

RESEARCH ARTICLE

Comprehensive single cell mRNA profiling reveals a detailed roadmap for pancreatic endocrinogenesis

Aimée Bastidas-Ponce^{1,2,3,4,*}, Sophie Tritschler^{1,5,6,*}, Leander Dony^{5,6,7}, Katharina Scheibner^{1,2,3,4}, Marta Tarquis-Medina^{1,2,3,4}, Ciro Salinno^{1,2,3,4}, Silvia Schirge^{1,2,3}, Ingo Burtscher^{1,2,3}, Anika Böttcher^{1,2,3}, Fabian J. Theis^{5,8,‡}, Heiko Lickert^{1,2,3,4,‡} and Mostafa Bakhti^{1,2,3,‡}

ABSTRACT

Deciphering mechanisms of endocrine cell induction, specification and lineage allocation *in vivo* will provide valuable insights into how the islets of Langerhans are generated. Currently, it is ill defined how endocrine progenitors segregate into different endocrine subtypes during development. Here, we generated a novel neurogenin 3 (Ngn3)-Venus fusion (NVF) reporter mouse line, that closely mirrors the transient endogenous Ngn3 protein expression. To define an *in vivo* roadmap of endocrinogenesis, we performed single cell RNA sequencing of 36,351 pancreatic epithelial and NVF⁺ cells during secondary transition. This allowed Ngn3^{low} endocrine progenitors, Ngn3^{high} endocrine precursors, Fev⁺ endocrine lineage and hormone⁺ endocrine subtypes to be distinguished and time-resolved, and molecular programs during the step-wise lineage restriction steps to be delineated. Strikingly, we identified 58 novel signature genes that show the same transient expression dynamics as Ngn3 in the 7260 profiled Ngn3-expressing cells. The differential expression of these genes in endocrine precursors associated with their cell-fate allocation towards distinct endocrine cell types. Thus, the generation of an accurately regulated NVF reporter allowed us to temporally resolve endocrine lineage development to provide a fine-grained single cell molecular profile of endocrinogenesis *in vivo*.

KEY WORDS: Endocrine progenitor-precursor, Neurog3, Single cell RNA sequencing, Endocrinogenesis, Endocrine cell allocation, Mouse

INTRODUCTION

In rodents, pancreatic endocrine cells are generated at two distinct stages (transitions) during embryonic development. The first transition [embryonic day (E) 9.0–12.5] produces mainly glucagon-producing α -cells, whereas all other endocrine cell types (insulin-producing β -cells, somatostatin-producing δ -cells,

pancreatic polypeptide-producing PP-cells and ghrelin-producing ϵ -cells) are mainly generated during the secondary transition (E12.5–15.5) and thereafter. At both stages, endocrine cells are derived from the endocrine progenitor-precursor (EP) pool (Bastidas-Ponce et al., 2017a). EPs are marked by the transient expression of the transcription factor (TF) neurogenin 3 (Neurog3, hereafter called Ngn3) and are derived from Sox9⁺ bipotent ductal/endocrine progenitors, which are located within the pancreatic epithelium (Gradwohl et al., 2000; Gu et al., 2002; Kopp et al., 2011; Seymour et al., 2007; Shih et al., 2012). Upon lineage priming and specification, EPs express low levels of Ngn3 and by receiving additional signals these mitotic and non-committed Ngn3^{low} cells (which we refer to as endocrine progenitors) increase the levels of Ngn3 to become determined Ngn3^{high} EPs (to which we refer to as endocrine precursors) (Bechard et al., 2016; Pan and Wright, 2011; Wang et al., 2010). Further differentiation of Ngn3^{high} precursors generates distinct endocrine subtypes of the aforementioned hormone-expressing lineages. Changes in the gene regulatory networks, extracellular signaling cues from the respective niche and epithelial states (non-polarized versus polarized and epithelialized) cooperatively orchestrate the specification and determination of EPs (Apelqvist, 1999; Arda et al., 2013; Bankaitis et al., 2015, 2018; Cortijo et al., 2012; Löf-Öhlin et al., 2017). Yet, during progressive lineage restriction from specified EP progenitors to determined EP precursors the molecular signature changes over time and the mechanisms underpinning lineage allocation towards a specific endocrine subtype fate have remained largely elusive. Although several recent studies (Byrnes et al., 2018; Krentz et al., 2018; Ramond et al., 2018; Scavuzzo et al., 2018; Sharon et al., 2019; Stanescu et al., 2016; Yu et al., 2018, 2019) have shed light on the transcriptional profiles of EPs, the importance of these cells for endocrine cell formation still necessitates a comprehensive temporally resolved and fine-grained mapping of gene expression on the single cell level. The low percentage of EPs within the embryonic pancreas together with the limitation of transgenic reporter lines to reflect accurately the transient expression of the Ngn3 protein have been major obstacles to establishing a detailed roadmap of endocrinogenesis.

Here, we generated a novel reporter mouse line, in which the endogenous Ngn3 is fused to a very bright Venus fluorescence reporter protein (Ngn3-Venus fusion; NVF). The NVF reporter mirrors transient endogenous Ngn3 protein levels and allows specific isolation of EPs in the narrow time window when they actually express Ngn3. Single cell RNA sequencing (scRNA-seq) analysis of EP-enriched pancreatic epithelial cells highlighted the step-wise dynamic changes in gene expression programs from bipotent cells to EP progenitors and precursors, and finally to differentiated endocrine cells as well as ductal and acinar cells. We found a set of 58 previously undescribed EP-signature genes that are

¹Institute of Diabetes and Regeneration Research, Helmholtz Zentrum München, D-85764 Neuherberg, Germany. ²German Center for Diabetes Research (DZD), D-85764 Neuherberg, Germany. ³Institute of Stem Cell Research, Helmholtz Zentrum München, D-85764 Neuherberg, Germany. ⁴Technical University of Munich, School of Medicine, 81675 Munich, Germany. ⁵Institute of Computational Biology, Helmholtz Zentrum München, D-85764 Neuherberg, Germany. ⁶Technical University of Munich, School of Life Sciences Weihenstephan, 85354 Freising, Germany. ⁷Max Planck Institute of Psychiatry, Kraepelinstr. 2–10, 80804 Munich, Germany. ⁸Technical University of Munich, Department of Mathematics, 85748 Garching b. Munich, Germany.

*These authors contributed equally to this work

[†]Authors for correspondence (fabian.theis@helmholtz-muenchen.de; heiko.lickert@helmholtz-muenchen.de; mostafa.bakhti@helmholtz-muenchen.de)

ID F.J.T., 0000-0002-2419-1943; H.L., 0000-0002-4597-8825; M.B., 0000-0002-2307-1122

Received 15 November 2018; Accepted 21 May 2019

specifically expressed in EPs in a similar transient manner as Ngn3. Furthermore, we found that several of these EP-signature genes are differentially expressed in endocrine precursors derived from different developmental stages that defines their lineage restriction towards specific endocrine subtypes. Altogether, our data provide a comprehensive single cell survey of stage-dependent and lineage-specific gene regulation during EP induction, specification and segregation.

RESULTS

NVF mirrors the endogenous Ngn3 protein expression pattern

We generated an NVF reporter mouse line, in which the *Ngn3* translational stop codon was removed and Venus was fused in-frame with *Ngn3* using a previously described strategy (Petrezselyova et al., 2015). Heterozygous intercrosses produced offspring with normal Mendelian distribution and homozygous *Ngn3*^{Venus/Venus} animals were fertile and indistinguishable from wild-type (WT) littermates (data not shown). To assess whether the NVF reporter reflects the endogenous Ngn3 expression pattern during development, we performed immunostaining of embryonic pancreata from heterozygous *Ngn3*^{Venus/+} mice at E12.5–18.5. Using antibodies against Ngn3 and Venus, we detected a similar nuclear expression pattern of NVF and endogenous Ngn3 in pancreatic sections from different developmental stages (Fig. 1A; Fig. S1A). Plot profile analysis of staining by Venus and two different Ngn3 antibodies also revealed a similar pattern of nuclear signals between NVF and Ngn3 in pancreatic sections (Fig. S1B). Furthermore, using markers of the pancreatic tubular epithelium, such as Sox9, Nkx6-1, Mucin-1 (Muc1), EpCAM and E-cadherin (cadherin 1), we found that NVF⁺ cells exist as both intra- and extra-epithelial cells (Fig. 1B). This data is in line with the idea that at the Ngn3-expressing state, EP progenitors get specified in the pancreatic epithelium and endocrine precursor delaminate into the surrounding mesenchyme to form proto-islet clusters (Gouzi et al., 2011; Sharon et al., 2019). During development, Ngn3 is expressed at low levels in progenitors, reach peak levels in precursors and the levels are reduced upon differentiation of the precursors into the endocrine hormone⁺ lineage (Bechard et al., 2016). To find out whether the NVF protein reflects the transient expression pattern of endogenous Ngn3 protein, we stained E18.5 pancreatic sections and quantified the number of differentiated hormone⁺ endocrine cells that still expressed Ngn3 protein. We observed a subtle and weak NVF signal in the nucleus or cytoplasm of a minor fraction of the hormone⁺ endocrine cells (less than 0.6%) (Fig. 1C,D), confirming that the NVF reporter protein shows a similar transient expression pattern to that of the endogenous Ngn3 protein.

It has been shown that haploinsufficiency of Ngn3 alters endocrine cell development and post-translational modifications of this protein affect its activity (Azzarelli et al., 2017; Krentz et al., 2017; Wang et al., 2010). To confirm that the fusion protein truly represents the activity of endogenous Ngn3 protein, we first checked the expression of several downstream target genes, including *Neurod1*, *Pax4* and *Arx*. Quantitative PCR (qPCR) analysis revealed no significant differences in the expression levels of these genes in isolated epithelial cells from WT and homozygous NVF pancreata at E15.5 (Fig. S1C). Next, we analyzed endocrine cell composition in WT and NVF reporter embryos. E15.5 and E18.5 pancreata from WT and homozygous NVF embryos showed comparable α-, β-, δ- and PP-cell composition (Fig. 1E,F; Fig. S1D). Furthermore, qPCR data demonstrated comparable

expression levels of key endocrine cell markers in isolated islets from adult WT and homozygous NVF mice (Fig. S1E). These data were supported by the detection of normal fasting blood glucose levels in adult homozygous *Ngn3*^{Venus/Venus} animals (Fig. S1F) and by comparable levels of glucose-stimulated insulin secretion of isolated WT and *Ngn3*^{Venus/Venus} islets (Fig. S1G). Altogether, our analyses indicate that the NVF reporter is transiently expressed in EPs and reflects a spatiotemporal expression pattern mirroring the endogenous Ngn3 protein. Furthermore, the fusion of Venus to Ngn3 does not have a significant impact on TF function during endocrine cell development.

Single cell RNA-seq of the embryonic EP-enriched pancreatic epithelial cells

To time-resolve at the single cell level the molecular changes during endocrinogenesis, we performed high-throughput scRNA-seq analysis of pancreatic epithelial cells during the secondary transition (E12.5, E13.5, E14.5 and E15.5). To obtain the rare cell population enriched for EPs (expressing NVF), we first flow-sorted embryonic pancreatic cells using the fluorescence property of the Venus protein in homozygous *Ngn3*^{Venus/Venus} pancreata. As we were specifically interested into how epithelial multipotent progenitors become lineage restricted into acinar and bipotent (ductal/endocrine) as well as unipotent progenitors, we isolated from the remaining cells that were negative for NVF (Venus⁻) the pancreatic epithelial fraction by using the epithelial marker EpCAM. This allowed us to deplete non-epithelial cells and highly enrich for epithelial progenitor, ductal and acinar cells. A fraction of these EpCAM⁺/Venus⁻ (NVF⁻) cells were added to the corresponding EP cells (NVF⁺) at each stage (2:1 ratio for NVF⁺:NVF⁻) (Fig. 2A). We used droplet-based scRNA-seq and transcriptionally profiled 10,790 single cells at E12.5, 5042 cells at E13.5, 9633 cells at E14.5 and 10,886 cells at E15.5 (in total 36,351 cells) (Fig. 2B). Of the total cells derived from four different stages, 7260 cells (~20%) expressed *Ngn3*. Unsupervised graph-based clustering revealed eight major cell clusters including multipotent pancreatic progenitors (MPCs), tip, trunk, acinar, ductal, EPs, Fev^{high} and endocrine cells (Fig. 2C; Table S1). These clusters represent distinct pancreatic lineages at different embryonic stages, when the pancreatic epithelium is composed of MPCs. The MPCs undergo step-wise lineage restrictions and segregation into two distinct tip and trunk domains, which further differentiate into acinar and duct/EPs, respectively (Nyeng et al., 2019; Zhou et al., 2007). EPs then differentiate into Fev^{high} cells that subsequently generate different types of endocrine hormone⁺ cells (Fig. 1D). Fev^{high} cells have been recently reported as a novel cluster of pancreatic cells (Byrnes et al., 2018). Clusters were annotated based on the expression of well-known marker genes with a described function during pancreas development and/or function, including *Dlk1* (MPCs) (Ramond et al., 2018); *Cpa1* and *Myc* (tip) (Zhou et al., 2007); *Notch2* (trunk) (Lee et al., 2005); *Ptf1a*, *Cpa1*, *Cel*, *Rbpj* (acinar), *Sox9*, *Anxa2* and *Bicc1* (ductal) (Lemaire et al., 2016); *Ngn3* and *Hes6* (EPs) (Ahnfelt-Rønne et al., 2007); *Fev*, *Cck* and *Neurod1* (Fev^{high}) (Byrnes et al., 2018); and *Rbp4*, *Pyy* and *Chgb* (endocrine) (Fig. 2E, F). As the expression of Ngn3 is a well-described characteristic of EPs, we annotated *Ngn3*-expressing cells (except those expressing *Fev*) in the epithelial populations as EPs (see Materials and Methods). Our data revealed the expression of novel marker genes in each cluster that are uncharacterized or have not been associated with pancreatic cell development and/or function yet, e.g. *Mdk* and *Btf3* (MPCs); *Vtn* and *Jam3* (tip); *Cbx3*, *Hmgn1* and *Ybx1* (trunk);

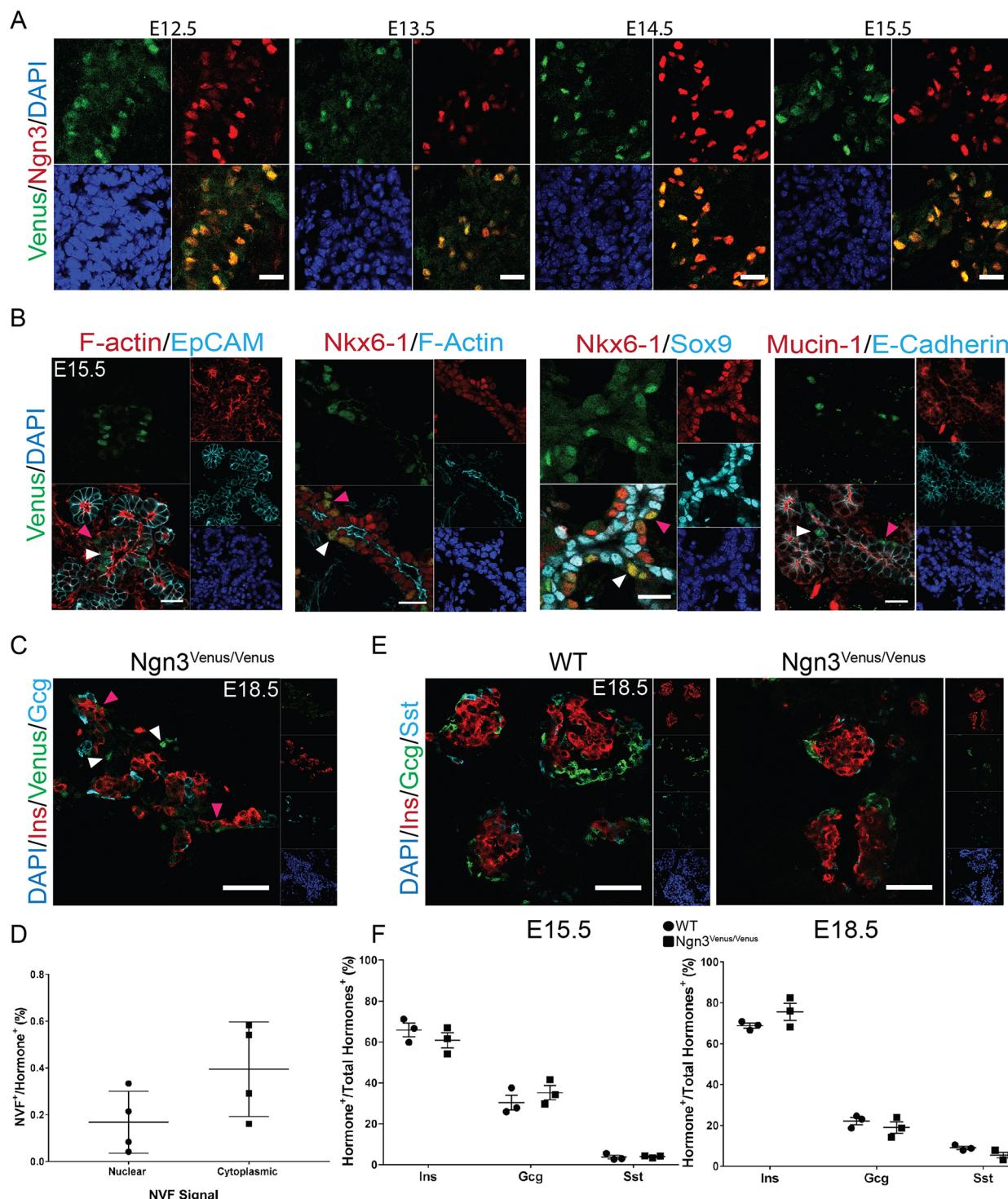


Fig. 1. Characterization of Ngn3-Venus fusion (NFn3) reporter mouse line. (A) Immunohistochemical analysis indicates the same spatiotemporal expression pattern of NFn3 and endogenous Ngn3 expression in heterozygous NFn3 (*Ngn3^{Venus/Venus}*) pancreatic sections from E12.5 to E15.5. (B) Staining of specific markers of pancreatic tubular epithelium on E15.5 homozygous NFn3 (*Ngn3^{Venus/Venus}*) sections indicates the presence of NFn3+ cells within (white arrowheads) and outside (pink arrowheads) the pancreatic epithelium. (C) Immunohistochemical analysis of co-expression of insulin and glucagon hormones with Venus in an E18.5 homozygous NFn3 section. White arrowheads indicate NFn3+ hormone- cells and pink arrowheads show low expression of Venus in hormone+ cells. (D) Quantification of the number of hormone-expressing endocrine cells that still express NFn3 at E18.5. (E) Immunostaining of insulin (Ins), glucagon (Gcg) and somatostatin (Sst) in E18.5 WT and homozygous NFn3 pancreatic sections. (F) Quantification of E15.5 and E18.5 homozygous NFn3 sections indicates normal endocrine cell formation compared with WT mice. Error bars represent s.d. Scale bars: 20 µm (A-C); 50 µm (E).

Reep5 (acinar); *Spp1* (duct); *Btbd17* and *Gadd45a* (EPs); *Vwa5b2* and *Tox3* (*Fev*^{high}); *Tmem97* and *Fam183b* (endocrine) (Fig. 2E,F). Thus, the identification of known markers for different clusters

validated our scRNA-seq approach. Furthermore, we identified many novel and functionally uncharacterized genes at different steps of pancreatic and endocrine lineage allocation.

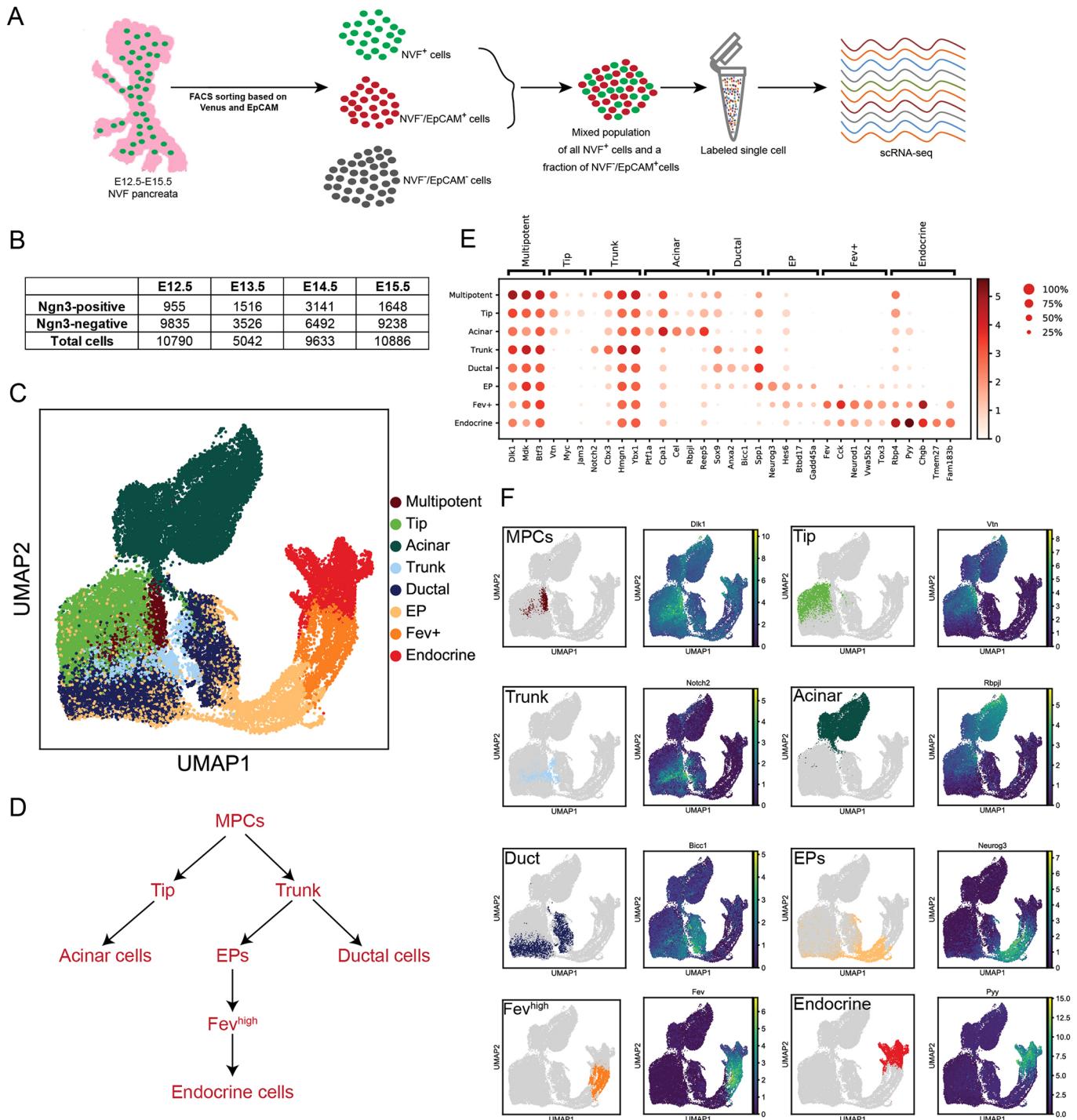


Fig. 2. Identification of embryonic pancreatic epithelial populations at single cell resolution. (A) Scheme of the procedure of pancreatic epithelial cell isolation, sorting and single cell RNA sequencing. (B) The number of cells positive for Ngn3 (*Neurog3* expression >0) and negative for Ngn3 (no *Neurog3* expression, non-EP pancreatic epithelial populations) as well as the total number of cells used in single cell data of each stage. (C) UMAP plot of 36,351 profiled single cells from four distinct stages. Colors highlight clustering into eight main cell types that include all embryonic pancreatic epithelial populations. (D) Diagram of the lineage relationship between pancreatic cells at early embryonic stages. (E) Dot plot showing expression of known and uncharacterized cell type-specific gene sets of pancreatic epithelial populations. Color intensity indicates mean expression (normalized) in a cluster, dot size indicates the proportion of cells in a cluster expressing the gene. (F) Representative gene expression and distribution of known marker genes for each epithelial population in UMAP plots. Normalized expression values are shown.

Dynamic molecular changes of pancreatic progenitors during step-wise lineage allocation

To define the step-wise lineage restriction and allocation and to explore the dynamic heterogeneity of pancreatic progenitors, we

performed stage-wise comparison of the molecular signatures of multipotent and bipotent progenitors as well as ductal, endocrine and acinar lineages. We found two subclusters of proliferating and non-proliferating cells for tip, trunk, acinar and ductal populations

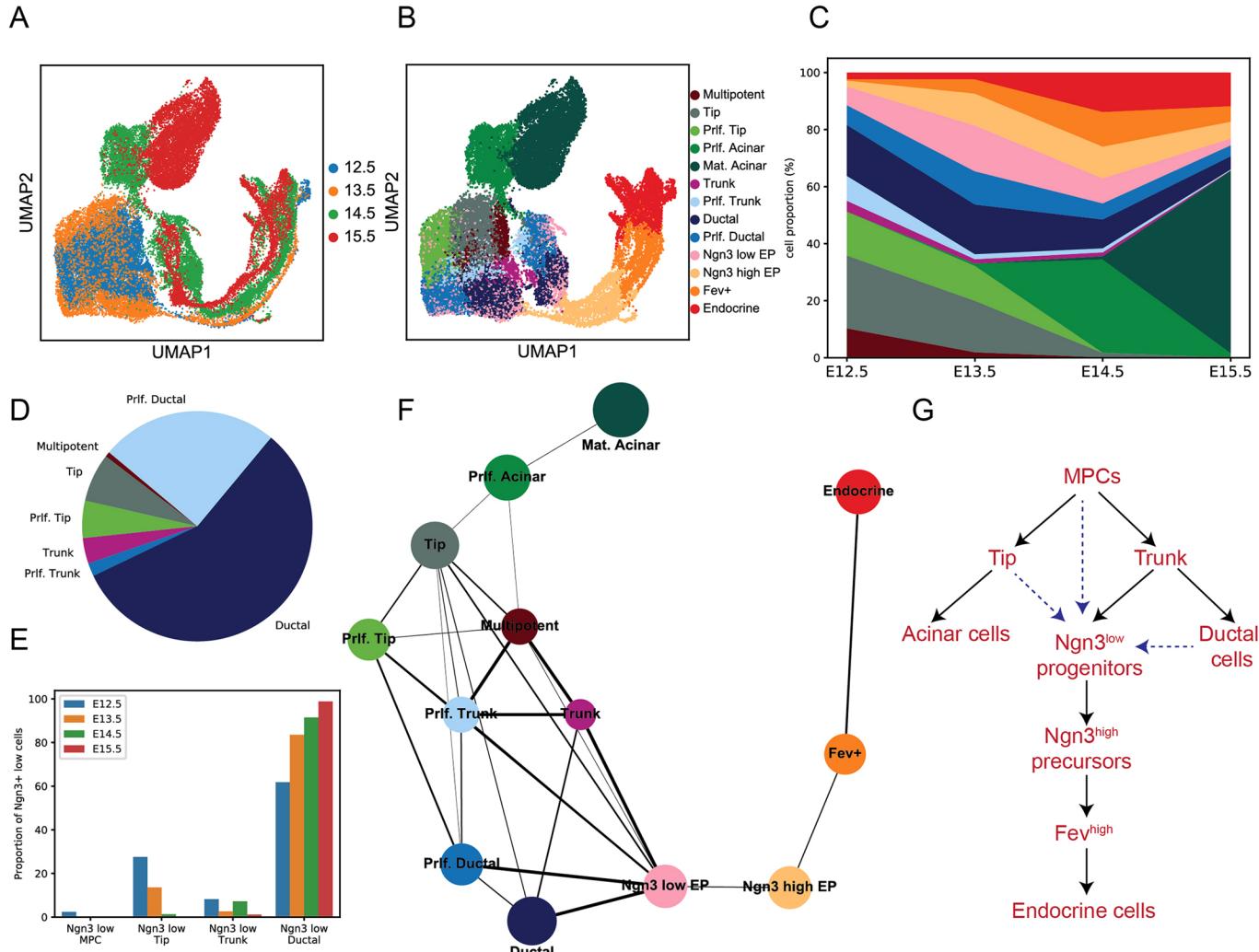


Fig. 3. Lineage segregation of early pancreatic epithelial cells. (A) UMAP plot of all pancreatic lineages from four different early stages of embryonic pancreas development. Cells from each stage are shown with different color codes. (B) UMAP plot showing the presence of proliferative and non-proliferative populations within acinar, tip, trunk and ductal clusters. (C) Area plot of pancreatic epithelial cell composition indicates rapid changes in different lineages during early pancreas development. The color codes are the same as in B. (D,E) The distribution of *Ngn3*^{low} progenitors within different pancreatic lineages (D) and by developmental stage (E). (F) Relationships of early pancreatic lineages inferred based on a measure for cluster connectivity using PAGA. Edges are weighted by significance. (G) *Ngn3*^{low} progenitors are generated from trunk bipotent progenitors. Dashed arrows indicate the possible contribution of MPCs, tip and ductal clusters to the generation of *Ngn3*^{low} progenitors. Mat., Mature; Prif., Proliferative.

(Fig. 3A,B; Fig. S2A). MPCs were mainly evident at E12.5 and expressed the multipotent progenitor and acinar marker *Cpa1* at high levels, whereas the tip and trunk clusters were prominently found at E12.5 and E13.5. Ductal cells were present throughout all stages and acinar cells appeared mainly at E14.5 and E15.5 (Fig. 3C). EP cells showed different expression levels of *Ngn3* as has been previously described (Bechard et al., 2016) and were separated into *Ngn3*^{low} progenitors and *Ngn3*^{high} precursors (see Materials and Methods).

When looking for the origin of *Ngn3*^{low} progenitors, we found that they were scattered within the MPCs, tip, trunk and ductal progenitor clusters, suggesting multiple niches and birthplaces for these progenitors (Fig. 3D,E,G). The contribution of distinct epithelial clusters to *Ngn3*^{low} progenitor formation varied at different stages. *Ngn3*^{low} progenitors were found within the MPCs only at E12.5. The tip cluster also contained these progenitors from E12.5 to E14.5, but with rapidly decreasing numbers. The *Ngn3*^{low} progenitors were detected in the trunk domain mainly until E14.5.

The presence of *Ngn3*^{low} progenitors from ductal cells was found at all stages and continuously increased from E12.5 to E15.5 (Fig. 3E). Thus, the MPCs and tip cells contain endocrine progenitors until E13.5, the existence of these cells within ductal cells persists until the end of the secondary transition. Moreover, the majority of *Ngn3*^{low} progenitors were derived from the ductal population (Fig. 3D,E), supporting the previous notion that this domain is the major epithelial source of endocrine progenitor-generating cells (Solar et al., 2009).

To infer the relationships between early pancreatic lineages, we performed partition-based graph abstraction (PAGA) analysis (Wolf et al., 2019) (see Materials and Methods). PAGA measures subgroup connectivity and draws a graph of possible differentiation paths. Consistent with the common belief (Sznurkowska et al., 2018; Zhou et al., 2007), we found that MPCs are mainly specified into tip and trunk domains. Tip cells then further connected to acinar cells, whereas trunk cells showed strong links to ductal cells. We found strong connections of *Ngn3*^{low} progenitors to the tip, trunk and ductal

cells, and a weaker connection of these progenitors with MPCs (Fig. 3F). This further supports the finding that all epithelial clusters contribute to *Ngn3*^{low} progenitor formation and that a lower number of *Ngn3*^{low} progenitors are derived from MPCs compared with other epithelial clusters. Taken together, these findings indicate that lineage restriction from multipotent to bipotent to unipotent fates occurs in a rapid step-wise manner during the secondary transition.

Identification of novel genes enriched in EPs

To obtain a dynamic and comprehensive picture of the molecular changes of the EP transcriptome during endocrinogenesis, we compared all *Ngn3*⁺ EPs (*Ngn3*^{low} and *Ngn3*^{high}) with all other (*Ngn3*⁻) cells except acinar cells (Fig. 4A; see Materials and Methods). Differential expression analysis revealed several genes significantly regulated in the *Ngn3*⁺ EPs (Fig. 4B,C; Table S2) and

provided a list of EP-enriched genes. We found increased expression of genes with known function in EPs, such as the endocrine-specification factor *Amotl2* (Scavuzzo et al., 2018), the cell-cycle inhibitor *Cdkn1a* (Miyatsuka et al., 2011), the EP differentiation factor *Sox4* (Xu et al., 2015) and the Notch inhibitor *Mfng* (Svensson et al., 2009) among many others (Fig. 4B,C). In addition, we identified several other upregulated genes that have not been associated with pancreas development before, including the cell-cycle inhibitors *Gadd45a* and *Btg2* (Krentz et al., 2018), the protein phosphatase *Ppp1r14a* (also known as *Cpi17*), mitochondrial carrier homolog 1 (*Mtch1*), Numb-like protein (*Numbl*), and the *Hes1* repressor *Hes6* (Fig. 4B,C). Among those, *Numbl* and *Hes6* inhibit Notch signaling highlighting the requirement of efficient suppression of this pathway by multiple mechanisms during endocrine induction.

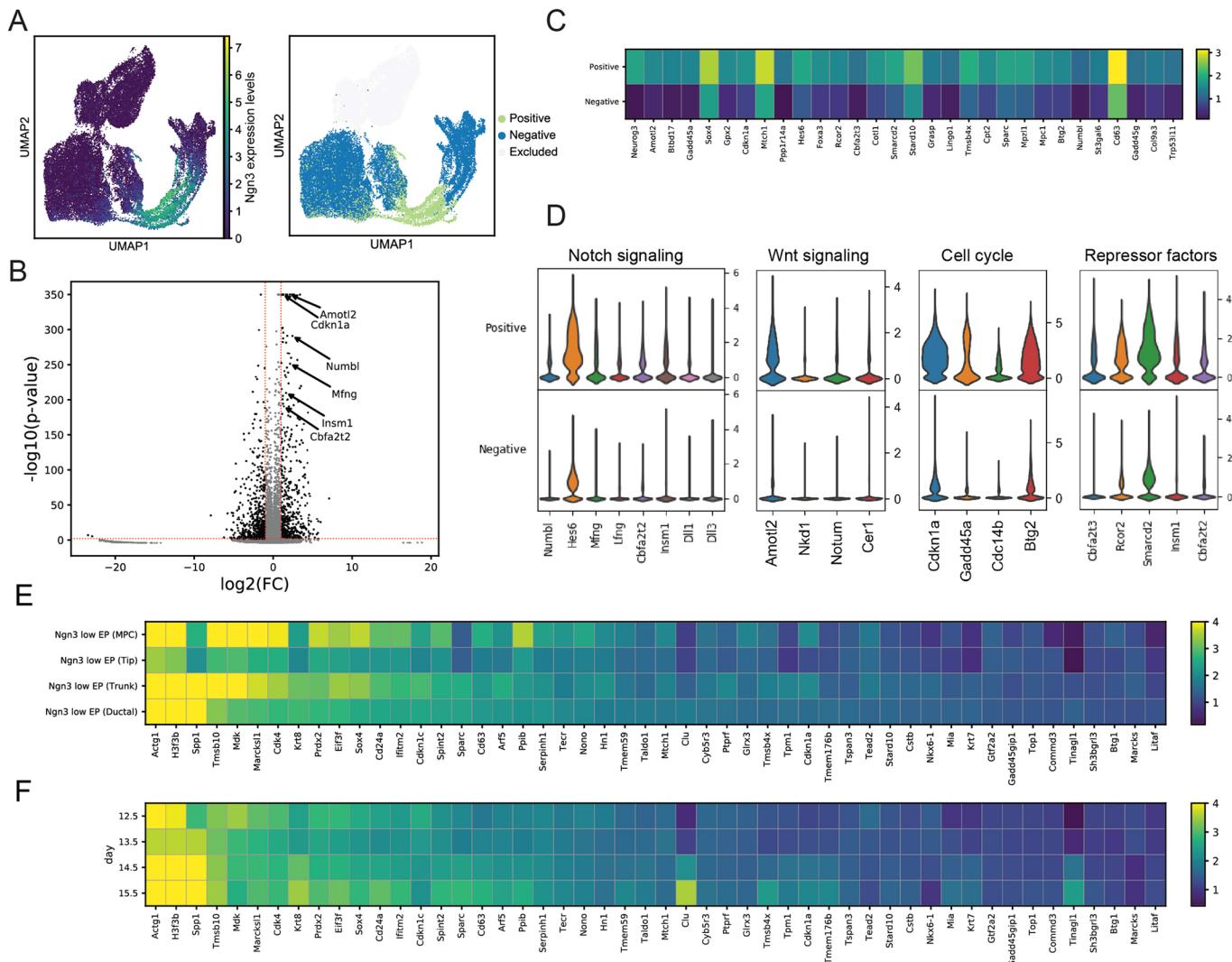


Fig. 4. Gene expression profiles of endocrine progenitors-precursors (EPs). (A) UMAP plots showing the distribution of cells considered as *Ngn3*⁺ cells. Left: *Neurog3* expression (normalized). Right: classification into *Ngn3*⁺ and *Ngn3*⁻ cells. Acinar cells were excluded here. (B) Volcano plot showing differential expression and its significance [$-\log_{10}(\text{adjusted } P\text{-Value})$, limma] for each gene in *Ngn3*⁺ EPs compared with *Ngn3*⁻ cells. Red line indicates thresholds used for significance level and gene expression change, and regulated genes are highlighted in black. Selected relevant genes are annotated. (C) Heatmap showing the mean expression in *Ngn3*⁺ and *Ngn3*⁻ cells of the top 30 upregulated genes in *Ngn3*⁺ from the data shown in B. Expression values are normalized and scaled to unit variance. (D) Pathway enrichment analyses indicate inhibition of Notch and Wnt signaling, increased expression of repressor factors and cell cycle exit in *Ngn3*⁺ cells. Distributions of gene expression (normalized) in *Ngn3*⁺ and *Ngn3*⁻ of relevant genes of these pathways are shown as violin plots. (E,F) Heatmap showing the mean expression of genes that are highly enriched in *Ngn3*^{low} progenitors between progenitors derived from different pancreatic epithelial populations (E) and derived at different developmental stages (F). Genes displayed are the 50 genes showing the highest mean expression in *Ngn3*^{low} progenitors from the top 250 upregulated genes in *Ngn3*⁺ cells from the data shown in B. Expression values are normalized.

To identify EP-specific signaling pathways and cellular processes, we analyzed the genes upregulated in *Ngn3*⁺ cells and identified Notch and Wnt inhibitors, further suggesting that these pathways are negatively modulated in EPs (Fig. 4D). Interestingly, we found increased expression of the Hippo pathway components *Amotl2* and *Nkd1*, which are also antagonists of the Wnt pathway. This highlights that crosstalk between these pathways might be involved in regulation of endocrine cell formation. Furthermore, we found an upregulation of cell-cycle inhibitors in EPs (Fig. 4D), supporting the idea that endocrine differentiation leads to lengthening of the cell cycle. We identified several transcriptional repressors or co-repressors in the EP population, such as the co-repressor *Cbfa2t3*, the histone-modifying factor *Rcor2*, the chromatin remolding factor *Smarcd2*, the endocrine-specification factor *Insm1* (Liu et al., 2006; Osipovich et al., 2014) and the Notch repressor *Cbfa2t2* (Fig. 4D), that might be important for the suppression of non-EP gene programs.

Next, we investigated whether *Ngn3*^{low} progenitors derived from different epithelial clusters or from different developmental stages (E12.5–15.5) are transcriptionally heterogeneous. To this end, using the list of genes enriched in the total EP population (*Ngn3*^{low} and *Ngn3*^{high}) (Table S2), we first identified those genes that were highly expressed in *Ngn3*^{low} progenitors. Then, we compared the expression of these genes in *Ngn3*^{low} cells derived from MPCs, tip, trunk and ductal cells (Fig. 4E), as well as from different developmental stages (Fig. 4F). Some genes, such as *Spp1*, *Tmsb10*, *Mdk*, *Marcks1*, *Cdk4* and *Sox4*, were differentially expressed in *Ngn3*^{low} progenitors; however, expression of these genes only reflect the origin of the *Ngn3*^{low} cells. For instance, *Ngn3*^{low} progenitors derived from trunk and ductal epithelium expressed high levels of *Spp1*, which is highly expressed in all trunk and ductal cells. In contrast, the expression of most of the highly enriched genes in *Ngn3*^{low} progenitors that could not be directly associated to their origin was not different between *Ngn3*^{low} progenitor populations derived from distinct epithelial clusters or from different stages (Fig. 4E,F).

Identification of transiently expressed EP-signature genes

During development, bipotent epithelial cells generate *Ngn3*^{low} mitotic progenitors that either reverse to a ductal fate or further generate *Ngn3*^{high} precursors (Bechard et al., 2016). *Ngn3*^{high} cells then differentiate into *Fev*^{high} cells that consequently produce hormone-expressing endocrine cells (Fig. 5A). Consistently, pseudotemporal ordering (Haghverdi et al., 2016) of the cells along this route, which was also inferred by PAGA (Fig. 3F), revealed a continuous differentiation trajectory (see Materials and Methods). Along this pseudotime *Ngn3* is transiently expressed as the levels of *Ngn3* increased from *Ngn3*^{low} to *Ngn3*^{high} cells and then decreased again in *Fev*^{high} cells (Fig. 5B,C). Next, we investigated whether there are other EP-specific signature genes that follow transient *Ngn3* expression in EPs. From the list of genes enriched in EPs we identified 58 genes (including *Ngn3*) that were expressed in EPs, but not or only at very low level in other pancreatic lineages (see Materials and Methods). Moreover, their expression was transient in EPs, i.e. their expression levels were gradually reduced upon differentiation of EPs into *Fev*^{high} cells (Fig. 5B; Fig. S2B). Among these genes were *Ppp1r14a*, the neuronal determination TF neurogenic differentiation factor 2 (*Neurod2*) (Gasa et al., 2008), sulfotransferase family cytosolic 2B member 1 (*Sult2b1*), uroplakin 3b-like protein 1 (*Upk3bl*), Ly6/neurotoxin-like protein 1 (*Lynx1*), the cell cycle regulator polo-like kinase 3 (*Plk3*), G protein γ subunit *Gy13* (*Gng13*), the Six1 transcriptional

co-activator *Eya2*, regulator of G-protein signaling 16 (*Rgs16*), semaphorin 3G (*Sema3g*), the cancer-associated cell-surface antigen six-transmembrane epithelial antigen of the prostate 1 (*Steap1*), and the uncharacterized gene *Gm8773* (Fig. 5C; Fig. S2B). Notably, for most of these genes no function associated with pancreas development has been reported so far. However, genes such as *Neurod2*, *Sult2b1* and *Lynx1* are involved in neuronal development and function, which highlights the similarities of the developmental programs required for the formation of EPs and neuronal cells. We also found comparable expression levels of several of these signature genes in E15.5 pancreatic epithelial cells from WT and homozygous NVF pancreata, which further supports proper functioning of the fusion protein (Fig. S2C). Furthermore, scRNA-seq analysis of human fetal pancreatic cells has indicated the expression of some of these genes, including *EYA2* and *RGS16* in human endocrine progenitors (Ramond et al., 2018).

Next, from the list of genes enriched in the *Fev*^{high} population we extracted several genes that followed a similar expression pattern to that of *Fev*. The expression of these genes started in *Ngn3*^{high} precursors, reached maximum levels in *Fev*^{high} cells and then decreased again in hormone-expressing endocrine cells. Such *Fev*^{high}-enriched genes included *Neurod1*, von Willebrand factor A domain-containing protein 5B2 (*Vwa5b2*), the paired box protein *Pax4*, cholecystokinin (*Cck*), the cancer suppressor gene *Tox3*, solute carrier family 35 member D3 (*Slc35d3*), the *Ngn3*-regulated transcriptional co-repressor *Runx1t1* (*Cbfa2t1*) (Benitez et al., 2014) and the diabetes-associated protein voltage-dependent calcium channel subunit $\alpha_2\delta$ -1 (*Cacna2d1*) (Mastrolia et al., 2017) and the uncharacterized gene transmembrane protein 185A (*BC023829*) (Fig. 5D,E; Fig. S2D). Among these, a high expression of *RUNX1T1* in *FEV*⁺ cells has also been reported in human fetal pancreas (Ramond et al., 2018). Moreover, the expression of genes, such as chromogranin-A (*Chga*) and -B (*Chgb*), carboxypeptidase E (*Cpe*), *Pax6*, the ciliary-localized protein *Fam183b* (Stauber et al., 2017) and the neuroendocrine protein ProSAAS endopeptidase inhibitor (*Pcsk1n*) was turned on at the *Fev*^{high} stage and persisted in fully differentiated endocrine cells (Fig. S2D,E). Altogether, these data suggest that the development from EPs to fully differentiated endocrine cells occurs in a step-wise manner and involves the expression of unique molecular signatures at each stage (Fig. 5E).

To resolve further the dynamic changes in gene expression during the step-wise endocrine induction, specification and differentiation, we compared the transcriptional profiles of ductal bipotent (as the main source of *Ngn3*^{low} progenitors), *Ngn3*^{low}, *Ngn3*^{high} and *Fev*^{high} populations (Fig. S3A; Table S3). We identified a large number of differentially expressed genes between ductal bipotent and *Ngn3*^{low}, *Ngn3*^{low} and *Ngn3*^{high} as well as *Ngn3*^{high} and *Fev*^{high} cells (Fig. 5F). Ordering the cells along pseudotime supported a continuous developmental trajectory and revealed stage-specific induction or repression of genes involved in signaling pathways, cell-cycle control and organ morphogenesis during EP formation and differentiation. The low expression of *Ngn3* in EPs was accompanied by increased expression levels of genes involved in activation of cAMP, Rap1 and Ras/MAPK signaling as well as Notch inhibitors (Fig. 5F; Fig. S3B). Furthermore, compared with ductal bipotent cells, *Ngn3*^{low} progenitors expressed decreased levels of genes involved in Wnt signaling activation, suggesting the requirement of Wnt signaling inhibition during EP development. This was supported by upregulation of the secreted TGF β and Wnt inhibitor *Cer1* (Piccolo et al., 1999) and Wnt-deacylation enzyme *Notum* (non-canonical inhibitor) (Kakugawa et al., 2015) in *Ngn3*^{low} progenitors compared with ductal bipotent cells (Fig. 5F).

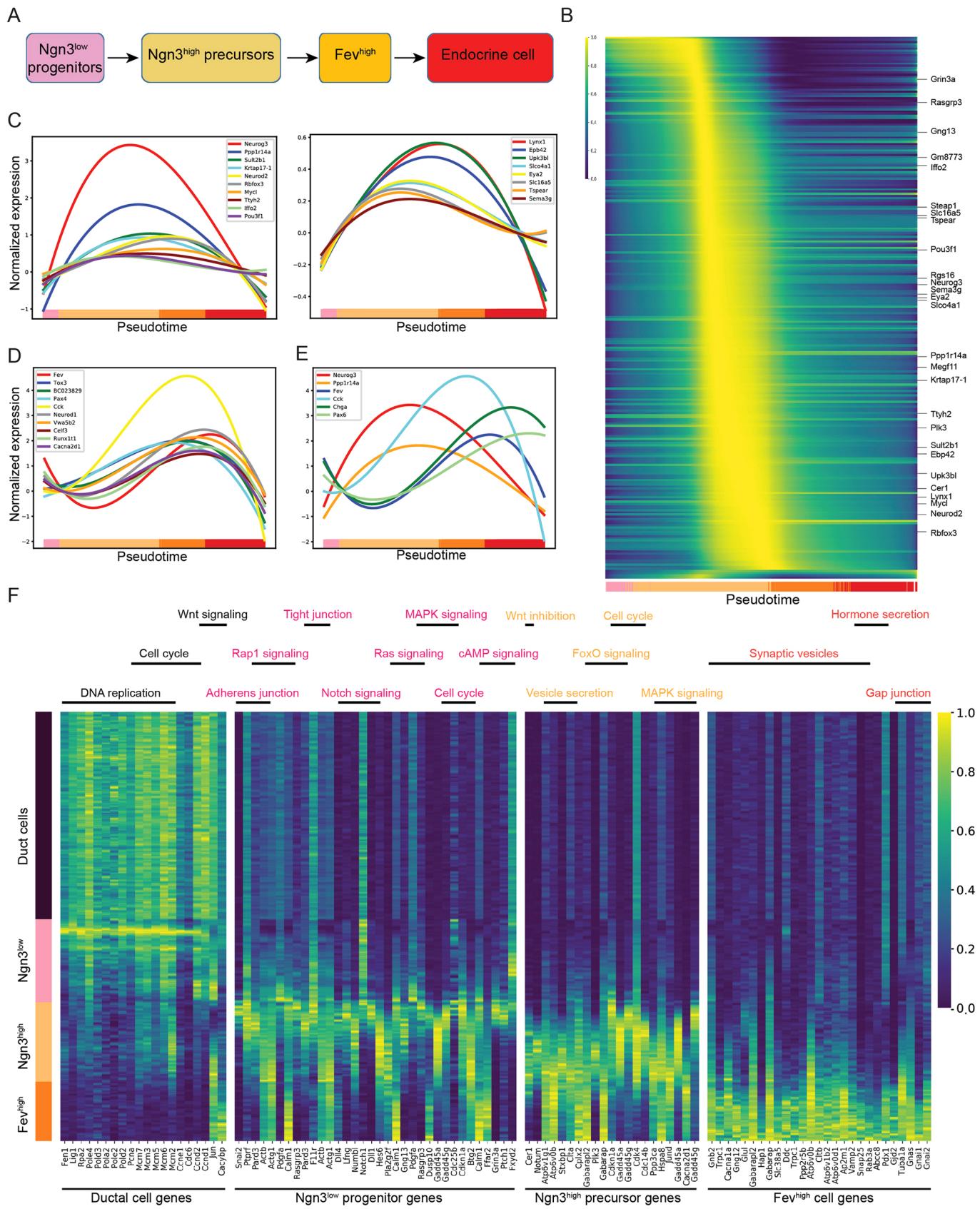


Fig. 5. See next page for legend.

Fig. 5. Distinct molecular signatures are expressed at different stages of endocrinogenesis. (A) Developmental scheme of differentiation of endocrine progenitors into endocrine cells. (B) Gene expression along a one-dimensional differentiation trajectory from *Ngn3*^{low} progenitors to endocrine cells. Each cell is assigned a value in diffusion pseudotime, which reflects the differentiation process. The top 250 genes upregulated in *Ngn3*⁺ EPs are shown (Table S2) and selected EP signature genes highlighted. Expression is normalized and approximated by polynominal regression fits along pseudotime. Fitted values of each gene are then scaled to the range between 0 and 1. Cluster membership of the cells is indicated at the bottom. Color codes are as in A. (C-E) Gene expression changes along the pseudotime from endocrine progenitors to endocrine cells. Lines are polynominal regression fits of normalized expression data. Cluster membership of the cells is indicated at the bottom. Color codes are as in A. (C) Several EP-signature genes follow a transient expression pattern in EPs similar to that of *Ngn3*. For simplicity the representative genes are presented in two different graphs. (D) Expression profile of several *Fev*^{high}-enriched genes. The expression of these genes reaches its peak in *Fev*^{high} cells and is reduced upon differentiation into endocrine cells. (E) Gradual increase or decrease of several EP-, *Fev*^{high}- and endocrine-specific genes during endocrinogenesis. (F) Heatmap showing subtype-specific gene expression profiles of duct cells, *Ngn3*^{low} progenitors, *Ngn3*^{high} precursors and *Fev*^{high} cells. Cells are ordered along pseudotime rooted in the duct cell population. Cluster membership of the cells is indicated on the left. Pathways associated with the gene sets are highlighted at the top. Expression is normalized and shown as the running average over 80 cells scaled to the maximum observed level per gene. Several genes that are involved in different cellular processes or signaling pathways have been included in the heatmap repeatedly.

Additionally, alterations in cell adhesion molecules and activation of the Ras/Rap1 pathway in *Ngn3*^{low} progenitors likely drive cell morphological changes, suggesting that programs of morphogenesis are initiated at the progenitor level and continue while EPs delaminate from the pancreatic epithelium and differentiate into distinct endocrine cell types. Moreover, the expression of genes involved in cell-cycle exit was initiated in *Ngn3*^{low} progenitors and continued in *Ngn3*^{high} precursors. The high expression levels of *Ngn3* in EPs initiated the expression of the secretory machinery during differentiation towards slowly cycling hormone-expressing cells (Fig. 5F; Fig. S3B). The expression of secretory components was further increased in the *Fev*^{high} cells and at this stage cells expressed the hormone secretion machinery (Fig. 5F; Fig. S3B). Interestingly, *Fev*^{high} cells highly expressed neurotransmitters and synaptic vesicle genes (Fig. 5F; Fig. S3B), again highlighting the molecular and cellular similarities between endocrine and neuronal cells.

Stage-dependent heterogeneity of EPs defines endocrine cell allocation

Although all endocrine cells are derived from EPs, at which stage these cells become unipotent to produce distinct endocrine cell types remains largely unknown. It has been shown that mouse endocrine cells are generated in a step-wise manner, in which EPs are consecutively specified towards α -, β -, PP-, δ - and ϵ -cells during the secondary transition (Johansson et al., 2007). Indeed, the proportions of endocrine subtypes obtained at each time point reflected such a step-wise induction of endocrine cell fates. At E12.5, we only found α -cells, whereas β -, δ - and ϵ -cells appear between E13.5 and E14.5 (Fig. 6A). Interestingly, a portion of α -cells was also generated at later stages, suggesting that formation of these cells is not restricted to the primary transition (Fig. 6A; Fig. S4B,C). To investigate further the timing of endocrine specification, we estimated the RNA velocity of each cell (see Materials and Methods) (La Manno et al., 2018). RNA velocity is a proxy for the rate of pre-mRNA to mRNA processing and, therefore, is a vector in time that can be used to infer to which state a cell is

moving in the high-dimensional transcriptome space. Thus, we can predict towards which endocrine subtype the EPs differentiate at a specific embryonic stage. As expected, we detected a movement of a portion of bipotent cells towards the endocrine lineage and a strong predicted directional flow from *Ngn3*^{low} to *Ngn3*^{high} to *Fev*^{high} cells. Interestingly, EPs and *Fev*^{high} cells at E12.5 and E13.5 pointed towards α -cells and to the few detected β -cells, whereas at E14.5 and E15.5 the majority of EPs showed a velocity towards β -cells (Fig. 6B). These data indicate that the α - and β -cell fate is specified in EPs mainly at early and late stages, respectively.

Furthermore, we found stage-dependent differential expression of several genes in differentiated α -cells (Fig. S4A). Genes such as *Arx* and *Slc38a5* showed similar expression levels in early (E12.5) and late (E15.5) differentiated α -cells (Fig. 6C), whereas *Sct*, *Mafb* and *Meis2* were differentially expressed. We found higher expression levels of *Sct* in early α -cells and higher expression levels of *Mafb* and *Meis2* in late α -cells (Fig. 6D,E). This differential gene expression of early and late α -cells might be related to the formation of these cells during the first and secondary transition. However, we could not find α -cells with a primary transition signature at E14.5 and E15.5 (Fig. S4B), suggesting that the α -cells derived from early stages either acquire a similar signature to the late-stage α -cells or they have been eliminated during pancreas development.

One of the proposed mechanisms for stage-dependent endocrine specification is a change in the epithelial status and the surrounding niche. In mouse, the formation of endocrine cells coincides with the ramification of epithelial plexus to a single-layer epithelial network that possibly exposes the EPs to different ECM components and defines their fate towards specific endocrine cell types (Bakhti et al., 2019; Bankaitis et al., 2015). However, it is unclear whether EP transcriptional heterogeneity also primes endocrine subtype specification. To this end, we compared the expression levels of the newly described EP-signature genes between EPs derived from different developmental stages. *Ngn3*^{low} progenitors derived from different stages showed low expression levels of most of the EP-signature genes and, except for a few genes, they expressed comparable levels of most of these genes (Fig. S4C). In contrast, we found differential expression of several signature genes in *Ngn3*^{high} precursors derived from different developmental stages (E12.5–15.5) (Fig. 7A). We discovered that several genes, including *Gng13*, *Steap1* and *Gm8773*, were expressed more highly in E12.5 *Ngn3*^{high} precursors compared with later stages, whereas other genes, including *Upk3bl*, *Sultb2l* and *Neurod2*, showed clearly increased expression in E15.5 *Ngn3*^{high} precursors compared with earlier stages (Fig. 7B-D). These data were further supported experimentally by performing qPCR analysis of isolated cells from E13.5 and E15.5 pancreata (Fig. 7E). We speculated that the differential expression of these genes at early and late stages is related to a fate decision of EPs towards distinct endocrine subtypes. In such a scenario, early precursors that express genes such as *Gng13* and *Gm8773* might be specified towards α -cells, whereas late precursors that express genes such as *Upk3bl* and *Neurod2* might allocate to the β -cell fate. Indeed, PAGA analysis revealed that *Ngn3*^{high} precursors expressing high levels of *Gng13*, *Steap1* and *Gm8773* are highly connected to α -cells and *Ngn3*^{high} precursors expressing high levels of *Upk3bl*, *Sultb2l* and *Neurod2* are highly connected to β -cells (Fig. 7F). This finding is in line with a recent study showing that *Ngn3*⁺ cells that express *Myt1* are biased toward β -cell fate (Liu et al., 2019). The balance between the expression of *Arx* and *Pax4* is the only currently known mechanism of endocrine cell subtype specification (Collombat et al., 2003, 2009). Although the expression of *Pax4* was started at late stages of

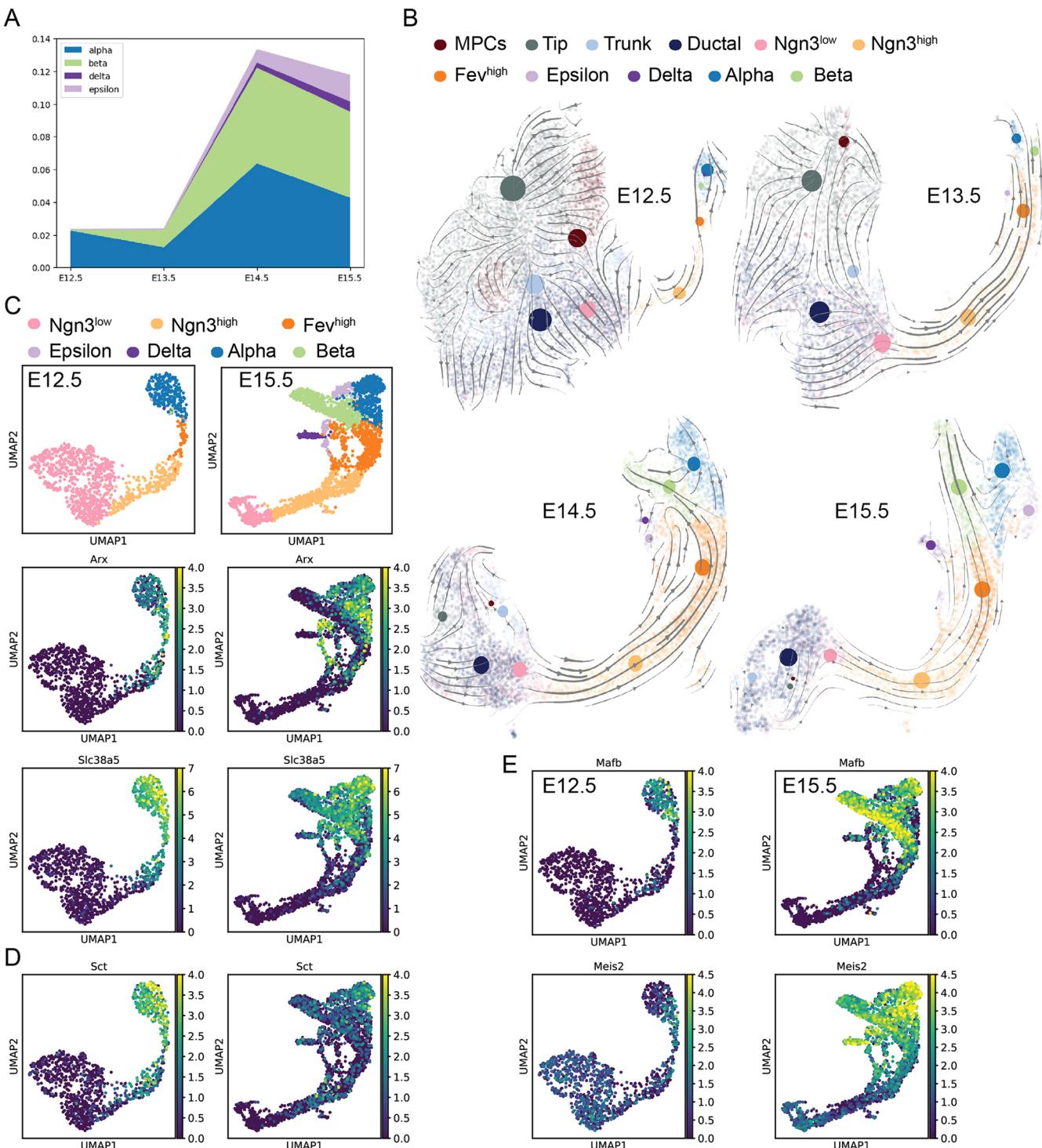


Fig. 6. Endocrine cells are produced at distinct developmental stages at different rates and with specific characteristics. (A) Area plot of endocrine cell type composition indicative of different differentiation rates of each cell type at distinct developmental stages. (B) Differentiation of EPs into endocrine subtypes at E12.5-E15.5. The direction of cell differentiation at each stage inferred from estimated RNA velocities are plotted as streamlines on the UMAP. Direction indicates transition towards the estimated future state of a cell. Dots are cluster centers of the indicated cell types. Dot size represents cluster size. Acinar clusters were removed from these analyses to better depict the endocrine lineage segregation. (C) Expression of different gene markers in primary (E12.5) and secondary (E15.5) α-cells. UMAP plots show that primary and secondary α-cells express comparable levels of Arx and Slc38a5. See also Fig. S4. (D) Primary α-cells (E12.5) express higher levels of Sct compared with the secondary α-cells (E15.5). (E) The expression levels of Mafb and Meis2 is higher in secondary α-cells (E15.5) compared with the primary α-cells (E12.5). The UMAP plots in C-E are calculated and presented only for Ngn3^{low} progenitors until differentiated endocrine cells. Normalized expression values are shown.

Ngn3^{high} precursors, the expression of *Arx* was initiated only in *Fev^{high}* cells (Fig. S4D), suggesting that the *Pax4-Arx* axis is likely not established in *Ngn3^{high}* precursors but rather in the *Fev^{high}* cells.

The fact that transcriptional heterogeneity of EPs exists in *Ngn3^{high}* precursors proposes that these cells might be already determined towards specific endocrine cell fates. Therefore, our findings,

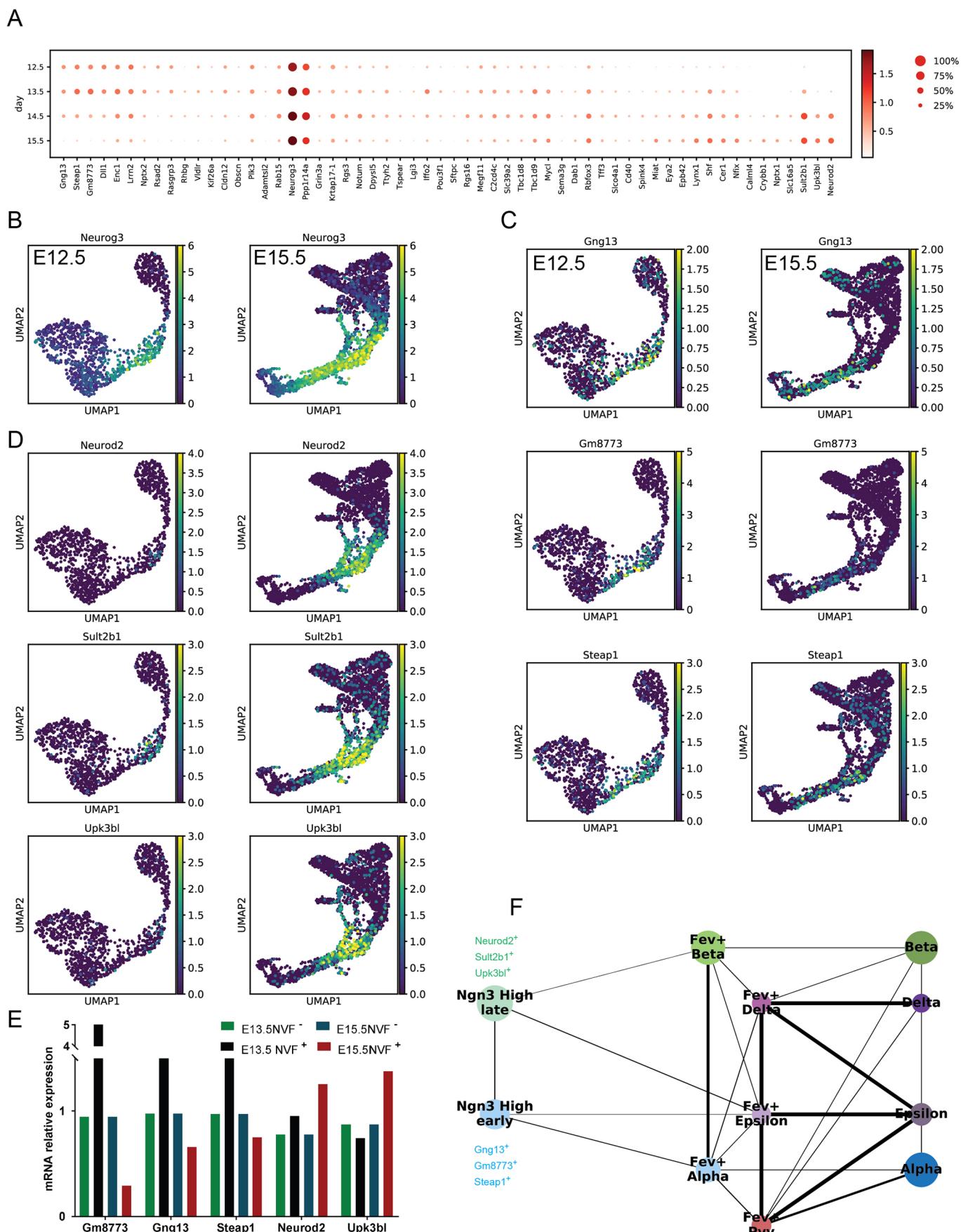


Fig. 7. See next page for legend.

Fig. 7. Stage-dependent heterogeneity of *Ngn3*^{high} precursors defines endocrine cell fate. (A) Dot plot of the expression levels of EP-signature genes in *Ngn3*^{high} precursors derived from different developmental stages. Although the expression levels of many of the EP-signature genes are comparable in cells derived from different stages, several genes show stage-dependent differential expression. Color intensity indicates mean expression (normalized) in a cluster, dot size indicates the proportion of cells in a cluster expressing the gene. (B) UMAP plots for the expression of *Ngn3* in the EP population at E12.5 and E15.5. Only clusters from *Ngn3*^{low} progenitors until differentiated endocrine cells are shown. (C) UMAP plots showing that expression levels of *Gng13*, *Gm8773* and *Steap1* are higher in E12.5 *Ngn3*^{high} precursors compared with those from E15.5. (D) UMAP plots indicating increased expression of *Neurod2*, *Sult2b1* and *Upk3bl* in E15.5 *Ngn3*^{high} precursors compared with those from E12.5. For B-D, normalized expression values are shown. (E) qPCR analysis of isolated cells from E13.5 and E15.5 homozygous NVF pancreata for several EP-signature genes ($n=1$ experiment derived from 8–10 embryos). (F) PAGA analysis indicating the association between the differential expressions of EP-signature genes in *Ngn3*^{high} precursors with their fate towards α - and β -cells. This analysis also shows two different *Fev*^{high} populations that are committed towards α - or β -cell fate.

together with other previous studies (Bechard et al., 2016; Desgraz and Herrera, 2009), support the idea that EPs are not unipotent progenitors at birth (*Ngn3*^{low}), but rather undergo cell-fate determination at the *Ngn3*^{high} state.

DISCUSSION

Improved therapy for diabetic patients could be achieved by islet cell replacement or *in vivo* regeneration of lost or dysfunctional β -cells. Both approaches rely on a detailed understanding of endocrinogenesis and the molecular programs driving endocrine lineage induction, specification and differentiation. Although gene regulatory networks and signaling cues that orchestrate endocrine cell differentiation have been well studied, the endogenous niche factors, morphogenetic cues and neighboring cell-type interactions priming endocrine cell specification are still only beginning to be understood. Uncovering mechanisms of endocrinogenesis is important for the generation of stem cell-derived islet-like clusters (ILCs) with defined cell composition and improved function (Bakhti et al., 2019).

NVF: a novel *Ngn3* reporter mouse line

Here, we generated a novel NVF reporter mouse line, which accurately reflects the spatiotemporal protein expression and regulation of the endogenous *Ngn3* protein. Previous *Ngn3* reporter mouse lines have been generated either by the replacement of *Ngn3* by a fluorescent reporter gene (Lee et al., 2002) to generate heterozygous knock-in/knockout reporter mice or by transgenic expression of fluorescent proteins (FPs) regulated by the *Ngn3* promoter (Bechard et al., 2016; Kim et al., 2015; Mellitzer et al., 2004). Although the onset of expression of the transgenic reporter FPs reflects the expression of endogenous *Ngn3* protein in these animals, the transient expression of the endogenous *Ngn3* protein is not reflected due to the relatively high stability of the FPs. Therefore, analysis of EPs using such reporter mice is affected by the existence of a portion of differentiated endocrine cells that still express FPs but not *Ngn3*, undermining the efficiency and accuracy of EP labeling/tracking and their isolation. In addition, one of these mouse lines (Lee et al., 2002) can be used only in the heterozygous condition, which impacts endocrine formation and differentiation due to heterozygosity (Wang et al., 2010). In contrast, our NVF reporter mouse line shows no detectable endocrine phenotype even in the homozygous state and the fusion of Venus to *Ngn3* results in a transient expression kinetic of NVF that is comparable to that of the endogenous *Ngn3* protein. These properties enabled us to isolate a large number of early and late

EPs to perform a high-throughput single cell transcriptomic analysis to decipher lineage relationships and molecular changes during endocrinogenesis.

Lineage relationship between embryonic pancreatic epithelial cells using scRNA-seq analysis

Using scRNA-seq during secondary transition, we identified eight main epithelial clusters and found the existence of *Ngn3*^{low} progenitors within MPCs, tip, trunk and ductal clusters. Among these, the presence of *Ngn3*^{low} progenitors within the MPCs and the tip cluster, although at lower numbers than in the ductal epithelial cells, was remarkable. Previously, a lineage-tracing study has shown that the pancreatic distal tip domain contains *Cpa1*⁺ MPCs that at E12.5 are tripotent and are able to give rise to exocrine, duct and endocrine cells, but only generate exocrine cells after E13.5 (Zhou et al., 2007). Moreover, another study has shown that the tip domain cells retain their tripotency until E12.5 and afterwards become restricted towards the acinar fate (Sznurkowska et al., 2018). Lineage-tracing analysis using *Ptf1a*^{CreERTM} mice has also indicated contribution of MPCs to all three pancreatic lineages until E14.0, when the number of MPCs significantly declines (Pan et al., 2013). Therefore, our finding supports these previous reports on the contribution of tip cells to EP formation at early stages before tip cells become unipotent, i.e. fate restricted towards the acinar fate. Although our data provide evidence of the existence of endocrine progenitors within the MPCs and tip cluster, they do not show whether EP lineage-primed progenitor cells are physically located within the tip MPC domain. Along this line, a recent study not only reported the contribution of tip cells to the EP pool at E14.5, but also showed the existence of *Ptf1a*⁺ EPs within the distal domain of the epithelium (Scavuzzo et al., 2018). Therefore, this study may provide the first evidence of the existence of EPs that are physically positioned within the tip domain at early stages before the tip cells become unipotent acinar cells.

We found that the major portion of endocrine progenitors existed within the ductal cluster, supporting a lineage-tracing study using *Hnf1* β ⁺ cells that demonstrated the appearance of EPs from the ductal epithelium until E16.5 (Solar et al., 2009). Furthermore, the appearance of high numbers of *Ngn3*^{low} progenitors at E15.5 indicates that the endocrine progenitor pool is not set at very early stages. Instead, it continues to be generated during secondary transition, as has been suggested previously (Bechard et al., 2016). It is important to note that we only detected *Ngn3*^{low} progenitors, but not *Ngn3*^{high} precursors, within the MPCs, tip, trunk and ductal clusters. Therefore, it still remains unclear whether all cells expressing low levels of *Ngn3* eventually differentiate into *Ngn3*^{high} precursors or if some of them differentiate back to the respective epithelial cells from which they are derived.

EP-enriched and -signature genes

The high number of isolated *Ngn3*⁺ cells (7260) allowed us to obtain not only a high-confidence EP-enriched gene list (such as *Amotl2*, *Sox4* and *Hes6*), but also provided a unique set of endocrine signature genes (such as *Neurod2*, *Ppp1r14a*, *Gng13* and *Sult2b1*) with a similar expression kinetic as *Ngn3* in EPs. To our knowledge, this is the first study reporting the transient expression of several genes other than *Ngn3* in EPs. This unique EP-signature offers new markers for the labeling and/or tracking of these cells. Now, the important questions are whether the expression of these genes is controlled by *Ngn3* and if they are upstream regulators or downstream effectors of *Ngn3* activity. Interestingly, a recent time-resolved single cell study in the intestine identified several of

these endocrine-enriched and -signature genes, such as *Rcor2*, *Smarcd2*, *Sox4*, *Dll1*, *Eya2*, *Mycl* and *Neurod2* (Gehart et al., 2019). This data set together with our analysis further highlight significant similarities between the developmental programs for pancreatic and intestinal enteroendocrine cells. Remarkably, the *Ngn3*^{low} progenitors derived from different epithelial sources or at different developmental stages, did not show transcriptional heterogeneity in regard to the expression levels of EP-enriched and EP-signature genes. These data propose that, although in the *Ngn3*^{low}-state progenitors are biased towards an endocrine fate (Bechard et al., 2016), they might not yet be committed towards the endocrine fate and a specific endocrine subtype, leading us to question the unipotency of EPs at birth (*Ngn3*^{low}). In contrast, *Ngn3*^{high} precursors generated at different developmental stages expressed distinct levels of several EP-signature genes, such as *Neurod2*, *Sult2b1*, *Gng13* and *Gm8773*. Moreover, PAGA analysis revealed that the differential expression of these EP-signature genes associates with the fate of *Ngn3*^{high} precursors towards α - or β -cells. These findings indicate that *Ngn3*^{high} precursors are not a homogenous population on the transcriptional level and suggest that endocrine subtype specification likely occurs in the *Ngn3*^{high} precursor state. This conclusion supports the previous *in vivo* clonal study that has shown the unipotency of EPs expressing high levels of *Ngn3* (Desgraz and Herrera, 2009). Future work should address the function of those differentially expressed signature genes in early and late *Ngn3*^{high} precursors to provide deeper insights into the mechanisms underlying endocrine cell allocation. Such functional analysis might uncover the upstream cues defining the fate of endocrine cells through the establishment of the Arx-Pax4 axis. In addition, profound analysis of potential receptor-ligand interactions between EPs and endocrine subtypes could reveal unknown (feedback) cues from differentiated endocrine cells that further regulate induction of specific endocrine subtypes. Moreover, how many of those EP-signature genes are also expressed in human endocrine progenitors-precursors and whether they have a conserved function in endocrinogenesis deserves more attention. Finally, integration of real experimental time into pseudotime and trajectory inference can provide further insights into the dynamics of endocrine differentiation and the evolution of *Ngn3*^{high} precursors and their descendants over time (Fischer et al., 2019; Schiebinger et al., 2019). We anticipate that further development of computational methods to include large asynchronicity as found here will enable investigators to track back the differentiated endocrine subtypes to their origin and further elucidate their differentiation trajectories.

Fev^{high} population link endocrine precursors to endocrine cells

The Fev^{high} population has been recently reported as a novel endocrine cell state during pancreas development in mouse and human (Byrnes et al., 2018; Krentz et al., 2018; Ramond et al., 2018). These cells express low levels of *Ngn3* (declining after the *Ngn3*^{high} precursor state), but also several early endocrine TFs and marker genes, such as *Pax4*, *Neurod1* and *Chga/b*. Lineage-tracing analyses have indicated that these cells are derived from *Ngn3*⁺ EPs and differentiate into endocrine cells (Byrnes et al., 2018). This finding is supported by our PAGA and RNA velocity analysis indicating that differentiation of EPs into endocrine cells passes through a Fev⁺ stage. Indeed, most of EP-signature genes, such as *Ppp1rl4a*, *Sult2b1* and *Sema3g*, are still expressed (though with lower levels compared with that of EPs) in Fev^{high} cells, whereas the expression of many endocrine-specific genes, such as *Chga/b*, *Pax6*

and *Cpe*, is initiated and increasing in these cells. This observation suggests that although each stage of endocrinogenesis involves the expression of a set of highly specific genes, the differentiation of EPs into endocrine cells does not occur through a rapid transcriptional switch but rather passes through gradual alterations in gene expression profiles (Fig. 5E). In addition, we found that the transcriptional heterogeneity of *Ngn3*^{high} EPs from different developmental stages is also partially reflected in Fev^{high} cells (data not shown), suggesting that similar to *Ngn3*^{high} precursors, Fev^{high} cells are also unipotent cells with a defined fate towards a specific endocrine subtype.

Conclusions

In summary, we provide here a comprehensive high-resolution single cell gene expression profile of distinct pancreatic and endocrine progenitors and lineages during secondary transition. This information can be used as a blueprint for the generation of ILCs from stem cells. We identified the formation of *Ngn3*^{low} progenitors from the heterogeneous populations of tip, trunk and ductal cells, suggesting that only a subset of specified pancreatic progenitors can be primed towards EPs. Defining this subpopulation and the involved pathways will help to expand adequate numbers of endocrine progenitors. Furthermore, our analysis indicated step-wise activation/suppression of different signaling pathways such as cAMP and Wnt signaling, as well as changes in morphogenesis programs during the induction and differentiation of EPs that will aid the efficient production of differentiated endocrine subtypes *in vitro*. Finally, we identified 58 transiently expressed endocrine signature genes that subdivide *Ngn3*^{high} precursors into distinct subpopulations. Such markers and factors might help to discriminate EP subpopulations and allow differentiation towards a specific endocrine cell fate thus improving current protocols to generate ILCs with defined cell-type composition. In addition, understanding how EPs differentiate towards non- β - and β -cells might uncover the molecular programs required for redifferentiation of dedifferentiated β -cells or transdifferentiation of non- β - into β -cells for regenerative diabetes therapy.

MATERIALS AND METHODS

Generation of the Ngn3-Venus fusion reporter mice

Mice were kept at the central facilities at Helmholtz Zentrum München German Research Center of Environmental Health in accordance with the German animal welfare legislation and acknowledged guidelines of the Society of Laboratory Animals (GV-SOLAS) and of the Federation of Laboratory Animal Science Associations (FELASA). Post-mortem examination of organs was not subject to regulatory authorization.

The Ngn3-Venus fusion (NVF) construct was generated as described previously (Petrozselyova et al., 2015). Mouse IDG3.2 embryonic stem cells were electroporated with pCAG195 Cas9v2D10A-bpA (encoding for the nuclease Cas9), pbs-U6-chimericRNA and linearized targeting vector. Using geneticin treatment, positive Neo-resistant clones were selected and the targeting of Ngn3-Venus was assessed by PCR genotyping. Next, the NVF-positive clones were aggregated with CD1 morulae and the resulting chimeric mice passed the NVF allele through their germline cells. Finally, the Neo cassette flanked by LoxP sites was deleted by intercrossing with the ROSACre mouse line. The excision of the Neo cassette was confirmed by genotyping PCR. The mouse colony was maintained with a mixed background (C57BL/6J×129/SvJ). We named this mouse line the Ngn3-Venus fusion (NVF) reporter line. The mouse line will be available to the research community upon direct request to H.L.

Pancreata dissection and cell sorting

Embryonic pancreata from NVF homozygous mice were dissected and pooled together for each stage as follows: 29 pancreata from E12.5, 35

pancreata from E13.5, ten pancreata from E14.5 and seven pancreata from E15.5. Next, pancreata were kept in 0.25% Trypsin for 5 min on ice and then incubated at 37°C for 10 min. The single cell samples were then centrifuged at 1700 rpm (290 g) for 5 min at 4°C. The supernatant was removed and cells were counted. 5 µl anti-mouse CD326 (EpCAM) PE (eBioscience, 12-5791-81) and rat IgG2a K isotype control (eBioscience, 12-4321-42) were used for 1×10^6 cells in 100 µl total volume. After staining for 30 min at 4°C, the cells were stained with DAPI to detect dead cells. The samples were then washed twice and resuspended in FACS buffer (PBS, 1% BSA, 0.5 mM EDTA) and loaded for FACS sorting. The gating strategy was as follows: main population > single cells > living cells (DAPI negative) > Ngn3⁺ (FITC⁺) and EpCAM⁺ (PE⁺) cells. The collected cells were enriched for Ngn3⁺ (~2/3) and EpCam⁺ cells (~1/3) in FACS buffer containing 1% BSA and 0.1 mM EDTA. Finally, cells were counted and the number of dead cells was identified by Trypan Blue staining. If less than 20% of the cells were dead, cells were processed for single cell RNA seq.

Single cell sequencing

Single cell libraries were generated using the Chromium Single Cell 3' library and gel bead kit v2 (PN #120237) from 10x Genomics. Briefly, to reach a target cell number of 10,000 cells per sample 16,000 cells per sample were loaded onto a channel of the 10x chip to produce Gel Bead-in-Emulsions (GEMs). This underwent reverse transcription to barcode RNA before cleanup and cDNA amplification followed by enzymatic fragmentation and 5' adaptor and sample index attachment. Libraries were sequenced on the HiSeq4000 (Illumina) with 150 bp paired-end sequencing of read2.

Quantitative PCR (qPCR)

qPCR was performed using TaqMan probes and 25 ng cDNA per reaction. Each reaction consisted of 4.5 µl cDNA in nuclease-free water, 5 µl TaqMan Advanced master mix (Life Technologies) and 0.5 µl TaqMan probe. The qPCR was performed using ViiA7 (Thermo Fisher Scientific). Ct-values were normalized among samples, transformed to linear expression values, normalized to reference genes and to control samples, i.e. relative expression (gene) = $(2^{Ct(\text{mean genes})} - Ct(\text{gene})) / (2^{Ct(\text{mean references})} - Ct(\text{reference}))$, and normalized expression (gene) = relative expression (gene)/relative expression_{control} (gene).

See supplementary Materials and Methods for details of the probes used.

Immunohistochemistry and microscopy

Embryonic pancreata were dissected and fixed in 4% paraformaldehyde (PFA) in PBS overnight at 4°C. The tissues were equilibrated in 10% and 30% sucrose-PBS solutions at RT (2 h each solution) followed by 1:1 tissue-freezing medium:30% sucrose overnight at 4°C. Afterwards, they were embedded in cryoblocks using tissue-freezing medium (Leica, 14020108926) and sections of 20 µm were cut. Next, samples were permeabilized (0.1% Triton X-100, 0.1 M glycine) for 15 min and incubated in blocking solution (10% fetal calf serum, 3% donkey serum, 0.1% BSA and 0.1% Tween-20 in PBS) for 1 h at room temperature (RT). Primary antibodies were diluted in blocking solution and incubated with the samples overnight at 4°C. After washing with PBS, samples were stained with secondary antibodies diluted in the blocking solution for 3–5 h at RT. The samples were then incubated with 4', 6-diamidino-2-phenylindole, followed by three washes with PBS and embedded in commercial medium ProLong Gold (Life Technologies). All images were obtained with a DMI 6000 Leica microscope using LAS AF software. Images were analyzed using LAS AF and ImageJ software programs. See supplementary Materials and Methods for details of the antibodies used.

Islet isolation

The isolation of islets was performed using collagenase P (Roche) digestion of adult pancreas as described previously (Bastidas-Ponce et al., 2017b). See supplementary Materials and Methods for further details.

Glucose-stimulated insulin secretion (GSIS)

For GSIS analysis, the isolated islets were cultured overnight before transferring to a 96-well plate containing modified Krebs Ringer phosphate Hepes (KRPH) buffer with 1 mM glucose for 1 h. Different glucose

concentrations (2.8 and 16.8 mM) were added to the islets (1 h for each). The supernatants were collected and used for insulin measurement. At the end, the islets were lysed for DNA measurements. Insulin concentrations were measured and quantified using an ultrasensitive insulin ELISA kit (CristalChem). The analysis was performed using a standard curve, and the data were normalized to the DNA content.

Blood glucose level measurement

Mice were maintained in standard conditions, and were starved 4 h before the measurements. Blood glucose values were determined from venous blood using an automatic glucose monitor (Glucometer Elite, Bayer).

Analysis of single cell RNA-seq data

The following data processing and computational analyses of single cell data were used: preprocessing of droplet-based single cell RNA-seq data, low dimensional embedding, visualization and clustering, marker gene identification and subtype characterization, cell cycle classification, reconstruction of lineage relationships and differentiation trajectories, directionality of endocrine differentiation using RNA velocity estimation and software specifications. See supplementary Materials and Methods for further details of each analysis.

Statistical analysis

Data are presented as mean ± s.d. Two-tailed unpaired t-tests were used to analyze the data.

Acknowledgements

We thank Jessica Jaki, Bianca Vogel, Kerstin Diemer and Julia Beckenbauer for technical support. We thank Maike Sander and Helena Edlund for providing Ngn3 antibodies.

Competing interests

The authors declare no competing or financial interests.

Author contributions

Conceptualization: M.B., H.L.; Methodology: A.B.-P., K.S., M.T.-M., C.S., S.S., I.B., A.B., M.B.; Software: F.J.T., S.T., L.D.; Validation: M.B., H.L., F.J.T., A.B.-P., S.T., L.D.; Investigation: M.B., H.L., A.B.-P., S.T., L.D.; Resources: H.L., F.J.T.; Data curation: M.B., F.T., A.B.-P., S.T., L.D.; Writing - original draft: M.B., H.L., A.B.-P., S.T.; Writing - review & editing: M.B., H.L., F.J.T., A.B.-P., S.T., L.D., K.S., M.T.-M., C.S., S.S., I.B., A.B.; Visualization: M.B., A.B.-P., S.T.; Supervision: M.B., H.L., F.J.T.; Project administration: M.B., H.L.; Funding acquisition: M.B., H.L., F.J.T.

Funding

This work was supported by the Helmholtz-Gemeinschaft (Helmholtz Portfolio Theme 'Metabolic Dysfunction and Common Disease'), Deutsche Forschungsgemeinschaft (DFG) (Collaborative Research Centre 1243, Subproject A17) and Deutsches Zentrum für Diabetesforschung (DZD). This project is supported by DZD NEXT Grant funding. H.J.T. acknowledges financial support by the Chan Zuckerberg Initiative Donor-Advised Fund (DAF) (grant 182835). S.T. is supported by a DFG Fellowship through the Graduate School of Quantitative Biosciences Munich (QBM). L.D. is supported by the Graduate Program of the International Max Planck Research School for Translational Psychiatry (IMPRS-TP).

Data availability

scRNA-seq data have been deposited in Gene Expression Omnibus under accession number GSE132188. All code to reproduce the results from this data as well as fully processed and annotated count matrices are available at <http://github.com/theislab/pancreatic-endocrinogenesis>.

Supplementary information

Supplementary information available online at <http://dev.biologists.org/lookup/doi/10.1242/dev.173849.supplemental>

References

- Ahnfelt-Rønne, J., Hald, J., Bødker, A., Yassin, H., Serup, P. and Hecksher-Sørensen, J. (2007). Preservation of proliferating pancreatic progenitor cells by Delta-Notch signaling in the embryonic chicken pancreas. *BMC Dev. Biol.* **7**, 63. doi:10.1186/1471-213X-7-63
- Apelqvist, A., Li, H., Sommer, L., Beatus, P., Anderson, D. J., Honjo, T., de Angelis, M. H., Lendahl, U. and Edlund, H. (1999). Notch signalling controls pancreatic cell differentiation. *Nature* **400**, 877–881. doi:10.1038/23716

- Arda, H. E., Benitez, C. M. and Kim, S. K.** (2013). Gene regulatory networks governing pancreas development. *Dev. Cell* **25**, 5-13. doi:10.1016/j.devcel.2013.03.016
- Azzarelli, R., Hurley, C., Sznurkowska, M. K., Rulands, S., Hardwick, L., Gamper, I., Ali, F., McCracken, L., Hindley, C., McDuff, F. et al.** (2017). Multi-site Neurogenin3 phosphorylation controls pancreatic endocrine differentiation. *Dev. Cell* **41**, 274-286.e5. doi:10.1016/j.devcel.2017.04.004
- Bakhti, M., Böttcher, A. and Lickert, H.** (2019). Modelling the endocrine pancreas in health and disease. *Nat. Rev. Endocrinol.* **15**, 155-171. doi:10.1038/s41574-018-0132-z
- Bankaitis, E. D., Bechard, M. E. and Wright, C. V. E.** (2015). Feedback control of growth, differentiation, and morphogenesis of pancreatic endocrine progenitors in an epithelial plexus niche. *Genes Dev.* **29**, 2203-2216. doi:10.1101/gad.267914.115
- Bankaitis, E. D., Bechard, M. E., Gu, G., Magnuson, M. A. and Wright, C. V. E.** (2018). ROCK-nmMyoII, Notch and *Neurog3* gene-dosage link epithelial morphogenesis with cell fate in the pancreatic endocrine-progenitor niche. *Development* **145**, dev162115. doi:10.1242/dev.162115
- Bastidas-Ponce, A., Scheibner, K., Lickert, H. and Bakhti, M.** (2017a). Cellular and molecular mechanisms coordinating pancreas development. *Development* **144**, 2873-2888. doi:10.1242/dev.140756
- Bastidas-Ponce, A., Roscioni, S. S., Burtscher, I., Bader, E., Sterr, M., Bakhti, M. and Lickert, H.** (2017b). Foxa2 and Pdx1 cooperatively regulate postnatal maturation of pancreatic β -cells. *Mol. Metab.* **6**, 524-534. doi:10.1016/j.molmet.2017.03.007
- Bechard, M. E., Bankaitis, E. D., Hipkens, S. B., Ustione, A., Piston, D. W., Yang, Y.-P., Magnuson, M. A. and Wright, C. V. E.** (2016). 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. doi:10.1101/gad.284729.116
- Benitez, C. M., Qu, K., Sugiyama, T., Pauerstein, P. T., Liu, Y., Tsai, J., Gu, X., Ghodasara, A., Arda, H. E., Zhang, J. et al.** (2014). An integrated cell purification and genomics strategy reveals multiple regulators of pancreas development. *PLoS Genet.* **10**, e1004645. doi:10.1371/journal.pgen.1004645
- Byrnes, L. E., Wong, D. M., Subramaniam, M., Meyer, N. P., Gilchrist, C. L., Knox, S. M., Tward, A. D., Ye, C. J. and Sneddon, J. B.** (2018). Lineage dynamics of murine pancreatic development at single-cell resolution. *Nat. Commun.* **9**, 3922. doi:10.1038/s41467-018-06176-3
- Collombat, P., Mansouri, A., Hecksher-sørensen, J., Serup, P., Krull, J., Gradwohl, G. and Gruss, P.** (2003). Opposing actions of Arx and Pax4 in endocrine pancreas development. *Genes Dev.* **17**, 2591-2603. doi:10.1101/gad.269003
- Collombat, P., Xu, X., Ravassard, P., Sosa-Pineda, B., Dussaud, S., Billestrup, N., Madsen, O. D., Serup, P., Heimberg, H. and Mansouri, A.** (2009). The Ectopic Expression of Pax4 in the Mouse Pancreas Converts Progenitor Cells into α and Subsequently β Cells. *Cell* **138**, 449-462. doi:10.1016/j.cell.2009.05.035
- Cortijo, C., Gouzi, M., Tissir, F. and Grapin-Botton, A.** (2012). Planar cell polarity controls pancreatic beta cell differentiation and glucose homeostasis. *Cell Rep.* **2**, 1593-1606. doi:10.1016/j.celrep.2012.10.016
- Desgraz, R. and Herrera, P. L.** (2009). Pancreatic neurogenin 3-expressing cells are unipotent islet precursors. *Development* **136**, 3567-3574. doi:10.1242/dev.039214
- Fischer, D. S., Fiedler, A. K., Kernfeld, E. M., Genga, R. M. J., Bastidas-Ponce, A., Bakhti, M., Lickert, H., Hasenauer, J., Maehr, R. and Theis, F. J.** (2019). Inferring population dynamics from single-cell RNA-sequencing time series data. *Nat. Biotechnol.* **37**, 461-468. doi:10.1038/s41587-019-0088-0
- Gasa, R., Mrejen, C., Lynn, F. C., Skewes-Cox, P., Sanchez, L., Yang, K. Y., Lin, C.-H., Gomis, R. and German, M. S.** (2008). Induction of pancreatic islet cell differentiation by the neurogenin-neuroD cascade. *Differentiation* **76**, 381-391. doi:10.1111/j.1432-0436.2007.00228.x
- Gehart, H., van Es, J. H., Hamer, K., Beumer, J., Kretzschmar, K., Dekkers, J. F., Rios, A. and Clevers, H.** (2019). Identification of enteroendocrine regulators by real-time single-cell differentiation mapping. *Cell* **176**, 1158-1173.e16. doi:10.1016/j.cell.2018.12.029
- Gouzi, M., Kim, Y. H., Katsumoto, K., Johansson, K. and Grapin-Botton, A.** (2011). Neurogenin3 initiates stepwise delamination of differentiating endocrine cells during pancreas development. *Dev. Dyn.* **240**, 589-604. doi:10.1002/dvdy.20544
- Gradwohl, G., Dierich, A., LeMeur, M. and Guillemot, F.** (2000). Neurogenin3 is required for the development of the four endocrine cell lineages of the pancreas. *Proc. Natl. Acad. Sci. USA* **97**, 1607-1611. doi:10.1073/pnas.97.4.1607
- Gu, G., Dubauskaite, J. and Melton, D. A.** (2002). Direct evidence for the pancreatic lineage: NGN3+ cells are islet progenitors and are distinct from duct progenitors. *Development* **129**, 2447-2457.
- Haghverdi, L., Büttner, M., Wolf, F. A., Buettner, F. and Theis, F. J.** (2016). Diffusion pseudotime robustly reconstructs lineage branching. *Nat. Methods* **13**, 845-848. doi:10.1038/nmeth.3971
- Johansson, K. A., Dursun, U., Jordan, N., Gu, G., Beermann, F., Gradwohl, G. and Grapin-Botton, A.** (2007). Temporal control of Neurogenin3 activity in pancreas progenitors reveals competence windows for the generation of different endocrine cell types. *Dev. Cell* **12**, 457-465. doi:10.1016/j.devcel.2007.02.010
- Kakugawa, S., Langton, P. F., Zebisch, M., Howell, S. A., Chang, T.-H., Liu, Y., Feizi, T., Bineva, G., O'Reilly, N., Snijders, A. P. et al.** (2015). Notum deacylates Wnt proteins to suppress signalling activity. *Nature* **519**, 187-192. doi:10.1038/nature14259
- Kim, Y. H., Larsen, H. L., Rué, P., Lemaire, L. A., Ferrer, J. and Grapin-Botton, A.** (2015). Cell cycle-dependent differentiation dynamics balances growth and endocrine differentiation in the pancreas. *PLoS Biol.* **13**, e1002111. doi:10.1371/journal.pbio.1002111
- Kopp, J. L., Dubois, C. L., Schaffer, A. E., Hao, E., Shih, H. P., Seymour, P. A., Ma, J. and Sander, M.** (2011). Sox9+ ductal cells are multipotent progenitors throughout development but do not produce new endocrine cells in the normal or injured adult pancreas. *Development* **138**, 653-665. doi:10.1242/dev.056499
- Krentz, N. A. J., van Hoof, D., Li, Z., Watanabe, A., Tang, M., Nian, C., German, M. S. and Lynn, F. C.** (2017). Phosphorylation of NEUROG3 links endocrine differentiation to the cell cycle in pancreatic progenitors. *Dev. Cell* **41**, 129-142. doi:10.1016/j.devcel.2017.02.006
- Krentz, N. A. J., Xu, E. E., Lynn, F. C., Sasaki, S., Lee, M. Y. Y., Maslova, A. and Sprout, S. L. J.** (2018). Single-cell transcriptome profiling of mouse and hESC-derived pancreatic progenitors. *Stem Cell Rep.* **11**, 1551-1564. doi:10.1016/j.stemcr.2018.11.008
- La Manno, G., Soldatov, R., Zeisel, A., Braun, E., Hochgerner, H., Petukhov, V., Lidschreiber, K., Kastriti, M. E., Lönnberg, P., Furlan, A. et al.** (2018). RNA velocity of single cells. *Nature* **560**, 494-498. doi:10.1038/s41586-018-0414-6
- Lee, C. S., Perreault, N., Brestelli, J. E. and Kaestner, K. H.** (2002). 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. doi:10.1101/gad.985002
- Lee, K. M., Yasuda, H., Hollingsworth, M. A. and Ouellette, M. M.** (2005). Notch2-positive progenitors with the intrinsic ability to give rise to pancreatic ductal cells. *Lab. Investig.* **85**, 1003-1012. doi:10.1038/labinvest.3700298
- Lemaire, L. A., Gouley, J., Kim, Y. H., Carat, S., Jacquemin, P., Rougemont, J., Constam, D. B. and Grapin-Botton, A.** (2016). Bicaudal C1 promotes pancreatic NEUROG3+ endocrine progenitor differentiation and ductal morphogenesis. *Development* **142**, 858-870. doi:10.1242/dev.114611
- Liu, W.-D., Wang, H.-W., Muguira, M., Breslin, M. B. and Lan, M. S.** (2006). INSM1 functions as a transcriptional repressor of the neuroD β 2 gene through the recruitment of cyclin D1 and histone deacetylases. *Biochem. J.* **397**, 169-177. doi:10.1042/BJ20051669
- Liu, J., Banerjee, A., Herring, C. A., Attalla, J., Hu, R., Xu, Y., Shao, Q., Simmons, A. J., Dadi, P. K., Wang, S. et al.** (2019). Neurog3-independent methylation is the earliest detectable mark distinguishing pancreatic progenitor identity. *Dev. Cell* **48**, 49-63.e7. doi:10.1016/j.devcel.2018.11.048
- Löf-Öhlin, Z. M., Nyeng, P., Bechard, M. E., Hess, K., Bankaitis, E., Greiner, T. U., Ameri, J., Wright, C. V. and Semb, H.** (2017). EGFR signalling controls cellular fate and pancreatic organogenesis by regulating apicobasal polarity. *Nat. Cell Biol.* **19**, 1313-1325. doi:10.1038/ncb3628
- Mastrolia, V., Flucher, S. M., Obermaier, G. J., Drach, M., Hofer, H., Renström, E., Schwartz, A., Striessnig, J., Flucher, B. E. and Tuluc, P.** (2017). Loss of c2d-1 calcium channel subunit function increases the susceptibility for diabetes. *Diabetes* **66**, 897-907. doi:10.2337/db16-0336
- Mellitzer, G., Martín, M., Sidhoum-Jenny, M., Orvain, C., Barths, J., Seymour, P. A., Sander, M. and Gradwohl, G.** (2004). Pancreatic islet progenitor cells in neurogenin 3-yellow fluorescent protein knock-add-on mice. *Mol. Endocrinol.* **18**, 2765-2776. doi:10.1210/me.2004-0243
- Miyatsuka, T., Kosaka, Y., Kim, H. and German, M. S.** (2011). Neurogenin3 inhibits proliferation in endocrine progenitors by inducing Cdkn1a. *Proc. Natl. Acad. Sci. USA* **108**, 185-190. doi:10.1073/pnas.1004842108
- Nyeng, P., Heilmann, S., Löf-Öhlin, Z. M., Pettersson, N. F., Hermann, F. M., Reynolds, A. B. and Semb, H.** (2019). p120ctn-mediated organ patterning precedes and determines pancreatic progenitor fate. *Dev. Cell* **49**, 31-47.e9. doi:10.1016/j.devcel.2019.02.005
- Osipovich, A. B., Long, Q., Manduchi, E., Gangula, R., Hipkens, S. B., Schneider, J., Okubo, T., Stoeckert, C. J., Takada, S. and Magnuson, M. A.** (2014). Insm1 promotes endocrine cell differentiation by modulating the expression of a network of genes that includes Neurog3 and Ripply3. *Development* **141**, 2939-2949. doi:10.1242/dev.104810
- Pan, F. C. and Wright, C.** (2011). Pancreas organogenesis: from bud to plexus to gland. *Dev. Dyn.* **240**, 530-565. doi:10.1002/dvdy.22584
- Pan, F. C., Bankaitis, E. D., Boyer, D., Xu, X., Van de Casteele, M., Magnuson, M. A., Heimberg, H. and Wright, C. V. E.** (2013). Spatiotemporal patterns of multipotentiality in Pit1a-expressing cells during pancreas organogenesis and injury-induced facultative restoration. *Development* **140**, 751-764. doi:10.1242/dev.090159
- Petrezselyova, S., Kinsky, S., Truban, D., Sedlacek, R., Burtscher, I. and Lickert, H.** (2015). Homology arms of targeting vectors for gene insertions and CRISPR/Cas9 technology: Size does not matter; Quality control of targeted clones does. *Cell. Mol. Biol. Lett.* **20**, 773-787. doi:10.1515/cmb-2015-0047
- Piccolo, S., Agius, E., Leyns, L., Bhattacharyya, S., Grunz, H., Bouwmeester, T. and De Robertis, E. M.** (1999). The head inducer cerberus is a multifunctional

- antagonist of Nodal, BMP and Wnt signals. *Nature* **397**, 707-710. doi:10.1038/17820
- Ramond, C., Beydag-Tasöz, B. S., Azad, A., van de Bunt, M., Petersen, M. B. K., Beer, N. L., Glaser, N., Berthault, C., Gloyn, A. L., Hansson, M. et al.** (2018). Understanding human fetal pancreas development using subpopulation sorting, RNA sequencing and single-cell profiling. *Development* **145**, dev165480. doi:10.1242/dev.165480
- Scavuzzo, M. A., Hill, M. C., Chmielowiec, J., Yang, D., Teaw, J., Sheng, K., Kong, Y., Bettini, M., Zong, C., Martin, J. F. et al.** (2018). Endocrine lineage biases arise in temporally distinct endocrine progenitors during pancreatic morphogenesis. *Nat. Commun.* **9**, 3356. doi:10.1038/s41467-018-05740-1
- Schiebinger, G., Shu, J., Tabaka, M., Cleary, B., Subramanian, V., Solomon, A., Gould, J., Liu, S., Lin, S., Berube, P. et al.** (2019). Optimal-transport analysis of single-cell gene expression identifies developmental trajectories in reprogramming. *Cell* **176**, 928-943. doi:10.1016/j.cell.2019.01.006
- Seymour, P. A., Freude, K. K., Tran, M. N., Mayes, E. E., Jensen, J., Kist, R., Scherer, G. and Sander, M.** (2007). SOX9 is required for maintenance of the pancreatic progenitor cell pool. *Proc. Natl. Acad. Sci. USA* **104**, 1865-1870. doi:10.1073/pnas.0609217104
- Sharon, N., Chawla, R., Mueller, J., Vanderhoof, J., Whitehorn, L. J., Rosenthal, B., Gürler, M., Estanboulieh, R. R., Shvartsman, D., Gifford, D. K. et al.** (2019). A peninsular structure coordinates asynchronous differentiation with morphogenesis to generate pancreatic islets. *Cell* **176**, 790-804.e13. doi:10.1016/j.cell.2018.12.003
- Shih, H. P., Kopp, J. L., Sandhu, M., Dubois, C. L., Seymour, P. A., Grapin-Botton, A. and Sander, M.** (2012). A Notch-dependent molecular circuitry initiates pancreatic endocrine and ductal cell differentiation. *Development* **139**, 2488-2499. doi:10.1242/dev.078634
- Solar, M., Cardalda, C., Houbracken, I., Martín, M., Maestro, M. A., De Medts, N., Xu, X., Grau, V., Heimberg, H., Bouwens, L. et al.** (2009). Pancreatic exocrine duct cells give rise to insulin-producing beta cells during embryogenesis but not after birth. *Dev. Cell* **17**, 849-860. doi:10.1016/j.devcel.2009.11.003
- Stanescu, D. E., Yu, R., Won, K.-J. and Stoffers, D. A.** (2016). Single cell transcriptomic profiling of mouse pancreatic progenitors. *Physiol. Genomics* **49**, 105-114. doi:10.1152/physiolgenomics.00114.2016
- Stauber, M., Weidemann, M., Dittrich-Breiholz, O., Lobschat, K., Alten, L., Mai, M., Beckers, A., Kracht, M. and Gossler, A.** (2017). Identification of FOXJ1 effectors during ciliogenesis in the foetal respiratory epithelium and embryonic left-right organiser of the mouse. *Dev. Biol.* **423**, 170-188. doi:10.1016/j.ydbio.2016.11.019
- Svensson, P., Bergqvist, I., Norlin, S. and Edlund, H.** (2009). MFng is dispensable for mouse pancreas development and function. *Mol. Cell. Biol.* **29**, 2129-2138. doi:10.1128/MCB.01644-08
- Sznurkowska, M. K., Hannezo, E., Azzarelli, R., Rulands, S., Nestorowa, S., Hindley, C. J., Nichols, J., Göttgens, B., Huch, M., Philpott, A. et al.** (2018). Defining lineage potential and fate behavior of precursors during pancreas development. *Dev. Cell* **46**, 360-375.e5. doi:10.1016/j.devcel.2018.06.028
- Wang, S., Yan, J., Anderson, D. A., Xu, Y., Kanal, M. C., Cao, Z., Wright, C. V. E. and Gu, G.** (2010). Neurog3 gene dosage regulates allocation of endocrine and exocrine cell fates in the developing mouse pancreas. *Dev. Biol.* **339**, 26-37. doi:10.1016/j.ydbio.2009.12.009
- Wolf, F. A., Hamey, F. K., Plass, M., Solana, J., Dahlin, J. S., Göttgens, B., Rajewsky, N., Simon, L. and Theis, F. J.** (2019). PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. *Genome Biol.* **20**, 59. doi:10.1186/s13059-019-1663-x
- Xu, E. E., Krentz, N. A. J., Tan, S., Chow, S. Z., Tang, M., Nian, C. and Lynn, F. C.** (2015). SOX4 cooperates with neurogenin 3 to regulate endocrine pancreas formation in mouse models. *Diabetologia* **58**, 1013-1023. doi:10.1007/s00125-015-3507-x
- Yu, X.-X., Qiu, W.-L., Yang, L., Li, L.-C., Zhang, Y.-W. and Xu, C.-R.** (2018). Dynamics of chromatin marks and the role of JMJD3 during pancreatic endocrine cell fate commitment. *Development* **145**, dev163162. doi:10.1242/dev.163162
- Yu, X.-X., Qiu, W.-L., Yang, L., Zhang, Y., He, M.-Y., Li, L.-C. and Xu, C.-R.** (2019). Defining multistep cell fate decision pathways during pancreatic development at single-cell resolution. *EMBO J.* **38**, e100164. doi:10.15252/embj.2018100164
- Zhou, Q., Law, A. C., Rajagopal, J., Anderson, W. J., Gray, P. A. and Melton, D. A.** (2007). A multipotent progenitor domain guides pancreatic organogenesis. *Dev. Cell* **13**, 103-114. doi:10.1016/j.devcel.2007.06.001

Supplementary materials and methods

Quantitative PCR (qPCR) probes

The used qPCR probes were from Life Technologies and listed as follow; Mm00440465_g1 for Neurod2, Mm00435565_m1 for Pdx1, Mm01950294_s1 for Insulin 1, Mm00731595_Gh for Insulin 2, Mm01269055_m1 for Glucagon, Mm00436671_m1 for Somatostatin, Mm01946604_s1 for Neurod1, Mm00545903_m1 for Arx, Mm01159036_m1 for Pax4, Mm00450550_m1 for Sult2b1, Mm01208248_m1 for Upk3bl, Mm01269371_m1 for Shf, Mm03053598_s1 for Mycl, Mm00459097_m1 for Steap1, Mm00803317_m1 for Rgs16, Mm04239638_m1 for Gm8773 and Mm00458153_g1 for Gng13.

Antibody list for immunohistochemistry

The following primary antibodies were used for staining: rabbit anti-Sox9 (Millipore, Cat. Nr: AB5535, Batch Nr: 3107073, dilution 1:300) (Carrasco et al., 2012), goat anti-Nkx6.1 (R&D systems, Cat. Nr: AF5857, dilution 1:300) (Bastidas-Ponce et al., 2017), rat anti-E-Cadherin (DECMA-1, a generous gift from Kremmer, dilution 1:500) (Bastidas-Ponce et al., 2017), rat anti-EpCAM (DSHB Hybridoma, Cat. Nr: 4G1, dilution 1:300), goat anti-Pancreatic polypeptide (Novus, Cat. Nr: NB100-1793, 1:300), hamster anti-Mucin-1(Invitrogen, Cat. Nr: HM1630P0, Batch Nr: 022017, dilution 1:100), guinea pig anti-glucagon (TAKARA, Cat. Nr: M182, dilution 1:3500) (Bastidas-Ponce et al., 2017), chicken anti-GFP (to detect Venus; Aves Labs, Cat. Nr: 1020, dilution 1:1000) (Bastidas-Ponce et al., 2017), rabbit anti-insulin (Cell signaling, Cat. Nr: 3014, dilution 1:300) (Bastidas-Ponce et al., 2017), rabbit anti-Ngn3 (a generous gift from Helena Edlund, dilution 1:800), guinea pig anti-Ngn3 (a generous gift from Maike Sander, dilution 1:1000) and rat anti-somatostatin (Invitrogen, Cat. Nr: MA5-16987, Batch Nr: 75739969, dilution 1:300). The secondary antibodies used were from Dianova and Invitrogen, (dilution 1:800) and Phalloidin Alexa Fluor 546 was from Invitrogen (Cat. Nr: A22283, dilution 1:200).

Islet isolation

The isolation of islets was performed using collagenase P (Roche) digestion of adult pancreas. First, 3 mL of collagenase P (1 mg/mL) was injected into the bile duct, then the perfused pancreas was dissected and placed into other 3 mL collagenase P for 15 min at 37 °C. To stop the digestion, 10 mL of G-solution (HBSS + 1% BSA) were added followed by centrifugation at 1600 rpm (421 g) at 4 °C. After another washing step with G-solution, the pellets were re-suspended in 5.5 mL of gradient solution (5 mL 10% RPMI + 3 mL 40% Optiprep/per sample), and placed on top of 2.5 mL of the same solution. To form a 3-layers gradient, 6 mL of G-solution were added on top of the abovementioned layers. Samples were then centrifuged at 1700 rpm (475 g). Finally, the interphase between the upper and the middle

layer of the gradient was recovered and filtered through a 70 µm Nylon filter, washed with G-solution and the islets were handpicked under the microscope.

Analysis of single cell RNAseq data

Preprocessing of droplet-based single-cell RNAseq data

Demultiplexing of raw base call (BCL) files, alignment, read filtering, barcode and UMI counting were performed using the CellRanger analysis pipeline (Version 2.1.1) provided by 10X Genomics. High quality barcodes were selected based on the overall UMI distribution using the standard CellRanger cell detection algorithm. All further analyses were run using the python-based Scanpy API (Wolf et al., 2018) except stated otherwise (for software specifications and code availability see below and in Materials and Methods). To further remove low quality cells, we filtered cells with a high fraction of counts from mitochondrial genes (20% or more) indicating stressed or dying cells, and cells expressing less than 1200 genes. In addition, genes with expression in less than 20 cells were excluded. Cell by gene count matrices of all samples were then concatenated to a single matrix and values log transformed. To account for differences in sequencing depth or cell size UMI counts were normalized using quantile normalization as previously described (Weinreb et al., 2018). Briefly, each cell is normalized by the total UMI count in the cell of genes that account for less than 5% of the total UMI counts across all cells. The top 4000 variable genes were selected based on normalized dispersion as described in (Zheng et al., 2017). This output matrix was input to all further analyses except for differential expression testing where all genes were used.

Low dimensional embedding, visualization and clustering

A single-cell neighborhood graph was computed on the 50 first principal components that sufficiently explain the variation in the data using 15 nearest neighbors. Uniform Manifold Approximation and Projection (UMAP) (McInnes and Healy, 2018) was run for visualization. For clustering and cell type identification louvain-based clustering (Blondel et al., 2008) at varying resolution in different parts of the data manifold was used as implemented in louvain-igraph (<https://github.com/vtraag/louvain-igraph>) and adopted by Scanpy. Cell types were annotated based on the expression of known marker genes. Clusters were merged if only reflecting further heterogeneity within a cell type not discussed in this manuscript. For the exact steps of clustering and annotation and the parameters used consult the available code. We removed clearly distinct clusters of cells highly expressing immune cell or mesenchymal marker genes, respectively. In addition, we found a major cluster and a small endocrine subcluster expressing a mixture of gene expression profiles of other cell types but no specific marker genes suggesting that these cells are doublets. We further checked this hypothesis and calculated a doublet score signifying the probability of being a doublet for each cell using the doublet detection package Scrublet (Wolock et al., 2019) using *scr.compute_doublet_scores* function with parameters as recommended in the Scrublet tutorial (min_counts = 2, min_cells = 3, vscore_percentile = 85, n_pc = 50, expected_doublet_rate = 0.08, sim_doublet_ratio = 3, n_neighbors = 15). Indeed, the majority of

these cells were predicted to be doublets even when setting a high threshold for the doublet score and we therefore removed these two clusters from further analyses. In all other subtypes we found only a small proportion of predicted doublet cells that did not form a separate cluster or misleadingly connect two clusters. We did not remove these cells as downstream analyses were not strongly influenced by them. In different figures varying resolutions of subtype clustering are shown. EP, Pre-endocrine and α -cell subtypes were annotated based on marker gene expression as well as the developmental stages they existed. For the specification of $Ngn3^{low}$ progenitors we referred to the expression of *Neurog3* and picked all cells with expression > 0 in the multipotent, tip, trunk and ductal cluster. Proliferating subtypes were annotated using the cell cycle classification described below.

Marker gene identification and subtype characterization

Characteristic gene signatures were identified by testing for differential expression of a subgroup against all other cells or between two subgroups as outlined in the text using a t-test with overestimated variance implemented in the *tl.rank_genes_groups* function of Scanpy. Testing was performed on the log-transformed quantile normalized data to account for confounding differences in sequencing depth between samples. It has been shown that a simple t-test performs comparable to differential expression tools adopting more complex models in a simple set-up as here (no additional confounding factors) (Soneson and Robinson, 2018). Gene set enrichment was performed with the gseapy (v0.9.3) implementation of EnrichR. To identify genes enriched in EPs, $Ngn3^+$ cells ($Ngn3^{low}$ progenitors and $Ngn3^{high}$ precursors) were compared against all other cells except for acinar cells. Top 250 upregulated genes were considered (Table S2). From this list, EP signature genes that are transiently expressed in EPs similar to Ngn3 but not or only lowly expressed in other populations were extracted by computing the mean expression of all 250 genes in each cell type after scaling the data from 0 to 1. All genes that showed a mean expression > 0.04 in $Ngn3^{high}$ precursors, < 0.04 in all other non-endocrine progenitor-precursor clusters and < 0.18 in Fev^{high} cells were defined as EP signature genes. Expression thresholds were chosen based on the mean expression levels of *Neurog3*. For characterization of the EP subtypes we compared ductal cells against $Ngn3^{low}$ progenitors, $Ngn3^{low}$ progenitors against ductal cells, $Ngn3^{high}$ precursors against $Ngn3^{low}$ progenitors and Fev^{high} against $Ngn3^{high}$ precursors to identify genes changing along the differentiation trajectory.

Cell cycle classification

To classify cells into cell cycle phases we used a cell scoring function as described by (Satija et al., 2015) and implemented in the *tl.score_genes_cell_cycle* function in Scanpy with default parameters. Shortly, the score is the average expression of the gene set subtracted with the average expression of a randomly sampled background set with expression values within the same range. Cell cycle genes as defined in (Tirosh et al., 2016) were used to distinguish between S, G2/M and G1 phase cells.

Reconstruction of lineage relationships and differentiation trajectories

To infer lineage relationships between clusters and predict potential routes of differentiation partition-based graph abstraction (PAGA) was performed (Wolf et al., 2019) using the *tl.paga* function of Scanpy with a threshold on edge significance of 0.11 for all subtypes and 0.1 for the endocrine lineage. Edge weights represent the confidence of a connection calculated based on a measure for connectivity. Paths in the PAGA graph signify cluster relationships indicating potential differentiation paths. To infer a pseudotemporal ordering of the cells along the predicted routes in the PAGA graph, diffusion pseudotime (dpt) (Haghverdi et al., 2016) was used as implemented in Scanpy (*tl.dpt*) setting a root cell within the starting population. Dpt is designed to learn such continuous cellular differentiation trajectories and projects the differentiating cells to a one-dimensional developmental trajectory. Before dpt inference the single cell neighborhood-graph and diffusion map was recalculated for the represented subset of cells. To test whether cell cycle effects mask other biological processes and affect the pseudotemporal ordering, dpt was recomputed after linearly regressing out S and G2M cell scores (using *sc.pp.regress_out*; data not shown). Removal of cell cycle effects did not considerably change the ordering of the cells. Also, the expression of genes described in the manuscript figures was not affected except for, as expected, genes involved in DNA replication and cell cycle.

Directionality of endocrine differentiation using RNA velocity estimation

To infer directionality of endocrine differentiation we performed RNA velocity estimation (La Manno et al., 2018) implemented as a stochastic version in the scVelo python package (<https://github.com/theislab/scvelo>). We extracted spliced and unspliced reads using the velocyto pipeline (<http://velocyto.org>), which generated a loompy file. The file was read into an AnnData object for downstream analysis with Scanpy. All analysis related to RNA velocity such as estimating velocities and obtaining the RNA velocity projection were done using scVelo. Here, we followed the recommended procedures described in scVelo, consisting of four main steps: preprocessing, computing first- and second-order moments, estimating velocities, and constructing a velocity graph. That is, we filtered genes with less than 10 spliced and 10 unspliced counts and normalized unspliced and spliced reads by their initial counts and log transformed the data. Then, for each stage separately, we recomputed the neighborhood graph on the represented cell subset using 15 neighbors and the 50 first PCs that sufficiently explain the variation in the data, and the UMAP embedding. Next, first-order and second-order moments for each cell were computed across its nearest neighbors. Velocities were estimated using a stochastic model of transcriptional dynamics, whereby we used a 95% quantile fit for more conservative velocity estimates. Finally, the velocity graph was obtained by computing the correlations between potential cell transitions and the predicted cell state change given by the velocity vector. Only genes with an $r^2 > 0.1$ were considered. The graph was then used to project the velocities into the low dimensional UMAP embedding.

Software specifications

All analyses from UMI count matrices were run with python 3 with the Scanpy API v.1.3.2 to v.1.3.4 and anndata v.0.6.11 (<https://github.com/theislab/scanpy>). Versions of packages required by Scanpy that might influence numerical results are indicated in the custom scripts. Unspliced and spliced reads were extracted using velocyto v0.17.7 and loompy v2.0.12 (<http://velocyto.org>). Velocities were calculated with scVelo v0.1.16 (<https://github.com/theislab/scvelo>). All figures were plotted with matplotlib and seaborn and data was exported to excel-sheets with xlsxwriter. Custom scripts of code for all analyses will be made available upon publication.

Supplementary references

- Bastidas-Ponce, A., Roscioni, S. S., Burtscher, I., Bader, E., Sterr, M., Bakhti, M. and Lickert, H.** (2017). Foxa2 and Pdx1 cooperatively regulate postnatal maturation of pancreatic β -cells. *Mol. Metab.* **6**, 524–534.
- Blondel, V. D., Guillaume, J. L., Lambiotte, R. and Lefebvre, E.** (2008). Fast unfolding of communities in large networks. *J. Stat. Mech. Theory Exp.* **10**, P10008.
- Carrasco, M., Delgado, I., Soria, B., Martín, F. and Rojas, A.** (2012). GATA4 and GATA6 control mouse pancreas organogenesis. *J. Clin. Invest.* **122**, 3504–3515.
- Haghverdi, L., Büttner, M., Wolf, F. A., Buettner, F. and Theis, F. J.** (2016). Diffusion pseudotime robustly reconstructs lineage branching. *Nat. Methods* **13**, 845–8.
- La Manno, G., Soldatov, R., Zeisel, A., Braun, E., Hochgerner, H., Petukhov, V., Lidschreiber, K., Kastriti, M. E., Lönnberg, P., Furlan, A., et al.** (2018). RNA velocity of single cells. *Nature* **560**, 494–498.
- McInnes, L. and Healy, J.** (2018). UMAP : Uniform Manifold Approximation and Projection for Dimension Reduction. *arXiv:1802.03426* 1–18.
- Satija, R., Farrell, J. A., Gennert, D., Schier, A. F. and Regev, A.** (2015). Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol.* **33**, 495–502.
- Soneson, C. and Robinson, M. D.** (2018). Bias, robustness and scalability in single-cell differential expression analysis. *Nat. Methods* **15**, 255–261.
- Tirosh, I., Izar, B., Prakadan, S. M., Wadsworth, M. H., Treacy, D., Trombetta, J. J., Rotem, A., Rodman, C., Lian, C., Murphy, G., et al.** (2016). Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science (80-.).* **352**, 189–96.
- Weinreb, C., Wolock, S. and Klein, A. M.** (2018). SPRING: A kinetic interface for visualizing high dimensional single-cell expression data. *Bioinformatics* **34**, 1246–1248.
- Wolf, F. A., Angerer, P. and Theis, F. J.** (2018). SCANPY: Large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15.
- Wolf, F. A., Hamey, F., Plass, M., Solana, J., Dahlin, J. S., Gottgens, B., Rajewsky, N., Simon, L. and Theis, F. J.** (2019). PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. *Genome Biol.* **20**, 59.
- Wolock, S. L., Lopez, R. and Klein, A. M.** (2019). Scrublet: Computational Identification of Cell Doublets in Single-Cell Transcriptomic Data. *Cell Syst.* **8**, 281–291.e9.
- Zheng, G. X. Y., Terry, J. M., Belgrader, P., Ryvkin, P., Bent, Z. W., Wilson, R., Ziraldo, S. B., Wheeler, T. D., McDermott, G. P., Zhu, J., et al.** (2017). Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* **8**, 14049.

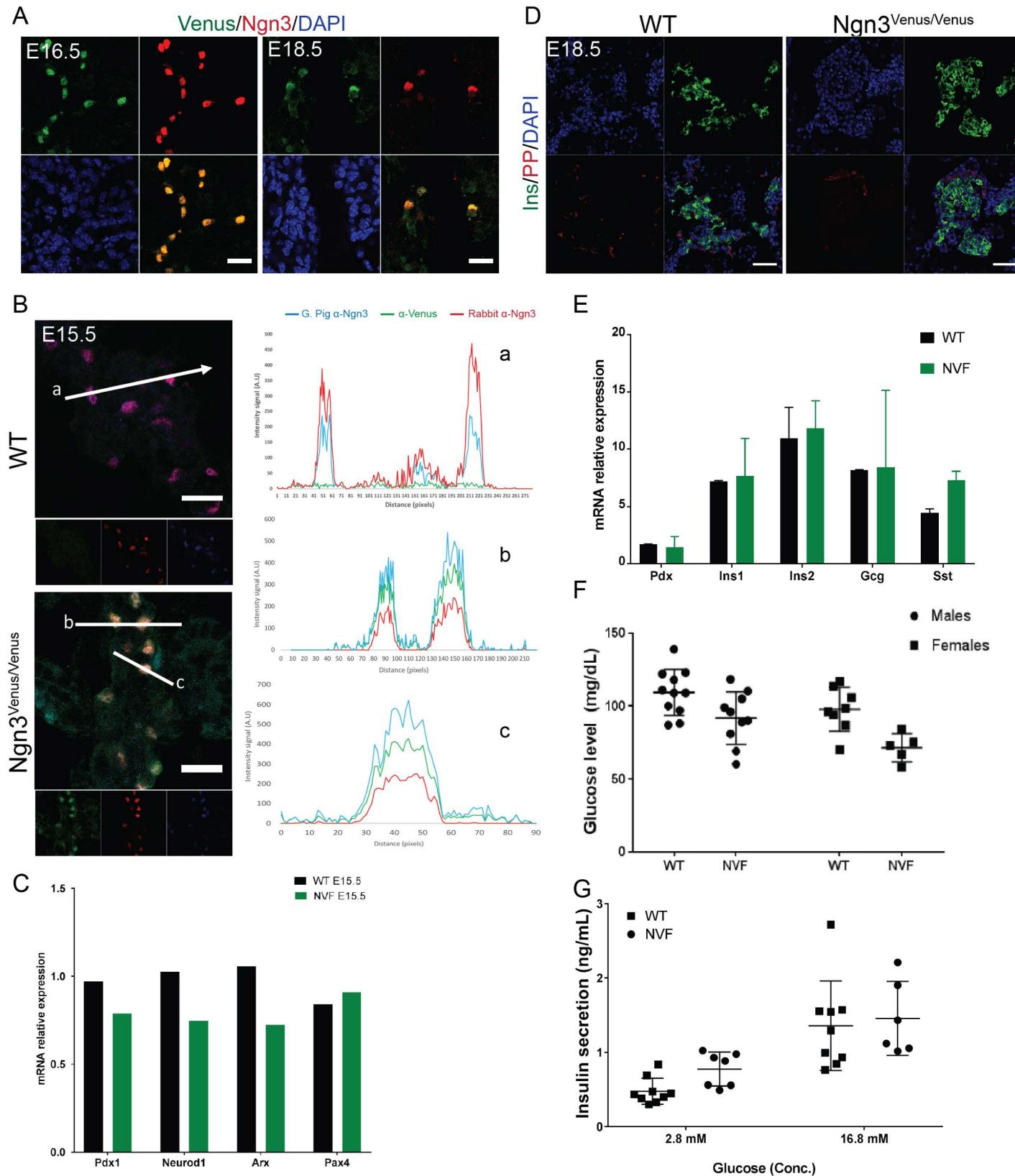
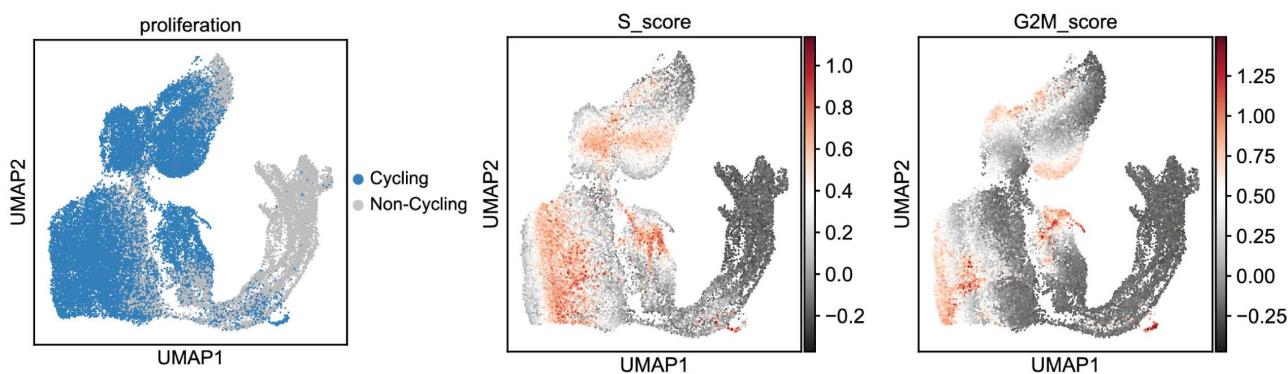
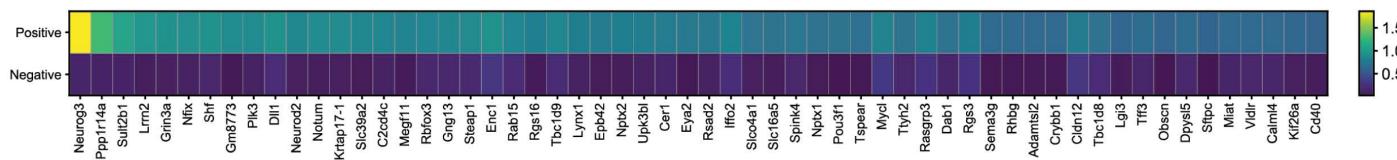


Figure S1. Characterization of endocrine cell formation and function in NVF mice. (A) Immunohistochemical analysis shows similar spatio-temporal expression pattern of NVF and endogenous Ngn3 expression at E16.5 and E18.5 heterozygous NVF pancreatic section. Scale bar, 20 μm . (B) Plot profile of Venus staining together with two different antibodies against endogenous Ngn3 in WT and homozygous NVF pancreatic sections. No Venus signal was detected in the WT sample. Scale bar, 20 μm . (C) qPCR analysis of isolated cells from E15.5 WT and homozygous NVF pancreata for expression of Ngn3 target genes (N= 1 experiment derived from 8-10 embryos). (D) Staining of PP-cells in E18.5 pancreatic sections from WT and homozygous NVF mice show no significant difference in the formation of these cells. Scale bar, 50 μm . (E) qPCR analysis of key pancreatic hormones from isolated islets from 3-month-old WT and homozygous NVF mice (N = 3). (F) Normoglycemic levels in 3-month-old homozygous NVF mice compared to WT. (G) Glucose-stimulated insulin secretion test shows no differences in insulin secretion from isolated islets from 3-month-old homozygous NVF and WT mice.

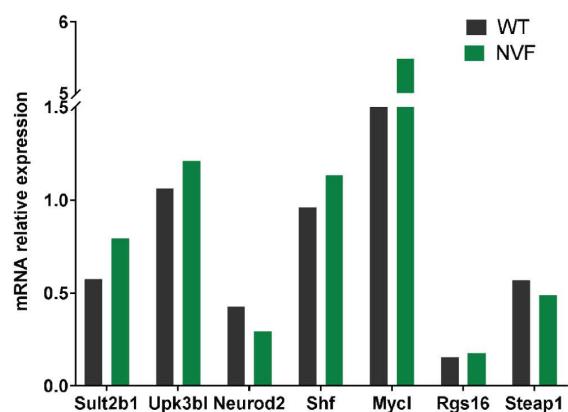
A



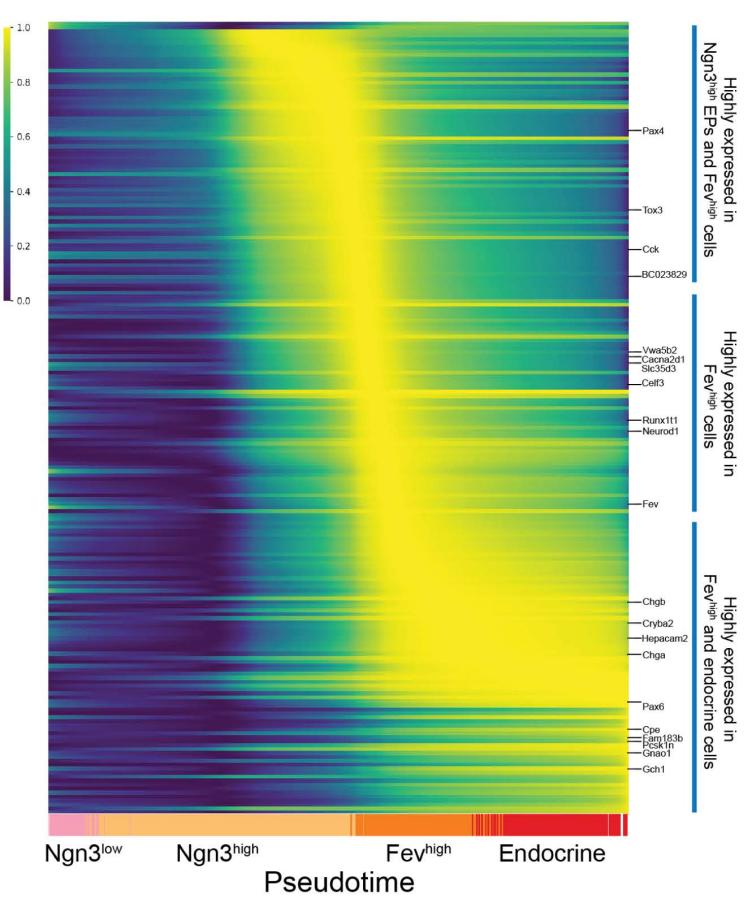
B



C



D



E

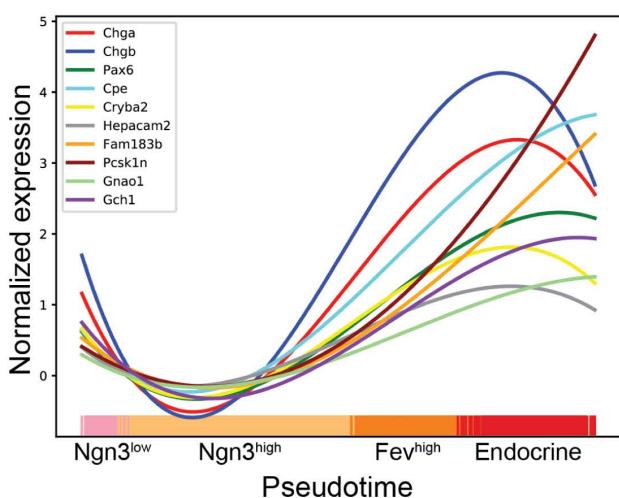


Figure S2. Cell cycle classification and *Fev*^{high} cell-enriched genes. (A) UMAP plots of proliferative status of all pancreatic epithelial populations. Classification into cycling and non-cycling cells (left) based on cell score levels for cell cycles phases S and G2/M calculated based on the expression of a set of genes involved in cell cycle progression (right). (B) Heatmap showing the mean expression in *Ngn3*⁺ and *Ngn3*⁻ cells of the 58 EP-signature genes. Expression values are normalized and scaled to unit variance. (C) qPCR analysis of E15.5 isolated pancreatic epithelial cells indicates no significant changes in the expression levels of several EP-signature genes in WT and homozygous NVF mice (N= 1 experiment derived from 8-10 embryos). (D) Gene expression along pseudotime from *Ngn3*^{low} progenitors to endocrine cells. The top 200 genes enriched in *Fev*^{high} cells are shown (Table S3) and selected genes highlighted. Expression is normalized and approximated by polynominal regression fits along pseudotime. Fitted values of each gene are then scaled to the range between 0 and 1. Cluster membership of the cells is indicated at the bottom. (E) Gene expression along pseudotime of selected genes starting expression in *Fev*^{high} cells and continuously expressed in differentiated endocrine cells. Lines are polynominal regression fits of normalized data. Cluster membership of the cells is indicated at the bottom.

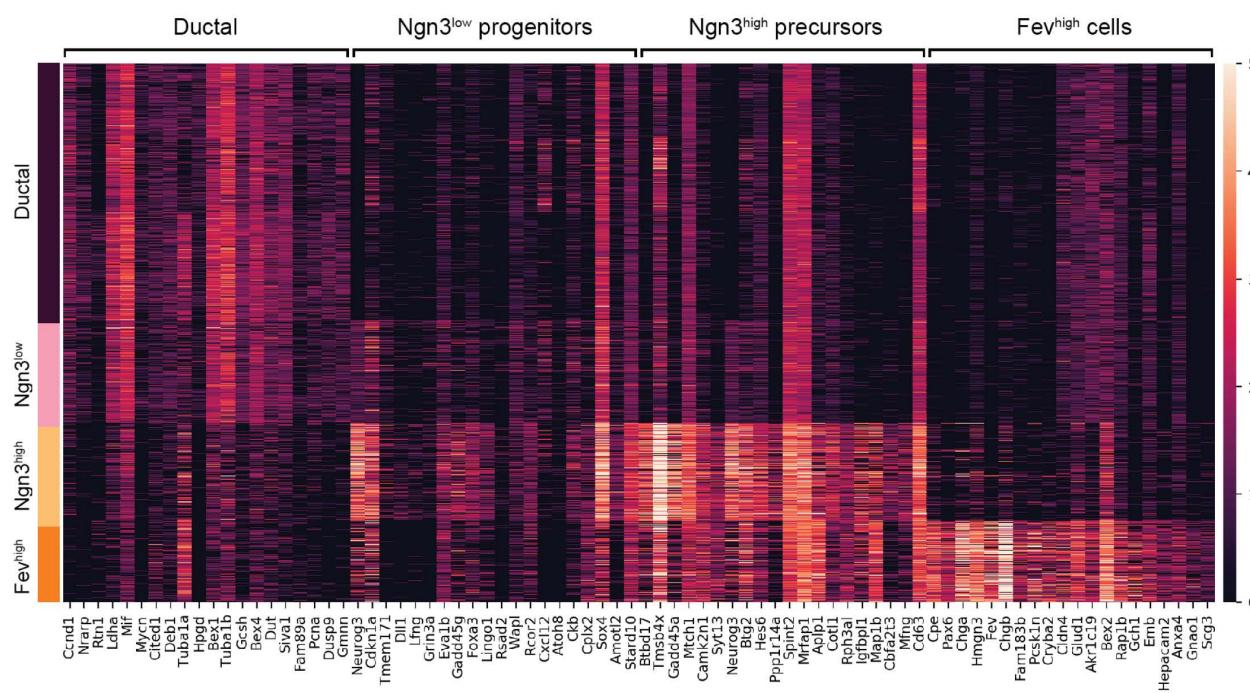
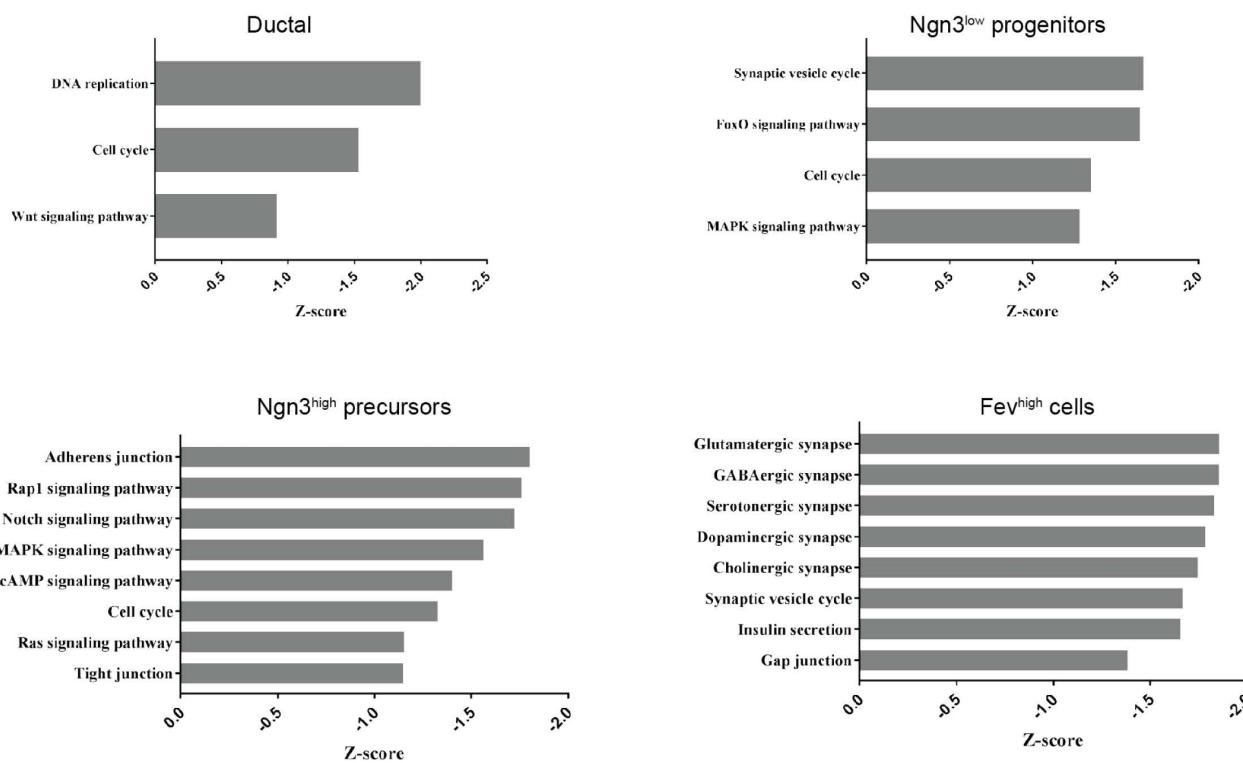
A**B**

Figure S3. Pathway analysis of EPs and *Fev^{high}* cells. (A) Heatmap showing the top 20 genes of subtype-specific expression profiles of ductal, *Ngn3^{low}* progenitor, *Ngn3^{high}* precursor and *Fev^{high}* cells. Normalized expression values are shown. (B) GO-term enrichment analysis of gene sets specific to ductal, *Ngn3^{low}* progenitors, *Ngn3^{high}* precursors and *Fev^{high}* populations.

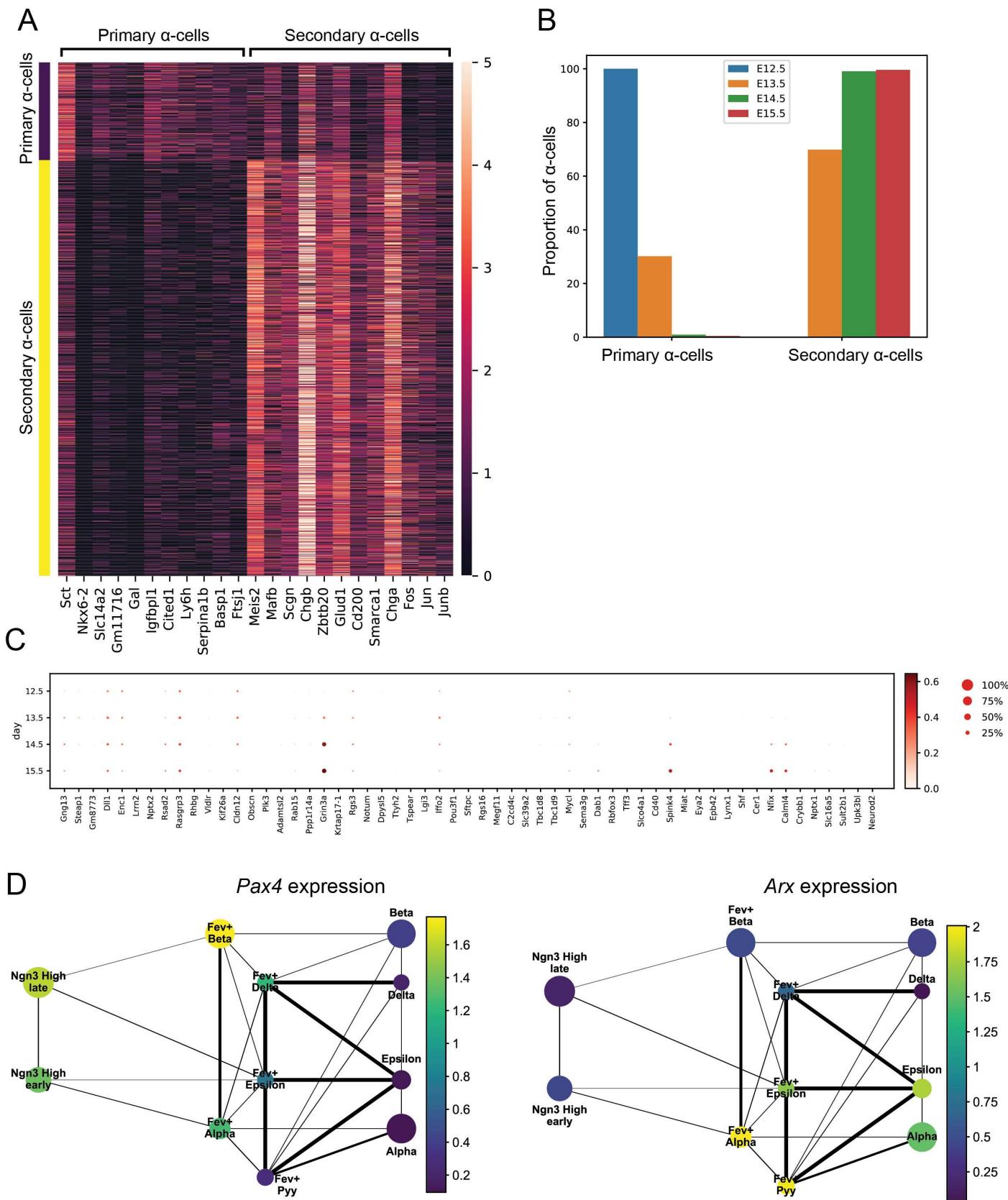


Figure S4. Early and late α -cells are transcriptionally heterogeneous. (A) Expression of selected differentially expressed genes between primary and secondary α -cells indicates stage-dependent α -cell heterogeneity. Normalized expression values are shown. (B) Proportion of primary and secondary α -cells from E12.5 to E15.5 developmental stages. (C) Dotplot of the expression levels of EP-signature genes in $Ngn3^{\text{low}}$ progenitors derived from different developmental stages. (D) The expression of *Pax4* and *Arx* in different pancreatic clusters overlaid onto the PAGA graph. The expression of *Pax4* starts in late $Ngn3^{\text{high}}$ precursors, while the expression of *Arx* is mainly evident in Fev^{high} alpha cells. Mean of normalized expression per cluster is shown.

Table S1. Marker genes of pancreatic epithelial populations. Differential gene expression analysis of 8 main pancreatic epithelial clusters from E12.5-E15.5 mouse embryonic pancreas. List of genes that are differentially upregulated in each cluster compared to all other clusters. This list is related to Figure 2.

[Click here to Download Table S1](#)

Table S2. Gene expression profile of *Ngn3*+ EPs. List of upregulated and downregulated genes in EPs (*Ngn3*^{low} progenitors and *Ngn3*^{high} precursors) compared to non-EP clusters. This list is related to Figure 4 and 5.

[Click here to Download Table S2](#)

Table S3. Gene expression changes during early endocrinogenesis. Differentially expressed genes by comparison of ductal cluster to *Ngn3*^{low} progenitors, *Ngn3*^{low} progenitors to ductal cluster, *Ngn3*^{high} precursors to *Ngn3*^{low} progenitors and *Fev*^{high} cells to *Ngn3*^{high} precursors. This list is related to Figure 5 and Figure S3.

[Click here to Download Table S3](#)