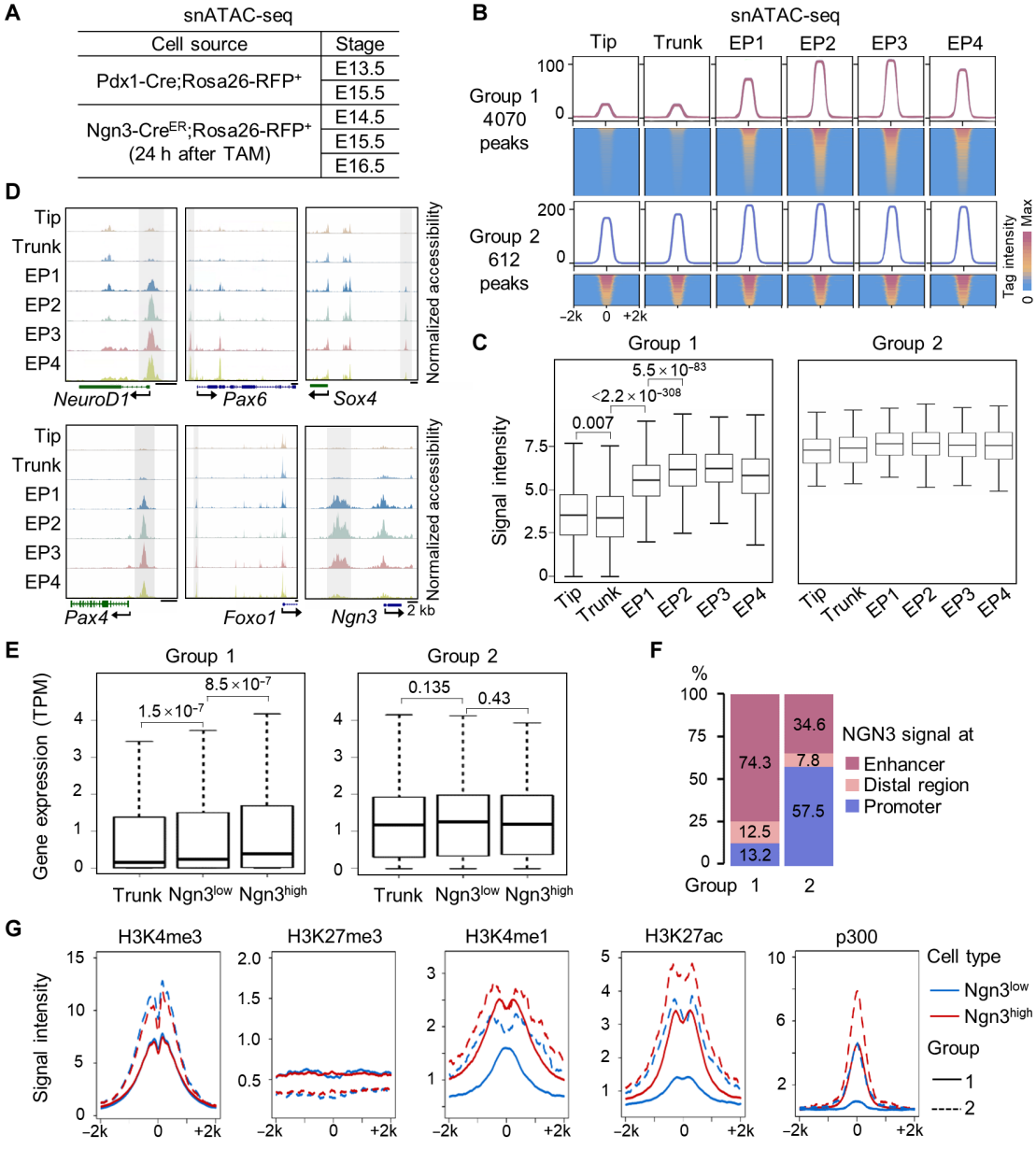

Fig. 2. Developmental dynamics of NGN3 target sites and their chromatin features. (**A**) Overview of the cell sources at different developmental time points analyzed using 10x snATAC-seq in this study. Ngn3-Cre^{ER};Rosa26-RFP^{+} cells were sorted from mice with a single tamoxifen injection 24 hours before harvest. (**B**) Metaplots and heatmaps illustrating the dynamics of chromatin accessibility at ±2 kb from the center of NGN3-binding sites during development from tip, trunk to EP4. Each row represents a peak, and each column represents a cell type. Colors ranging from blue to red indicate low to high chromatin accessibility in each cell type. (**C**) Boxplots showing the intensity levels of chromatin accessibility of group 1 and group 2 sites during development from tip, trunk to EP1 to EP4 cells. *P* values were calculated by Wilcoxon rank-sum test. (**D**) Bedgraph plots showing chromatin accessibility at gene loci in group 1. Regions identified as NGN3 target sites are highlighted in gray. The scale bars indicate 2 kb. (**E**) Boxplots showing the mRNA expression levels of genes with NGN3 binding in group 1 and group 2 during development from trunk to Ngn3^{low} and Ngn3^{high} cells. *P* values were calculated by Wilcoxon rank-sum test. (**F**) Genomic distribution of NGN3-binding sites in group 1 and group 2 as identified in (B). (**G**) Composite plots showing the distribution of H3K4me3 and H3K27me3 marks around TSSs (±2 kb) and H3K4me1, H3K27ac, and p300 marks around enhancers (±2 kb).

Ngn3-KO embryos revealed a chromatin accessibility profile similar to that of ductal cells, highlighting the potential pioneering role of NGN3 in endocrine specification (*24*). To precisely delineate NGN3's function, we performed bulk-cell ATAC-seq on Pdx1^{high} cells at E11.5 as well as Ngn3^{low}, Ngn3^{high}, and *Ngn3*-KO GFP^{+} cells at E13.5 (Fig. 3B and fig. S2K). At NGN3-bound sites, chromatin accessibility at group 1 loci in *Ngn3*-KO cells was highly similar to that of Pdx1^{high} cells but significantly lower compared to Ngn3^{high} cells and even Ngn3^{low} cells (Fig. 3, B and C), as evident at loci such as *Pax4*, *Ngn3*, and *NeuroD1* (Fig. 3D). In contrast, group 2 loci presented similar levels of chromatin accessibility across Pdx1^{high}, Ngn3^{low}, Ngn3^{high}, and *Ngn3*-KO cells, indicating that NGN3 does not play a pioneering role at these sites (Fig. 3B). These observations confirm the role of NGN3 in facilitating chromatin accessibility during the

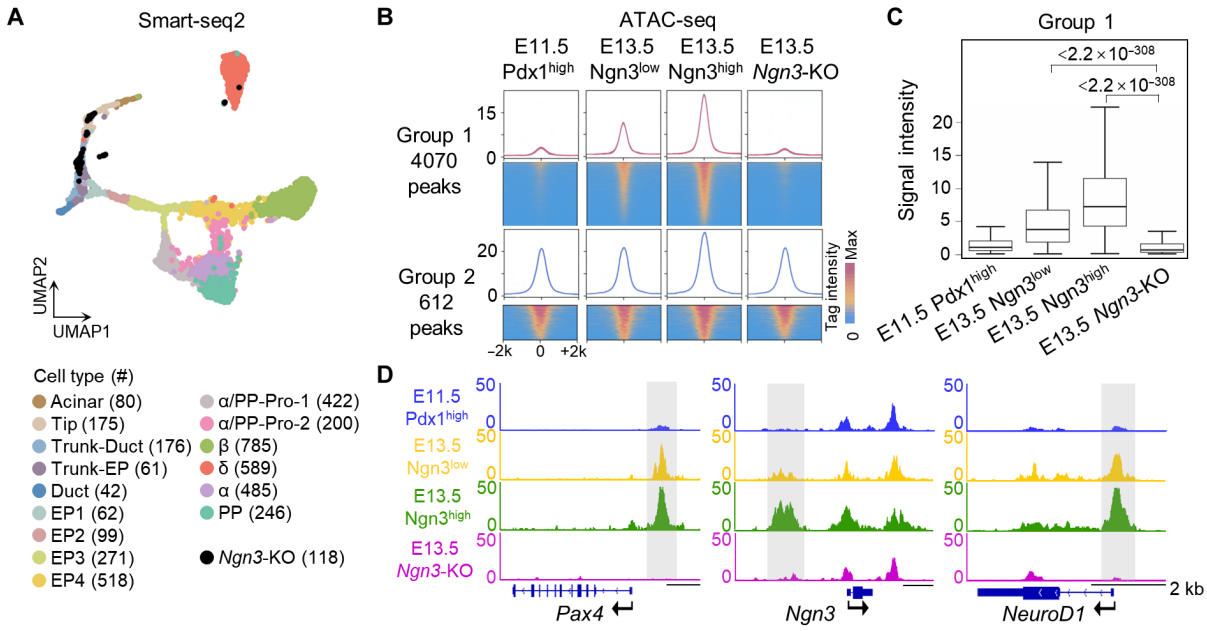

Fig. 3. Functional consequences of NGN3 deletion on chromatin accessibility and gene regulation. (**A**) UMAP plot displaying the distribution of E13.5 *Ngn3-GFP/GFP* (*Ngn3*-KO) single cells across distinct cell types identified by the published scRNA-seq dataset (*14*). Cell counts are labeled in brackets. (**B**) Metaplots and heatmaps illustrating the chromatin accessibility of bulk E11.5 Pdx1^{high}, E13.5 Ngn3^{low}, Ngn3^{high} and E13.5 *Ngn3*-KO cells at ±2 kb from the center of NGN3-binding sites in group 1 and group 2 from Fig. 2B. Each line represents a peak, with colors ranging from blue to red indicating low to high levels of chromatin accessibility. (**C**) Boxplots showing the intensity levels of chromatin accessibility of group 1 in E11.5 Pdx1^{high}, E13.5 Ngn3^{low}, Ngn3^{high}, and E13.5 *Ngn3*-KO cells. *P* values were calculated by Wilcoxon rank-sum test. (**D**) Bedgraph plots showing the normalized ATAC signal of representative genes in group 1 and group 2. Regions identified as NGN3 target sites are highlighted in gray. Scale bars, 2 kb.

trunk-to-Ngn3^{low} and Ngn3^{high} transitions, highlighting its function in initiating low-level chromatin accessibility at group 1 sites in Ngn3^{low} cells, even where NGN3 binding may not reach the threshold of detection by CUT&RUN (Fig. 1D). This finding is consistent with the theoretical properties attributed to pioneer factors (*37*).

Therefore, the combined analyses of CUT&RUN, ChIP-seq, and ATAC-seq data indicate that NGN3 plays dual roles, both pioneering and nonpioneering, during EP development.

NGN3 dose tolerance in EP differentiation

To determine whether increased NGN3 dosage promotes its pioneering function, we generated a conditional overexpression (COE) model using an engineered mouse strain, *Ngn3-mCherry-OE*, driven by *Ngn3-Cre* (*Ngn3-Cre;Ngn3*-COE) (Fig. 4A). In this model, overexpressed cells presented mCherry fluorescence and elevated NGN3 protein levels, with the average NGN3 protein level quantified as 1.37-fold higher than in control cells (Fig. 4, B and C). Increased NGN3 expression resulted in a greater proportion of NGN3^{+} cells compared to MUC1^{+} trunk cells (Fig. 4D).

To investigate the effects of elevated NGN3 dosage on developmental pathways and gene expression, we performed 10x Genomics scRNA-seq analysis on cells isolated from E14.5 and E17.5 *Ngn3-Cre;Ngn3*-COE and control pancreata. Since control pancreata, including wild-type (WT), *Ngn3-Cre*, or *Ngn3-OE*/+ genotypes, lacked fluorescent markers, we used epithelial cell adhesion molecule (EpCAM) as a marker for fluorescence-activated cell sorting (FACS) to enrich pancreatic epithelial cells in both control and *Ngn3-Cre;Ngn3*-COE samples (fig. S3A). To minimize batch effects, we pooled control samples from one sex and *Ngn3-Cre;Ngn3*-COE samples from the opposite sex into the same 10x Genomics run, distinguishing cellular resources by sex-specific genes (fig. S3, B to G). Uniform manifold approximation and projection (UMAP) analysis revealed consistent cell clusters and developmental pathways between control and *Ngn3-Cre;Ngn3*-COE samples at E14.5 (Fig. 4E and table S3). Elevated NGN3 protein levels in the *Ngn3-Cre;Ngn3*-COE samples corresponded to increased *Ngn3* expression from the EP1 stage onward (Fig. 4, C and F), indicating that a higher NGN3 dosage does not significantly impact EP or endocrine cell states.

In addition, at E17.5, transcriptional profiles of α, β, and δ cells were nearly identical between control and *Ngn3-Cre;Ngn3*-COE cells, with observed differences limited to several sex-related genes (fig. S3, H and I, and tables S3 and S4). Intriguingly, despite increased *Ngn3* expression in differentiated endocrine cells (fig. S3J), NGN3 protein remained undetectable in insulin (INS)$^{+}$ and glucagon (GCG)$^{+}$ cells (fig. S3K).

To modulate *Ngn3* expression at lower levels, we created a conditional knockout (CKO) in a specialized mouse strain, *Ngn3-mCherry-flox/flox*, via *Ngn3-Cre* (*Ngn3-Cre;Ngn3*-CKO). In this model, Cre recombinase expression, driven by the *Ngn3* promoter, excises the *Ngn3* CDS flanked by *loxP* sites (Fig. 5A). This process requires transcriptional activation of *Ngn3* and occurs gradually, leading to a delayed loss of NGN3 protein.

IF revealed significantly lower NGN3 levels in the E14.5 *Ngn3-Cre;Ngn3*-CKO pancreata than in the *Ngn3-mCherry-flox/flox* pancreata (Fig. 5, B and C), resulting in a reduced proportion of EPs in the *Ngn3-Cre;Ngn3*-CKO mice (Fig. 5D). Using a strategy similar to that for the *Ngn3-Cre;Ngn3*-COE mice, we conducted 10x Genomics scRNA-seq on E14.5 EpCAM^{+} cells from the control (WT, *Ngn3-Cre*, or

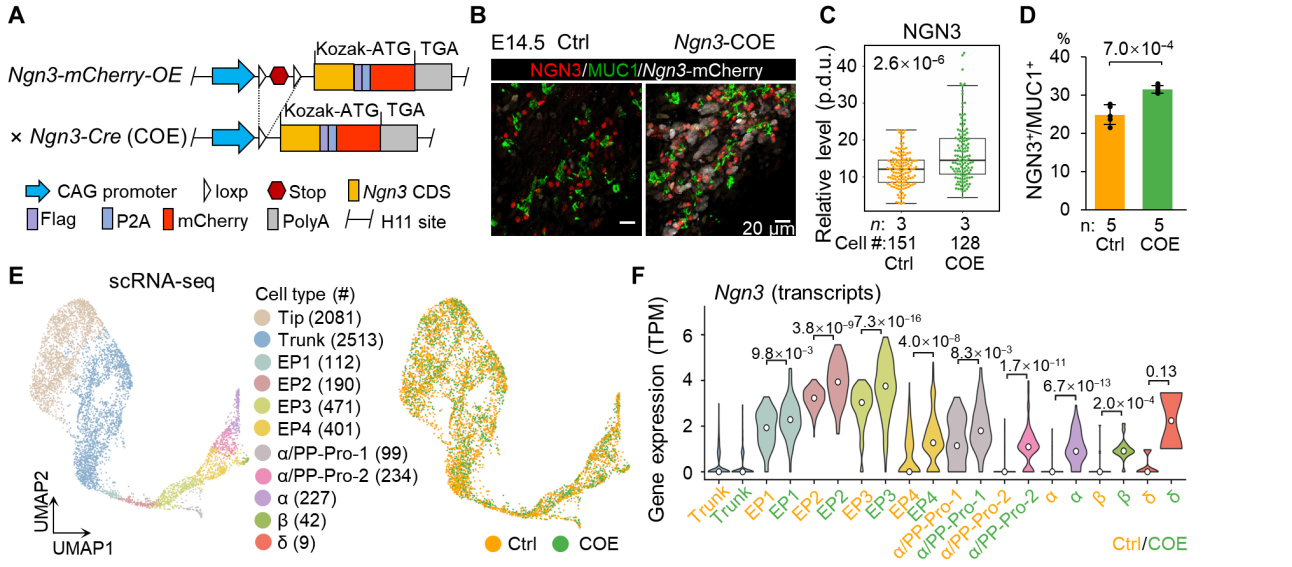

Fig. 4. Tolerance to higher doses of NGN3 in EP differentiation. (**A**) Generation of the *Ngn3-mCherry-OE* strain via the insertion of the CAG-LSL-*Ngn3* CDS-Flag-P2A-mCherry-PolyA sequence at the H11 site. Upon *Ngn3-Cre*–mediated excision of the stop sequence, the CDS of *Ngn3* is expressed under the CAG promoter. (**B**) *Ngn3*-mCherry expression and IF staining of NGN3 and MUC1 in E14.5 pancreata from the control (Ctrl: WT, *Ngn3-Cre*, or *Ngn3-OE/+*) and *Ngn3-Cre;Ngn3-OE/+* (*Ngn3*-COE) embryos. Scale bars, 20 μm. (**C**) Statistical analysis of NGN3 IF intensity in E14.5 pancreata from the control (Ctrl) and *Ngn3-Cre;Ngn3*-COE (COE) embryos. *n* denotes the number of embryos analyzed. "Cell #" indicates the total number of the analyzed NGN3$^+$ cells. *P* values were calculated via the Wilcoxon rank-sum test. p.d.u., program-defined unit. (**D**) Statistical analysis of the proportion of NGN3$^+$ cells relative to MUC1$^+$ trunk cells in the E14.5 control (Ctrl) and *Ngn3-Cre;Ngn3*-COE (COE) pancreata. The data are presented as the means ± SEMs. *n* denotes the number of embryos analyzed. *P* values were calculated via the Wilcoxon rank-sum test. (**E**) UMAP plots of 10x scRNA-seq data displaying the distribution of cell types (left) from the E14.5 control (Ctrl) and *Ngn3-Cre;Ngn3*-COE (COE) pancreata (right). The control and COE samples each had two replicates. The cell counts are labeled in brackets. (**F**) Violin plots illustrating the increased expression level of *Ngn3* in different cell types from the *Ngn3-Cre;Ngn3*-COE (COE) pancreata compared with the control (Ctrl). The white dots indicate the average expression levels. *P* values were calculated via the Wilcoxon rank-sum test.

Ngn3- flox/flox) (male) or *Ngn3-Cre*;*Ngn3*-CKO (female) pancreata combined in the same sample. The analysis revealed consistent cell clusters and developmental pathways between the control and *Ngn3-Cre;Ngn3*-CKO pancreata (Fig. 5E; fig. S4, A to E; and table S3 and S4), with a notable decrease in *Ngn3* expression level in EPs from the *Ngn3-Cre;Ngn3*-CKO mice (Fig. 5F). In addition, IF revealed lower percentages of INS$^+$ and GCG$^+$ cells in the pancreata of the *Ngn3-Cre*;*Ngn3*-CKO mice than in those of the control mice (fig. S4, F and G).

To assess the impact of a reduced NGN3 dosage on chromatin accessibility, we performed snATAC-seq on E14.5 EpCAM$^+$ cells from the control or *Ngn3-Cre*;*Ngn3*-CKO mice. The results revealed consistent cell clusters with no significant differences in chromatin accessibility between the control and *Ngn3-Cre*;*Ngn3*-CKO pancreata (fig. S4, H to J, and tables S3 and S5).

Comparing NGN3 expression levels across CKO, control, and COE samples revealed significant differences, with COE samples displaying an average threefold higher NGN3 level compared to CKO samples (Figs. 4C and 5C). Despite these variations, analysis showed no differences in cell clusters or developmental trajectories between COE and CKO groups (Figs. 4E and 5E).

Overall, our findings demonstrate that endocrine cell differentiation exhibits NGN3 dose tolerance. Relatively low NGN3 levels are sufficient to regulate chromatin accessibility and drive endocrine specification, whereas higher NGN3 doses do not enhance transcriptomic outcomes or developmental trajectories.

Association of NEUROD1 and NGN3 targeting

To elucidate the regulatory function of NEUROD1, a downstream target of NGN3, during endocrine development, we conducted CUT&RUN assays on E14.5 Ngn3^{high} cells with an anti-NEUROD1 antibody (fig. S5A). Following quality control, we identified 11,051 peaks in these cells. Notably, NGN3 and NEUROD1 share a core binding motif, consistent with previous findings (*31*), but their flanking sequence preferences are distinct (Fig. 1F and fig. S5B), indicating distinct binding patterns and functional roles in pancreatic endocrine differentiation.

Integration of snATAC-seq data across tip, trunk, EP, and embryonic α and β cells revealed that NEUROD1-bound regions were either pre-accessible in tip/trunk cells or became accessible during the trunk-to-EP1 transition, preceding *NeuroD1* expression at the EP2 stage, and remained accessible at later stages (Fig. 6A, fig. S2E, and table S1).

By comparing NEUROD1 and NGN3-binding sites, we identified 2201 NGN3-specific, 8571 NEUROD1-specific, and 2477 overlapping regions (Fig. 6B and table S1). Genes bound by NGN3, whether NGN3-specific or shared, showed dynamic expression during differentiation, peaking in EP2 and down-regulating at EP3 (Fig. 6C). These included key TFs for endocrine specification, such as *Ngn3*, *NeuroD1*, *Pax4*, *Gata4*, and *St18* (fig. S5C) (*38*, *39*). In contrast, NEUROD1-specific genes were up-regulated later, at EP3, and remained active in mature endocrine cells. These genes—including *Fev*, *Mnx1*, *Vdr*, and *Nkx6-1*—are critical for β cell maintenance and insulin secretion (fig. S5C) (*38*, *40*, *41*). Notably, NEUROD1-specific binding at *Arx*, which antagonizes *Pax4* to promote α cell fate, highlights NEUROD1's functional specificity (fig. S5C).

In addition to their distinct peak regions, both NGN3 and NEUROD1 exhibit relatively high signal intensities in each other's binding regions, even when these regions fall below peak-calling

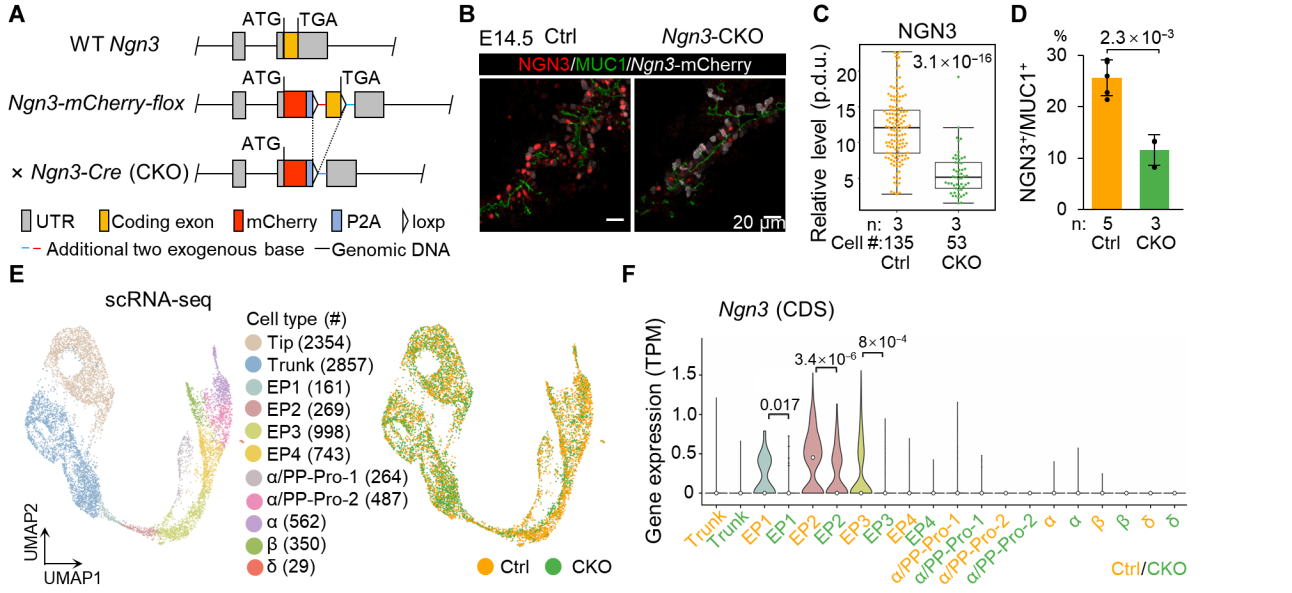

Fig. 5. Sensitivity to lower doses of NGN3 in EP differentiation. (**A**) Generation of the *Ngn3-mCherry-flox/flox* strain by replacing the CDS of *Ngn3* with the mCherry-P2A-Loxp-*Ngn3* CDS-Loxp sequence. Excision of the *Ngn3* CDS is mediated by *Ngn3-Cre*. (**B**) *Ngn3*-mCherry expression and IF staining of NGN3 and MUC1 in E14.5 pancreata from the control (Ctrl: *Ngn3-flox/flox*) and *Ngn3-Cre;Ngn3-flox/flox* (*Ngn3*-CKO) embryos. Scale bars, 20 μm. (**C**) Statistical analysis of NGN3 IF intensity in E14.5 pancreata from the control (Ctrl) and *Ngn3-Cre;Ngn3*-CKO (CKO) embryos. *n* denotes the number of embryos analyzed. "Cell #" indicates the total number of the analyzed $NGN3^+$ cells. *P* values were calculated via the Wilcoxon rank-sum test. (**D**) Statistical analysis of the proportion of $NGN3^+$ cells relative to $MUC1^+$ trunk cells in the pancreata of the E14.5 control (Ctrl) and *Ngn3-Cre;Ngn3*-CKO (CKO) mice. The data are presented as the means ± SEMs. *n* denotes the number of embryos analyzed. P values were calculated via the Wilcoxon rank-sum test. (**E**) UMAP plots of 10x scRNA-seq data displaying the distribution of cell types (left) from the E14.5 control (Ctrl) and *Ngn3-Cre;Ngn3*-CKO (CKO) pancreata (right). The control and CKO samples each had two replicates. The cell counts are labeled in brackets. (**F**) Violin plots illustrating the decreased expression level of the *Ngn3* CDS in different cell types from the *Ngn3-Cre;Ngn3*-CKO (CKO) pancreata compared with the control (Ctrl). The white dots indicate the average expression levels. *P* values were calculated via the Wilcoxon rank-sum test.

thresholds (fig. S5, C and D). This suggests potential cooperative or sequential binding dynamics. However, conventional peak-based overlap analysis does not fully capture the dynamic nature of NGN3 and NEUROD1 binding during differentiation. To investigate their synergistic regulatory roles during endocrine differentiation, we redefined NGN3- and NEUROD1-targeted regions based on signal intensities in $Ngn3^{high}$ cells (fig. S5, D and E). This reanalysis identified 8138 overlapping sites and reclassified NGN3-specific (435) and NEUROD1-specific (4676) sites (fig. S5F and table S1). This approach more accurately reflects the dynamic interplay between NGN3 and NEUROD1 and highlights their distinct functional roles.

Overlapping regions, identified based on signal intensities, were enriched for T cell factor (TCF) (e.g., TCF3 and TCF12) and Forkhead-Box family motifs (e.g., FOXA2, FOXA3, FOXO3, and FOXP1), which are critical for cell fate specification (fig. S5G). These regions were associated with genes involved in cell development, morphogenesis, and ion transport, emphasizing their role in endocrine specification (fig. S5H and table S2). In contrast, NEUROD1-specific regions showed significant enrichment for ETS family motifs (e.g., ELK1, ELF1, ELK4, ETV1, and ELF4), which are associated with tissue morphogenesis and transcriptional activation (fig. S5G) (*42*). NEUROD1-specific regions were further enriched for genes involved in endocrine cell physiological processes, such as protein, organelle, and RNA metabolism (fig. S5H and table S2). Notably, the NEUROD1-specific loci were predominantly associated with promoter regions (Fig. 6D and fig. S5I), in contrast to NGN3's preference for enhancer regions (Figs. 1G and 6D). Together, these findings underscore NGN3's role in chromatin priming and NEUROD1's complementary function in transcriptional activation, ensuring robust endocrine lineage commitment.

To gain deeper insights into enhancer-promoter interactions, we performed Hi-C analysis (*43*) on E14.5 $Ngn3^{high}$ cells. This analysis revealed the spatial organization of regulatory elements and their corresponding target genes, enabling us to identify topological associating domains (TADs) enriched for NGN3- and NEUROD1-bound regions. Notably, 85.7% of genomic regions where both enhancer and promoter sites are occupied by NGN3 or NEUROD1 were located within the same TADs (fig. S6A). This strong correlation highlights the existence of enhancer-promoter interactions critical for regulating endocrine cell specification. Furthermore, we identified specific TADs encompassing key TFs essential for endocrine differentiation, such as *NeuroD1*, *Tead1*, *Mnx1*, *Nkx6-1*, *Glis3*, and *Myt1* (fig. S6B). These findings emphasize the spatial alignment of regulatory elements with their target genes, offering a more comprehensive understanding of how NGN3 and NEUROD1 contribute to transcriptional regulation during endocrine development.

In addition, dispersion plots revealed that ATAC-seq signals correlated more strongly with NEUROD1 binding than with NGN3 binding (Fig. 6E), suggesting that NEUROD1 binding depends on preestablished chromatin accessibility. These observations indicate that NEUROD1 functions as a conventional TF rather than a pioneer TF during pancreatic endocrine development.

Nonpioneering role of NEUROD1 in EP development

To investigate the role of NEUROD1 in transcription and chromatin remodeling during EP development, we generated a *NeuroD1-*

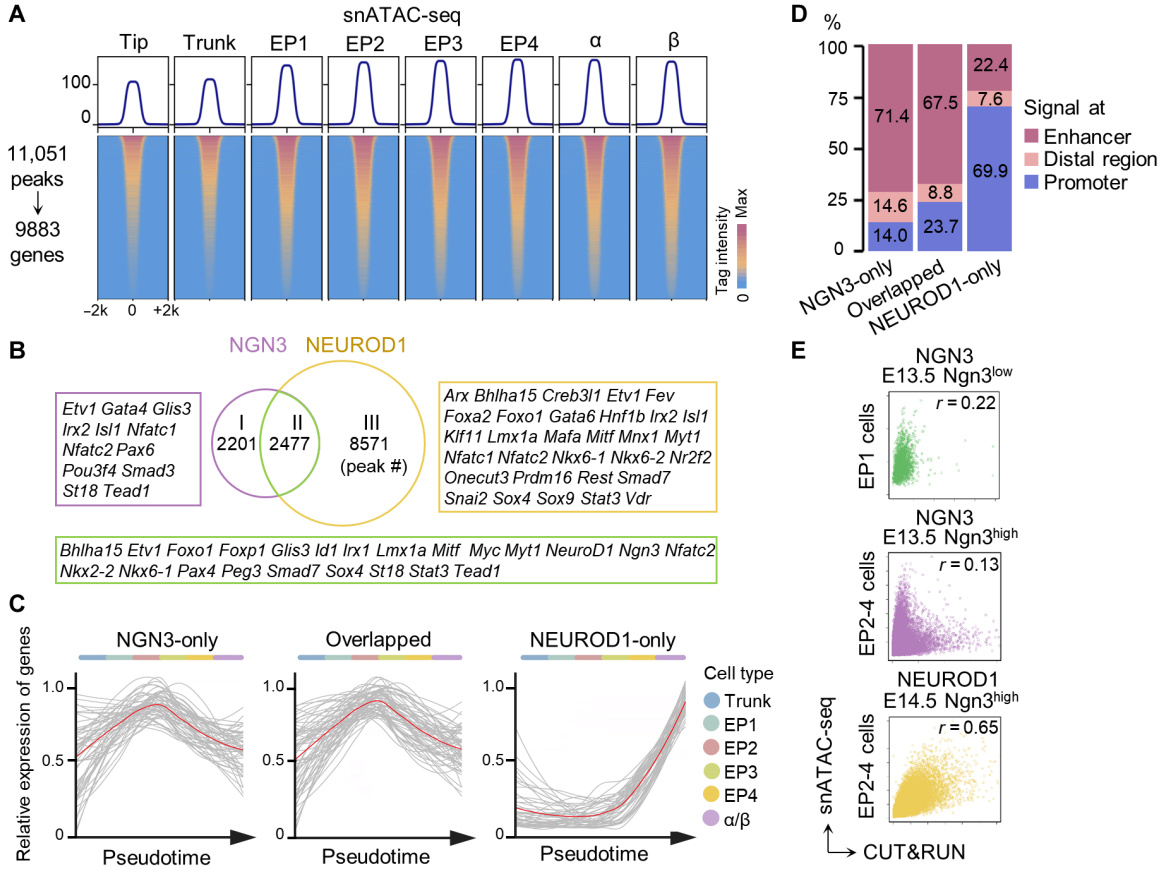

Fig. 6. Targeting patterns of NEUROD1 and NGN3 in endocrine differentiation. (**A**) Metaplots and heatmaps illustrating the dynamics of chromatin accessibility at ±2 kb from the center of NEUROD1-binding sites during development from tip to EP and endocrine (α and β) cells. Each line represents a peak, with colors ranging from blue to red indicating low to high levels of chromatin accessibility. (**B**) Overlap of NGN3- and NEUROD1-targeted regions based on peak regions. Selected TFs critical for pancreas development are listed beside the Venn diagrams. (**C**) Expression dynamics of NGN3-specific genes (left), overlapping genes (middle), and NEUROD1-specific genes (right) along the developmental trajectory from trunk to EP1 to EP4 cells from the dataset of Yu *et al.* (*14*). Gray lines represent individual genes, and the red lines indicate the average trend. (**D**) Genomic distribution of the NGN3-specific, NGN3-NEUROD1-overlapped, and NEUROD1-specific binding sites in Ngn3^{high} cells. (**E**) Dispersion plots showing the correlations between chromatin accessibility and NGN3 or NEUROD1 binding intensities in Ngn3^{low} or Ngn3^{high} cells. Pearson's correlation coefficient (*r*) is provided.

Venus-flox/flox strain in which the Venus fluorescent protein is activated by the *NeuroD1* promoter. This strain allows CKO of *NeuroD1* via *Pdx1-Cre*^{ER} with tamoxifen injection at E12.0 (*Pdx1-Cre*^{ER};*NeuroD1*-CKO) (Fig. 7A). We performed 10x Genomics scRNA-seq analysis on E15.5 EpCAM^{+} cells from control (WT, *Pdx1-Cre*^{ER}, or *NeuroD1-flox/flox*) and *Pdx1-Cre*^{ER};*NeuroD1*-CKO mice of opposite sexes pooled into the same sample (fig. S7, A to D). UMAP analysis revealed distinct developmental pathways for α and β cells starting from EP4 in the *Pdx1-Cre*^{ER};*NeuroD1*-CKO pancreata (Fig. 7B and table S3), leading to altered transcriptional characteristics of EP4, α, and β cells (fig. S7E and table S4). Compared with those from the control mice, EP4, α, and β cells from the *Pdx1-Cre*^{ER};*NeuroD1*-CKO mice presented down-regulation of the expression of endocrine genes, including the TFs *Insm1*, *Isl1*, *Fev*, *Nkx2-2*, *Etv1*, *Rfx6*, *Pax4*, *Arx*, and *Irx2*, along with up-regulation of earlier trunk progenitor genes such as *Sox4*, *Sox9*, and *Nkx6-1* (fig. S7F).

While previous studies have shown that deletion of *NeuroD1* affects α and β cell properties and production (*27*, *44*), our single-cell analyses indicate that these effects begin with an altered state of EP4. Furthermore, the proportions of α and β cells were significantly reduced in *Pdx1-Cre*^{ER};*NeuroD1*-CKO mice compared to controls (Fig. 7C). IF confirmed decreased percentages of INS^{+} and GCG^{+} cells in *Pdx1-Cre*^{ER};*NeuroD1*-CKO pancreata (Fig. 7, D and E), consistent with prior findings in *NeuroD1*-deficient models (*27*, *44*–*46*).

We further assessed changes in chromatin accessibility due to *NeuroD1* deficiency by performing bulk-cell ATAC-seq on *NeuroD1*-Venus^{+} cells sorted from the E15.5 and E17.5 *NeuroD1-flox/flox* control and *Pdx1-Cre*^{ER};*NeuroD1*-CKO mice. Linear discriminant analysis (LDA) and heatmap analyses revealed no significant differences in the chromatin accessibility of the majority of NEUROD1-bound regions between the *Pdx1-Cre*^{ER};*NeuroD1*-CKO and control mice (Fig. 7F and fig. S7G). Consistently, snATAC-seq analysis revealed consistent chromatin accessibility patterns across all cell types between EpCAM^{+} cells from the *Pdx1-Cre*^{ER};*NeuroD1*-CKO and control mice (Fig. 7G; fig. S7, H and I; and tables S3 and S5) despite a reduced proportion of α and β cells observed in *Pdx1-Cre*^{ER};*NeuroD1*-CKO pancreata (Fig. 7H). These findings suggest that while *NeuroD1* KO affects gene expression, it does not significantly alter the chromatin structure. In summary, NEUROD1 functions as a conventional TF during pancreatic endocrine differentiation, relying on chromatin

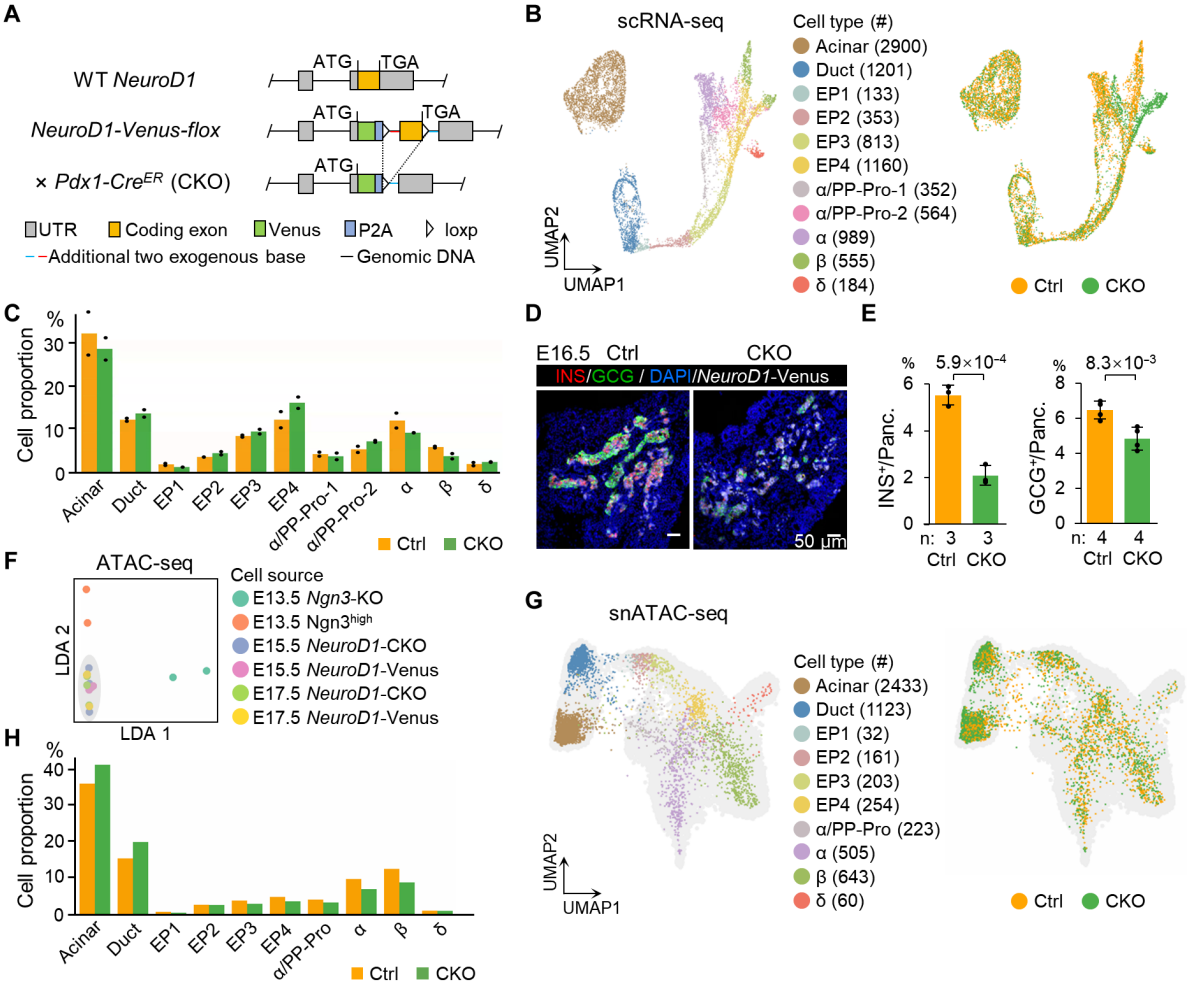

Fig. 7. Nonpioneering role of NEUROD1 in EP development. (**A**) Generation of the *NeuroD1-Venus-flox/flox* strain by replacing the CDS of *NeuroD1* with the Venus-P2A-Loxp-*NeuroD1* CDS-Loxp sequence. Excision of the *NeuroD1* CDS is mediated by *Pdx1-Cre*^{ER}. (**B**) UMAP plots of 10x scRNA-seq data displaying the distribution of cell types (left) from the E15.5 control (Ctrl: WT, *Pdx1-Cre*^{ER}, or *NeuroD1-flox/flox*) and *Pdx1-Cre*^{ER}*;NeuroD1-flox/flox* (CKO) pancreata (right). The control and CKO samples each had two replicates. The cell counts are labeled in brackets. (**C**) Bar plots quantifying the proportions of different cell types across Ctrl and CKO scRNA-seq samples, with two replicates for each condition. Black dots represent the proportion of each cell type in individual replicates. (**D**) *NeuroD1*-Venus expression and IF staining of INS and GCG in the E16.5 pancreata. Scale bars, 50 μm. (**E**) Statistical analysis of the proportions of INS$^+$ and GCG$^+$ cells in E16.5 pancreata from control (Ctrl) and *Pdx1-Cre*^{ER}*;NeuroD1*-CKO (CKO) embryos. The data are presented as the means ± SEMs. *n* denotes the number of embryos analyzed. *P* values were calculated via the Wilcoxon rank-sum test. (**F**) LDA plot showing differences in chromatin accessibility among the *Ngn3*-KO, Ngn3^{high}, *NeuroD1*-CKO, and *NeuroD1*-Venus$^+$ control cells. (**G**) UMAP plots showing snATAC-seq data of the control (Ctrl) and *Pdx1-Cre*^{ER}*;NeuroD1*-CKO (CKO) cells projected onto the UMAP plot in fig. S2A. The cell counts are labeled in brackets. (**H**) Bar plots displaying the quantified proportions of different cell types across Ctrl and CKO snATAC-seq samples.

states established by upstream factors such as NGN3 rather than initiating chromatin remodeling itself.

The pioneering role of NEUROD1 without NGN3

To explore the regulatory role of NEUROD1 in the absence of *Ngn3*, we generated *NeuroD1*-overexpressing mice (*NeuroD1-EGFP-OE*) and used *Ngn3-Cre* to induce *NeuroD1* overexpression in both the *Ngn3*-+/+ and *Ngn3-GFP/GFP* knock-in (*Ngn3*-KO) backgrounds (Fig. 8A). IF analysis revealed elevated NEUROD1 levels in the pancreata of the *Ngn3-Cre*;*NeuroD1*-COE;*Ngn3*-+/+ mice, with the highest intensity in the *Ngn3-Cre*;*NeuroD1*-COE;*Ngn3*-KO mice (fig. S8, A and B). Notably, more than 60% of cells in *Ngn3-Cre*;*NeuroD1*-COE;*Ngn3*-KO mice exhibited NEUROD1 protein levels comparable to those in control embryos (fig. S8C), suggesting that a substantial proportion of cells in our overexpression system maintained NEUROD1 levels near endogenous conditions. Flow cytometry categorized GFP$^+$ cells from the *Ngn3-Cre*;*NeuroD1*-COE;*Ngn3*-KO pancreata into GFP^{low} and GFP^{high} populations (fig. S8D). GFP^{low} cells displayed fluorescence intensity similar to that of GFP$^+$ cells from the *Ngn3*-KO pancreata, with no *NeuroD1* expression detected by single-cell reverse transcription quantitative polymerase chain reaction. However, *NeuroD1* expression was confirmed in the GFP^{high} population (fig. S8, D and E). These findings suggest that *NeuroD1* expression may compensate for the loss of NGN3 by promoting endocrine differentiation.

To assess the effect of *NeuroD1* overexpression on endocrine differentiation in the context of normal *Ngn3* expression, we isolated E15.5 EpCAM$^+$ pancreatic epithelial cells from control (WT, *Ngn3-Cre*,

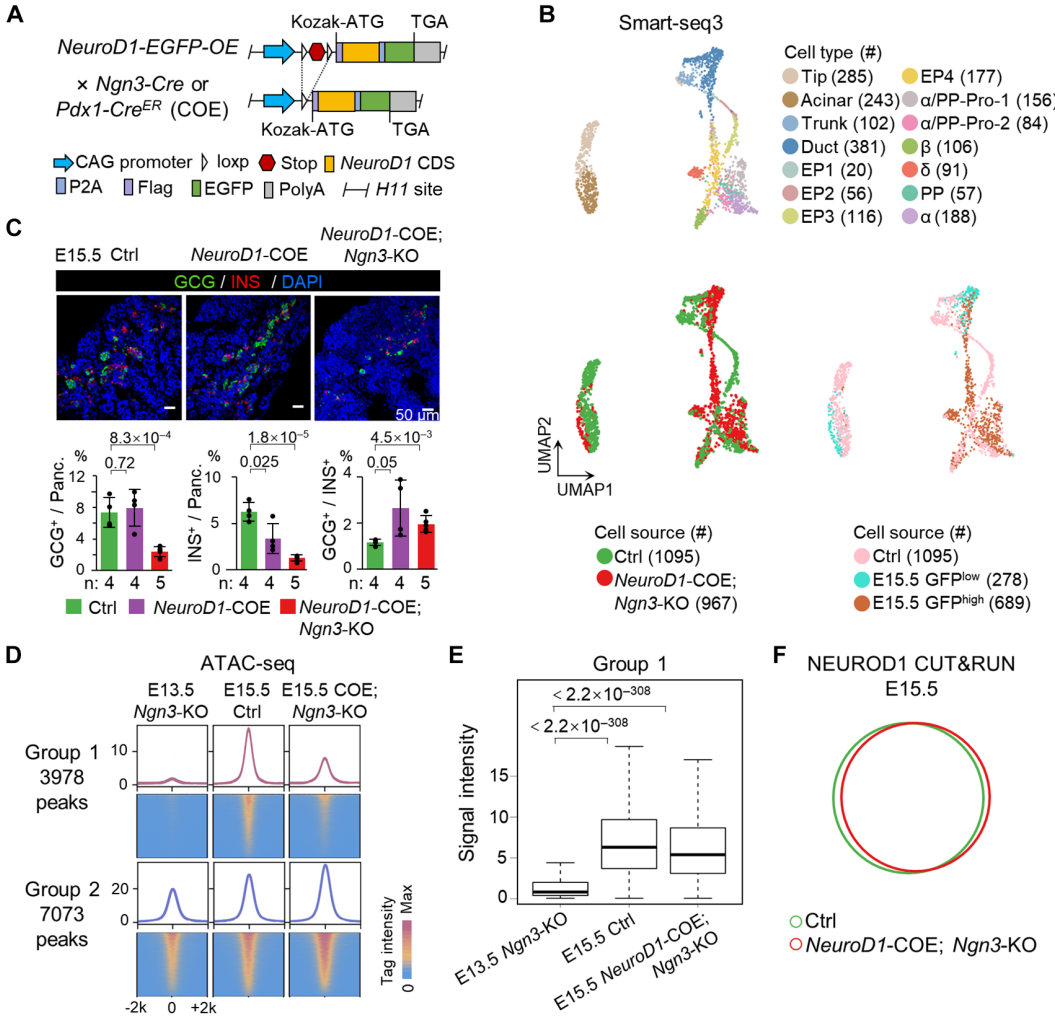

Fig. 8. The pioneering role of NEUROD1 without NGN3. (**A**) Generation of the *NeuroD1-EGFP-OE* strain via the insertion of the CAG-LSL-Flag-*NeuroD1* CDS-P2A-EGFP-PolyA sequence at the H11 site. Upon *Ngn3-Cre*- or *Pdx1-Cre*^{ER}–mediated excision of the STOP sequence, the CDS of *NeuroD1* is expressed under the CAG promoter. (**B**) UMAP plots of Smart-seq3 scRNA-seq data displaying the distribution of cell types from the control (Ctrl: WT, *Ngn3-Cre*, *Ngn3-GFP*/+, or *NeuroD1-OE*/+) and *Ngn3-Cre*;*NeuroD1*-COE;*Ngn3*-KO pancreata. The contribution of the cell source is color-coded. The cell counts are labeled in brackets. (**C**) IF staining of INS and GCG in the E15.5 pancreata (top). Scale bars: 50 μm. Statistical analysis of the proportions of GCG$^+$ and INS$^+$ cells and the GCG$^+$/INS$^+$ cell ratio in the pancreata at E15.5 (bottom). The data are presented as the means ± SEMs. *n* denotes the number of embryos analyzed. *P* values were calculated via the Wilcoxon rank-sum test. (**D**) Metaplots and heatmaps illustrating chromatin accessibility in the E13.5 *Ngn3*-KO cells, E15.5 *NeuroD1*-Venus$^+$ control (Ctrl) cells, and GFP^{high} cells from the *Ngn3-Cre*;*NeuroD1*-COE;*Ngn3*-KO pancreata. Each line represents a peak, with colors ranging from blue to red indicating low to high levels of chromatin accessibility at ±2 kb from the center. (**E**) Boxplots showing the intensity levels of the chromatin accessibility of group 1 sites in the E13.5 *Ngn3*-KO cells, E15.5 *NeuroD1*-Venus$^+$ control (Ctrl) cells, and GFP^{high} cells from the *Ngn3-Cre*;*NeuroD1*-COE;*Ngn3*-KO pancreata. *P* values were calculated via the Wilcoxon rank-sum test. (**F**) Venn diagram showing the overlap of NEUROD1-binding sites in the control (Ctrl) pancreata and NEUROD1-binding sites in the *Ngn3-Cre*;*NeuroD1*-COE;*Ngn3*-KO pancreata.

Ngn3-GFP/+, or *NeuroD1-OE*/+) mice and EGFP$^+$ cells from *Ngn3-Cre*;*NeuroD1*-COE mice for scRNA-seq (fig. S8D). We used the highly precise Smart-seq3 method (*47*), which is comparable to Smart-seq2 for effectively capturing developmental trajectories in complex cell differentiation systems. We found that EP and endocrine cells in both the control and *Ngn3-Cre*;*NeuroD1*-COE mice followed similar developmental pathways and presented comparable transcriptomic profiles (fig. S8, F to H, and tables S3 and S6).

To investigate the function of NEUROD1 without NGN3, we performed Smart-seq3 scRNA-seq on E15.5 EpCAM$^+$ cells from the control and GFP$^+$ cells from the *Ngn3-Cre*;*NeuroD1*-COE;*Ngn3*-KO pancreata. We also included E13.5 EpCAM$^+$ pancreatic epithelial cells from the WT and GFP$^+$ cells from the *Ngn3-GFP*/+ pancreata for reference developmental trajectory construction (Fig. 8B; fig. S8, D and F; and table S3). Compared with the control cells, the *NeuroD1*-overexpressing cells transitioned from EPs to endocrine cells via a divergent pathway (Fig. 8B). Differential gene expression analysis revealed that EP3 and EP4 cells in the *Ngn3-Cre*;*NeuroD1*-COE;*Ngn3*-KO embryos had increased expression of TFs related to α and δ cell specification, such as *Etv1* and *Hhex*, and decreased expression of TFs related to β cell specification, such as *Pax4* and *Lhx1*, and the endocrine TF *Pax6* (fig. S8, I and J, and table S6).

Intriguingly, the ratio of GCG$^+$/INS$^+$ cells increased with NEUROD1 overexpression, regardless of *Ngn3* expression, despite a decrease in the overall number of endocrine cells under *Ngn3*-KO conditions (Fig. 8C).

We performed bulk-cell ATAC-seq on E15.5 *NeuroD1*-Venus$^+$ cells from *NeuroD1-flox/flox* control embryos and GFP^{high} cells from *Ngn3-Cre*;*NeuroD1*-COE;*Ngn3*-KO embryos. Our analysis focused on NEUROD1-binding sites and revealed that, even without NGN3, forced NEUROD1 expression could open the group 1 chromatin regions closed in E13.5 *Ngn3*-KO trunk cells but open in E15.5 endocrine cells (Fig. 8, D and E). CUT&RUN analysis of NEUROD1 in E15.5 GFP^{high} cells from the *Ngn3-Cre*;*NeuroD1*-COE;*Ngn3*-KO embryos revealed NEUROD1 binding patterns consistent with those in the E15.5 Ngn3^{high} cells (Fig. 8F and table S1). Thus, in the absence of *Ngn3*, NEUROD1 can function as a pioneer TF, initiating chromatin accessibility and binding to regions essential for endocrine differentiation and promoting the generation of endocrine cells, particularly α cells.

NeuroD1 preceding Ngn3 drives α cell generation

Previous studies have shown that *Pdx1*-driven ectopic expression of *NeuroD1* predominantly results in α cells (*17*). *Pdx1* is expressed early during the endoderm stages and is present in both the pancreas and parts of the gastrointestinal tract, making the origin and timing of α cell differentiation unclear. To investigate the impact of sequential *Ngn3* and *NeuroD1* expression on endocrine cell generation, we induced *NeuroD1* overexpression at E12.0 prior to *Ngn3* expression via the *Pdx1-Cre^{ER}*;*NeuroD1*-COE model (Fig. 8A) in an *Ngn3-mCherry-flox/+* background (Fig. 5A). Given the known regulation of *Ngn3* expression by NEUROD1 binding (fig. S5C), the colocalization of the NGN3 and NEUROD1 proteins was observed in E12.5 *Pdx1-Cre^{ER}*;*NeuroD1*-COE pancreata, in contrast to the barely detectable NGN3 expression in WT tissues at E12.5 (fig. S9A).

Smart-seq3 scRNA-seq analysis revealed that at E12.5, EGFP$^+$ mCherry$^-$ cells were MP cells, whereas EGFP$^+$mCherry$^+$ cells were primarily α cells, as confirmed by IF (Fig. 9A; fig. S9, B to D; and table S3). Notably, EP1 to EP4 cells were absent in the *Pdx1-Cre^{ER}*;*NeuroD1*-COE pancreata, indicating that α cells bypass the typical EP generation pathway (Fig. 9A) (*14*, *15*). We also analyzed the *Pdx1-Cre^{ER}*;*NeuroD1*-COE model in an *Ngn3*-CKO background using *Ngn3-mCherry-flox/flox* (fig. S9, A to C). At E15.5, most EGFP$^+$mCherry$^+$ cells retained their α cell identity, with 128 genes down-regulated compared with those in the control mice; these genes are involved primarily in protein synthesis, translation, peptide biosynthetic processes, and ribosome assembly, suggesting reduced cell function (Fig. 9A; fig. S9, E to H; and tables S2 and S6). In contrast, the E15.5 EGFP$^+$mCherry$^-$ cells were partially retained in the acinar and duct states (Fig. 9A). These findings suggest that *NeuroD1* expression preceding *Ngn3* drives α cell generation independently of NGN3.

Bulk-cell ATAC-seq analysis of E15.5 *NeuroD1*-Venus$^+$ cells from *NeuroD1-flox/flox* control and EGFP$^+$mCherry$^+$ cells from *Pdx1-Cre^{ER}*;*NeuroD1*-COE pancreata revealed that group 1 regions, which were closed in E13.5 trunk cells, became accessible in *Pdx1-Cre^{ER}*;*NeuroD1*-COE cells even without *Ngn3* expression (Fig. 9, B and C). However, the accessibility of these group 1 regions in

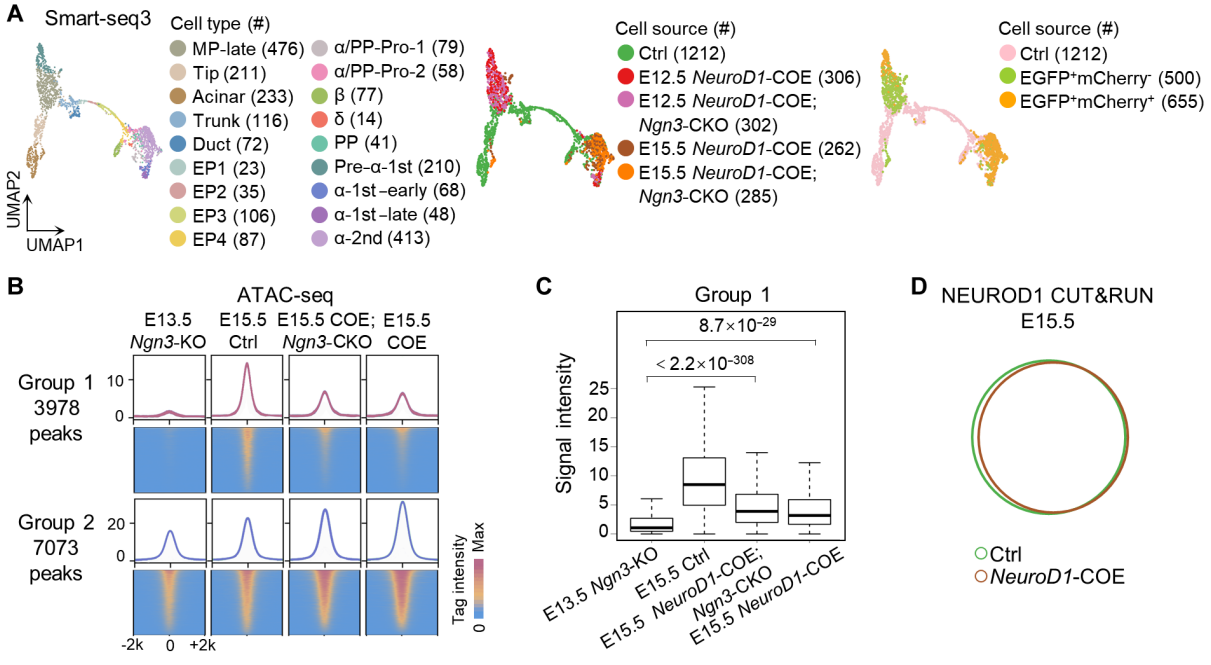

Fig. 9. *NeuroD1* preceding *Ngn3* drives α cell generation. (**A**) UMAP plots of Smart-seq3 scRNA-seq data displaying the distribution of cell types (left) from the control (Ctrl) and *Pdx1-Cre^{ER}*;*NeuroD1*-COE pancreata (middle). The contribution of the cell source is color-coded (right). The cell counts are labeled in brackets. (**B**) Metaplots and heatmaps illustrating chromatin accessibility in the E13.5 *Ngn3*-KO cells, E15.5 *NeuroD1*-Venus$^+$ control (Ctrl) cells, EGFP$^+$mCherry$^+$ cells from the *Pdx1-Cre^{ER}*;*NeuroD1*-COE;*Ngn3*-CKO pancreata, and EGFP$^+$mCherry$^+$ cells from the *Pdx1-Cre^{ER}*;*NeuroD1*-COE pancreata. Each line represents a peak, with colors ranging from blue to red indicating low to high levels of chromatin accessibility at ±2 kb from the center. (**C**) Boxplots showing the intensity levels of chromatin accessibility at group 1 sites in the E13.5 *Ngn3*-KO cells, E15.5 *NeuroD1*-Venus$^+$ control (Ctrl) cells, EGFP$^+$mCherry$^+$ cells from the *Pdx1-Cre^{ER}*;*NeuroD1*-COE;*Ngn3*-CKO pancreata, and EGFP$^+$mCherry$^+$ cells from the *Pdx1-Cre^{ER}*;*NeuroD1*-COE pancreata. *P* values were calculated via the Wilcoxon rank-sum test. (**D**) Venn diagram showing the overlap of NEUROD1-binding sites in the control (Ctrl) pancreata and NEUROD1-binding sites in the *Pdx1-Cre^{ER}*;*NeuroD1*-COE pancreata.

NeuroD1-COE cells did not reach normal levels. In addition, the NEUROD1-binding sites in the EGFP^{+}mCherry^{+} cells were similar to those in the E15.5 Ngn3^{high} cells (Fig. 9D and table S1). These findings indicate that *NeuroD1* expression preceding *Ngn3* initiates the endocrine differentiation program but preferentially drives α cell production.

In humans, *NGN3* expression in the embryonic pancreas precedes *NEUROD1* expression, as shown by reanalysis of scRNA-seq data from the human fetal pancreas (fig. S9I) (*48*). Approximately 70% of the cells in nascent islets are β cells (*49*–*51*). However, in the in vitro human embryonic stem cell–derived system, the proportion of induced β cells is significantly lower than that of α cells (*52*–*55*). An analysis of scRNA-seq data from an in vitro induction system (*52*) revealed that, contrary to the in vivo expression sequence, *NEUROD1* expression precedes *NGN3* expression along the differentiation pathway in vitro (fig. S9J). These findings suggest that the in vitro differentiation process does not accurately replicate in vivo development and that early *NeuroD1* expression increases the production of α over β cells during differentiation.

DISCUSSION

Our study elucidates the distinct yet complementary roles of NGN3 and NEUROD1 in pancreatic endocrine differentiation, highlighting their temporal dynamics, binding specificities, and functional interplay in chromatin regulation and gene expression. Through a combination of in vivo genetic models, chromatin accessibility assays, and single-cell transcriptomics, we provide evidence supporting NGN3 as a pioneer factor and NEUROD1 as a conventional TF, while also revealing their functional adaptability under altered conditions.

By utilizing a genetically engineered *Ngn3-Flag-P2A-EGFP* mouse line and CUT&RUN assays, we overcame prior technical limitations to precisely map NGN3-binding sites. These regions were inaccessible prior to NGN3 binding, as confirmed by the absence of H3K4me1 and H3K27ac modifications in pre-EP stages, consistent with the theoretical properties of pioneer factors. Conversely, group 2 loci exhibited pre-established accessibility, indicating that NGN3 also binds to already open chromatin, where it fine-tunes transcriptional regulation without remodeling chromatin structure. This dual functionality reflects NGN3's versatility in orchestrating gene activation across distinct chromatin contexts. Further analysis confirmed that NGN3's role extends beyond its pioneering ability. Positive autofeedback regulation of NGN3 was supported by in vivo evidence showing reduced *Ngn3* expression in *Ngn3-KO* embryos compared to heterozygous controls. At the posttranslational level, phosphorylation-dependent ubiquitination fine-tunes NGN3 protein stability, preventing excessive or prolonged activity.

To assess the impact of NGN3 levels on its pioneering function, we generated COE and CKO models. Remarkably, low NGN3 levels in CKO mice were sufficient to drive EP differentiation and establish chromatin accessibility, suggesting a threshold effect where NGN3 function saturates beyond a critical level. Elevated NGN3 in COE models did not further promote differentiation or alter transcriptomic trajectories, reinforcing the idea that NGN3's activity is tightly regulated and dosage tolerant.

NEUROD1, a downstream TF of NGN3, has been hypothesized to play a pioneering role in pancreatic endocrine development (*27*, *28*). However, our study revealed that NEUROD1 functions primarily as a conventional TF during endocrine differentiation and relies on pre-established chromatin accessibility. In E14.5 Ngn3^{high} cells, NEUROD1 binds predominantly to promoter regions, unlike the enhancer binding of NGN3. NEUROD1 binding depends on NGN3-established chromatin accessibility, fine-tuning gene expression rather than initiating chromatin remodeling. In the absence of *Ngn3*, NEUROD1 can assume pioneering functions. The overexpression of *NeuroD1* in *Ngn3-Cre*;*NeuroD1*-COE;*Ngn3*-KO mice demonstrated that NEUROD1 can open chromatin regions typically established by NGN3. This compensatory effect underscores NEUROD1's latent ability to perform pioneer-like functions in altered contexts, such as NGN3 deficiency.

Temporal misexpression experiments further revealed that early NEUROD1 expression can bypass the NGN3-dependent EP pathway, driving α cell generation at the expense of β cell differentiation. Our findings emphasize the necessity of sequential NGN3 and NEUROD1 expression during endocrine differentiation. In the *Pdx1-Cre^{ER}*;*NeuroD1*-COE model, early NEUROD1 expression disrupted normal progenitor differentiation, resulting in transcriptional shifts favoring α cell fates. This highlights NGN3's role in priming chromatin for β cell specification, with NEUROD1 executing downstream activation programs.

The cooperative and sequential binding dynamics of NGN3 and NEUROD1 are further reflected in their shared genomic targets. Although both TFs recognize the CANNTG motif, differences in motif flanking sequences, temporal expression, and chromatin context account for their distinct binding patterns. NGN3 primarily targets enhancer regions, consistent with its role as a pioneer factor, initiating chromatin remodeling to activate regulatory elements critical for lineage specification. In contrast, NEUROD1 predominantly binds promoter regions, fine-tuning gene activation during later stages of endocrine differentiation. In overlapped binding regions, NEUROD1 may adopt a pioneer-like function in the absence of NGN3, highlighting its functional versatility and adaptability.

Reanalysis of human fetal pancreatic scRNA-seq data corroborated the sequential *Ngn3*-to-*NeuroD1* expression observed in mice. Intriguingly, in in vitro differentiation of human pluripotent stem cells, *NEUROD1* precedes *NGN3*, leading to predominant α cell generation. This discrepancy underscores the importance of TF timing and chromatin priming in determining cell fate and highlights a critical limitation of current in vitro protocols.

Our study provides insights into the dynamic interplay between pioneer factors and conventional TFs, demonstrating how specific TFs can adapt their roles under altered conditions. While NGN3 primarily functions as a pioneer factor, NEUROD1 can acquire pioneering capabilities in NGN3-deficient contexts. This suggests that the prevalence of pioneer-like activity among TFs may be underappreciated and context dependent.

Collectively, these findings advance our understanding of how NGN3 and NEUROD1 cooperatively regulate chromatin landscapes and gene expression to orchestrate pancreatic endocrine differentiation. By distinguishing the roles of NGN3 as a chromatin-priming factor and NEUROD1 as a transcriptional activator, we demonstrate how the sequential and cooperative actions of these TFs ensure robust lineage commitment. This study provides broader insights into transcriptional regulation during development and highlights the critical importance of TF timing and functional interplay in determining cell fate. This study has crucial implications for understanding disease development, identifying therapeutic targets, and assessing the efficacy and quality of

in vitro cell differentiation processes, offering a framework for improving protocols in regenerative medicine and stem cell therapies.

MATERIALS AND METHODS

Mouse strains

All experiments were conducted on mice housed in a specific pathogen–free facility under a 12-hour light/12-hour dark cycle at 23° ± 2°C. The mice were provided with an autoclaved standard diet and water ad libitum. All procedures were approved by the Institutional Animal Care and Use Committee of the Animal Center of Peking University Health Science Center and adhered strictly to ethical standards for animal experimentation.

The mouse strains used in this study included *Pdx1-Cre* (*56*), *Pdx1-Cre^{ER}* (*12*), *Pdx1-GFP* (*35*), *Ngn3-Cre* (*13*), *Ngn3-Cre^{ER}* (*12*), *Ngn3-GFP* (*36*), and *Rosa26-RFP*, as well as *Ngn3-Flag-P2A-EGFP*, *Ngn3-mCherry-flox/flox*, *Ngn3-mCherry-OE*, *NeuroD1-Venus-flox/flox*, and *NeuroD1-EGFP-OE*, which were generated via the CRISPR-Cas9 system for targeted genomic insertion. The *Ngn3-Flag-P2A-EGFP* line was generated by inserting a Flag-P2A-EGFP cassette before the translation stop codon of the *Ngn3* gene. *Ngn3-mCherry-flox/flox* and *NeuroD1-Venus-flox/flox* lines were created by replacing the CDS with the fluorescence-P2A-Loxp-CDS-Loxp sequence. *Ngn3-mCherry-OE* and *NeuroD1-EGFP-OE* lines were produced by inserting a CAG-LSL-*Ngn3* CDS-Flag-P2A-mCherry-PolyA or CAG-LSL-Flag-*NeuroD1* CDS-P2A-EGFP-PolyA cassette at the H11 site. The sgRNA sequences used were as follows: *Ngn3-Flag-P2A-EGFP* (5′-CTCCCGGGAGCAGATAGGAT-3′), *Ngn3-mCherry-flox/flox* (5′-AGGGATGAGGCGCCATCCTG-3′, 5′-ATGAGGCGCCATCCTGCGGT-3′, 5′-CTGAGAACACCAGTGCTCCC-3′, and 5′-CTATCTGCTCCCGGGAGCAC-3′), *NeuroD1-Venus-flox/flox* (5′-CCCGCTCTCGCTGTATGATT-3′, 5′-ACCAAATCATACAGCGAGAG-3′, 5′-ATGCCATCTTTCACGATTAG-3′, and 5′-GTGCCTCTAATCGTGAAAGA-3′), and *Ngn3-mCherry-OE* and *NeuroD1-EGFP-OE* (5′-CTGAGCCAACAGTGGTAGTA-3′).

The presence of a vaginal plug was considered E0.5. For induction of *Cre* expression in *Ngn3-Cre^{ER}* embryos, pregnant mice received a single intraperitoneal injection of tamoxifen (Sigma-Aldrich, T5648) at a dose of 0.1 mg/g body weight, which was prepared from a stock solution of corn oil (20 mg/ml), 1 day before euthanizing. For the induction of *Cre* expression in *Pdx1-Cre*^{ER} embryos, tamoxifen was injected at E12.0.

Cell suspension preparation

Pancreata from E12.5 to E15.5 embryos were dissociated into single-cell suspensions using 0.25% trypsin-EDTA (Gibco, 25200072) for 5 min at 37°C. The reaction was terminated by adding 0.4 volumes of fetal bovine serum (FBS). For E16.5 to E17.5 pancreata, the tissues were first digested with collagenase P (0.5 mg/ml; Roche, 11213873001) for 2 min at 37°C, followed by treatment with 0.25% trypsin-EDTA for 3 min. Following digestion, the cell suspensions were filtered through a 35-μm cell strainer (Corning, 352235) for FACS.

Flow cytometry

The cells were sorted via a BD FACSAria III cell sorter. Sorting was based either on specific fluorescent proteins from fluorescent-labeled knock-in mice or through staining with an anti-EpCAM antibody. For EpCAM staining, the cell suspensions were stained with a CD326 (EpCAM) monoclonal antibody conjugated to APC (Invitrogen, 17-5791-82) for 10 min on ice. After staining, the cells were washed twice with FACS buffer [Hanks' balanced salt solution buffer containing 1% FBS (pH 7.4)] and then prepared for sorting.

Single-cell RNA-seq

For Smart-seq2 scRNA-seq, single cells were manually picked from FACS-purified bulk cells via a mouth pipette. cDNA synthesis and amplification followed the protocol of Picelli *et al.* (*29*). Libraries were prepared with 2 ng of cDNA via the TruePrep DNA Library Prep Kit V2 for Illumina (Vazyme, TD502) and sequenced on the Illumina HiSeq 4000 system, generating 150–base pair (bp) paired-end reads.

For Smart-seq3 scRNA-seq, single cells were directly sorted into 384-well plates containing lysis buffer. cDNA synthesis and amplification were performed according to the Smart-seq3 protocol (*47*). Libraries were constructed via the TruePrep DNA Library Prep Kit V2 for Illumina (Vazyme, TD501) and sequenced on the MGIseq-2000 platform, generating 100-bp paired-end reads.

For 10x Genomics scRNA-seq, the procedure was based on the manufacturer's instructions for the Single Cell 3′ Reagent Kit v3.1. For the *Ngn3*-CKO samples, E14.5 EpCAM$^+$ pancreatic cells were pooled per replicate: Replicate 1 consisted of two control male embryos and one CKO female embryo, and replicate 2 consisted of three control male embryos and three CKO female embryos. For the *Ngn3*-COE samples, E14.5 EpCAM$^+$ pancreatic cells were pooled from four control female embryos and four COE male embryos, and E17.5 EpCAM$^+$ pancreatic cells were pooled from two control female embryos and three COE male embryos. For *NeuroD1*-CKO samples, E15.5 EpCAM$^+$ pancreatic cells were pooled per replicate: Replicate 1 consisted of one control male embryo and one CKO female embryo, and replicate 2 consisted of two control female embryos and one CKO male embryo. The sex of the embryos was determined by genotyping for the Y chromosome–associated gene *Sry*. Libraries were sequenced as 150 bp paired-end reads on an Illumina HiSeq 4000 system.

CUT&RUN

Cleavage Under Targets and Release Using Nuclease (CUT&RUN) was performed following the published protocol (*30*) with modifications to the DNA purification and library construction steps. We used 30,000 cells per CUT&RUN assay with primary antibodies against FLAG [1:25; Cell Signaling Technology (CST), 2368S], p300 (1:100; CST, 54062S), and NEUROD1 (1:100; Millipore, ABE991). DNA was purified via a spin column (Vistech, DC2005) and processed for library construction with the KAPA Hyper Prep Kit (KAPA Biosystems, KK8502). Briefly, after end repair and adaptor ligation, the mixture was purified via VAHTS DNA Clean Beads (Vazyme, N411-03-AA) and subjected to PCR amplification with 2× KAPA HiFi HotStart ReadyMix (KAPA Biosystems, KM2602). The PCR program was as follows: 98°C for 45 s, followed by 19 cycles of 98°C for 15 s, 60°C for 10 s, and 72°C for 1 min, with a final extension at 72°C for 5 min. The final DNA was purified via 1× VAHTS DNA Clean Beads. The CUT&RUN libraries were subjected to 150-bp paired-end sequencing or 51-bp single-end sequencing on the Illumina NovaSeq 6000 or Illumina HiSeq 2500 platform.

ULI-NChIP sequencing

ULI-NChIP was performed following a published protocol (*34*). Approximately 1.5×10^4 cells were used per ChIP, with 1 ng of primary antibody specific for H3K4me1 (Millipore, 07-436, lot: 2343581).

Libraries were prepared using the Ultra DNA Library Prep Kit for Illumina [New England Biolabs (NEB), E7370] according to the manufacturer's instructions. The libraries were subjected to 51-bp single-end sequencing on the Illumina NovaSeq 6000 platform.

ATAC sequencing

Bulk-cell ATAC-seq was performed following a published protocol (*57*) with modifications to cell sorting and nuclear preparation. Specifically, 2000 cells per ATAC assay were directly sorted into 50 μl of lysis buffer in a 200-μl tube for nuclear preparation. Libraries were constructed via the TruePrep DNA Library Prep Kit V2 for Illumina. PCR amplification was carried out under the following conditions: 72°C for 3 min, 98°C for 30 s, followed by 12 cycles of 98°C for 15 s, 60°C for 30 s, and 72°C for 3 min, with a final extension at 72°C for 5 min. The final DNA was purified via 1× VAHTS DNA Clean Beads and sequenced as 150-bp paired-end reads on the Illumina NovaSeq 6000 platform.

For snATAC-seq, the procedure was performed according to the manufacturer's instructions via the Chromium Next GEM Single Cell ATAC Library & Gel Bead Kit v1.1 (10x Genomics). For the *Ngn3*-CKO samples, E14.5 EpCAM$^+$ pancreatic cells from two control male embryos and two CKO female embryos were pooled and loaded into one sample. Extra pancreatic cells from control embryos were included at a ratio of 1:4 to EpCAM$^+$ cells to ensure experimental success and an adequate number of nuclei. For the *NeuroD1*-CKO samples, E15.5 EpCAM$^+$ pancreatic cells from two control female embryos and three CKO male embryos were pooled and loaded into one sample. Libraries were sequenced as 50-bp paired-end reads on the NovaSeq 6000 system.

Hi-C

Hi-C was performed following a published protocol (*43*) with modifications to the DNA fragmentation and library construction steps. Briefly, E14.5 Ngn3^{high} cells were fixed with 1% formaldehyde at room temperature (RT) for 10 min, and the crosslinking reaction was quenched with glycine for 10 min at RT. Cells were washed with 1 × phosphate-buffered saline and lysed in 50 μl of lysis buffer [10 mM tris-HCl (pH 7.4), 10 mM NaCl, 0.1 mM EDTA, 0.5% NP-40, and protease inhibitors] on ice for 50 min. After centrifugation at 3000 rpm at 4°C for 5 min, the supernatant was carefully removed.

Chromatin was solubilized in 10 μl of 0.5% SDS and incubated at 62°C for 10 min, followed by quenching with 5 μl of 10% Triton X-100 at 37°C for 30 min. The nuclei were digested with 50 U of Mbo I (NEB, R0147L) at 37°C overnight with rotation in a final volume of 50 μl. Mbo I was inactivated at 62°C for 20 min. For biotin labeling, 1.5 μl of 1 mM 2′-deoxyadenosine 5′-triphosphate, 1.5 μl of 1 mM 2′-deoxyguanosine 5′-triphosphate, 1.5 μl of 1 mM 3′-deoxythymidine 5′-triphosphate, 3.75 μl of 0.4 mM biotin-14-dCTP (Thermo Fisher Scientific, 19518018), and 10 U of Klenow enzyme (NEB, M0210) were added, and the reaction was incubated at 37°C for 1.5 hours with rotation.

Ligation was performed by adding 60 μl of ligation mix [38.8 μl of water, 12 μl of 10 × NEB T4 DNA ligase buffer, 7 μl of 10% Triton X-100, 1.2 μl of bovine serum albumin (10 mg/ml), and 1 μl of T4 DNA ligase (400 U/μl; NEB, M0202)] and incubating at RT for 6 hours with rotation. Crosslinking was reversed, and the DNA was purified. DNA was sheared to 300 to 500 bp using a Q800R3 sonicator. Biotin-labeled DNA fragments were captured with 10 μl of Dynabeads MyOne Streptavidin C1 (Life Technologies, 65001).

Sequencing libraries were prepared using the KAPA Hyper Prep Kit (KAPA Biosystems, KK8502). The libraries were subjected to 150-bp paired-end sequencing on the Illumina NovaSeq 6000 platform.

IF and microscopy

E12.5 to E17.5 pancreata were fixed in 4% paraformaldehyde at 4°C for 8 to 12 hours and then dehydrated with 15 and 30% sucrose solutions. The tissues were embedded in optimum cutting temperature compound (Epredia, 6502) and sectioned at a thickness of 5 μm. Sections were stained with primary antibodies against FLAG (1:100; CST, 14793S), NGN3 (1:100; DSHB, F25A1B3), INS (1:500; CST, 3014S), GCG (1:200; R&D Systems, MAB1249), MUC1 (1:500; Abcam, ab15481), mCherry (1:100; SICGEN, AB8181-200), and NEUROD1 (1:100; Invitrogen, PA5-47381). For antigen retrieval, sections stained for FLAG, NGN3, MUC1, and mCherry were boiled in antigen unmasking solution (VectorLabs, H-3300) for 3 min in a microwave oven. The secondary antibodies used were donkey anti-rabbit Alexa Fluor 594 (Thermo Fisher Scientific, A21207), donkey anti-rabbit Alexa Fluor 488 (Thermo Fisher Scientific, A21206), donkey anti-goat Alexa Fluor 488 (Thermo Fisher Scientific, A11055), and donkey anti-mouse Alexa Fluor 647 (Thermo Fisher Scientific, A31571). Images were captured via a Leica TCS SP8 confocal microscope.

CUT&RUN data analysis

Overrepresented sequences were initially identified via FastQC (v0.11.3). These sequences—along with sequencing adaptors, primers, and other overrepresented sequences—were removed via Trimmomatic (v0.38). The remaining sequencing reads were then aligned to the *Mus musculus* genome (mm10) via Bowtie2 (v2.2.5) with default parameters. PCR duplicate reads were removed via SAMtools (v1.3.1), and only reads with a mapping quality score ≥ 30 were retained for downstream analysis.

Peak calling was performed on individual samples via MACS2 (v2.2.6), and peaks with P values less than 10^{-3} were retained. The reproducibility between two biological replicates was assessed via idr software (v2.0.4.2). Different irreproducible discovery rate (IDR) thresholds were applied depending on the CUT&RUN targets. For CUT&RUN, which targets NGN3 in Ngn3^{low} cells, peaks with an IDR greater than 250 were retained for downstream analysis. For CUT&RUN, which targets NGN3 in Ngn3^{high} cells, and NEUROD1, the threshold was set to 400. Genome regions overlapping with the ENCODE blacklist were filtered out from the final identified peaks.

The NGN3 or NEUROD1 peaks located within 2 kb from the TSSs were defined as promoter-associated peaks. Peaks located beyond the 2-kb regions with $\log_2$ H3K4me1 tag intensities greater than 0.5 and $\log_2$ H3K4me3 tag intensities less than 2 were defined as enhancer-associated peaks. All the remaining peaks were defined as distal peaks.

Tag intensities were calculated as tags per 10 million (TP10M) values and stored in bedGraph format using Homer (v4.7.2). Genomic track plots were generated using IGV (v2.19.1). To minimize the effects of background differences, the average background tag intensities were calculated from regions outside NGN3 and NEUROD1-binding sites, and specific background levels were subtracted for each condition: (background intensities: NGN3^{low}: 0.8; Ngn3^{high}: 3; and NEUROD1: 13).

For the analysis of overlaps between NGN3 and NEUROD1-binding sites, as well as comparisons of NEUROD1-binding sites between E15.5 Ngn3^{high} cells and *NeuroD1*-OE-EGFP$^+$ cells, we first

combined the peak regions from both counterparts. Signal intensities were then calculated separately for each sample within these combined regions. To capture the dynamic continuum of occupancy at NGN3 and NEUROD1-binding sites, we applied a lower background threshold for signal detection. Regions with NGN3 signal intensities greater than 3 were defined as NGN3-binding sites, regions with NEUROD1 signal intensities greater than 3 were defined as NEUROD1-binding sites, and regions where both NGN3 and NEUROD1 signal intensities exceeded three were classified as synergistic NGN3-NEUROD1–binding sites. For the comparison of NEUROD1-binding sites between E15.5 $Ngn3^{high}$ cells and *NeuroD1*-OE-$EGFP^{+}$ cells, regions with log 2–transformed NEUROD1 signal intensities greater than 1 were defined as NEUROD1-binding sites.

Motif enrichment analysis was conducted using the "findMotifsGenome.pl" function in Homer (version 4.7.2) with parameters "mm10 -len 8,10,12" and "-bg" for respective regions. Motifs were filtered based on log*P* values (>50 for overlapped regions and >20 for NEUROD1-only regions), a target percentage > 10%, and logTPM > 2 (average TF gene expression in EP1 to EP4 cells).

ChIP-seq data analysis

The sequencing reads were aligned to the *M. musculus* genome (mm10) using Bowtie2 (v2.2.5) with default parameters, and only reads with a mapping quality score ≥ 30 were retained for downstream analysis. PCR duplicate reads were removed using SAMtools (v1.3.1) with the commands: samtools sort in_bam | samtools rmdup -s - out_bam. Reads from biological replicates were pooled to increase coverage. TP10M values were calculated and stored in the bedGraph format using Homer with the following parameters: makeUCSCfile out_dir –o out.bdg –name sample_name -color track_color –fragLength 150 –avg -fsize 1e20. The input TP10M signal values were subtracted from the ChIP TP10M to generate normalized signal intensity tracks for downstream analyses.

Bulk-cell ATAC-seq data analysis

Overrepresented sequences were first identified via FastQC (v0.11.3) and subsequently removed along with sequencing adaptors, primers, and other overrepresented sequences via Trimmomatic (v0.38). The remaining sequencing reads were aligned to the *M. musculus* genome (mm10) via Bowtie2 (v2.2.5) with default parameters. PCR duplicate reads were removed via SAMtools (v1.3.1), and only reads with a mapping quality score ≥ 30 were retained for downstream analysis.

Before peak calling, reads generated from biological replicates were combined. ATAC-seq peaks were then identified via MACS2 (v2.2.6), and peaks with false discovery rates of less than 0.01 were retained. Genome regions overlapping with the ENCODE blacklist were excluded from further analysis. Tag intensities were calculated as TP10M values and stored in bedGraph files via Homer (v4.7.2).

10x Genomics snATAC-seq data analysis

Preprocessing was conducted via cellranger-atac (v1.2.0) with default parameters against the refdata-cellranger-atac-mm10-1.2.0 genome. Cells with detected unique molecular identifier (UMI) counts greater than 1000 and reads in promoter regions between 15 and 45% were retained for downstream analysis. The mm10 genome was segmented into discrete 5-kb intervals, which served as bins for each cell. The tag intensities for each bin were calculated on the basis of the frequency of sequence tags and imported into the R package Signac (v1.3.0) for downstream analysis. The cell and tag intensity matrix were normalized via the "RunTFIDF" function, and highly variable features (HVFs) were identified via the "FindTopFeatures" function.

For the reference datasets, where Pdx1-Cre^{ER};Rosa26-RFP^{+} and Ngn3-Cre^{ER};Rosa26-RFP^{+} cells were sorted, dimensional reduction was performed via singular value decomposition (SVD) on the basis of HVFs. Batch effects were corrected via the "reducedMNN" function in the R package batchelor (v1.8.1) applied to the SVD coordinates. Subsequently, UMAP and SNN clustering algorithms were applied to the batch-corrected SVD coordinates. Cell cluster identities were defined by chromatin accessibility patterns at loci known to represent specific lineages, as shown in fig. S2B. For definition of chromatin accessible regions for each cell population, single-cell labeled tip, trunk, EP1, EP2, EP3, EP4, α, and β cells were combined into pseudobulk datasets. The peaks of these pseudobulk datasets were identified via MACS2 (v2.2.6). Peaks with "neg_log10qvalue_summit" values greater than 40 were regarded as chromatin accessible regions for these cell populations. Metaplots were generated from the pseudobulk datasets for each cell population using the "computeMatrix reference-point" and "plotHeatmap" functions of deeptools (v3.2.1).

For the control, *Ngn3-Cre;Ngn3*-CKO, and *Pdx1-*Cre^{ER}*;NeuroD1*-CKO datasets, control and CKO cells in the same sample were initially distinguished on the basis of tags per 0.1 million (TP0.1M) values of all reads located on the Y chromosome. These values were $\log_{10}$-transformed, and cells with values greater than 2.5 were classified as male, while the others were classified as female. Control and CKO cells were then projected onto the reference datasets via the "FindTransferAnchors" function in Seurat. The UMAP coordinates and cell cluster identities were predicted via the "TransferData" function.

To identify differentially accessible regions (DARs) between control and CKO cells, we employed the following approaches: (i) Chromatin-accessible regions were independently identified for control and CKO cells via MACS2 on their respective pseudobulk datasets, retaining regions with "neg_log10qvalue_summit" values greater than 10 and merging them to form candidate regions. (ii) A region was considered "open" in a cell if its tag intensity exceeded zero, and the proportion of cells with an open status for each peak was calculated for both control and CKO cells, resulting in open rate vectors. For correction for potential biases between groups, these vectors were normalized via linear regression, standardizing the skewness to 1 and setting the intercept to zero. (iii) A nonparametric *Z* test was applied to the normalized open rate vectors to identify significant differences in chromatin accessibility, with significance thresholds set at a Bonferroni-adjusted *P* value of 0.05.

Classification of NGN3-binding sites based on chromatin accessibility

Chromatin accessibility at NGN3-binding sites was assessed across tip, trunk, EP1 to EP4, α, and β cells using reference pseudobulk datasets generated from snATAC-seq. Peaks were grouped using the SNN clustering algorithm implemented in the R package Seurat (v4.1.1) with the parameters "k.param = 20" and "resolution = 1.2". Fold changes of aggregated peak signals were calculated between the EP and trunk stages as well as between the EP stage and the *Ngn3*-KO group. Clusters with minimal fold changes (below 10% of the maximum observed) were classified as "group 2," representing consistently open regions, while the remaining clusters were categorized

as "group 1," representing newly opened regions. This approach provides a systematic framework for classifying NGN3-binding sites based on their chromatin accessibility profiles across different cell types, offering insights into their functional roles during pancreatic endocrine differentiation.

Smart-seq2 scRNA-seq analysis

The sequencing reads were aligned to the mm10 mouse genome assembly via TopHat (v2.1.0). Gene-level quantification of mapped reads was performed with HTSeq (v0.6.6) using the parameters "-s no -a 30." Cells with fewer than 0.2 million mapped reads or fewer than 4000 detected genes were excluded from further analysis. Gene expression levels were quantified as transcripts per million (TPM) values.

To analyze the Smart-seq2 scRNA-seq datasets, we used our previously published Smart-seq2 dataset of endocrine development trajectories as a reference (*14*). HVGs were identified from the reference dataset via coexpression analysis. Initially, HVG candidates were identified via the "FindVariableGenes" function in the Seurat package. These candidates were then used to construct a Pearson correlation matrix for the entire dataset. The final HVGs were filtered by applying the following criteria: (i) Each gene must exhibit significant coexpression, defined by a Pearson correlation coefficient greater than 0.3, with at least five other HVGs, and (ii) each gene must not be detectable in more than 80% of the cells. On the basis of the identified HVGs, the Smart-seq2 scRNA-seq datasets were subsequently imported into the Seurat package (version 4.1.1). Principal components analysis and UMAP were conducted for dimensional reduction, followed by SNN clustering.

Smart-seq3 scRNA-seq analysis

Trimmomatic (v0.39) was used to remove primer and adaptor sequences from the reads. The trimmed forward reads (read 1) were then aligned to the mm10 mouse genome via HISAT2 (v2.1.0). The aligned reads were annotated to gene features via featureCounts (v2.0.0), considering both the 5′ and internal fragments. Cells with fewer than 4000 detected genes were filtered out. Gene expression levels were quantified via TPM.

For identification of cell types in the Smart-seq3 datasets, the datasets were projected onto the Smart-seq2 reference trajectories via the previously identified HVGs. This projection was performed via the "FindIntegrationAnchors" and "IntegrateData" functions in the Seurat package. The cell cluster identities were then predicted via the *k*-nearest neighbors algorithm with the parameter "$k = 10$". Owing to the relatively low numbers of ductal cells, 2nd-α cells, and MP-late cells in the Smart-seq2 reference dataset, the predicted cell counts of these cell types in the Smart-seq3 dataset might be underestimated. Therefore, we refined the cell boundaries of these populations in the Smart-seq3 datasets via SNN clustering. The adjustments involved the following cell populations: (i) trunk and ductal cells; (ii) E12.5-predicted EP cells derived from *Pdx1-Cre^{ER};NeuroD1*-COE cells were renamed pre-α-1st cells; (iii) E12.5-predicted tip and trunk cells derived from *Pdx1-Cre^{ER};NeuroD1*-COE cells were renamed MP-late cells; (iv) E15.5-predicted α-1st cells derived from *Pdx1-Cre^{ER};NeuroD1*-COE cells were renamed α-2nd cells.

10x Genomics scRNA-seq analysis

Preprocessing was conducted via Cell Ranger (v3.0.1) with default parameters aligned to the refdata-cellranger-mm10-1.2.0 reference genome. The "filtered_feature_bc_matrix" output from Cell Ranger was imported into the Seurat package (version 4.1.1) for downstream analysis in R. Within each sample, control and CKO or COE cells were initially distinguished on the basis of the transcripts per 10,000 (TP10K) values of sex-specific genes: *Uty*, *Eif2s3y*, and *Ddx3y* for male cells and *Xist* for female cells. These values were log-transformed, with cells exhibiting male-related values greater than 0.2 and female-related values equal to 0 classified as male, whereas cells with female-related values greater than 1 and male-related values equal to 0 were classified as female.

The downstream analysis—including dimensionality reduction, cell clustering, and differential expression analysis—was conducted via the standard Seurat pipeline. To mitigate batch effects, HVGs were identified separately within each biological replicate when a condition had two replicates and then combined for subsequent analysis.

Hi-C data analysis

Preprocessing of Hi-C data was performed using juicer (v1.6) with default parameters aligned to the mm10 reference genome. TADs were identified using the "hicFindTADs" function in HiCExplorer (v2.1.1). Genomic regions where both promoter and enhancer elements of a gene were located within a single TAD were identified as potential promoter-enhancer loops. Hi-C contact maps were normalized using the "coverage" method in juicer and visualized to assess chromatin interactions.

Dynamics of the expression levels of *Ngn3* and *NeuroD1*

The relevant cells from the scRNA-seq datasets (*14*, *48*, *52*) were extracted and ordered via UMAP. The pseudotimes were calculated from principal curves. For the in vivo human fetal pancreas, the trunk cells were posited to be located 0.25 units ahead of the EP1 cells. The dynamic curves were fitted via loess regression.

GO analysis

GO enrichment analysis was conducted via the R package GOstats (v2.58.0) against the org.Mm.eg.db (v3.13.0) database.

Statistical analysis

For the IF staining of independent biological replicates, the SEM and *P* value from an unpaired two-tailed Wilcoxon rank-sum tests were calculated. Chi-square tests were used to assess the significance of NGN3 peak enrichment in promoter and enhancer regions. For differential gene expression analysis, the SEM and *P* value from unpaired two-tailed Wilcoxon rank-sum tests were calculated. A *P* value less than 0.05 was considered statistically significant.

Supplementary Materials

The PDF file includes:

Figs. S1 to S9

Legends for tables S1 to S6

Other Supplementary Material for this manuscript includes the following:

Tables S1 to S6

REFERENCES AND NOTES

1. L. A. Cirillo, F. R. Lin, I. Cuesta, D. Friedman, M. Jarnik, K. S. Zaret, Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. *Mol. Cell* **9**, 279–289 (2002).
2. S. Sekiya, A. Suzuki, Direct conversion of mouse fibroblasts to hepatocyte-like cells by defined factors. *Nature* **475**, 390–393 (2011).

3. O. L. Wapinski, T. Vierbuchen, K. Qu, Q. Y. Lee, S. Chanda, D. R. Fuentes, P. G. Giresi, Y. H. Ng, S. Marro, N. F. Neff, D. Drechsel, B. Martynoga, D. S. Castro, A. E. Webb, T. C. Südhof, A. Brunet, F. Guillemot, H. Y. Chang, M. Wernig, Hierarchical mechanisms for direct reprogramming of fibroblasts to neurons. *Cell* **155**, 621–635 (2013).
4. A. Mayran, K. Sochodolsky, K. Khetchoumian, J. Harris, Y. Gauthier, A. Bemmo, A. Balsalobre, J. Drouin, Pioneer and nonpioneer factor cooperation drives lineage specific chromatin opening. *Nat. Commun.* **10**, 3807 (2019).
5. R. Gao, X. Liang, S. Cheedipudi, J. Cordero, X. Jiang, Q. Zhang, L. Caputo, S. Günther, C. Kuenne, Y. Ren, S. Bhattacharya, X. Yuan, G. Barreto, Y. Chen, T. Braun, S. M. Evans, Y. Sun, G. Dobreva, Pioneering function of Isl1 in the epigenetic control of cardiomyocyte cell fate. *Cell Res.* **29**, 486–501 (2019).
6. C. van Oevelen, S. Collombet, G. Vicent, M. Hoogenkamp, C. Lepoivre, A. Badeaux, L. Bussmann, J. L. Sardina, D. Thieffry, M. Beato, Y. Shi, C. Bonifer, T. Graf, C/EBPα activates pre-existing and de novo macrophage enhancers during induced pre-B cell transdifferentiation and myelopoiesis. *Stem Cell Rep.* **5**, 232–247 (2015).
7. H. P. Shih, A. Wang, M. Sander, Pancreas organogenesis: From lineage determination to morphogenesis. *Annu. Rev. Cell Dev. Biol.* **29**, 81–105 (2013).
8. X.-X. Yu, C.-R. Xu, Understanding generation and regeneration of pancreatic β cells from a single-cell perspective. *Development* **147**, dev179051 (2020).
9. H. L. Larsen, A. Grapin-Botton, The molecular and morphogenetic basis of pancreas organogenesis. *Semin. Cell Dev. Biol.* **66**, 51–68 (2017).
10. A. Bastidas-Ponce, K. Scheibner, H. Lickert, M. Bakhti, Cellular and molecular mechanisms coordinating pancreas development. *Development* **144**, 2873–2888 (2017).
11. F. C. Pan, C. Wright, Pancreas organogenesis: From bud to plexus to gland. *Dev. Dyn.* **240**, 530–565 (2011).
12. G. Gu, J. Dubauskaite, D. A. Melton, Direct evidence for the pancreatic lineage: NGN3+ cells are islet progenitors and are distinct from duct progenitors. *Development* **129**, 2447–2457 (2002).
13. S. E. Schonhoff, M. Giel-Moloney, A. B. Leiter, Neurogenin 3-expressing progenitor cells in the gastrointestinal tract differentiate into both endocrine and non-endocrine cell types. *Dev. Biol.* **270**, 443–454 (2004).
14. X.-X. Yu, W.-L. Qiu, L. Yang, Y. Zhang, M.-Y. He, L.-C. Li, C.-R. Xu, Defining multistep cell fate decision pathways during pancreatic development at single-cell resolution. *EMBO J.* **38**, e100164 (2019).
15. X.-X. Yu, W.-L. Qiu, L. Yang, Y.-C. Wang, M.-Y. He, D. Wang, Y. Zhang, L.-C. Li, J. Zhang, Y. Wang, C.-R. Xu, Sequential progenitor states mark the generation of pancreatic endocrine lineages in mice and humans. *Cell Res.* **31**, 886–903 (2021).
16. G. Gradwohl, A. Dierich, M. LeMeur, F. Guillemot, Neurogenin3 is required for the development of the four endocrine cell lineages of the pancreas. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 1607–1611 (2000).
17. V. M. Schwitzgebel, D. W. Scheel, J. R. Conners, J. Kalamaras, J. E. Lee, D. J. Anderson, L. Sussel, J. D. Johnson, M. S. German, Expression of neurogenin3 reveals an islet cell precursor population in the pancreas. *Development* **127**, 3533–3542 (2000).
18. M. E. Bechard, E. D. Bankaitis, S. B. Hipkens, A. Ustione, D. W. Piston, Y. P. Yang, M. A. Magnuson, C. V. E. Wright, Precommitment low-level *Neurog3* expression defines a long-lived mitotic endocrine-biased progenitor pool that drives production of endocrine-committed cells. *Genes Dev.* **30**, 1852–1865 (2016).
19. E. D. Bankaitis, M. E. Bechard, C. V. E. Wright, Feedback control of growth, differentiation, and morphogenesis of pancreatic endocrine progenitors in an epithelial plexus niche. *Genes Dev.* **29**, 2203–2216 (2015).
20. L. E. Byrnes, D. M. Wong, M. Subramaniam, N. P. Meyer, C. L. Gilchrist, S. M. Knox, A. D. Tward, C. J. Ye, J. B. Sneddon, Lineage dynamics of murine pancreatic development at single-cell resolution. *Nat. Commun.* **9**, 3922 (2018).
21. M. A. Scavuzzo, M. C. Hill, J. Chmielowiec, D. Yang, J. Teaw, K. Sheng, Y. Kong, M. Bettini, C. Zong, J. F. Martin, M. Borowiak, Endocrine lineage biases arise in temporally distinct endocrine progenitors during pancreatic morphogenesis. *Nat. Commun.* **9**, 3356 (2018).
22. A. Bastidas-Ponce, S. Tritschler, L. Dony, K. Scheibner, M. Tarquis-Medina, C. Salinno, S. Schirge, I. Burtscher, A. Bottcher, F. J. Theis, H. Lickert, M. Bakhti, Comprehensive single cell mRNA profiling reveals a detailed roadmap for pancreatic endocrinogenesis. *Development* **146**, dev173849 (2019).
23. L. van Gurp, M. J. Muraro, T. Dielen, L. Seneby, G. Dharmadhikari, G. Gradwohl, A. van Oudenaarden, E. J. P. de Koning, A transcriptomic roadmap to α- and β-cell differentiation in the embryonic pancreas. *Development* **146**, dev173716 (2019).
24. E. Duvall, C. M. Benitez, K. Tellez, M. Enge, P. T. Pauerstein, L. Li, S. Baek, S. R. Quake, J. P. Smith, N. C. Sheffield, S. K. Kim, H. E. Arda, Single-cell transcriptome and accessible chromatin dynamics during endocrine pancreas development. *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2201267119 (2022).
25. H. P. Huang, M. Liu, H. M. El-Hodiri, K. Chu, M. Jamrich, M. J. Tsai, Regulation of the pancreatic islet-specific gene BETA2 (neuroD) by neurogenin 3. *Mol. Cell. Biol.* **20**, 3292–3307 (2000).
26. R. Gasa, C. Mrejen, F. C. Lynn, P. Skewes-Cox, L. Sanchez, K. Y. Yang, C. H. Lin, R. Gomis, M. S. German, Induction of pancreatic islet cell differentiation by the neurogenin-neuroD cascade. *Differentiation* **76**, 381–391 (2008).
27. R. Bohuslavova, V. Fabriciova, O. Smolik, L. Lebrón-Mora, P. Abaffy, S. Benesova, D. Zucha, L. Valihrach, Z. Berkova, F. Saudek, G. Pavlinkova, NEUROD1 reinforces endocrine cell fate acquisition in pancreatic development. *Nat. Commun.* **14**, 5554 (2023).
28. X.-X. Yu, W.-L. Qiu, L. Yang, L.-C. Li, Y.-W. Zhang, C.-R. Xu, Dynamics of chromatin marks and the role of JMJD3 during pancreatic endocrine cell fate commitment. *Development* **145**, dev163162 (2018).
29. S. Picelli, O. R. Faridani, A. K. Bjorklund, G. Winberg, S. Sagasser, R. Sandberg, Full-length RNA-seq from single cells using Smart-seq2. *Nat. Protoc.* **9**, 171–181 (2014).
30. P. J. Skene, J. G. Henikoff, S. Henikoff, Targeted in situ genome-wide profiling with high efficiency for low cell numbers. *Nat. Protoc.* **13**, 1006–1019 (2018).
31. V. Schreiber, R. Mercier, S. Jimenez, T. Ye, E. Garcia-Sanchez, A. Klein, A. Meunier, S. Ghimire, C. Birck, B. Jost, K. H. de Lichtenberg, C. Honore, P. Serup, G. Gradwohl, Extensive NEUROG3 occupancy in the human pancreatic endocrine gene regulatory network. *Mol. Metab.* **53**, 101313 (2021).
32. R. Pantier, K. Chhatbar, G. Alston, H. Y. Lee, A. Bird, High-throughput sequencing SELEX for the determination of DNA-binding protein specificities in vitro. *STAR Protoc.* **3**, 101490 (2022).
33. H. P. Shih, J. L. Kopp, M. Sandhu, C. L. Dubois, P. A. Seymour, A. Grapin-Botton, M. Sander, A Notch-dependent molecular circuitry initiates pancreatic endocrine and ductal cell differentiation. *Development* **139**, 2488–2499 (2012).
34. J. Brind'Amour, S. Liu, M. Hudson, C. Chen, M. M. Karimi, M. C. Lorincz, An ultra-low-input native ChIP-seq protocol for genome-wide profiling of rare cell populations. *Nat. Commun.* **6**, (2015).
35. G. Gu, J. M. Wells, D. Dombkowski, F. Preffer, B. Aronow, D. A. Melton, Global expression analysis of gene regulatory pathways during endocrine pancreatic development. *Development* **131**, 165–179 (2004).
36. C. S. Lee, N. Perreault, J. E. Brestelli, K. H. Kaestner, Neurogenin 3 is essential for the proper specification of gastric enteroendocrine cells and the maintenance of gastric epithelial cell identity. *Genes Dev.* **16**, 1488–1497 (2002).
37. M. L. Bulyk, J. Drouin, M. M. Harrison, J. Taipale, K. S. Zaret, Pioneer factors — Key regulators of chromatin and gene expression. *Nat. Rev. Genet.* **24**, 809–815 (2023).
38. R. Dassaye, S. Naidoo, M. E. Cerf, Transcription factor regulation of pancreatic organogenesis, differentiation and maturation. *Islets* **8**, 13–34 (2015).
39. C. Henry, A.-F. Close, J. Buteau, A critical role for the neural zinc factor ST18 in pancreatic β-cell apoptosis. *J. Biol. Chem.* **289**, 8413–8419 (2014).
40. Y. Ohta, Y. Kosaka, N. Kishimoto, J. Wang, S. B. Smith, G. Honig, H. Kim, R. M. Gasa, N. Neubauer, A. Liou, L. H. Tecott, E. S. Deneris, M. S. German, Convergence of the insulin and serotonin programs in the pancreatic β-cell. *Diabetes* **60**, 3208–3216 (2011).
41. R. Vangoitsenhoven, H. Wolden-Kirk, K. Lemaire, A. Verstuyf, L. Verlinden, Y. Yamamoto, S. Kato, L. Van Lommel, F. Schuit, B. Van der Schueren, C. Mathieu, L. Overbergh, Effect of a transcriptional inactive or absent vitamin D receptor on beta-cell function and glucose homeostasis in mice. *J. Steroid Biochem. Mol. Biol.* **164**, 309–317 (2016).
42. B. Kandemir, I. A. Kurnaz, The role of Pea3 transcription factor subfamily in the nervous system. *Mol. Neurobiol.* **62**, 3293–3304 (2025).
43. Z. Du, H. Zheng, B. Huang, R. Ma, J. Wu, X. Zhang, J. He, Y. Xiang, Q. Wang, Y. Li, J. Ma, X. Zhang, K. Zhang, Y. Wang, M. Q. Zhang, J. Gao, J. R. Dixon, X. Wang, J. Zeng, W. Xie, Allelic reprogramming of 3D chromatin architecture during early mammalian development. *Nature* **547**, 232–235 (2017).
44. R. Bohuslavova, O. Smolik, J. Malfatti, Z. Berkova, Z. Novakova, F. Saudek, G. Pavlinkova, NEUROD1 is required for the early α and β endocrine differentiation in the pancreas. *Int. J. Mol. Sci.* **22**, 6713 (2021).
45. F. J. Naya, H.-P. Huang, Y. Qiu, H. Mutoh, F. J. DeMayo, A. B. Leiter, M.-J. Tsai, Diabetes, defective pancreatic morphogenesis, and abnormal enteroendocrine differentiation in BETA2/neuroD-deficient mice. *Genes Dev.* **11**, 2323–2334 (1997).
46. A. I. Romer, R. A. Singer, L. Sui, D. Egli, L. Sussel, Murine perinatal β-cell proliferation and the differentiation of human stem cell-derived insulin-expressing cells require NEUROD1. *Diabetes* **68**, 2259–2271 (2019).
47. M. Hagemann-Jensen, C. Ziegenhain, R. Sandberg, Scalable single-cell RNA sequencing from full transcripts with Smart-seq3xpress. *Nat. Biotechnol.* **40**, 1452–1457 (2022).
48. Z. Ma, X. Zhang, W. Zhong, H. Yi, X. Chen, Y. Zhao, Y. Ma, E. Song, T. Xu, Deciphering early human pancreas development at the single-cell level. *Nat. Commun.* **14**, 5354 (2023).
49. C. Ionescu-Tirgoviste, P. A. Gagniuc, E. Gubceac, L. Mardare, I. Popescu, S. Dima, M. Militaru, A 3D map of the islet routes throughout the healthy human pancreas. *Sci. Rep.* **5**, 14634 (2015).
50. D. Bosco, M. Armanet, P. Morel, N. Niclauss, A. Sgroi, Y. D. Muller, L. Giovannoni, G. Parnaud, T. Berney, Unique arrangement of α- and β-cells in human islets of langerhans. *Diabetes* **59**, 1202–1210 (2010).

51. A. Pisania, G. C. Weir, J. J. O'Neil, A. Omer, V. Tchipashvili, J. Lei, C. K. Colton, S. Bonner-Weir, Quantitative analysis of cell composition and purity of human pancreatic islet preparations. *Lab. Invest.* **90**, 1661–1675 (2010).
52. H. Zhu, G. Wang, K.-V. Nguyen-Ngoc, D. Kim, M. Miller, G. Goss, J. Kovsky, A. R. Harrington, D. C. Saunders, A. L. Hopkirk, R. Melton, A. C. Powers, S. Preissl, F. M. Spagnoli, K. J. Gaulton, M. Sander, Understanding cell fate acquisition in stem-cell-derived pancreatic islets using single-cell multiome-inferred regulomes. *Dev. Cell* **58**, 727–743.e11 (2023).
53. P. Augsornworawat, K. G. Maxwell, L. Velazco-Cruz, J. R. Millman, Single-cell transcriptome profiling reveals β cell maturation in stem cell-derived islets after transplantation. *Cell Rep.* **32**, 108067 (2020).
54. A. Veres, A. L. Faust, H. L. Bushnell, E. N. Engquist, J. H. Kenty, G. Harb, Y. C. Poh, E. Sintov, M. Gurtler, F. W. Pagliuca, Q. P. Peterson, D. A. Melton, Charting cellular identity during human in vitro beta-cell differentiation. *Nature* **569**, 368–373 (2019).
55. C. Weng, J. Xi, H. Li, J. Cui, A. Gu, S. Lai, K. Leskov, L. Ke, F. Jin, Y. Li, Single-cell lineage analysis reveals extensive multimodal transcriptional control during directed beta-cell differentiation. *Nat. Metab.* **2**, 1443–1458 (2020).
56. S. R. Hingorani, E. F. Petricoin, A. Maitra, V. Rajapakse, C. King, M. A. Jacobetz, S. Ross, T. P. Conrads, T. D. Veenstra, B. A. Hitt, Y. Kawaguchi, D. Johann, L. A. Liotta, H. C. Crawford, M. E. Putt, T. Jacks, C. V. E. Wright, R. H. Hruban, A. M. Lowy, D. A. Tuveson, Preinvasive and invasive ductal pancreatic cancer and its early detection in the mouse. *Cancer Cell* **4**, 437–450 (2003).
57. J. Wu, J. Xu, B. Liu, G. Yao, P. Wang, Z. Lin, B. Huang, X. Wang, T. Li, S. Shi, N. Zhang, F. Duan, J. Ming, X. Zhang, W. Niu, W. Song, H. Jin, Y. Guo, S. Dai, L. Hu, L. Fang, Q. Wang, Y. Li, W. Li, J. Na, W. Xie, Y. Sun, Chromatin analysis in human early development reveals epigenetic transition during ZGA. *Nature* **557**, 256–260 (2018).

Acknowledgments: We thank the members of the Xu laboratory for their advice and comments. We also thank the Peking-Tsinghua Center for Life Science High Performance Computing Platform. **Funding:** This work was supported by the National Key Research and Development Program of China (2024YFA1802900 and 2023YFA1800600 to C.-R.X. and 2024YFA1802900 to L.Y.), the National Natural Science Foundation of China (32125014 and 32030034 to C.-R.X.), and the Peking-Tsinghua Center for Life Sciences to C.-R.X. **Author contributions:** Conceptualization: C.-R.X. Methodology: L.Y., X.-X.Y., and C.-T.J. Software: X.W. and C.-T.J. Validation: L.Y. and X.-X.Y. Formal analysis: C.-R.X., L.Y., X.-X.Y., X.W., and C.-T.J. Investigation: L.Y. and X.-X.Y. Resources: C.-R.X., L.Y., X.-X.Y., and C.-T.J. Data curation: L.Y., X.-X.Y., X.W., and C.-T.J. Writing—original draft: C.-R.X., L.Y., X.-X.Y., and X.W. Writing—review and editing: C.-R.X., L.Y., X.-X.Y., and C.-T.J. Visualization: C.-R.X., L.Y., X.-X.Y., and C.-T.J. Supervision: C.-R.X., L.Y., and X.-X.Y. Project administration: C.-R.X., L.Y., and X.-X.Y. Funding acquisition: C.-R.X. and L.Y. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** The data associated with this article have been deposited in the Gene Expression Omnibus (GEO) under the identifiers GSE271546 (www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE271546) and GSE270915 (www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE270915). All other data required to evaluate the conclusions in this article are provided within the paper or the Supplementary Materials.

Submitted 27 September 2024
Accepted 20 February 2025
Published 26 March 2025
10.1126/sciadv.adt4770