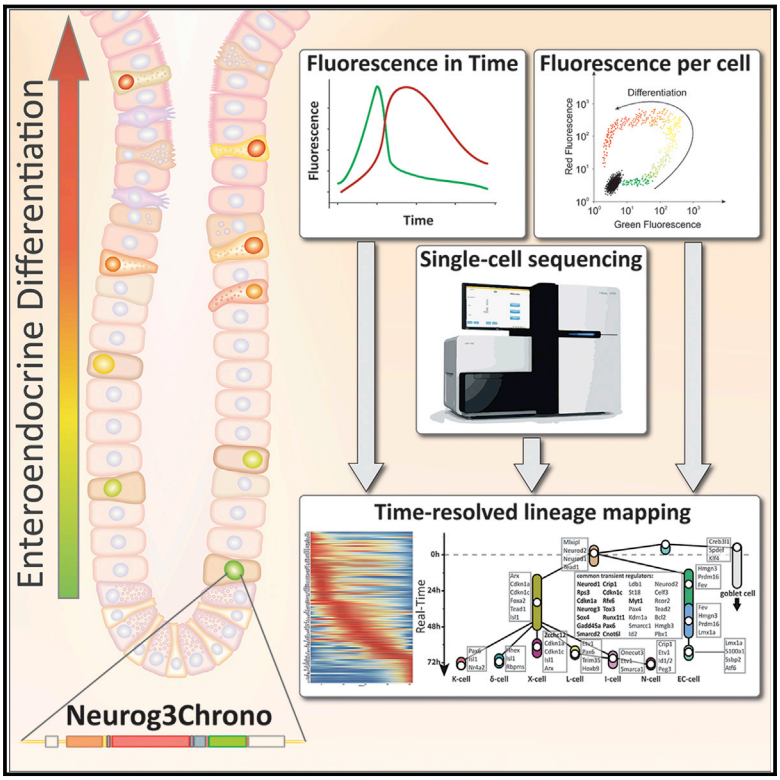
Article

Cell

Identification of Enteroendocrine Regulators by Real-Time Single-Cell Differentiation Mapping

Graphical Abstract



Authors

Helmuth Gehart, Johan H. van Es, Karien Hamer, ..., Johanna F. Dekkers, Anne Rios, Hans Clevers

Correspondence

h.clevers@hubrecht.eu

In Brief

The hierarchical lineage of intestinal enteroendocrine cells is defined at a spatiotemporal single-cell manner and validated using organoid and *in vivo* models.

Highlights

- Neurog3Chrono arranges enteroendocrine single-cell transcriptomes on real-time axis
- Enteroendocrine cells show hormonal plasticity in the course of their maturation
- A time-resolved map characterizes fate specification of all enteroendocrine lineages
- Individual knockout of 6 identified regulators gives robust enteroendocrine phenotypes

Data Resources

GSE113561

Gehart et al., 2019, Cell *176*, 1158–1173
February 21, 2019
https://doi.org/10.1016/j.cell.2018.12.029

CellPress

Cell

Article

Identification of Enteroendocrine Regulators by Real-Time Single-Cell Differentiation Mapping

Helmuth Gehart,[1] Johan H. van Es,[1,2] Karien Hamer,[1] Joep Beumer,[1] Kai Kretzschmar,[1,2] Johanna F. Dekkers,[1,3] Anne Rios,[3] and Hans Clevers[1,2,3,4,*]

[1]Hubrecht Institute, Royal Netherlands Academy of Arts and Sciences (KNAW) and University Medical Centre (UMC) Utrecht, 3584 CT Utrecht, the Netherlands

[2]Oncode Institute, Hubrecht Institute, 3584 CT Utrecht, the Netherlands

[3]Princess Máxima Centre for Paediatric Oncology, 3584 CT Utrecht, the Netherlands

[4]Lead Contact

*Correspondence: h.clevers@hubrecht.eu

https://doi.org/10.1016/j.cell.2018.12.029

SUMMARY

Homeostatic regulation of the intestinal enteroendocrine lineage hierarchy is a poorly understood process. We resolved transcriptional changes during enteroendocrine differentiation in real time at single-cell level using a novel knockin allele of *Neurog3*, the master regulator gene briefly expressed at the onset of enteroendocrine specification. A bi-fluorescent reporter, Neurog3Chrono, measures time from the onset of enteroendocrine differentiation and enables precise positioning of single-cell transcriptomes along an absolute time axis. This approach yielded a definitive description of the enteroendocrine hierarchy and its sub-lineages, uncovered differential kinetics between sub-lineages, and revealed time-dependent hormonal plasticity in enterochromaffin and L cells. The time-resolved map of transcriptional changes predicted multiple novel molecular regulators. Nine of these were validated by conditional knockout in mice or CRISPR modification in intestinal organoids. Six novel candidate regulators (Sox4, Rfx6, Tox3, Myt1, Runx1t1, and Zcchc12) yielded specific enteroendocrine phenotypes. Our time-resolved single-cell transcriptional map presents a rich resource to unravel enteroendocrine differentiation.

INTRODUCTION

Single cell sequencing is rapidly evolving to become an indispensable tool to investigate cellular composition of tissues (Haber et al., 2017; Halpern et al., 2017). By employing algorithms such as diffusion pseudotime (Haghverdi et al., 2016) or StemID (Grün et al., 2016), single-cell datasets can also be brought in pseudo-temporal order to investigate continuous changes in cellular identity (e.g., differentiation). This approach has proven useful but has inherent limitations: (1) pseudo-temporal order is relative, so no information on actual duration is available; (2) a densely populated dataset is needed because transition states between cellular identities have to be observed for high-confidence pseudo-temporal relations; and (3) the order is not based on a transcriptome-independent variable and is therefore easily biased by the applied method. Thus, especially rare cell populations with complex differentiation programs are difficult to study based on pseudo-time alone.

The intestinal epithelium renews completely every 3–5 days (Darwich et al., 2014). This rapid turnover necessitates the coexistence of stem cells, progenitors, and mature cells at any given time. All epithelial cells in the intestine originate from continuously cycling LGR5+ stem cells at the bottom of the crypt (Barker et al., 2007). Among their progeny are enteroendocrine (EE) cells, a scarce (<1% of the epithelium) but essential hormone-producing population scattered throughout the gastrointestinal epithelium. Their roles in metabolism and appetite control have put EE hormones at the forefront of the battle against metabolic syndrome and type II diabetes. Additionally, they also control intestinal motility and orchestrate mucosal immunity (Gribble and Reimann, 2017; Worthington et al., 2018). EE cell types are usually classified based on their hormone production: L cells (Glucagon-like peptide 1, GLP1), I cells (Cholecystokinin, Cck), Enterochromaffin (EC) cells (Serotonin, 5-HT), X cells (Ghrelin, GHRL), S cells (Secretin, SEC), K cells (Gastric inhibitory peptide, Gip), delta cells (Somatostatin, Sst), and N cells (Neurotensin, Nts) were originally distinguished, but reports on multihormonal cells make the classification significantly more complicated with up to 20 different cell types (Haber et al., 2017; Habib et al., 2012). A limited number of regulators of EE development, such as *Neurog3* (Mellitzer et al., 2010), *Neurod1* (Naya et al., 1997), or *Arx* (Beucher et al., 2012), have already been identified in knockout studies, but a complete description of the differentiation process of individual lineages is lacking. Understanding the developmental program that controls EE differentiation is of particular interest, as specific subtypes (e.g., L cells and K cells) harbor significant therapeutic potential.

In this study, we combine single-cell RNA sequencing, a fluorescent timer construct and organoid technology to generate a real-time resolved, lineage-specific map of EE differentiation on a single-cell level. In doing so, we identify significant differences in the differentiation speed of individual lineages, changing hormone production within lineages along the temporal

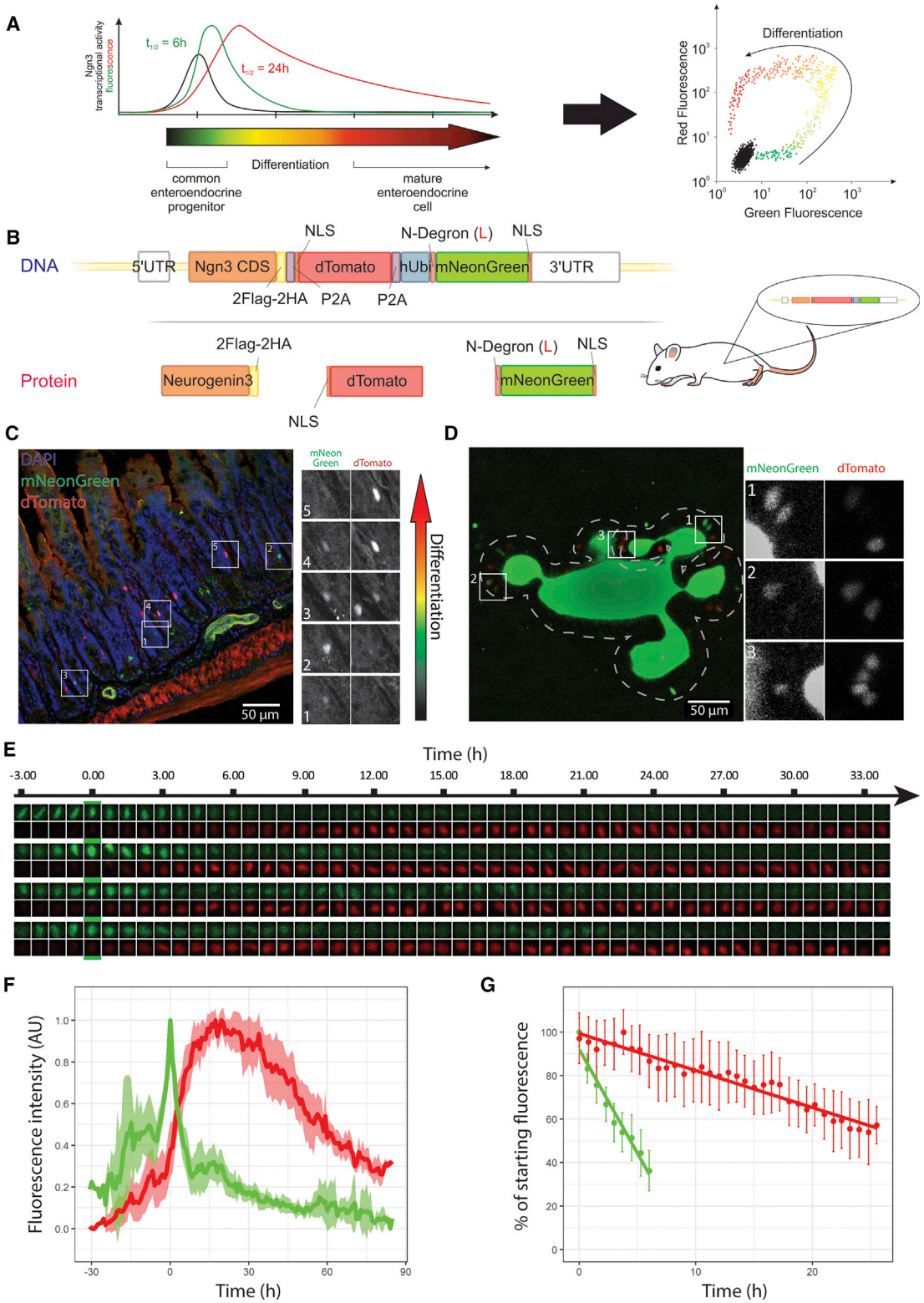


Figure 1. Generation and Characterization of the Neurog3Chrono Reporter

(A) Schematic depiction of reporter strategy.

(B) Structure of Neurog3Chrono reporter on DNA level and protein products.

(legend continued on next page)

trajectory, and novel transcriptional regulators of EE differentiation. Finally, we prove the relevance of our findings by knocking out 9 general and lineage specific candidate regulators in mice or organoids and demonstrate EE phenotypes in two-thirds of all tested genes.

RESULTS

Generation of an EE Real-Time Reporter

Due to the limited time frame from birth in the crypt to shedding at the tips of villi, differentiation of intestinal cells is a highly choreographed process. Neurogenin-3 (*Neurog3*) is transiently expressed in the common EE progenitor that gives rise to all EE cells in the intestine and to a majority of EE cells in the stomach. Thus, the transcriptional pulse of *Neurog3* expression in the course of EE differentiation used in conjunction with a fluorescent pulse-chase reporter can clearly identify a cell's progress on its way from common EE progenitor to mature EE cell.

Classic fluorescent reporters (e.g., EGFP) are of limited use as time-resolved reporters, due to their long half-life and the ambiguity of fluorescence intensity versus time. Previous studies have tried to overcome this limitation with fluorescence shifting fluorophores, such as DsRed-E5 (Miyatsuka et al., 2009). However, especially early gene activation was difficult to detect, due to the low brightness of the protein. Alternatively, coupling an unstable, fast folding fluorescent protein (green) that indicates acute transcriptional activity, with a second, stable, slow folding fluorophore (red) allows cell-tracking long after cessation of reporter gene transcription (Figures 1A and 1B). To generate a reporter with the highest possible sensitivity, accuracy, and temporal resolution, we employed the brightest available fluorescent proteins (mNeonGreen [Shaner et al., 2013] and dTomato [Shaner et al., 2004]) and destabilized mNeonGreen based on the N-end-rule by N-terminally fusing it to a single ubiquitin followed by an N-degron (Tasaki et al., 2012). Since co-translational cleavage of a single ubiquitin in the polypeptide chain exposes leucine at the N terminus of mNeonGreen, the protein is subject to active removal. The construct was inserted into the *Neurog3* locus at the endogenous stop codon to maintain potential regulatory functions of the 3′ UTR (Figure 1B). The knockin gene generates three independent proteins (NEUROG3, dTomato, and destabilized mNeonGreen) from a single polypeptide chain at a 1:1:1 ratio, which hinges changes in green and red fluorescence intensity exclusively on protein stability and thus time.

Neurog3Chrono animals were viable and even homozygous reporter mice showed normal EE differentiation (Figure S1A). The endogenous fluorescence showed the expected spectrum of cells from green over yellow (green + red) to red (Figure 1C). To establish a relationship between fluorescence signal and real time, we tracked individual cells during the differentiation process. Mini-guts (aka organoids) mimic the intestinal epithelium almost perfectly in structure and function and generate all cell types of the epithelial lining, including EE cells (Basak et al., 2017; Sato et al., 2009). We isolated intestinal organoids from homozygous Neurog3Chrono mice and followed spontaneous EE differentiation by live imaging (Figures 1D–1E and Video S1). Fluorescent cells changed sequentially from green over yellow to red (Figures 1E and 1F) indicating that *Neurog3* was—as expected—expressed in a pulse during EE differentiation. The calculated half-lives of mNeonGreen and dTomato were 4.39 ± 0.45 h and 29.78 ± 1.54 h (95% confidence intervals), respectively, which clearly indicates that mNeonGreen is highly destabilized (Figure 1G). We could reliably detect and distinguish cells from 24 h before to around 80 h after peak green fluorescence (our 0 h time point).

Global Transcriptional Changes during EE Differentiation

Having confirmed correct reporter behavior *in vivo* and *in vitro*, we proceeded to separate reporter-positive from reporter-negative cells by FACS of small intestinal crypts of homozygous Neurog3Chrono mice (Figure 2A). EE-specific genes were strongly expressed in reporter positive cells, whereas markers of epithelial stem cells (*Lgr5*) and other intestinal lineages (*Alpi*, *Muc2*, *Lyz1*, and *Dclk1*) were strongly enriched in the reporter-negative population (Figure 2B). Thus, Neurog3Chrono was correctly expressed and labeled EE cells with high fidelity. Subsequently, we separated the reporter-positive cells into early (green), differentiating (yellow), and mature EE cells (red) (Figure 2A). The chosen gates corresponded approximately to −24 h to 5 h, −5 h to 24 h, and older than 24 h (Figure 2A). Since our system allows us to highlight transcripts that show temporal modulation during the differentiation process, we focused our analysis on genes that are differentially expressed between different stages of EE maturation. A total of 1,418 genes showed significant changes during the differentiation process (Figure 2C).

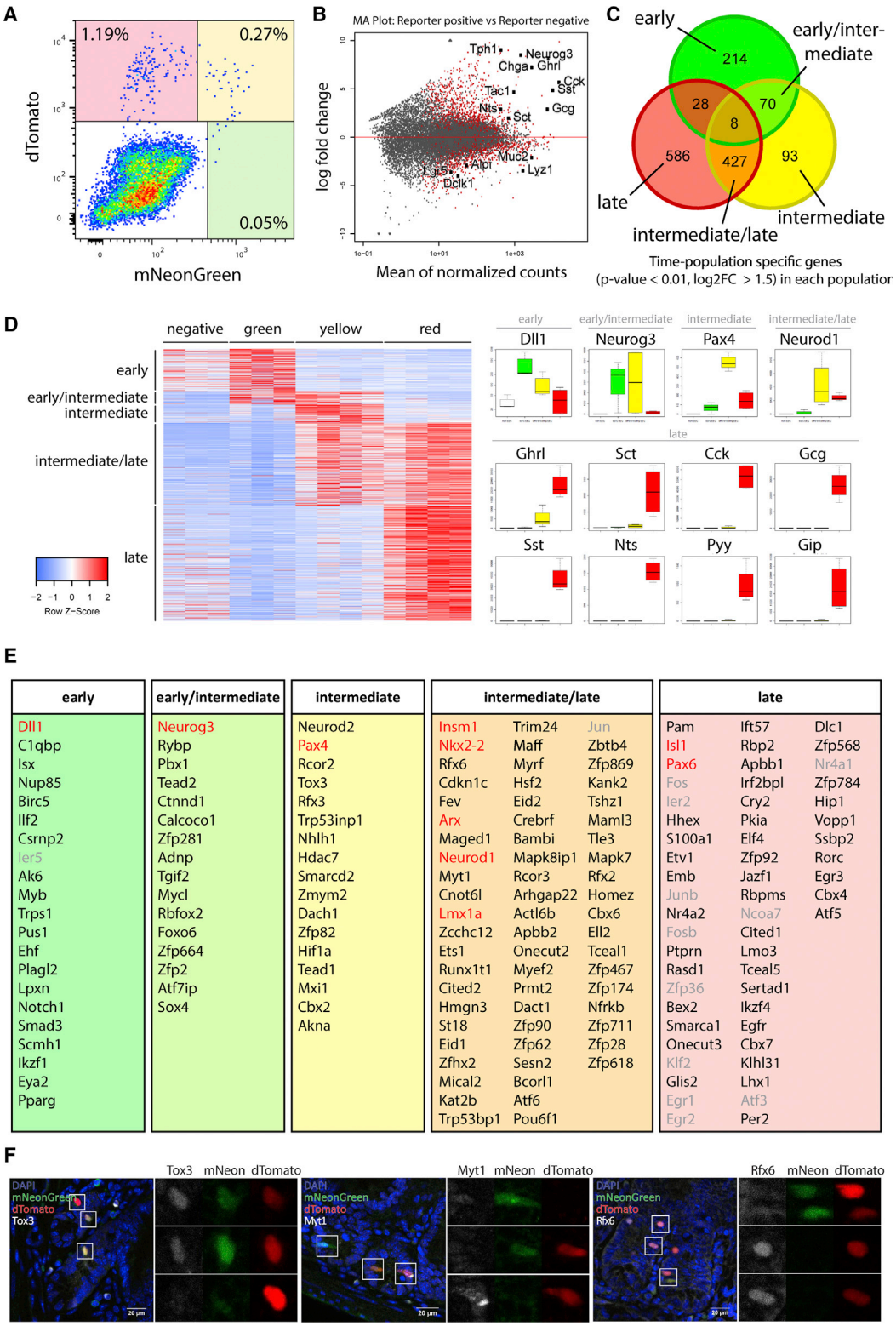
Based on their temporal expression pattern, we classified transcripts into early, early+intermediate, intermediate, intermediate+late, and late expression genes. Within these categories, we found known markers of the EE differentiation process (Figure 2D). Additionally, we identified a wide array of time-specific genes (Table S1). Among these were 172 transcriptional regulators that showed specific temporal expression (Figure 2E). 54 of these were only transiently upregulated during the maturation process. To validate our findings, we selected several candidate genes in the list of transcription factors and confirmed their temporal expression pattern at protein level *in situ*. TOX3, MYT1, and RFX6 were successfully detected in the small intestine in EE cells of the expected Neurog3Chrono fluorescence (Figure 2F). Likewise, staining for SOX4 protein overlapped widely (but not completely) with expression of NEUROG3, which indicates that expression peaks of both genes in EE cells are only slightly offset

(C and D) Detection of reporter fluorescence in (C) cryosections of homozygous Neurog3Chrono small intestine and (D) homozygous small intestinal Neurog3Chrono organoids in standard culture medium.

(E) Examples of fluorescence tracking of individual cells in Neurog3Chrono organoids over time; aligned on maximum green fluorescence.

(F) relative mean fluorescence intensities of mNeonGreen (green) and dTomato (red) over time (n = 25, mean ± 95% confidence interval [CI]).

(G) Fluorescence decay of the Neurog3Chrono fluorophores over time (n = 25, mean ± 95% CI).



E

early	early/intermediate	intermediate	intermediate/late			late		
Dll1	Neurog3	Neurod2	Insm1	Trim24	Jun	Pam	Ift57	Dlc1
C1qbp	Rybp	Pax4	Nkx2-2	Maff	Zbtb4	Isl1	Rbp2	Zfp568
Isx	Pbx1	Rcor2	Rfx6	Myrf	Zfp869	Pax6	Apbb1	Nr4a1
Nup85	Tead2	Tox3	Cdkn1c	Hsf2	Kank2	Fos	Irf2bpl	Zfp784
Birc5	Ctnnd1	Rfx3	Fev	Eid2	Tshz1	Ier2	Cry2	Hip1
Ilf2	Calcoco1	Trp53inp1	Arx	Crebrf	Maml3	Hhex	Pkia	Vopp1
Csrnp2	Zfp281	Nhlh1	Maged1	Bambi	Tle3	S100a1	Elf4	Ssbp2
Ier5	Adnp	Hdac7	Neurod1	Mapk8ip1	Mapk7	Etv1	Zfp92	Rorc
Ak6	Tgif2	Smarcd2	Myt1	Rcor3	Rfx2	Emb	Jazf1	Egr3
Myb	Mycl	Zmym2	Cnot6l	Arhgap22	Homez	Junb	Rbpms	Cbx4
Trps1	Rbfox2	Dach1	Lmx1a	Actl6b	Cbx6	Nr4a2	Ncoa7	Atf5
Pus1	Foxo6	Zfp82	Zcchc12	Apbb2	Ell2	Fosb	Cited1	
Ehf	Zfp664	Hif1a	Ets1	Onecut2	Tceal1	Ptprn	Lmo3	
Plagl2	Zfp2	Tead1	Runx1t1	Myef2	Zfp467	Rasd1	Tceal5	
Lpxn	Atf7ip	Mxi1	Cited2	Prmt2	Zfp174	Zfp36	Sertad1	
Notch1	Sox4	Cbx2	Hmgn3	Dact1	Nfrkb	Bex2	Ikzf4	
Smad3		Akna	St18	Zfp90	Zfp711	Smarca1	Egfr	
Scmh1			Eid1	Zfp62	Zfp28	Onecut3	Cbx7	
Ikzf1			Zfhx2	Sesn2	Zfp618	Klf2	Klhl31	
Eya2			Mical2	Bcorl1		Glis2	Lhx1	
Pparg			Kat2b	Atf6		Egr1	Atf3	
			Trp53bp1	Pou6f1		Egr2	Per2	

Figure 2. Global Transcriptional Changes during EE Differentiation

(A) Flow cytometry of isolated small intestinal crypt cells from homozygous Neurog3Chrono animals. Colored gates indicate the sorting windows for early (green), intermediate (yellow), and late (red) cells. The combination of early, intermediate, and late gate is considered reporter positive.

(B) MA-Plot of reporter positive versus reporter negative cells. (n = 4 mice).

(legend continued on next page)

from each other (Figure S1B). Taken together, time-resolved bulk sequencing of Neurog3Chrono crypts confirmed the fidelity of the reporter system and identified a wide array of new EE genes with distinct temporal expression profiles.

Generation of a Time-Resolved Dataset of EE Differentiation at Single-Cell Level

Bulk RNA sequencing is a highly sensitive method to assay global transcriptional changes. However, it does not distinguish specific gene-expression in individual EE subspecies. Thus, we performed single cell sequencing on reporter-positive cells from small intestinal crypts and villi. A total number of 6906 cells were sorted and processed following the SORT-Seq method (Muraro et al., 2016). This method is based on CEL-Seq2, which combines immediate barcoding with linear amplification to ensure reliable detection of expression differences even in lowly expressed genes (Hashimshony et al., 2016). Since the majority of reporter-positive cells were mature (Figure 2A), we specifically enriched for earlier (green and yellow) cells to cover the differentiation time frame evenly.

Cell transcriptomes were analyzed with RaceID2 (Grün et al., 2016). After filtering (minimal threshold of 2,000 unique transcripts/cell, see STAR Methods), we retained 2,281 cells for analysis (Figure 3A). Expression of *Dll1* marked the earliest stages of EE differentiation, in line with our bulk dataset (Figure 3B). High levels of *Neurog3* identified cells in the central cluster as progenitors (Figure 3B). In contrast, *Neurod1* and *Isl1* marked late progenitors and mature EE cells (Figure 3B). Two "common" markers of mature EE cells, *Chga* and *Reg4* (Grün et al., 2016), were most abundantly expressed in enterochromaffin (EC) cells, the most numerous EE cell type in the intestine. Whereas *Chga* was present in other EE subtypes, albeit at lower expression levels, *Reg4* appeared highly specific for the EC lineage (Figure 3B). All mature cell clusters were identified based on their hormone expression profile (Figure 3F). We observe well-separated clusters of I cells (*Cck*), L cells (*Gcg*), Delta cells (*Sst*), X cells (*Ghrl*), EC cells (*Tac1*/*Tph1*), N cells (*Nts*), and K cells (*Gip*). EE cells from proximal, medial, and distal small intestine clustered according to their EE subtype and not according to regional origin within the small intestine (Figure 3C). The distribution of mature cells between crypt and villus varied strongly based on cell type. L cells were almost exclusively crypt derived, whereas nearly all N cells stemmed from the villus (Figures 3D and 3E). S cells did not form a separate cluster, but *Sct*-high cells could be found in most other mature cell clusters, particularly among villus-derived I and N cells (Figure S1C). In addition to the expected clusters of EE cells, we also discerned a Goblet cell cluster (*Agr2*, *Tff3*, *Spink4*, and *Muc2*) and a small cluster of Paneth cells (*Lyz1* and *Defa17*) (Figure S1D), which was in line with previous observations (Schonhoff et al., 2004).

By integrating the recorded fluorescence with the established dynamics of the green and red fluorophore, we could determine the amount of time that had passed since peak NEUROG3 expression for each individual cell (Figures 3G and 3H, see STAR Methods). When projected onto the tSNE map (Figure 3A), the time information correlated perfectly with the established cluster identities (Figure 3I). Thus, we could distinguish EE lineages and link differentiation time with single-cell transcriptomes.

EE Cells Display Hormonal Plasticity

Despite the clear separation of hormone-expression on the tSNE map, only 28.7% of all mature EE cells dedicated more than 90% of their hormone-encoding mRNA to a single gene product. However, the number varied strongly by lineage. Whereas more than 70% of K, X, and Delta cells expressed essentially only their primary hormone, less than 10% of L cells did (Figures S2A and S2B). The majority of L, I, N, and EC cells were bi- or trihormonal. The expression of several groups of hormones (e.g., *Nts*, *Pyy*, and *Sct*) was clearly correlated, which matched previous reports of multihormonal cells (Egerod et al., 2012; Habib et al., 2012) (Figure 4A). The strongest positive correlation was found between *Gcg* and *Cck* in L and I cells and between *Tac1* and *Tph1* in EC cells. *Tac1* and *Tph1* showed low correlation with other hormones, emphasizing the difference of the EC lineage to all other lineages.

Given the wide spectrum of observed co-expression in individual cells (Figure S2B), it appeared unlikely that each combination constituted an independent lineage. High variability could also be a sign of hormonal plasticity that allows mature cells to modify their hormonal repertoire based on extrinsic cues. One such cue can be the changing environment along the crypt-villus axis during the journey of EE cells toward the villus tips. In fact, when separated by crypt versus villus origin, EC cells showed striking differences in their hormonal expression (Figure 4B). While crypt EC cells expressed *Tac1* and *Tph1*, villus EC cells expressed *Sct* instead of *Tac1* and even higher levels of *Tph1*. Previously, it had been suggested that different sub-lineages of EC cells exist in parallel. Alternatively, there could be a single lineage, which changes hormone expression in the course of maturation. By increasing our minimal transcript threshold, we resolved a *Tac1*-high and *Sct*-negative and a *Tac1*-low and *Sct*-positive sub-cluster within the EC population (Figure S2C). When arranging these clusters in time, we clearly noticed that these populations arose subsequently, and not in parallel (Figure 4C, top). The median age of *Tac1*+ EC cells was 43.98 h after peak NEUROG3 expression, whereas *Sct*+ EC cells were on average 64.82 h old. The decrease of *Tac1*+ EC cells was concomitant with the appearance of *Sct*+ EC cells, which suggested efficient cell type conversion. This observation was further supported by plotting hormone expression in the EC lineage versus time (Figure 4C, bottom) and confirmed in a

(C) Venn diagram of population-specific genes ($p < 0.01$, log2FC > 1.5).

(D) Heatmap of cluster specific genes ($p < 0.01$, log2FC > 1.5, row *Z* scores) and examples for genes in each group (y axis = normalized counts). For a complete list, see Table S1.

(E) Time-resolved list of transcriptional regulators. Genes with known EE function are in red. Grayed-out genes have been reported to be associated with cell dissociation (van den Brink et al., 2017).

(F) Immunofluorescent stainings for TOX3, MYT1, and RFX6 on vibratome sections of homozygous Neurog3Chrono small intestine.

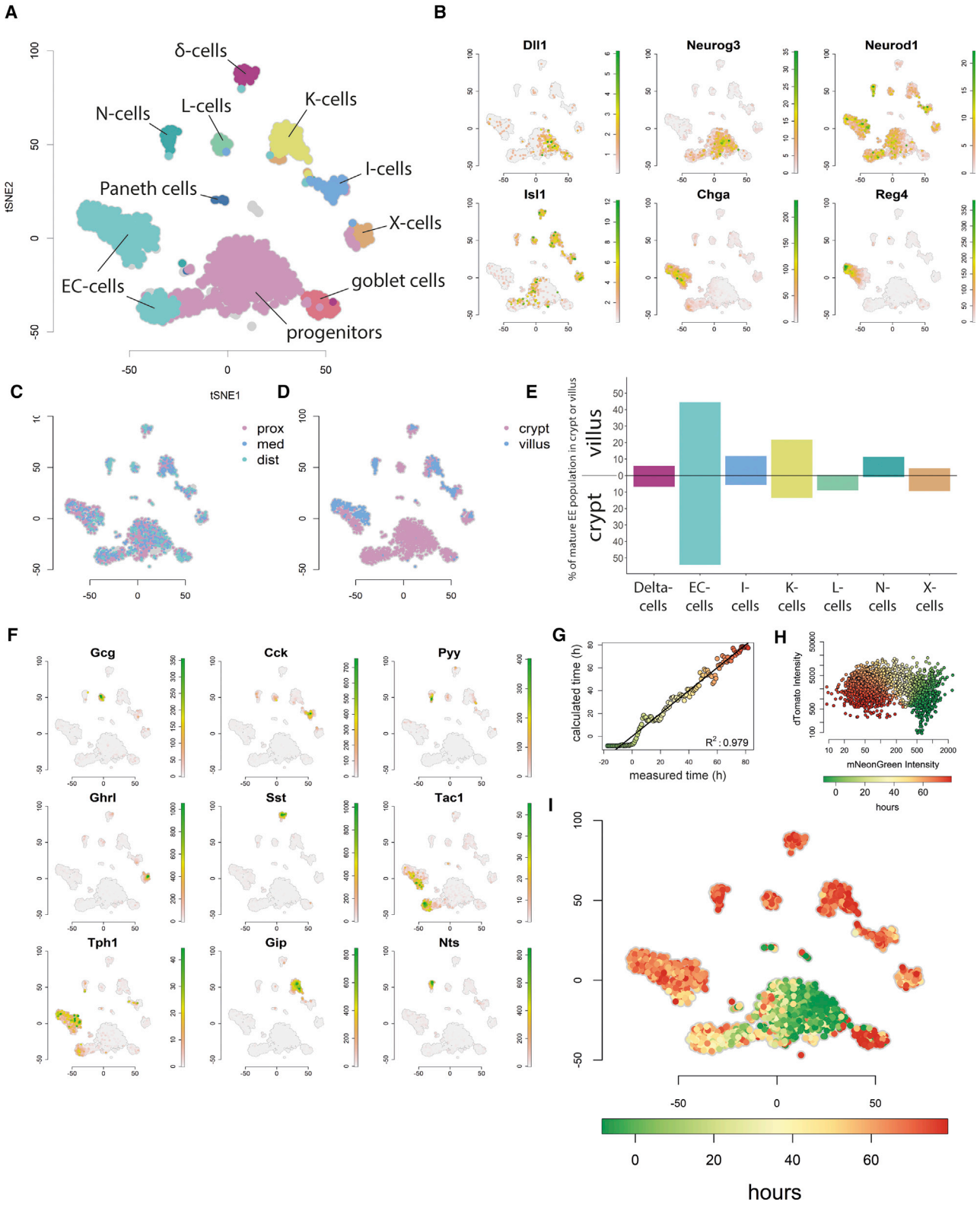


Figure 3. Generation of a Time-Resolved Dataset of EE Differentiation on a Single-Cell Level

(A) tSNE plot of sorted Neurog3Chrono cells (min. 2,000 unique transcripts/cell, 2,281 cells, n = 17 sorts of 2–3 mice each).

(B) General marker expression of EE cells (normalized unique transcript counts).

(legend continued on next page)

reporter-independent manner in organoids (Figure S3C). Tracking of *Tac1*+ cells on an individual and population level in Tac1Cre-LSL-tdTomato organoids excluded premature cell death of *Tac1*+ cells (Figures S3A and S3B). Plasticity, on the other hand, was clearly supported by lineage analysis with the "destiny/dpt" package (Angerer et al., 2016; Haghverdi et al., 2016). The EC lineage formed a continuous trajectory in diffusion space from progenitors over *Tac1*+ EC cells (blue) to *Sct*+ EC cells (green) (Figures 4D and S4B). This illustrates that *Tac1*+ EC cells represent the link between progenitors and Sct+ EC cells, and not a separate lineage. As a complementary approach, we also employed RNA velocity, a method that establishes differentiation trajectories in single-cell datasets based on the ratio of unspliced to spliced mRNA (La Manno et al., 2018). Transcripts of *Chgb*, a marker strongly expressed in the later EC population, showed high ratios of retained introns in the earlier EC cluster (Figure 4E). This indicated the start of a transition from early to late EC cell identity. The same behavior could be seen slightly later along the EC maturation path for the gene *Reg4* (Figure S2D). Thus, based on Neurog3Chrono time, the lack of premature cell death in the *Tac1*+ EC population, the differentiation trajectory in diffusion space and RNA velocity information we conclude that the two EC sub-populations are not independent parallel lineages, but subsequent stages in EC-cell maturation.

However, plasticity is not limited to EC cells. A similar relationship was observed for the closely linked L, I, and N cells (Figures 4F–4H and S2E). The number of L cells decreased from 70 h onward, while the number of I cells and N cells increased concomitantly (Figure 4G, top). This behavior was also reflected by the transient nature of *Gcg* expression in the LIN cell population (Figure 4G, bottom), which indicated that L cells start to acquire transcriptional I- or N-cell identity around 70 h into their lifetime. The ILN lineage trajectory in diffusion space visualizes the progression from L over I to N cell (Figure 4D), which was further corroborated by RNA velocity analysis for *Cck* (Figure 4H) and *Nts* (Figure S2F). It is important to note, however, that L cells that acquire transcriptional I- or N-cell identity still retain GLP1/2 protein for a significant time. This explains why an overlap of GLP1 (highest RNA expression in L cell cluster) and PYY (highest RNA expression in N cell cluster) is observed in cells on protein level.

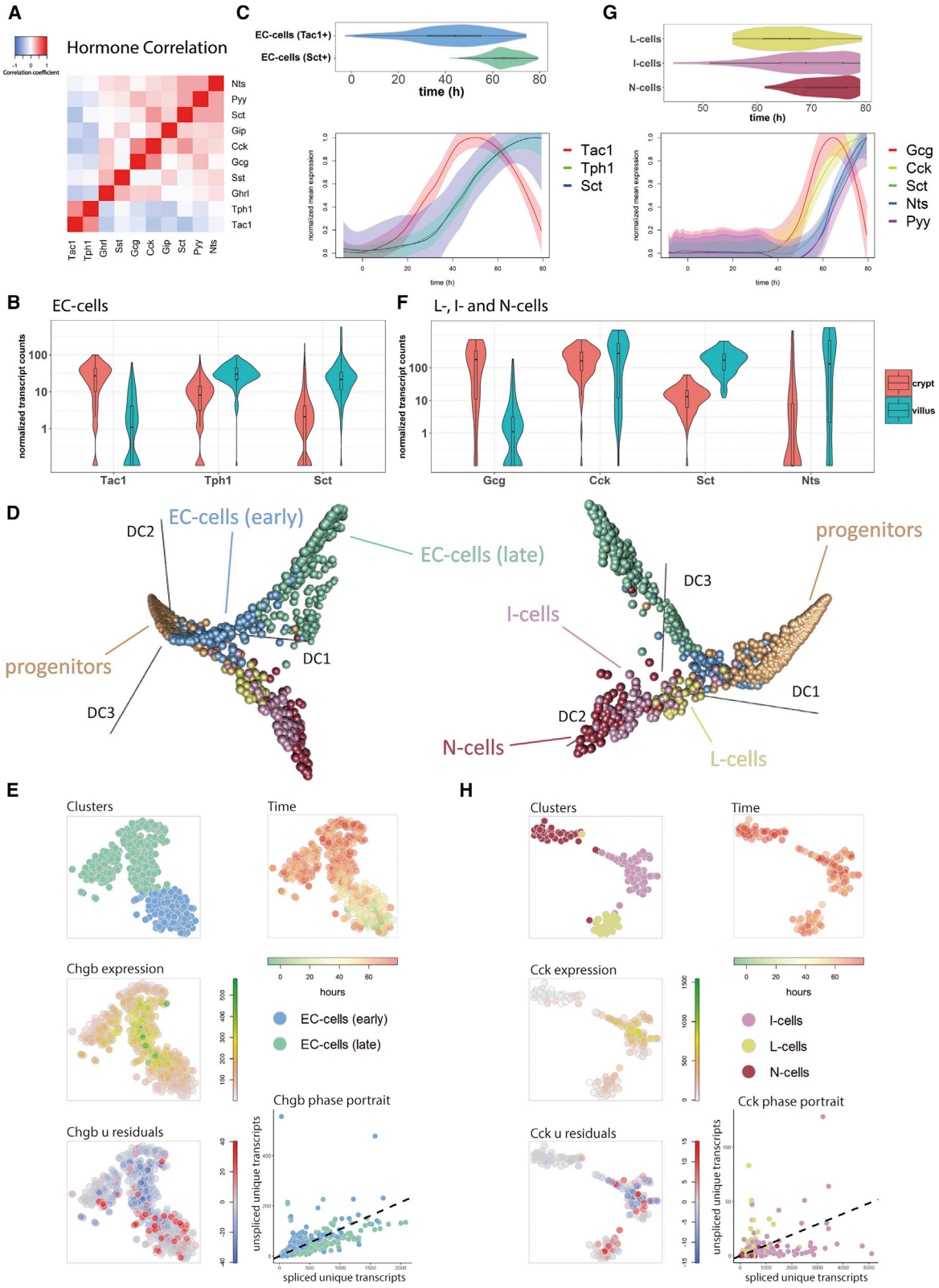
EE Cells Do Not Move Freely with the Intestinal "Conveyor Belt"

For both EC cells and LIN cells, the switch in identity occurred at around 70 h after the peak of NEUROG3 expression. Thus, we wondered whether this time coincided with movement of EE cells along the crypt-villus axis. We recorded confocal images of Neurog3Chrono small intestine and reconstructed the crypts in 3D (Video S2, Figure S3E). Subsequently, we measured the fluorescence intensities of EE cells and correlated thereby their distance from the +4 position with differentiation time (Figure S3D). As expected, we observed the youngest EE cells in the lower half of the crypt close to the +4 position. Unexpectedly, however, EE cells did not immediately move up the crypt-villus axis. The first EE cells were observed outside the crypt at around 60 h, after which their number steadily increased. Nevertheless, a significant portion of EE cells remained in the crypt even at 80 h. Since enterocytes travel from the bottom of the crypt to the tip of villi within 72 h regardless of small intestinal region (Darwich et al., 2014), it was obvious that EE cells do not travel freely with the epithelial "conveyor belt". 97.2% ± 2.2% (n = 145 cells in 9 sections) of all red (old) cells in crypts expressed CHGA either at high (EC cell) or low levels (non-EC cell) (Figures S3F and S3G). Thus, the vast majority of these cells was mature. Notably, the time the first EE cells reached the base of villi did coincide with hormonal switches in EE lineages (e.g., EC and L cells). This is well in line with a recent publication that described changes in hormonal expression in EE cells in response to BMP signaling, which is known to form a gradient of increasing signal strength from the crypt to the villus (Beumer et al., 2018). Thus, our data strongly support the ability of EE cells to change their hormonal repertoire dynamically in response to environmental cues, such as position along the crypt-villus axis.

Establishing Differentiation Trajectories with Real-Time Information

To follow the process of EE differentiation and subsequently identify its regulators, we established the order of events from common progenitor to mature cell. By further subclustering progenitor cells, we resolved 7 populations with differing temporal profiles (Figures 5A and 5B). The earliest population (early progenitor) covered a tight time span around the −10 h position. It was followed by two clusters of immediate offspring. One of these (goblet cell progenitors) showed a clear bias for the goblet cell fate with increased expression of markers such as *Fcgbp* and *Agr2* (Figure 5C). The other cluster (common EE progenitor) expressed markers of EE differentiation, e.g., *Neurod1*, and was distinctly positive of *Olfm1* (Figure 5C). Following the common EE progenitor were two equi-temporal populations of biased progenitors: one expressing markers of early EC cells (e.g., *Fev*, *Chgb*), the other markers of non-EC EE populations (e.g., *Isl1*) (Figure S4A). Each of these was followed by a cluster of the earliest mature cells of the two respective lineages: *Tac1*+ early EC cells and X cells (Figures 5A and 5B). To visualize the branching point between EC and non-EC lineage and identify potential actors in the decision process, we analyzed all progenitors of the EE branch together with early EC and non-EC cells (younger than 48 h) in diffusion space (Figure 5D). The diffusion

(C) Proximal (pink), medial (blue), or distal (turquoise) origin of cells along the proximal-distal axis. Gray cells stem from whole SI preparations.
(D) Crypt (pink) or villus (blue) origin of cells on tSNE map. Gray cells are of mixed origin.
(E) Relative frequency of individual mature EE cell types in crypt and villus.
(F) Hormone expression projected on the tSNE map (normalized unique transcript counts).
(G) Correlation between measured (real) time and calculated time based on fluorescence intensities. Colored dots correspond to reference values measured by live-imaging.
(H) Projection of calculated differentiation time on flow-cytometry data.
(I) Projection of calculated differentiation time on tSNE map.



(legend on next page)

map clearly illustrated the branching point and identified *Hmgn3*, *Prdm16*, and *Fev* as the highest enriched transcriptional regulators at the EC side of the branch (Figure 5E). *Isl1*, *Cdkn1a*, and *Arx* clearly marked cells on the non-EC trajectory (Figure 5E). To establish relations from progenitors to mature cell populations, we transferred cell identities established in the progenitor subclustering to our high-sensitivity dataset of all EE cells (Figure 5F). Cluster distribution along the time axis showed clear differences in the maturation speed of individual lineages. Peaking at 44 h after maximal NEUROG3 expression, *Tac1*+ early EC cells were the first hormone-producing cells to develop (Figure 5G). This was in line with a previous study that utilized BrdU incorporation to demonstrate that TAC1 is the earliest detectable hormone (Aiken and Roth, 1992). At 60 h, *Ghrl*+ X cells were the first non-EC cells to appear in high numbers, followed by L cells. Around 70 h, all other EE lineages were readily detectable. Based on their expression of transcription factors such as *Arx*, *Cdkn1a*, and *Isl1*, X cells, I cells, L cells, and K cells were derived from non-EC biased EE progenitors (Figure S4A). Early EC cells, on the other hand, arose from EC-biased progenitors (based on low *Arx*, *Isl1*, and *Cdkn1a* and high levels of *Fev* and *Chgb*) (Figure S4A). This was further corroborated by the fact that the number of non-EC biased progenitors decreased in time concomitantly with the rise of early EC cells (Figure 5G). While delta cells shared *Cdkn1a* and *Isl1* expression with the non-EC biased progenitor, they did not express *Arx* (similar to the EC biased progenitor) (Figure S4A). The temporal profile favored the non-EC-biased progenitor as origin because very few cells remained in the EC-biased progenitor state at the time delta cells arose. Indeed, *Arx* knockout causes a strong increase in delta-cell numbers at the expense of I cells, L cells, and K cells. EC cells are not affected and arise in normal numbers, while X cells increase slightly (Beucher et al., 2012; Du et al., 2012). If *Arx* was the main regulator at the decision point between EC-biased (low *Arx*) and non-EC-biased (high *Arx*) progenitor, we would have expected increased numbers of EC cells upon knockout and complete loss of all cells that arise from non-EC progenitors. Therefore, it was more likely that *Arx* controls a secondary decision within the non-EC lineage between I cells, L cells, and K cells (high *Arx*) and delta cells (low *Arx*). Finally, when projected in diffusion space, delta cells follow a similar trajectory as other non-EC cells (Figure 5H) and are distinct from the ILN lineage (Figure S4B). Thus, due to expression of *Isl1* and *Cdkn1a*, the fitting temporal profile, the *Arx* knockout phenotype and the diffusion map trajectory, we also assigned delta cells to the non-EC lineage.

Transcriptional Regulators of Lineage Specification

Having established the relationship between clusters, we proceeded to map the transcriptional order of events during lineage specification. By ordering peak expression of highly modulated transcriptional regulators in time, we generated a comprehensive description of each lineage and its maturation stages (Figures 6A and S5A–S5F). Among the modulated genes of the EC lineage, we noticed known regulators of EE differentiation (e.g., *Neurog3*, *Pax4*, *Pax6*, *Neurod1*, and *Lmx1a*) (Figure 6A). Whereas the modulated expression pattern of genes like *Sox4*, *Neurog3*, *Tox3*, or *Myt1* did not differ between lineages, others (such as *Hmgn3*, *Fev*, *Cdkn1a*, *Etv1*, or *Crip1*) deviated between lineages at specific time points during the maturation process (Figures 6B and 6D).

Having established the expression order of transcriptional regulators, we proceeded to determine which factors define individual cell types. Because differential gene expression analysis among mature populations alone (Figure S5G) would miss the observed transient lineage differences (Figure 6B), we opted to examine differences in gene expression at three time intervals. The first two intervals (10–20 and 30–50 h) contained genes that predominantly separate EC from non-EC lineage (e.g., *Hmgn3*, *Fev*, *Cdkn1a*, *Isl1*, *Atf6*, *Arx*), while the third interval (>50 h) highlighted genes that were specific for individual non-EC lineages. Among these, we identified *Hhex* as delta-cell-specific transcription factors. *Hhex* has been previously described to be necessary for differentiation of SST-producing delta cells in the pancreas (Zhang et al., 2014) and was thus a strong candidate to fulfill the same role in the intestine. *Zcchc12* was specifically expressed in X cells but has so far not been functionally linked to endocrine development. *Lmx1a*, *Atf6*, *Gtf2f2*, and *Taf1* were specific for EC cells. *Lmx1a* has been recently described to control serotonin biosynthesis in the intestine (Gross et al., 2016). In addition to single-lineage factors, we also found various regulators with specific expression in two or more lineages. For example, expression of *Onecut3* separated I and N cells from all other lineages, while *Parp1* was specifically active in L, delta, and EC cells. Likewise, *Etv1* was enriched in L, I, and N cells, whereas *Pax6* was highly expressed in L and K cells.

Figure 4. Hormonal Plasticity in EC and LIN Cells

(A) Heatmap of correlation coefficients of log transformed normalized hormone transcript counts.

(B) Normalized unique hormone transcript counts in crypt- versus villus-derived EC cells.

(C) temporal profiles of *Tac1*+ and *Sct*+ subclusters of EC cells and Loess-smoothed normalized mean expression of *Tac1* (red), *Tph1* (green), and *Sct* (blue) within the EC lineage over time. Shaded regions denote 95% CIs.

(D) Diffusion map showing all EE progenitors plus all cells of the EC and ILN-lineages (min. 4,000 unique transcripts/cell, 2 viewing angles).

(E) tSNE map of 461 early and late EC cells (min. 4,000 unique transcripts/cell) (top-left), time projection on tSNE (top-right), expression levels of *Chgb* (middle-left, normalized unique transcript counts), phase portrait (bottom-right) showing regions of increasing (over dotted line) or decreasing (under dotted line) expression based on unspliced/spliced mRNA balance (colors correspond to clusters), unspliced unique transcript count residuals projected on the tSNE map (red indicates high relative levels of unspliced mRNA = begin of gene expression, blue indicates low relative levels of unspliced mRNA = downregulation of gene expression, bottom left) (related to Figure S2D).

(F) Normalized unique hormone transcript counts in crypt- versus villus-derived combined L, I, and N cells.

(G) Temporal profiles of L, I, and N cell clusters and Loess-smoothed normalized mean expression of *Gcg* (red), *Cck* (yellow), *Sct* (green), *Nts* (blue), and *Pyy* (purple) within the combined LIN lineage over time. Shaded regions denote 95% CIs.

(H) As (E) but for *Cck* mRNA in 176 ILN cells (related to Figure S2F).

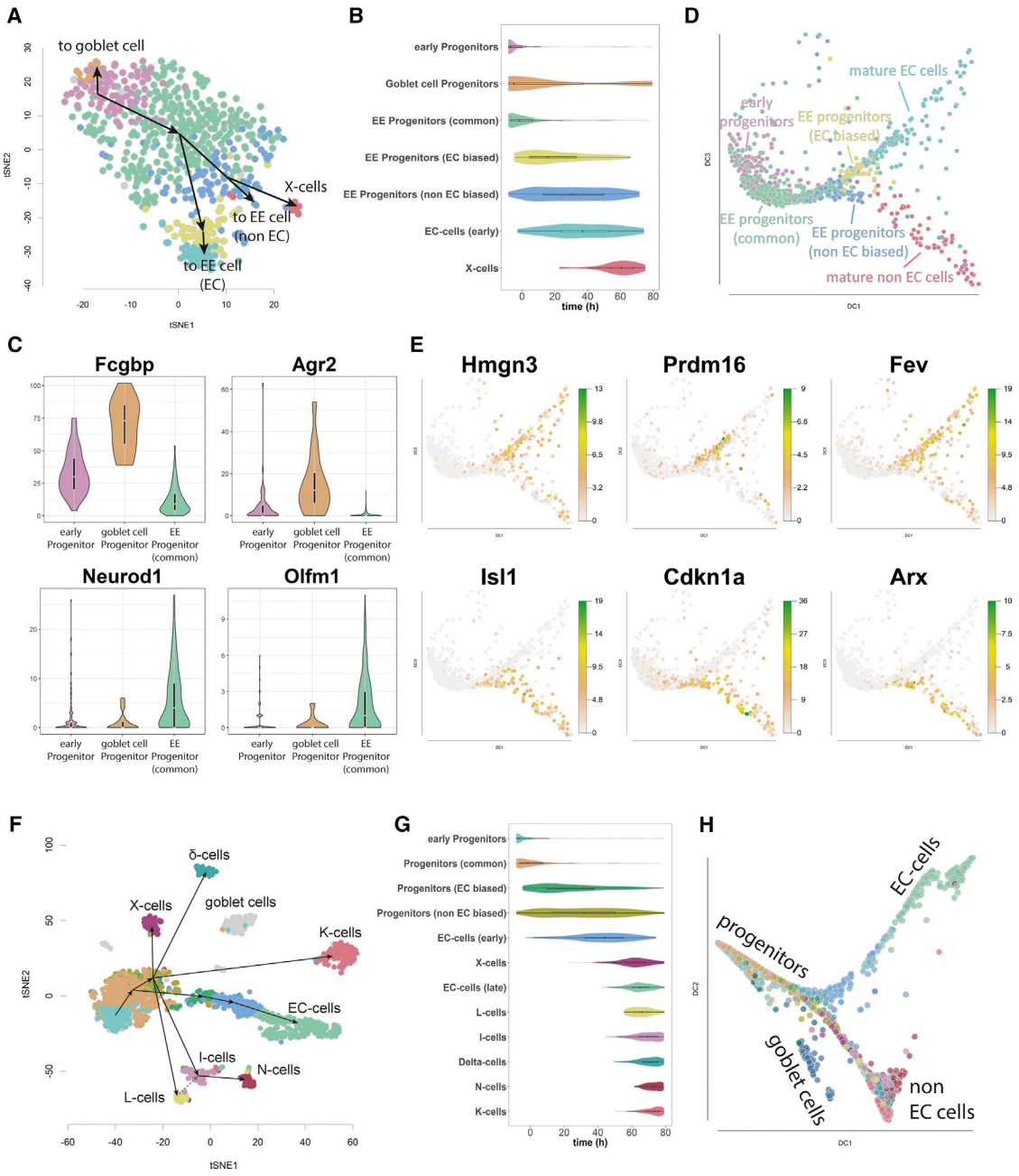


Figure 5. Establishing Differentiation Trajectories with Real-Time Information

(A) tSNE map of 613 EE progenitor cells (min. 5,000 unique transcripts/cell). Arrows indicate lineage relationships. Cells in gray could not be clearly identified.

(B) Violin/boxplots of time distribution within progenitor clusters in Figure 5A.

(C) Violin/boxplot of selected differentially expressed genes between early progenitors (light purple), goblet cell progenitors (orange) and common EE progenitors (green). Numbers indicate normalized unique transcript counts.

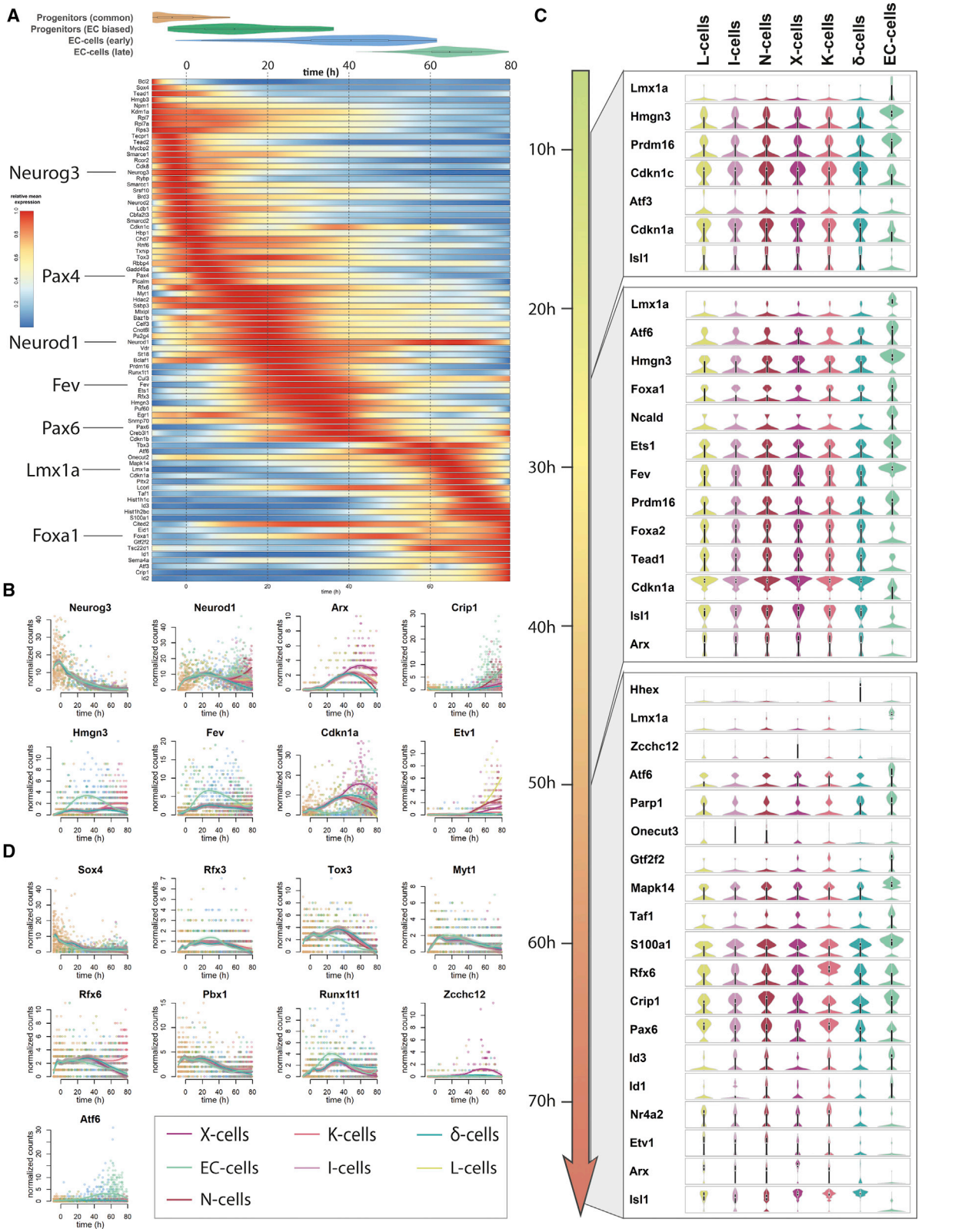
(D) Diffusion map of all EE progenitor cells with cells of mature EE cluster (<48 h of age).

(E) Differentially expressed transcriptional regulators at the branching region between EC and non-EC cells, projected on a diffusion map (normalized unique transcript counts).

(F) High sensitivity dataset of 1,750 cells (min. 4,000 unique transcripts/cell). Arrows indicate lineage relationships. Cells in gray mark non-EE cells.

(G) Violin/boxplots of time distribution within clusters in Figure 5E.

(H) Diffusion map of all cells in the high sensitivity dataset, illustrating the points of lineage decisions.



(legend on next page)

Knockout of Candidate Genes

To corroborate the validity of our findings, we chose 9 candidate genes (Figure 6D) for loss-of-function experiments. We chose *Sox4* due to the similarity of its expression profile to *Neurog3*. *Rfx3*, *Tox3*, *Myt1*, *Pbx1*, and *Runx1t1* represented genes that were transiently activated in progenitors but mostly lost in mature populations. *Rfx6*, has been reported to be a K-cell-specific gene (Suzuki et al., 2013). Indeed, we observed that *Rfx6* expression was only maintained in mature K cells (Figure 6D), but all lineages expressed *Rfx6* highly at earlier stages. Thus, we selected *Rfx6* to investigate whether its function was broader than conventional analysis would suggest. Finally, we chose *Zcchc12* and *Atf6*—two lineage-specific transcription factors for X and EC cells, respectively—which have not yet been linked to EE differentiation.

As confirmed by immunohistochemistry, SOX4 expression is limited to rare cells close to the +4 position (Figure S6A). This corresponds well to the short time frame of *Sox4* expression at the start of EE differentiation (Figure 6D, S1B, and S6B). To study *Sox4* function, we conditionally deleted the gene in the intestine by using the beta-naphthoflavone-inducible AhCre transgene crossed into $Sox4^{loxp/loxp}$ mice (Penzo-Méndez et al., 2007) (Figures 7A and S6A). Upon *Sox4* deletion, the animals lost all GLP1-positive cells in duodenum and jejunum (Figure 7B). In the ileum, a reduced number of GLP1-positive cells was observed, most likely a result of the incomplete deletion by AhCre in the distal small intestine (Ireland et al., 2004) (Figures S6B and S6C). Concomitant with loss of GLP1, CCK-, SST-, and GIP-positive cells were significantly reduced, while the number of GHRL-expressing cells increased (Figures 7B and S6D). This phenotype was further corroborated by microarray analysis (Figures S6E and S6F). These findings are well in line with another very recently published study that explored the role of *Sox4* in *Atoh1* independent lineage allocation of Tuft and EE cells (Gracz et al., 2018). Thus, *Sox4* plays a broad but essential role for correct fate specification during EE differentiation.

Because generation and analysis of conditional knockout animals is a technique with extremely low throughput, it was not compatible with the number of candidate genes we aimed to analyze. Intestinal organoids are a well-established system to study intestinal epithelial biology. However, for organoids to serve as a faithful model for EE lineage specification, the EE differentiation process *in vitro* and *in vivo* needs to be comparable. Therefore, we induced EE differentiation in homozygous Neurog3Chrono organoids (Figure S7A) and investigated the transcriptome of reporter positive. 950 organoid cells containing a minimum of 4,000 unique transcripts/cell were combined with the 1,750 cells from our primary tissue dataset. Cells clustered according to cell type irrespective of tissue or organoid origin (Figure S7 and S7E). In fact, we could detect organoid cells in all previously identified EE progenitor and mature cell clusters (Figure S7D), and they maintained the same differentiation dynamics as primary cells (Figure S7C). Consequently, organoids represent an excellent tool to study EE lineage allocation *in vitro* due to their faithful representation of EE fate specification.

To create loss-of-function organoids, we generated a mouse line with constitutive expression of Cas9 from the *Rosa26* locus. In contrast to previously published Cas9 animals (Platt et al., 2014), these animals do not express EGFP, which enables fluorescence-based readouts. Organoids generated from Rosa-Cas9 animals were transiently transfected with *in vitro*-transcribed gRNA (Figure 7C). Subsequently, we picked and genotyped clones to identify organoids carrying homozygous loss-of-function alleles. Knockout clones were expanded and differentiated under EE-inducing conditions (Basak et al., 2017). Phenotypes in EE differentiation were then assessed by hormone-specific quantitative real-time PCR and whole-mount immunofluorescent staining of multiple independent knockout clones. As proof of principle, we first knocked out *Neurog3* and confirmed loss of all hormone expression (Figure S7F).

Among the 6 transiently expressed candidate regulators (*Rfx6, Rfx3*, *Tox3*, *Myt1*, *Pbx1*, and *Runx1t1*), only *Rfx3* and *Pbx1* showed no EE phenotype. Loss of *Rfx6* induced severe loss of K, X, and L cells and significant reduction in I and EC lineages (Figures 7D and 7E). Despite the strong reduction in EE cells, there was no detectable decrease in *Neurog3* expression, which indicated that *Rfx6* acts downstream of *Neurog3*. This order of events is further supported by our own temporal map of transcription factor activation (Figure 6A) and *Rfx6* knockout studies in the endocrine pancreas (Soyer et al., 2010). Thus, despite being maintained exclusively in adult K cells (Figure 6D), *Rfx6* controls the differentiation of multiple EE lineages due to its transient expression in all lineages during maturation.

Tox3 is a member of the HMG-box protein family and has so far not been linked to endocrine development. Knockout of *Tox3* caused a strong decrease in *Tph1* and thus in serotonin (Figures 7F and 7G). However, *Tox3*-deficient organoids produced significantly higher numbers of X cells. A similar phenotype was observed in *Myt1* knockouts, where EC cells were also strongly reduced (Figures 7G and 7H). Even though *Myt1* has been proposed to act in a reciprocal feed-forward loop with *Neurog3* in the endocrine pancreas (Wang et al., 2008), we could not find evidence to support a similar mechanism in EE cells. *Neurog3* levels did not decrease upon loss of *Myt1*,

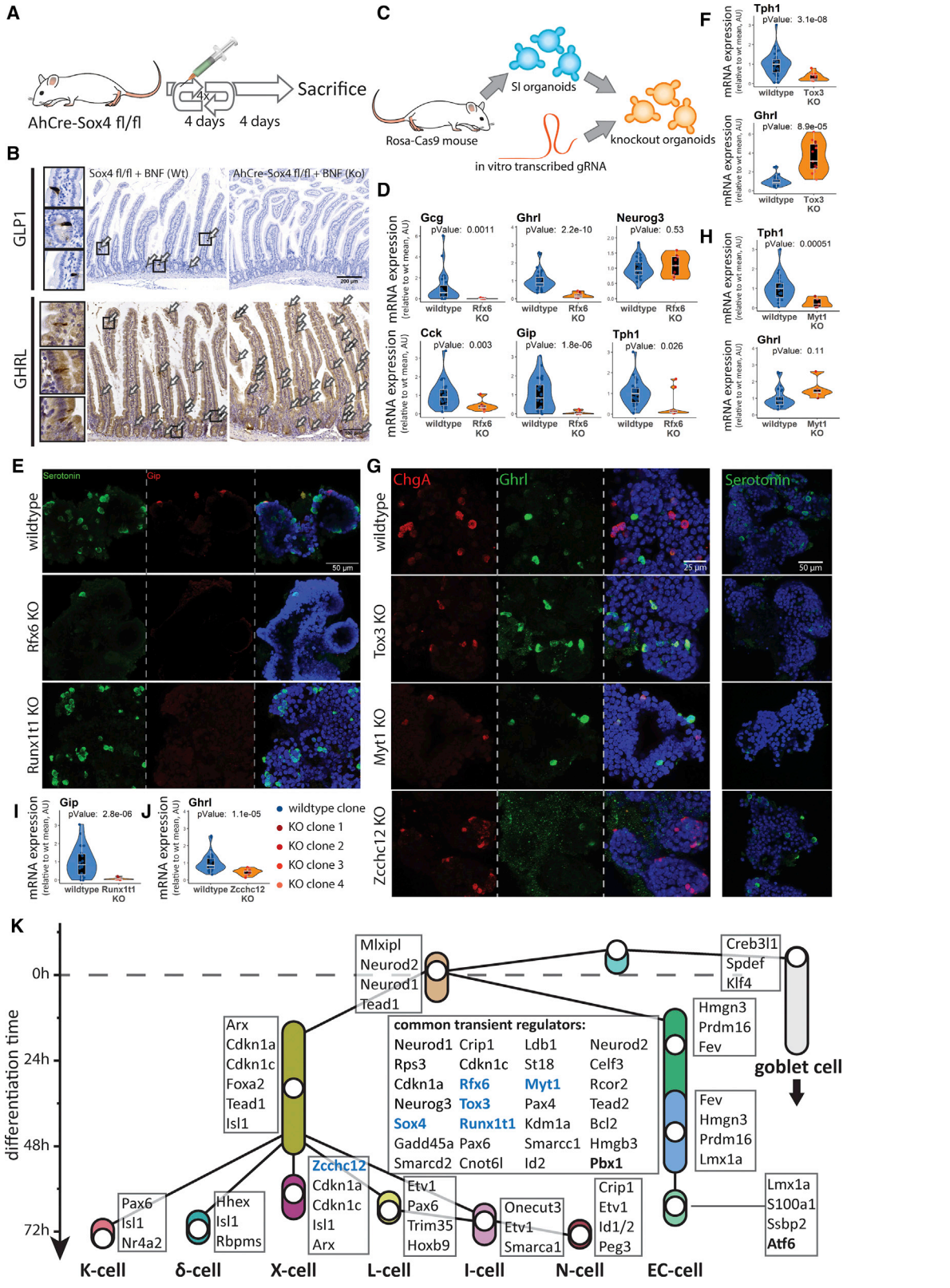
Figure 6. Identification of General and Lineage-Specific Regulators of EE Differentiation

(A) Modulated transcriptional regulators (UniProt: KW-0805) in the EC lineage ordered by peak relative mean transcript count along the time axis (mean modulation > 80%, mean transcripts > 0.9). Colors indicate Loess-smoothed mean expression relative to maximum expression in time. Graphs for other lineages can be found in Figures S5A–S5E.

(B) Lineage-resolved expression profiles of selected genes. Colored lines represent lineage-specific Loess-smoothed expression means. Numbers indicate normalized unique transcript counts.

(C) Differential gene expression between lineages at 3 depicted time intervals ($p < 0.01$, min. FC > 3). Violin/boxplots depict log-transformed normalized unique transcript counts + 0.1.

(D) Lineage-resolved expression profiles of genes chosen for knockout confirmation. Colored lines represent lineage-specific Loess-smoothed expression means. Numbers indicate normalized unique transcript counts.



(legend on next page)

and the temporal expression profile indicated that *Myt1* peaks after *Neurog3* starts to taper off (Figures S7G and S7H). It was thus more likely that *Myt1* acts downstream of *Neurog3* in EE differentiation.

In line with its broad expression during EE lineage specification, loss of *Runx1t1* affected K, I, and delta cells (Figures 7E, 7I, and S7I). The fact that it is also expressed in other lineages (e.g., EC cells), suggests that it is not the primary driver for these fates, but a necessary co-factor. Among the lineage-specific candidates, *Atf6* did not show an EE phenotype. However, knockout of *Zcchc12*, a highly lineage-specific transcription factor in X cells, led to strongly reduced levels of *Ghrl* expression and significant loss of X cells (Figures 7G and 7J).

Thus, we demonstrate that 6 out of 9 candidate genes identified by real-time-resolved, lineage-specific, single-cell RNA sequencing show significant EE phenotypes upon knockout in mice or organoids. This result showcases the ability of our method to correctly identify important transiently expressed genes that would not be highlighted in a conventional differential expression analysis. Finally, we summarized our results in a time-resolved EE differentiation tree that depicts common and lineage-specific transcriptional regulators (Figure 7K).

DISCUSSION

Our method of real-time resolved single-cell transcriptomics is applicable for all cellular processes that display transient activation of a marker gene. Due to the choice of fluorescent proteins, our reporter construct provides high signal-to-noise ratio even for lowly expressed genes without interfering with gene function. Furthermore, the observed time frame is tunable by exchanging the first amino-acid after the N-degron of mNeonGreen. This makes the technology suitable for a wide range of applications from studying short-term oscillations of cellular signals to long differentiation processes lasting more than a week. The system is especially well suited to study rare cell types with complex differentiation dynamics, such as the EE lineage.

Recently, two other high-profile publications surveyed the EE compartment at a single-cell level (Haber et al., 2017; Yan et al., 2017). Using a conventional differential gene expression analysis, Haber et al. (2017) identified many genes that we see upregulated in mature populations. However, due to the lack of time information, these authors did not pick up transiently expressed regulators such as *Tox3*, *Runx1t1*, *Rfx6*, or *Myt1*, all of which yield EE phenotypes when knocked out. Additionally, Haber et al. (2017) created a new lineage nomenclature based on observed hormone co-expression. Our method demonstrates, however, that some of the observed subtypes (e.g., "EC" and "EC Reg4") are not separate lineages but are consecutive stages in EE cell maturation. The study of Yan et al. (2017) showed that *Prox1*+ and *Bmi1*+ EE cells possess stem cell potential. *Prox1* and *Bmi1* were, according to our dataset, already expressed in early EE progenitors (Figure S5H). Thus, it is not clear whether the crypt repopulation capacity that Yan et al. (2017) observed rests indeed within mature EE cells or in early progenitors. The cluster that Yan et al. (2017) identified as common EE precursor shows a transcriptional profile that our data, based on transcriptome and temporal profile, clearly identified as mature I cells. Due to the appearance of I cells long after *Tac1*+ EC cells, simultaneous with most other EE lineages, this relationship is highly unlikely. This highlights the difficulty of inferring lineage relationships based on transcriptomic data alone. The dataset generated in our study thus appears to be an excellent tool to train a new generation of algorithms for higher-accuracy predictions of pseudo-time and lineage relationships.

In primary tissue, we saw that EE cells do not move with the intestinal "conveyor" belt. Given that most EE cell types appeared only around 60 h after the NEUROG3 pulse, this is likely a biological necessity due to the otherwise exceedingly narrow window of cell activity. On the transcriptome level, our method highlighted not only intrinsic differences in the maturation speed of EE cell types but showcased unexpected plasticity in mature EE populations. Hormonal plasticity is of interest for therapeutic application, as it may present an accessible route to modulate endogenous levels of specific hormones. A proof-of-concept study has already demonstrated that BMP levels can control the hormonal repertoire of EE cells *in vivo* (Beumer et al., 2018). In summary, our study describes a new technique to link real-time with single-cell sequencing information and provides a rich resource to understand and eventually manipulate EE differentiation for scientific and medical purposes.

Figure 7. Confirmation of Candidate Genes

(A) Schematic representation of *Sox4* deletion in AhCre-Sox4$^{loxp/loxp}$ animals. The syringe indicates injection of beta-naphthoflavone (BNF).

(B) Immunohistochemical detection of GLP1 and GHRL in the proximal intestine of the indicated genotypes.

(C) Schematic representation of the knockout strategy in intestinal organoids.

(D) mRNA expression of indicated genes measured by quantitative real-time PCR (qPCR) relative to mean expression of all wild-type clones (3 independent knockout clones, distinguished by different point colors).

(E) Whole-mount immunofluorescent staining of the indicated hormones in wild-type and knockout organoids.

(F) mRNA expression of indicated genes measured by qPCR relative to mean expression of all wild-type clones (4 independent knockout clones, marked by different point colors).

(G) Whole-mount immunofluorescent staining of the indicated hormones in wild-type and knockout organoids.

(H–J) mRNA expression of indicated genes measured by qPCR relative to mean expression of all wild-type clones. (H) 2, (I) 3, and (J) 3 independent knockout clones, marked by different point colors.

(K) Time-resolved EE differentiation tree. Begin and end of colored boxes indicate first and third quantile of time-distribution within the cluster. White circles mark the median. Cluster-specific transcriptional regulators are indicated next to the cluster and common regulators are listed in the middle. Bold Regulators were knocked out, bold blue regulators gave an EE phenotype.

Statistical significance for (D, F, H, I and J) was determined by unpaired Student's t-test.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Mice
- METHODS DETAILS
  - Isolation of single cells from Neurog3Chrono animals or organoids
  - Flow Cytometric Purification
  - Establishing and imaging intestinal Organoid cultures
  - Preparation of intestinal tissue for imaging
  - *In vitro*-transcription of gRNAs
  - CRISPR-Cas9 mediated knockout in intestinal organoids
  - Enteroendocrine differentiation of intestinal Organoids
  - Whole-mount staining of intestinal organoids
  - Deletion of Sox4 in AhCre x Sox4$^{loxp/loxp}$ animals
  - Messenger RNA Sequencing
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Quantification of decay dynamics of Neurog3Chrono
  - Fluorescence time conversion
  - Quantification of Tac1Cre-LSL-tdTomato organoids
  - Quantification of enteroendocrine cell migration
  - Bioinformatics Analysis
  - Statistics
- DATA AND SOFTWARE AVAILABILITY
  - Data Resources

SUPPLEMENTAL INFORMATION

Supplemental Information includes seven figures, three tables, and two videos and can be found with this article online at https://doi.org/10.1016/j.cell.2018.12.029.

ACKNOWLEDGMENTS

We thank Stefan vd Elst and Reinier vd Linden for flow cytometry assistance, Harry Begthel for immunohistochemistry help, Judith Vivié for SORT-seq support, the Utrecht Sequencing Facility (USF) and Genome Scan for sequencing, Anna Alemany Arias and Alexander van Oudenaarden for mapping RNA velocity data, Anko de Graaff and the Hubrecht Imaging Centre (HIC) for microscopy assistance, Jeroen Korving for help with mouse generation, Cayetano Pleguezuelos Manzano for experimental help, and Zoraide Granchi for careful reading of the manuscript. Tac1Cre-LSL-tdTomato organoids were kindly provided by Hongkui Zeng. This work was supported by an EMBO long-term fellowship (ALTF 240-2014 to H.G.) and a VENI grant from the Netherlands Organization for Scientific Research (NWO-ZonMW, 016.166.119 to H.G.) as well as an HFSP fellowship (LT771/2015 to K.K.) and VENI fellowship (NWO-ZonMW, 016.166.140 to K.K.).

AUTHOR CONTRIBUTIONS

H.G. and H.C. conceived the project. H.G. generated the Neurog3Chrono mouse model, designed experiments, interpreted results, and performed bioinformatics analysis. H.G. designed experiments and interpreted results. H.G. and J.H.v.E performed animal experiments. H.G., J.F.D., and A.R. performed histology and imaging experiments. H.G. and K.H. performed all organoid experiments. J.B. provided reagents and conceptional input, K.K. mapped and managed data, and J.H.v.E. generated the *Rosa26-Cas9* mouse model. H.G. and H.C. acquired funding and wrote the manuscript with input from all other authors.

DECLARATION OF INTERESTS

H.C. holds several patents on organoid technology.

Received: April 21, 2018
Revised: September 21, 2018
Accepted: December 17, 2018
Published: January 31, 2019

REFERENCES

Aiken, K.D., and Roth, K.A. (1992). Temporal differentiation and migration of substance P, serotonin, and secretin immunoreactive enteroendocrine cells in the mouse proximal small intestine. Dev. Dyn. *194*, 303–310.

Angerer, P., Haghverdi, L., Büttner, M., Theis, F.J., Marr, C., and Buettner, F. (2016). destiny: diffusion maps for large-scale single-cell data in R. Bioinformatics *32*, 1241–1243.

Barker, N., van Es, J.H., Kuipers, J., Kujala, P., van den Born, M., Cozijnsen, M., Haegebarth, A., Korving, J., Begthel, H., Peters, P.J., and Clevers, H. (2007). Identification of stem cells in small intestine and colon by marker gene Lgr5. Nature *449*, 1003–1007.

Basak, O., Beumer, J., Wiebrands, K., Seno, H., van Oudenaarden, A., and Clevers, H. (2017). Induced Quiescence of Lgr5+ Stem Cells in Intestinal Organoids Enables Differentiation of Hormone-Producing Enteroendocrine Cells. Cell Stem Cell *20*, 177–190.

Beucher, A., Gjernes, E., Collin, C., Courtney, M., Meunier, A., Collombat, P., and Gradwohl, G. (2012). The homeodomain-containing transcription factors Arx and Pax4 control enteroendocrine subtype specification in mice. PLoS ONE *7*, e36449.

Beumer, J., Artegiani, B., Post, Y., Reimann, F., Gribble, F., Nguyen, T.N., Zeng, H., Van den Born, M., Van Es, J.H., and Clevers, H. (2018). Enteroendocrine cells switch hormone expression along the crypt-to-villus BMP signalling gradient. Nat. Cell Biol. *20*, 909–916.

Darwich, A.S., Aslam, U., Ashcroft, D.M., and Rostami-Hodjegan, A. (2014). Meta-analysis of the turnover of intestinal epithelia in preclinical animal species and humans. Drug Metab. Dispos. *42*, 2016–2022.

Du, A., McCracken, K.W., Walp, E.R., Terry, N.A., Klein, T.J., Han, A., Wells, J.M., and May, C.L. (2012). Arx is required for normal enteroendocrine cell development in mice and humans. Dev. Biol. *365*, 175–188.

Egerod, K.L., Engelstoft, M.S., Grunddal, K.V., Nøhr, M.K., Secher, A., Sakata, I., Pedersen, J., Windeløv, J.A., Füchtbauer, E.M., Olsen, J., et al. (2012). A major lineage of enteroendocrine cells coexpress CCK, secretin, GIP, GLP-1, PYY, and neurotensin but not somatostatin. Endocrinology *153*, 5782–5795.

Gracz, A.D., Samsa, L.A., Fordham, M.J., Trotier, D.C., Zwarycz, B., Lo, Y.H., Bao, K., Starmer, J., Raab, J.R., Shroyer, N.F., et al. (2018). SOX4 Promotes ATOH1-independent Intestinal Secretory Differentiation Toward Tuft and Enteroendocrine Fates. Gastroenterology *155*, 1508–1523.

Gribble, F.M., and Reimann, F. (2017). Signalling in the gut endocrine axis. Physiol. Behav. *176*, 183–188.

Gross, S., Garofalo, D.C., Balderes, D.A., Mastracci, T.L., Dias, J.M., Perlmann, T., Ericson, J., and Sussel, L. (2016). The novel enterochromaffin marker Lmx1a regulates serotonin biosynthesis in enteroendocrine cell lineages downstream of Nkx2.2. Development *143*, 2616–2628.

Grün, D., Lyubimova, A., Kester, L., Wiebrands, K., Basak, O., Sasaki, N., Clevers, H., and van Oudenaarden, A. (2015). Single-cell messenger RNA sequencing reveals rare intestinal cell types. Nature *525*, 251–255.

Grün, D., Muraro, M.J., Boisset, J.C., Wiebrands, K., Lyubimova, A., Dharmadhikari, G., van den Born, M., van Es, J., Jansen, E., Clevers, H., et al. (2016). De Novo Prediction of Stem Cell Identity using Single-Cell Transcriptome Data. Cell Stem Cell *19*, 266–277.

Haber, A.L., Biton, M., Rogel, N., Herbst, R.H., Shekhar, K., Smillie, C., Burgin, G., Delorey, T.M., Howitt, M.R., Katz, Y., et al. (2017). A single-cell survey of the small intestinal epithelium. Nature *551*, 333–339.

Habib, A.M., Richards, P., Cairns, L.S., Rogers, G.J., Bannon, C.A., Parker, H.E., Morley, T.C., Yeo, G.S., Reimann, F., and Gribble, F.M. (2012). Overlap of endocrine hormone expression in the mouse intestine revealed by transcriptional profiling and flow cytometry. Endocrinology *153*, 3054–3065.

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.

Halpern, K.B., Shenhav, R., Matcovitch-Natan, O., Toth, B., Lemze, D., Golan, M., Massasa, E.E., Baydatch, S., Landen, S., Moor, A.E., et al. (2017). Single-cell spatial reconstruction reveals global division of labour in the mammalian liver. Nature *542*, 352–356.

Harris, J.A., Hirokawa, K.E., Sorensen, S.A., Gu, H., Mills, M., Ng, L.L., Bohn, P., Mortrud, M., Ouellette, B., Kidney, J., et al. (2014). Anatomical characterization of Cre driver mice for neural circuit mapping and manipulation. Front. Neural Circuits *8*, 76.

Hashimshony, T., Senderovich, N., Avital, G., Klochendler, A., de Leeuw, Y., Anavy, L., Gennert, D., Li, S., Livak, K.J., Rozenblatt-Rosen, O., et al. (2016). CEL-Seq2: sensitive highly-multiplexed single-cell RNA-Seq. Genome Biol. *17*, 77.

Ireland, H., Kemp, R., Houghton, C., Howard, L., Clarke, A.R., Sansom, O.J., and Winton, D.J. (2004). Inducible Cre-mediated control of gene expression in the murine gastrointestinal tract: effect of loss of beta-catenin. Gastroenterology *126*, 1236–1246.

Kassambara, A. (2018). ggpubr: 'ggplot2' Based Publication Ready Plots. Available online: https://cran.r-project.org/web/packages/ggpubr/index.html.

La Manno, G., Soldatov, R., Zeisel, A., Braun, E., Hochgerner, H., Petukhov, V., Lidschreiber, K., Kastriti, M.E., Lönnerberg, P., Furlan, A., et al. (2018). RNA velocity of single cells. Nature *560*, 494–498.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550.

Mellitzer, G., Beucher, A., Lobstein, V., Michel, P., Robine, S., Kedinger, M., and Gradwohl, G. (2010). Loss of enteroendocrine cells in mice alters lipid absorption and glucose homeostasis and impairs postnatal survival. J. Clin. Invest. *120*, 1708–1721.

Miyatsuka, T., Li, Z., and German, M.S. (2009). Chronology of islet differentiation revealed by temporal cell labeling. Diabetes *58*, 1863–1868.

Muraro, M.J., Dharmadhikari, G., Grun, D., Groen, N., Dielen, T., Jansen, E., van Gurp, L., Engelse, M.A., Carlotti, F., de Koning, E.J., et al. (2016). A Single-Cell Transcriptome Atlas of the Human Pancreas. Cell Syst. *3*, 385–394.

Naya, F.J., Huang, H.P., Qiu, Y., Mutoh, H., DeMayo, F.J., Leiter, A.B., and Tsai, M.J. (1997). Diabetes, defective pancreatic morphogenesis, and abnormal enteroendocrine differentiation in BETA2/neuroD-deficient mice. Genes Dev. *11*, 2323–2334.

Penzo-Méndez, A., Dy, P., Pallavi, B., and Lefebvre, V. (2007). Generation of mice harboring a Sox4 conditional null allele. Genesis *45*, 776–780.

Platt, R.J., Chen, S., Zhou, Y., Yim, M.J., Swiech, L., Kempton, H.R., Dahlman, J.E., Parnas, O., Eisenhaure, T.M., Jovanovic, M., et al. (2014). CRISPR-Cas9 knockin mice for genome editing and cancer modeling. Cell *159*, 440–455.

Sato, T., Vries, R.G., Snippert, H.J., van de Wetering, M., Barker, N., Stange, D.E., van Es, J.H., Abo, A., Kujala, P., Peters, P.J., and Clevers, H. (2009). Single Lgr5 stem cells build crypt-villus structures in vitro without a mesenchymal niche. Nature *459*, 262–265.

Scheele, C.L., Hannezo, E., Muraro, M.J., Zomer, A., Langedijk, N.S., van Oudenaarden, A., Simons, B.D., and van Rheenen, J. (2017). Identity and dynamics of mammary stem cells during branching morphogenesis. Nature *542*, 313–317.

Schonhoff, S.E., Giel-Moloney, M., and Leiter, A.B. (2004). Neurogenin 3-expressing progenitor cells in the gastrointestinal tract differentiate into both endocrine and non-endocrine cell types. Dev. Biol. *270*, 443–454.

Shaner, N.C., Campbell, R.E., Steinbach, P.A., Giepmans, B.N., Palmer, A.E., and Tsien, R.Y. (2004). Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp. red fluorescent protein. Nat. Biotechnol. *22*, 1567–1572.

Shaner, N.C., Lambert, G.G., Chammas, A., Ni, Y., Cranfill, P.J., Baird, M.A., Sell, B.R., Allen, J.R., Day, R.N., Israelsson, M., et al. (2013). A bright monomeric green fluorescent protein derived from Branchiostoma lanceolatum. Nat. Methods *10*, 407–409.

Soyer, J., Flasse, L., Raffelsberger, W., Beucher, A., Orvain, C., Peers, B., Ravassard, P., Vermot, J., Voz, M.L., Mellitzer, G., and Gradwohl, G. (2010). Rfx6 is an Ngn3-dependent winged helix transcription factor required for pancreatic islet cell development. Development *137*, 203–212.

Suzuki, K., Harada, N., Yamane, S., Nakamura, Y., Sasaki, K., Nasteska, D., Joo, E., Shibue, K., Harada, T., Hamasaki, A., et al. (2013). Transcriptional regulatory factor X6 (Rfx6) increases gastric inhibitory polypeptide (GIP) expression in enteroendocrine K-cells and is involved in GIP hypersecretion in high fat diet-induced obesity. J. Biol. Chem. *288*, 1929–1938.

Tasaki, T., Sriram, S.M., Park, K.S., and Kwon, Y.T. (2012). The N-end rule pathway. Annu. Rev. Biochem. *81*, 261–289.

van den Brink, S.C., Sage, F., Vértesy, Á., Spanjaard, B., Peterson-Maduro, J., Baron, C.S., Robin, C., and van Oudenaarden, A. (2017). Single-cell sequencing reveals dissociation-induced gene expression in tissue subpopulations. Nat. Methods *14*, 935–936.

Wang, S., Hecksher-Sorensen, J., Xu, Y., Zhao, A., Dor, Y., Rosenberg, L., Serup, P., and Gu, G. (2008). Myt1 and Ngn3 form a feed-forward expression loop to promote endocrine islet cell differentiation. Dev. Biol. *317*, 531–540.

Warnes, G.R., Bolker, B., Bonebakker, L., Gentleman, R., Liaw, W.H.A., Lumley, T., Maechler, M., Magnusson, A., Moeller, S., Schwartz, M., et al. (2016). gplots: Various R Programming Tools for Plotting Data. Available online: https://cran.r-project.org/web/packages/gplots/index.html.

Wickham, H. (2009). Ggplot2: elegant graphics for data analysis (New York: Springer).

Wickham, H. (2011). The Split-Apply-Combine Strategy for Data Analysis. J. Stat. Soft. *40*. https://doi.org/10.18637/jss.v040.i01.

Worthington, J.J., Reimann, F., and Gribble, F.M. (2018). Enteroendocrine cells-sensory sentinels of the intestinal environment and orchestrators of mucosal immunity. Mucosal Immunol. *11*, 3–20.

Yan, K.S., Gevaert, O., Zheng, G.X.Y., Anchang, B., Probert, C.S., Larkin, K.A., Davies, P.S., Cheng, Z.F., Kaddis, J.S., Han, A., et al. (2017). Intestinal Enteroendocrine Lineage Cells Possess Homeostatic and Injury-Inducible Stem Cell Activity. Cell Stem Cell *21*, 78–90.

Zhang, J., McKenna, L.B., Bogue, C.W., and Kaestner, K.H. (2014). The diabetes gene Hhex maintains δ-cell differentiation and islet function. Genes Dev. *28*, 829–834.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Ghrelin antibody (C-18)	Santa Cruz	Cat # Sc-10368; RRID:AB_2232479
Chga antibody	Labned.com	Cat # LN1401487
Serotonin antibody	Abcam	Cat # Ab66047; RRID:AB_1142794
Gip antibody	Abcam	Cat # Ab22624; RRID:AB_2109683
Cck antibody (C-20)	Santa Cruz	Cat # Sc-21617; RRID:AB_2072464
Glp1 antibody (C-17)	Santa Cruz	Cat # Sc-7782; RRID:AB_2107325
Sst antibody	Novus biologicals	Cat # NB100-91966; RRID:AB_1217955
Alexa Fluor 568 donkey anti-rabbit IgG (H+L)	Thermo Scientific	Cat # A10042; RRID:AB_2534017
Alexa Fluor 568 donkey anti-goat IgG (H+L)	Thermo Scientific	Cat # A11057; RRID:AB_2534104
Alexa Fluor 488 donkey anti-goat IgG (H+L)	Thermo Scientific	Cat # A11055; RRID:AB_2534102
Alexa Fluor 488 donkey anti-rabbit IgG (H+L)	Thermo Scientific	Cat # A21206; RRID:AB_2535792
Dako EnVision+ System- HRP Labeled Polymer Anti-Rabbit	Dako	Cat # K4003; RRID:AB_2630375
Rabbit anti-Goat bridging antibody	Southern Biotech	Cat # 6160-01
Alexa Fluor 647 phalloidin	Thermo Scientific	Cat # A22287; RRID:AB_2620155
Sox4 antibody	Atlas antibodies	Cat # AMAb91380; RRID:AB_2716661
Neurog3 antibody	Developmental Studies Hybridoma Bank	Cat # F25A1B3; RRID:AB_528401
Tox3 antibody	Sigma Aldrich	Cat # HPA040376; RRID:AB_10795522
Myt1 antibody	Sigma Aldrich	Cat # HPA006303; RRID:AB_1079446
Rfx6 antibody	Millipore	Cat # ABD28; RRID:AB_11205418
Chemicals, Peptides, and Recombinant Proteins		
Advanced DMEM/F12	Thermo Scientific	Cat # 12634-010
EGF	Peprotech	Cat # AF-100-15
Noggin conditioned medium	U-Protein Express	Custom order
R-spondin 1 conditioned medium	In-house production	-
N-Acetylcysteine	Sigma Aldrich	Cat # A9165
IWP-2	Stemgent	Cat # 130-105-335
DAPT	Sigma Aldrich	Cat # D5942
Y-27632	Selleckchem	Cat # S1049
PD0325901	Sigma Aldrich	Cat # PZ0162
BME	Amsbio	Cat # 3533-005-02
DAPI	Thermo Scientific	Cat # D1306
TRIzol	Thermo Scientific	Cat # 15596026
SORT-seq reagents	Muraro et al., 2016	-
HEPES	Thermo Scientific	Cat # 15630-56
Glutamax	Thermo Scientific	Cat # 35050-038
Penicillin/Streptomycin	Thermo Scientific	Cat # 15140-122
B27 Supplement	GIBCO	Cat # 17504-44
TrypLE Express	GIBCO	Cat # 12605036
10x TrypLE Select (10x)	GIBCO	Cat # A1217701
Tissue Freezing Medium	Leica	Cat # 14020108926

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Dispase II	Thermo Scientific	Cat # 17105041
EGFP mRNA	Stemgent	Cat # 05-0020
β-Naphthoflavone	Sigma Aldrich	Cat # N3633
Triton X-100	Sigma Aldrich	Cat # X100-100ML
CS&T Research Beads	BD Biosciences	Cat # 650621
ProLong Gold Antifade Mounting Medium with DAPI	Thermo Scientific	Cat # P36935
Critical Commercial Assays		
Thermo Scientific reagents for CEL-Seq2	Hashimshony et al., 2016	N/A
Reagents for library preparation from CEL-Seq2	Hashimshony et al., 2016	N/A
Megashortscript T7 Transcription Kit	Ambion	Cat # AM1354
TransIT mRNA transfection kit	Mirus	Cat # MIR 2225
pGEM-T easy cloning kit	Promega	Cat # A1360
Alexa Fluor 488 Tyramide Superboost kit	invitrogen	Cat # B40941
iQ SybrGreen Supermix	Bio-Rad	Cat # 1708887
Experimental Models: Organisms/Strains		
Neurog3Chrono	This study	Hans Clevers, h.clevers@hubrecht.eu
AhCre	(Ireland et al., 2004)	MGI:3052655
Sox4$^{loxp/loxp}$	(Penzo-Méndez et al., 2007)	Veronique Lefebvre, lefebvv@ccf.org
Tac1Cre LSL-tdTomato	(Harris et al., 2014)	Hongkui Zeng, hongkuiz@alleninstitute.org
Rosa26-Cas9	J.v.Es and H. Clevers, to be published separately	Hans Clevers, h.clevers@hubrecht.eu
Software and Algorithms		
DESeq2 algorithm	(Love et al., 2014)	http://bioconductor.org/packages/release/bioc/html/DESeq2.html
RaceID2 (StemID) algorithm	(Grün et al., 2016)	https://github.com/dgrun/StemID
ImageJ	NIH	https://imagej.nih.gov/ij/
ggplot2	(Wickham, 2009)	https://cran.r-project.org/web/packages/ggplot2/
ggpubr	Kassambara, 2018	https://cran.r-project.org/web/packages/ggpubr/
plyr	(Wickham, 2011)	https://cran.r-project.org/web/packages/plyr/
gplots	Warnes et al., 2016	https://cran.r-project.org/web/packages/gplots/
R	R Foundation	http://www.r-project.org
destiny & dpt	(Angerer et al., 2016; Haghverdi et al., 2016)	https://bioconductor.org/packages/release/bioc/html/destiny.html
RNA velocity (velocyto)	(La Manno et al., 2018)	http://velocyto.org/
RStudio	RStudio	https://www.rstudio.com
Other		
Appliances needed for SORT-seq	Muraro et al., 2016	N/A
SP8 or SP8X confocal microscope	Leica	N/A
Olympus FV3000 microscope	Leica	N/A
FACSAria II	BD	N/A
Complete Dataset	This study	GEO: GSE113561

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for reagents and resources should be directed to the Lead Contact, Hans Clevers (h.clevers@hubrecht.eu).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Mice

Neurog3Chrono mice were generated by homologous recombination in mouse embryonic stem cells. The Chrono-reporter cassette (Figure 1B) was inserted just before the *Neurog3* stop-codon to maintain normal expression of the transcription factor and retain potential regulatory effects of the 3′ UTR. Genotyping primers for Neurog3Chrono can be found in the Table S4. Generation and genotyping of AhCre (Ireland et al., 2004), Sox4$^{loxp/loxp}$ (Penzo-Méndez et al., 2007) and Rosa26-Cas9 (J.H.v.E. et al., unpublished data) and Tac1Cre LSL-tdTomato (Harris et al., 2014) animals is/will be described elsewhere. All alleles were maintained on a mixed C57BL/6 background. All mouse experiments were conducted under a project license granted by the Dier Experiment Commissie / Animal Experimentation Committee (DEC) or Central Committee Animal Experimentation (CCD) of the Dutch government and approved by the Hubrecht Institute Animal Welfare Body (IvD). In all experiments animals were aged between 8 and 24 weeks and littermates were used as controls. Homozygous Neurog3Chrono animals/organoids were used for all experiments. Both male and female mice were used, except for isolation of organoids from Rosa26-Cas9 mice. Only male Rosa26-Cas9 organoids were isolated to simplify the knockout of genes located on the X chromosome.

METHODS DETAILS

Isolation of single cells from Neurog3Chrono animals or organoids

Mouse small intestines were harvested, cleaned, flushed with PBS and separated into proximal, medial and distal parts. Pieces were cut open and villi were scraped off with glass slides and processed separately. After repeated washes, the crypt fraction was incubated with 2 mM EDTA in PBS (without Ca^{2+}, Mg^{2+}) for 30 min to separate crypts from the muscle layer. Subsequently, crypts were mechanically detached, strained, washed and pelleted. The crypt pellet and the washed and pelleted villus fraction were then resuspended in warm TrypLE Express (GIBCO) and digested to single cells at 37°C. Samples were repeatedly taken to avoid over- or under-digestion. Alternatively, pellets were digested for 1.5 min in 10xTrypLE select diluted 1:1 with HBSS with continuous mechanic disruption by pipetting, which generally improves yields but generates larger numbers of doublets. Single cell suspensions were pelleted, washed, strained and resuspended in cold FACS buffer (Advanced DMEM/F12 + 4 mM EDTA).

Neurog3Chrono organoids at 0h, 24h, 48h and 72h after induction of EE differentiation (see ''Enteroendocrine differentiation of intestinal Organoids'' below) were harvested in cold medium, washed and digested to single cells with 10xTrypLE select diluted 1:1 with HBSS with continuous mechanic disruption. Single cell suspensions were pelleted, washed, strained and resuspended in cold FACS buffer (Advanced DMEM/F12 + 4 mM EDTA).

Flow Cytometric Purification

Flow cytometers were calibrated with CS&T beads, to insure reproducibility between experiments. 4',6-diamidino-2-phenylindole (DAPI) was added just before flow sorting. DAPI-negative living cells were sorted into TRIzol reagent (Thermo Scientific) for bulk mRNA-sequencing or into 384-well plates containing 96 or 384 unique molecular identifier (UMI) barcode primer-sets, ERCC92 spike-ins (Agilent) and dNTPs (Promega) for single-cell mRNA-sequencing (SORT-seq) (Muraro et al., 2016) using a FACSAria II (BD). Fluorescence values for all sorted cells were recorded for later analysis. Samples in Trizol or plates were stored at −80°C until further processing.

Establishing and imaging intestinal Organoid cultures

To establish organoids, crypts were isolated as described above, however, crypt pellets were not digested, but resuspended in cold BME and plated at approximately 100 crypts/50 ul drop of BME (see also (Sato et al., 2009)). After 10-20 min in the incubator, full growth medium was added. Growth medium consists of Advanced DMEM/F12 with 50 ng/mL EGF (Peprotech), 5% R-spondin 1 conditioned medium (made in-house), 1% Noggin conditioned medium (U Protein Express), 1x B27 (GIBCO) and 1.25 mM n-Acetylcysteine (Sigma Aldrich). Organoids were maintained at 37°C and medium was changed twice a week. Cultures were split weekly at 1:4 ratio by mechanical disruption of organoids.

For imaging spontaneous enteroendocrine differentiation in Neurog3Chrono organoids, cultures were split and seeded in BME at low density on glass bottom plates three days prior to image acquisition. To reduce potential background, we used phenol-free growth medium. Organoids were imaged for 96 on a Leica SP8 confocal microscope. Data for the fluorescence-time conversion dataset were collected in intervals of 45 min between recorded z stacks of individual organoids in 3 independent experiments.

Imaging of Tac1Cre-LSL-tdTomato mice was performed in the same fashion but imaging was started 3h after induction of EE differentiation and continued for 120h at 30 min intervals.

Preparation of intestinal tissue for imaging

For detection of reporter fluorescence, intestines from Neurog3Chrono animals were harvested and flushed with cold PBS. Subsequently, freshly prepared 4% PFA was flushed through the intestine and the tissue was fixed for 15 min at room temperature. Subsequently, tissue was washed with PBS, flushed with diluted Tissue Freezing Medium (Leica) and frozen in undiluted tissue Freezing Medium (Leica). Sections were cut on a cryotome and mounted with ProLong Gold Antifade Mounting Medium with DAPI (Thermo

Scientific). Alternatively, fixed intestines were embedded in UltraPure Low Melting Point Agarose (Thermo Fischer Scientific) and cut on a Microm HM 650V vibratome. Sections were blocked, incubated with primary antibody overnight and stained with secondary antibodies for 2h on the following day. Finally, sections were mounted with ProLong Gold Antifade Mounting Medium with DAPI (Thermo Scientific) and imaged on a Leica SP8 confocal microscope.

For crypt reconstruction the intestines were dissected, rinsed twice with ice-cold 4% paraformaldehyde, and incubated for 2h in 4% paraformaldehyde. The intestine was then washed with PBT (PBS, 0.1% Tween), dissected into pieces of 5 × 7 mm, and incubated overnight at 4°C with Alexa Fluor-647 Phalloidin (Thermo Fischer Scientific). The following day, the pieces were washed and embedded in 4% UltraPure Low Melting Point Agarose (Thermo Fischer Scientific) before 3D imaging using an Olympus FV3000 confocal microscope.

For immunohistochemistry intestines were fixed overnight in Formalin, dehydrated using ethanol and embedded in paraffin. 4 μm-sections were processed using standard methods and antigen retrieval was performed using citrate buffer (pH 6.0). Dewaxed sections were blocked, incubated with primary antibody overnight at 4°C and then incubated for 1 hour at room temperature with BrightVision poly-HRT anti-rabbit (Immunologic). For primary goat antibodies, a rabbit-anti-goat bridging antibody (1h, room temperature) was used in an intermediate step. Immunoreactivity was visualized using 3,3′-diaminobenzidine (DAB).

For immunofluorescent co-staining of SOX4 and NEUROG3 on paraffin sections, samples were processed as other paraffin samples above with the following changes: primary antibody incubation of highly diluted Neurog3 antibody (1:500) was followed by signal amplification with the Alexa Fluor 488 Tyramide Superboost kit (Invitrogen). Subsequently, SOX4 staining was performed as other immunofluorescent stainings above. This step was necessary since Neurog3 and Sox4 antibody are of mouse origin. Absence of false-positive SOX4 signal in NEUROG3 positive cells was confirmed in tissue sections from *Sox4* KO mice (see also Figure S7B).

In vitro-transcription of gRNAs

Two independent gRNAs targeting candidate genes were designed with the CRISPR Design Tool (http://crispr.mit.edu/). Oligos (Table S4) were synthetized and annealed with a common oligo containing a T7 promoter. Strands were filled in with T4 DNA polymerase (NEB), PCR amplified with Phusion high-fidelity polymerase (NEB) and purified. The resulting amplified DNA was sequenced to confirm correct assembly. 200 ng PCR product were *in vitro* transcribed using the Megashortscript T7 transcription kit (Ambion) following the manufacturer's instructions. Finally, gRNAs were purified using an RNeasy mini kit (QIAGEN) and a modified protocol for small RNAs (QIAGEN Supplementary Protocol: Purification of total RNA containing miRNA from animal cells using the RNeasy Plus Mini Kit).

CRISPR-Cas9 mediated knockout in intestinal organoids

Proximal small intestinal organoids were isolated from male Rosa26-Cas9 animals as described above and cultured for several weeks. 4 days after the last split organoids were harvested and digested to single cells with TripLE (GIBCO). *In vitro* transcribed gRNA (170 ng) was mixed with GFP mRNA (80 ng, Stemgent) in OptiMem (GIBCO) and TransIT mRNA transfection (Mirus) components according to the manufacturer's instructions. RNA complexes were added to organoid single-cell suspensions in growth medium supplemented with 10 uM Y-27632 (Selleckchem) in 96-well plates (approximately 50.000 cells/well) and incubated for 4h at 37°C. After the incubation time, cells were resuspended, strained and GFP positive cells were isolated by FACS. GFP positive single-cells were spun down, seeded at low density in BME and grown into organoids by addition of 50% Wnt conditioned medium (in-house production) and 10 uM Y-27632 (Selleckchem) to the full growth medium for the first 4 days. For every gRNA 16 clones were picked, expanded and genotyped by amplification of the region of interest by PCR and cloning of the PCR products into pGEM-T easy (Promega). Resulting vectors were analyzed by Sanger sequencing to identify potential indels in both Allels of each clone. Clones with confirmed out-of-frame insertions or deletions in both Allels were selected and used for further experiments. All clones were compared to mock-transfected wild-type organoids that underwent the same cloning procedure and were derived from the same organoid isolation.

Enteroendocrine differentiation of intestinal Organoids

Intestinal organoids were switched to enteroendocrine differentiation medium (Basak et al., 2017) 4 days after their last split. Enteroendocrine differentiation medium contains normal growth medium supplemented with 5 uM IWP-2 (Stemgent), 10 uM DAPT (Sigma Aldrich) and 1 uM MEK inhibitor (PD0325901, Sigma Aldrich). 48h after start of differentiation medium was renewed.

For RNA isolation organoids were lysed in 350ul Buffer RLT (RNeasy mini kit (QIAGEN)) and processed following the manufacturer's instructions. For whole-mount stainings, BME was digested by addition of 2 U/mL Dispase II (Thermo) to growth medium for 1h and removed by subsequent repeated washes with cold medium. Then, organoids were incubated in 4% freshly prepared PFA for 30 min, permeabilized with 0.5% Triton X-100 (Sigma) for 20 min and stored in PBS with 1% BSA at 4°C until stained.

Whole-mount staining of intestinal organoids

Organoids were blocked with 1% BSA at room temperature for 1 hour and incubated with primary antibodies in blocking solution overnight. The next day, organoids were washed and incubated with secondary antibody for 1h at room temperature. After washing, organoids were placed in glass-bottom 96-well plates in PBS and imaged on an inverted SP8 confocal microscope (Leica).

Deletion of Sox4 in AhCre x Sox4$^{loxp/loxp}$ animals

For conditional deletion of *Sox4*, AhCre x Sox4$^{loxp/loxp}$ and Cre-negative Sox4$^{loxp/loxp}$ littermates were treated with 4 injections of 100 ul (10 mg/mL) β-Naphthoflavone (Sigma Aldrich) at 4 day intervals. 3 days after the last injection, mouse intestines were harvested and fixed in Formalin or processed for RNA isolation.

Messenger RNA Sequencing

Bulk or single cell samples were lysed, barcoded and processed according to the CEL-Seq2 technique (Grün et al., 2016; Hashimshony et al., 2016; Muraro et al., 2016). In short, bulk samples were sorted into TRIzol and RNA was isolated following the manufacturer's instruction with the exception of adding 2 μg GlycoBlue (Ambion) overnight at −80°C to precipitate RNA. After removal of supernatants RNA precipitates were dissolved in reverse transcription reaction mix (Invitrogen), UMI barcode primers and dNTPs (Promega) were added and the reaction was incubated at 70°C for 2 min. For single-cell mRNA-sequencing, cells were sorted into 384 well plates, containing UMI barcode primers, frozen and incubated at 65°C for 5 min to ensure lysis. First and second strand synthesis (Invitrogen) was performed and all wells of a single plate were pooled. After *in vitro* transcription (Ambion), the amplified RNA was reverse transcribed and amplified for 10-12 cycles with Illumina Truseq primers. Finally, libraries were analyzed on an Illumina NextSeq500 using 75-bp pair-end sequencing.

QUANTIFICATION AND STATISTICAL ANALYSIS

Quantification of decay dynamics of Neurog3Chrono

z stacks taken by confocal live imaging of Neurog3Chrono organoids (25 cells in 8 organoids in 3 independent experiments with identical settings) were summed and reporter positive cells were identified and semi-automatically traced and measured over time in ImageJ. Cells that could not be clearly separated from fluorescent neighbors were excluded from the analysis. Fluorescence raw data was imported into R and normalized per cell after background subtraction (based on the background signal of each individual frame). Cell data were aligned in time based on peak mNeonGreen fluorescence. Time-dependent mean fluorescence and 95% confidence intervals were calculated and depicted in Figure 1F. To compare protein stability of each color, mean fluorescence decay rates after peak fluorescence of each color were compared and approximated by linear regression.

Fluorescence time conversion

Reference fluorescence values (see "Quantification of decay dynamics of Neurog3Chrono") for mNeonGreen and dTomato from live-imaged Neurog3Chrono organoids were normalized relative to maximum intensity in each channel and log transformed. Fluorescence values from flow cytometry (see "Flow Cytometric Purification") were equally normalized and log transformed and subsequently overlaid and aligned with the reference data. On the resulting mNeonGreen versus dTomato intensity graph (which resembles the original mNeonGreen versus dTomato flow cytometry plot with overlaid reference data) the central point in the reference value distribution is chosen. The angle between the vector (centerpoint to datapoint) and the vector (−1,-1) was determined for each reference datapoint and each point of flow cytometry data. Since the time value of each reference datapoint was known, a correlation between angle and time could be established and fitted by non-linear regression to a logistic curve. All unknown time values for flow cytometry datapoints were then interpolated.

Quantification of Tac1Cre-LSL-tdTomato organoids

z stacks for each time point were analyzed with the ImageJ Plugin 3D Object counter (included in Fiji). The object detection cutoff was set to 400 to exclude small fluorescent bodies, such as apoptotic remnants. The volume of tdTomato positive voxels was summed per time point, normalized to maximum values and plotted against time.

Quantification of enteroendocrine cell migration

Crypt axes were established and the relative position of Neurog3Chrono-positive cells between crypt bottom and crypt top along the crypt axes was semi-automatically measured in ImageJ. Fluorescence values (mNeonGreen and dTomato) of each individual cell were quantified and used to calculate time as described above.

Bioinformatics Analysis

For detailed information on DNA library preparation, sequencing, mapping to the mouse reference genome and quantification of transcript abundance please refer to (Muraro et al., 2016).

Bulk sequencing libraries were analyzed using the DESeq2 package (Love et al., 2014). Time-gate specific genes were determined by differential gene expression analysis ($p < 0.01$, log2FC > 1.5) comparing each time-gate against all others and including potential batch-effects in the model design.

For single cell analysis, cell clusters and expression comparisons between clusters were generated using the RaceID2 algorithm (Grün et al., 2016). For initial cell type analysis (Figure 3A), single-cell libraries were normalized by downsampling to a minimum number of 2,000 unique transcripts/cell. After an initial RaceID2 run cell clusters enriched for *Kcnq1ot1*, an indicator for necrotic cells and low quality RNA, were removed from the dataset. Likewise, cells with high auto-fluorescence (according to Index sort files) or raw

transcript counts higher than 1.5 times the interquartile range above the upper quartile of the rest of the population were removed as potential doublets. Mitochondrial genes, ERCC92 spike-ins as well as genes associated with clustering artifacts (*Rn45s*, *Malat1*, *Kcnq1ot1*, *A630089N07Rik*, and *Gm17821*) were excluded from the final dataset (Grün et al., 2015; Scheele et al., 2017). RaceID2 was re-run with gap statistics to generate the 2281 cell dataset presented in Figure 3A (RaceID2 settings: MinThresh = 2000, minexpr = 5, minnumber = 2, outminc = 5, probthr = 10^{-4}, outlg = 4). For presentation in the overview figure clusters of the same celltype were merged. For detailed analysis of subpopulations the following clustering parameters were used: Figure 5A (MinThresh = 5000, minexpr = 5, minnumber = 2, outminc = 5, probthr = 10^{-3}, outlg = 2), Figure 5F (MinThresh = 4000, minexpr = 5, minnumber = 2, outminc = 5, probthr = 10^{-4}, outlg = 4). For clustering the combined tissue-organoid dataset the following settings were used: Figure S7B (MinThresh = 4000, minexpr = 5, minnumber = 2, outminc = 5, probthr = 10^{-4}, outlg = 4)

Differentially expressed genes between clusters or cell subsets were determined as described in (Muraro et al., 2016). In short, a negative binomial distribution was calculated reflecting the gene expression variability within each cluster based on the background model for the expected transcript count variability. With these distributions p values were calculated and corrected for multiple testing by the Benjamini-Hochberg method. Cutoffs for adjusted p values and fold changes are given in the corresponding figure legend.

For diffusion map analysis, data were normalized via RaceID2 (as above, min. Threshold of 4000 unique transcripts/cell) and analyzed by principal component analysis. Diffusion components were then calculated from principal components and displayed in 2- or 3-dimensional diffusion maps.

For gene expression analysis in a lineage over time (e.g., Figure 6A), cells of all clusters associated with the lineage in question were arranged according to their age after peak NEUROG3 expression. Temporal outliers were identified (all cells of the previous cluster older than the 33% time quantile of the subsequent cluster) and excluded from the analysis. The expression profile of an individual gene in a particular lineage over time was then generated by Loess regression on normalized unique transcript counts of each individual cell along the time axis. Transcriptional regulators were identified by comparison to all genes associated with UniProt: KW-0805. Only genes that showed at least 80% modulation in amplitude of the Loess smoothed mean and had a minimal mean expression level of 0.9 were included in the list of modulated transcriptional regulators.

For RNA velocity analysis, Fastq files were trimmed with TrimGalore-0.4.3 and mapped with STAR-2.5.3a to the mouse reference genome (mm10). Introns and exons were extracted from the corresponding annotation files. Analysis was performed according to (La Manno et al., 2018). In short, spliced and unspliced unique transcript counts of the gene (e.g., *Chgb*) and cell population (e.g., EC cells) in question were extracted and plotted as a phase diagram. The gamma parameter was fit by least-squares using an extreme quantile fit. Unspliced (u) residuals were then calculated for every cell and plotted on a tSNE map.

All bioinformatics analysis was performed using R version 3.4.0 (R Foundation, https://www.r-project.org) and RStudio version 1.0.143 (https://www.rstudio.com).

Statistics

No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to the sample allocation during experiments and outcome assessment. No animals were excluded from analysis. All data are presented as mean ± standard deviation (SD), unless stated otherwise. Statistically significant differences between wild-type and knockout clones were determined using an unpaired two-tailed Student's t test. Data was obtained from at least two independent clones per knocked out gene and from at least two independent experiments to ensure reproducibility. All statistical analyses were performed using R version 3.4.0 (R Foundation, https://www.r-project.org) and RStudio version 1.0.143 (https://www.rstudio.com).

DATA AND SOFTWARE AVAILABILITY

Data Resources

Differentially expressed genes between sorting gates (Figure 2A) are provided in Table S1. A chronologically ordered list of transcriptional regulators in each lineage is provided in Table S2. Differentially expressed genes between mature cell clusters are provided in Table S3. RNA-sequencing data, time information, and fluorescence data are available under Gene Expression Omnibus ID GEO: GSE113561.

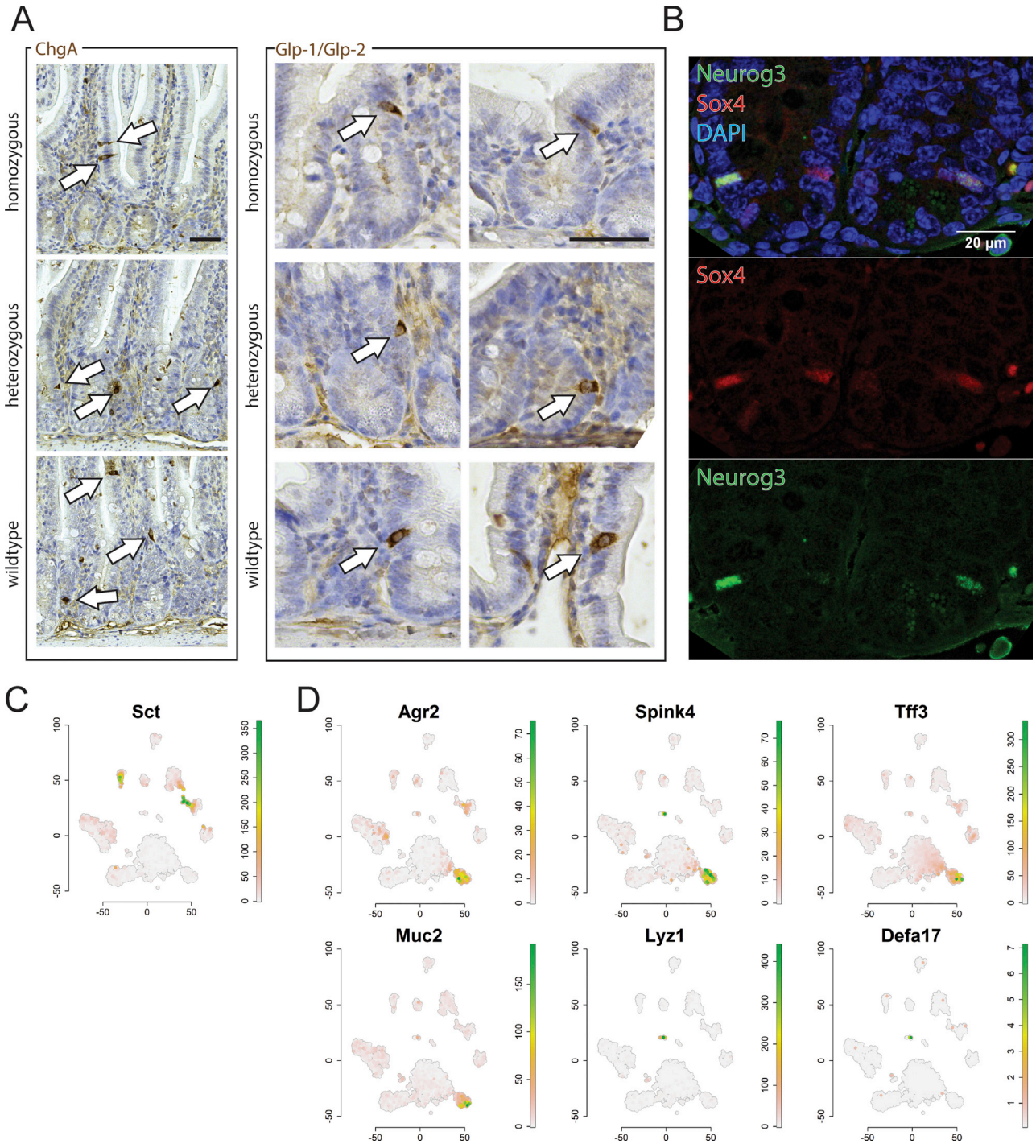


Figure S1. Confirmation of Neurog3 Function and Candidate Gene Identification, Related to Figures 1, 2, and 3

(A) Immunohistochemistry for CHGA and GLP1 in the proximal intestine of wild-type, heterozygous and homozygous Neurog3Chrono animals. Arrows indicate stained cells. (scale bar, 50 µm).

(B) Immunofluorescent co-staining of NEUROG3 and SOX4 in mouse small intestine.

(C) tSNE map indicating expression of Secretin (*Sct*). Numbers represent normalized unique transcript counts.

(D) tSNE map indicating expression of goblet cell and Paneth cell markers. Numbers represent normalized unique transcript counts.

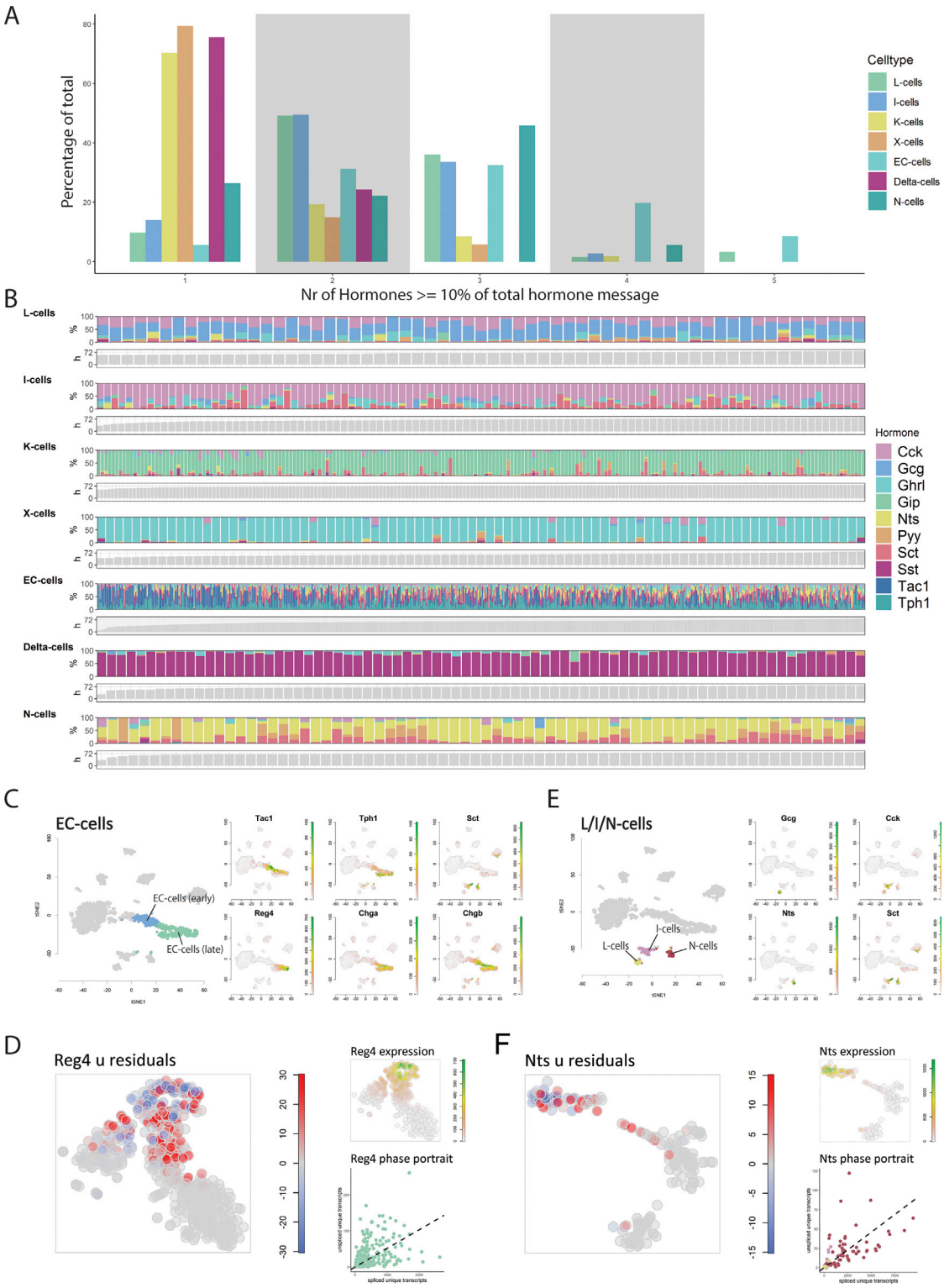


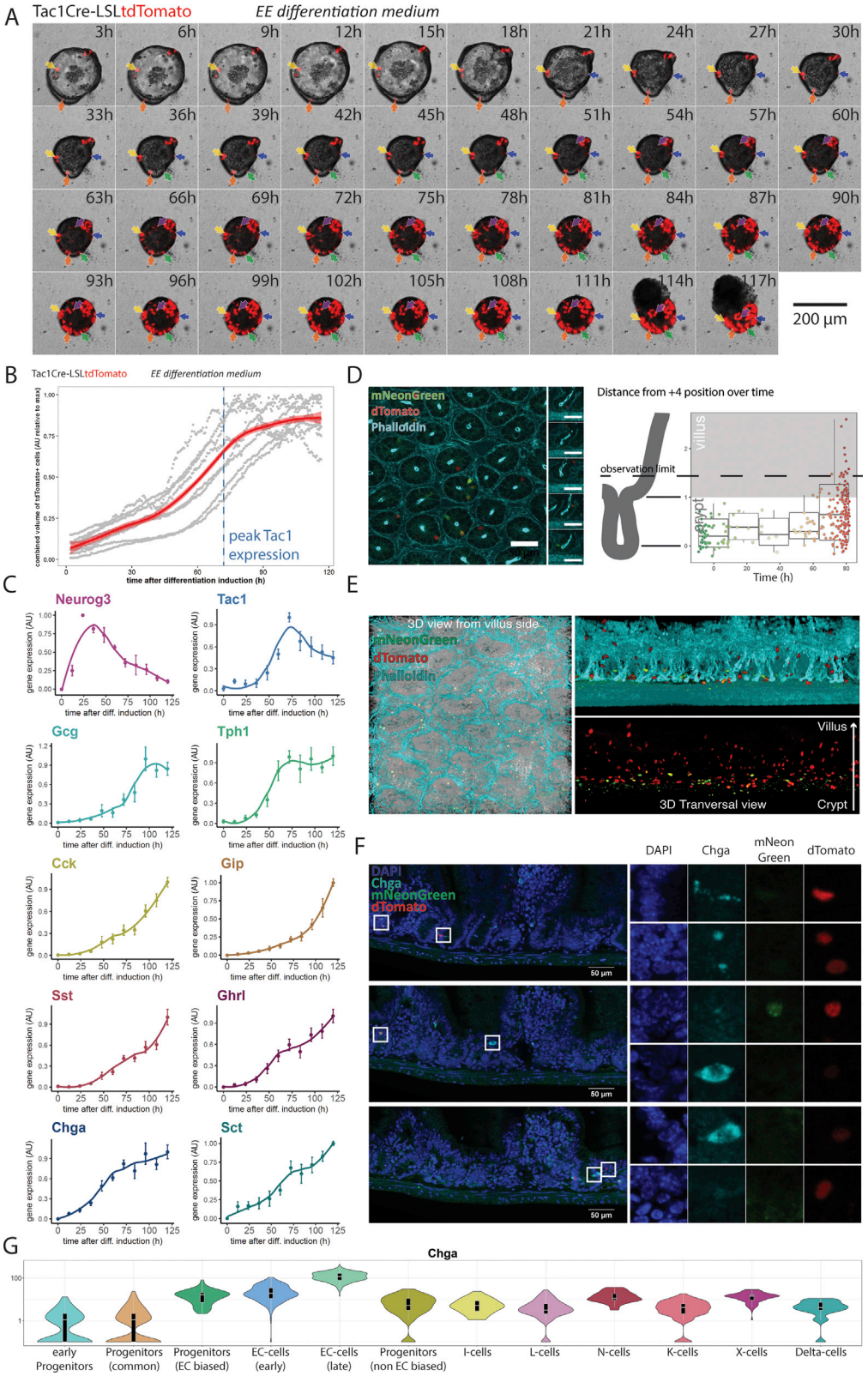
Figure S2. Hormonal Plasticity in the Course of EE Maturation, Related to Figure 4

(A) Quantification of hormonal co-expression in EE cells. The x-Axis indicates the number of different hormones with a contribution of at least 10% to the combined hormonal transcripts in a cell. The y-Axis indicates the percentage of all cells of the same lineage.

(B) Relative contribution of individual hormones to the total hormonal transcript pool in individual cells ordered in time. Please note that apparent higher fluctuations in the EC lineage are due to the generally lower expression of *Tac1*/*Tph1* when compared to other peptide hormones.

(legend continued on next page)

(C) tSNE map of EC clusters and EC markers after clustering with a raised threshold of 4000 normalized unique transcripts/cell.
(D) tSNE map of 461 early and late EC cells (min. 4000 unique transcripts/cell) (top-left, related to Fig. 4E), expression levels of *Reg4* (top-right, normalized unique transcript counts), Phase portrait (bottom-right) showing regions of increasing (over dotted line) or decreasing (under dotted line) expression based on unspliced/spliced mRNA balance (colors correspond to clusters), Unspliced unique transcript count residuals projected on the tSNE map (red indicates high relative levels of unspliced mRNA = begin of gene expression, blue indicates low relative levels of unspliced mRNA = downregulation of gene expression, left).
(E) tSNE map of L-, I- and N-cell clusters and hormonal products after clustering with a raised threshold of 4000 normalized unique transcripts/cell.
(F) as (D) but for *Nts* mRNA in LIN cells (related to Fig. 4H).



(legend on next page)

Figure S3. Organoid Differentiation Dynamics and Migratory Behavior of EE cells *In Vivo*, Related to Figure 4

(A) Live imaging of Tac1Cre LSL-tdTomato small intestinal organoids. EE differentiation was started at t = 0h. Arrows track individual tdTomato positive cells over time.

(B and C) (B) Combined volume of Tac1Cre LSL-tdTomato positive cells over time during EE differentiation; the red line indicates loess smoothed means with 99% confidence interval; the blue line marks the time point of maximal *Tac1* expression according to (C) hormone expression during a 120h, 12h interval EE differentiation time course in organoids (qPCR, mean values relative to max +/- SEM).

(D) (Left) Confocal image of cleared Neurog3Chrono SI crypts. Green and red mark reporter fluorescence. Phalloidin stain outlines crypt structure in turquoise. Small images represent 3D reconstructed side view for positional quantification. (right) Quantification of the distance of Neurog3Chrono positive cells from the bottom of the crypt. 0 indicates the inner crypt bottom, 1 the transition point to the villus (n = 253 cells).

(E) 3D reconstruction of Neurog3Chrono crypts.

(F) Immune-fluorescent staining for CHGA in small intestine vibratome sections of homozygous Neurog3Chrono mice.

(G) Expression of CHGA in all EE populations of the high sensitivity EE single cell dataset (min. 4000 unique transcripts/cell). Numbers indicate unique transcripts + 0.1.

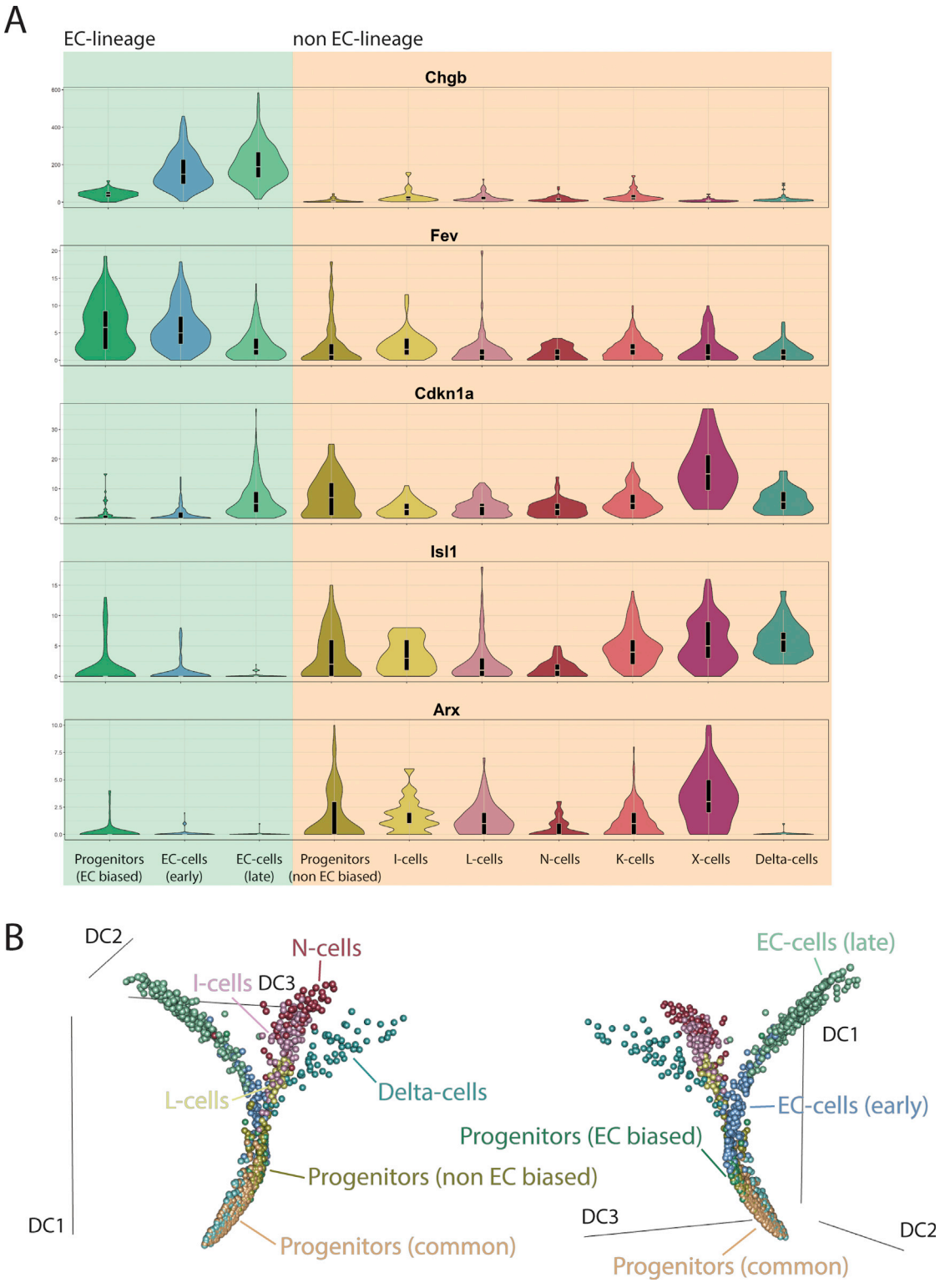


Figure S4. Characteristics of Individual EE Lineages, Related to Figure 5

(A) Violin/Boxplots of genes differentially expressed between EC and non-EC lineages. Numbers indicate normalized unique transcript counts.

(B) Diffusion map visualization showing all EE progenitors in combination with all cells of the EC, ILN- and Delta-cell lineages (min. 4000 unique transcripts/cell).

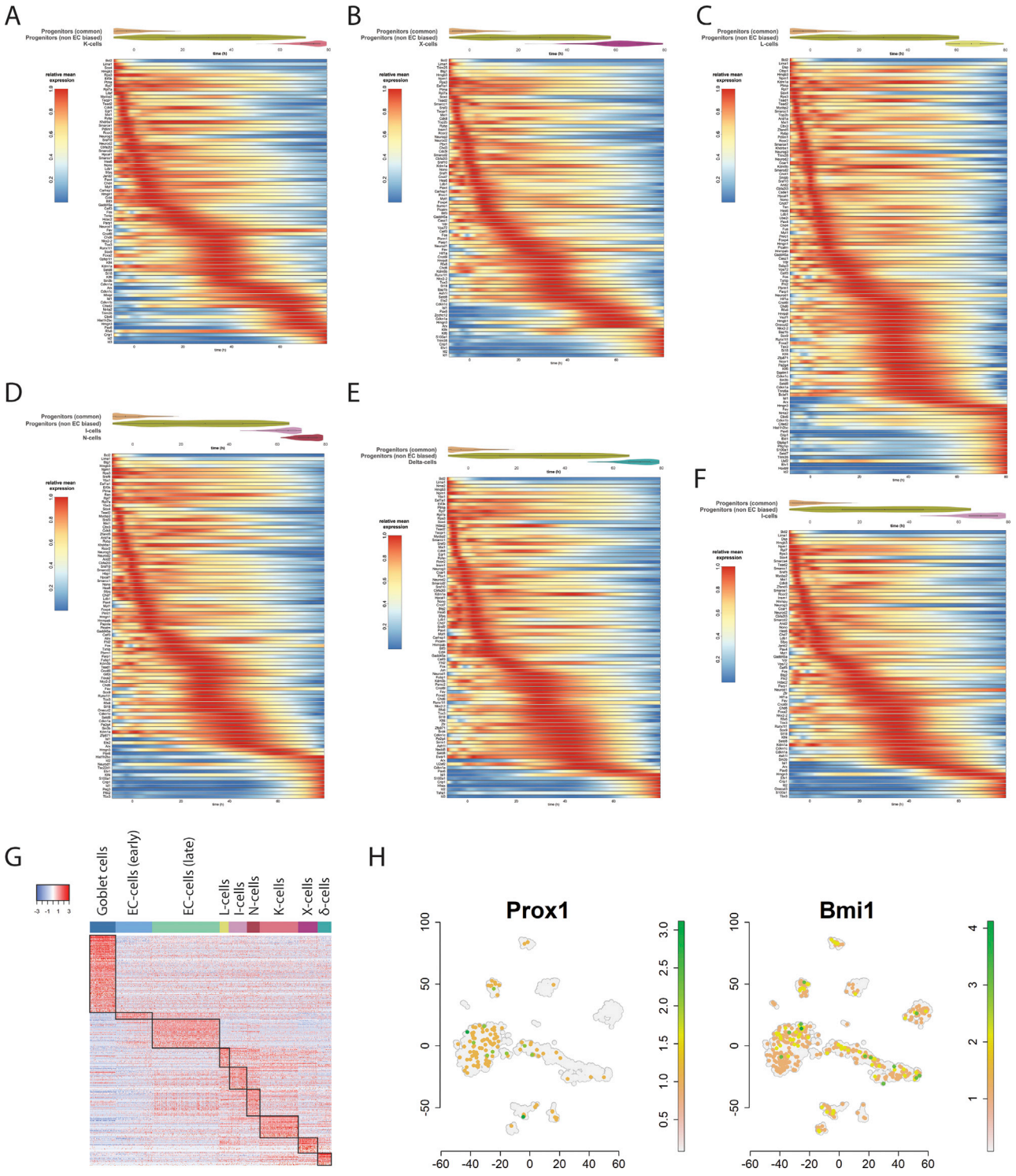


Figure S5. Sequential Expression of Transcriptional Regulators during EE Differentiation, Related to Figure 6)

(A–F) Modulated transcriptional regulators (UniProt: KW-0805) in the (A) K-cell, (B) X-cell, (C) L-cell, (D) N-cell, (E) Delta-cell or (F) I-cell lineage ordered by peak relative mean transcript count along the time axis (mean modulation > 80%, mean transcripts > 0.9). Colors indicate Loess-smoothed mean expression relative to maximum expression in time.

(G) Differential gene expression analysis between mature clusters. Relative expression level (row-wise *Z* score of log2 transformed normalized unique transcript counts + 0.1, $p < 0.05$, fc > 2) across cells (columns). A list of identified genes can be found as Table S3.

(H) tSNE map indicating expression of *Prox1* and *Bmi1*. Numbers represent normalized unique transcript counts.

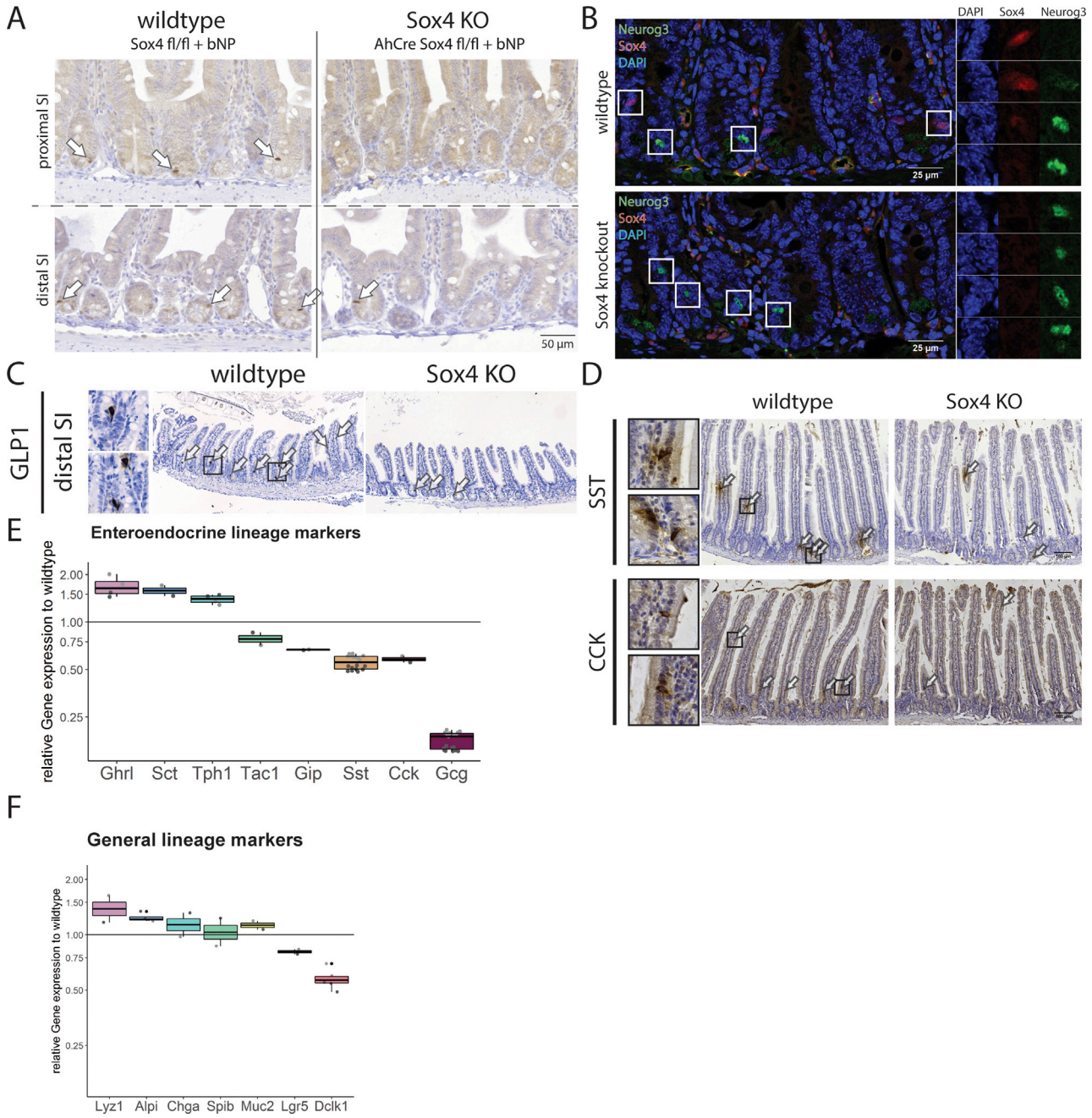


Figure S6. Confirmation of *Sox4* as EE Regulator, Related to Figure 7

(A) Immunohistochemistry staining for SOX4 in proximal and distal SI in wild-type and *Sox4* KO mice.

(B) Immunofluorescent co-staining of NEUROG3 and *Sox4* in proximal small intestine of wild-type and *Sox4* KO mice.

(C) Immunohistochemistry staining for GLP1 in distal SI of wild-type and *Sox4* KO mice.

(D) Immunohistochemical detection of SST and CCK in the proximal intestine of the indicated genotypes.

(E and F) (E) EE and (F) general marker expression measured in microarray of wild-type and *Sox4* knockout small intestinal RNA. Expression levels are denoted relative to wild-type. Dots mark independent probes; dot color distinguishes independent experiments.

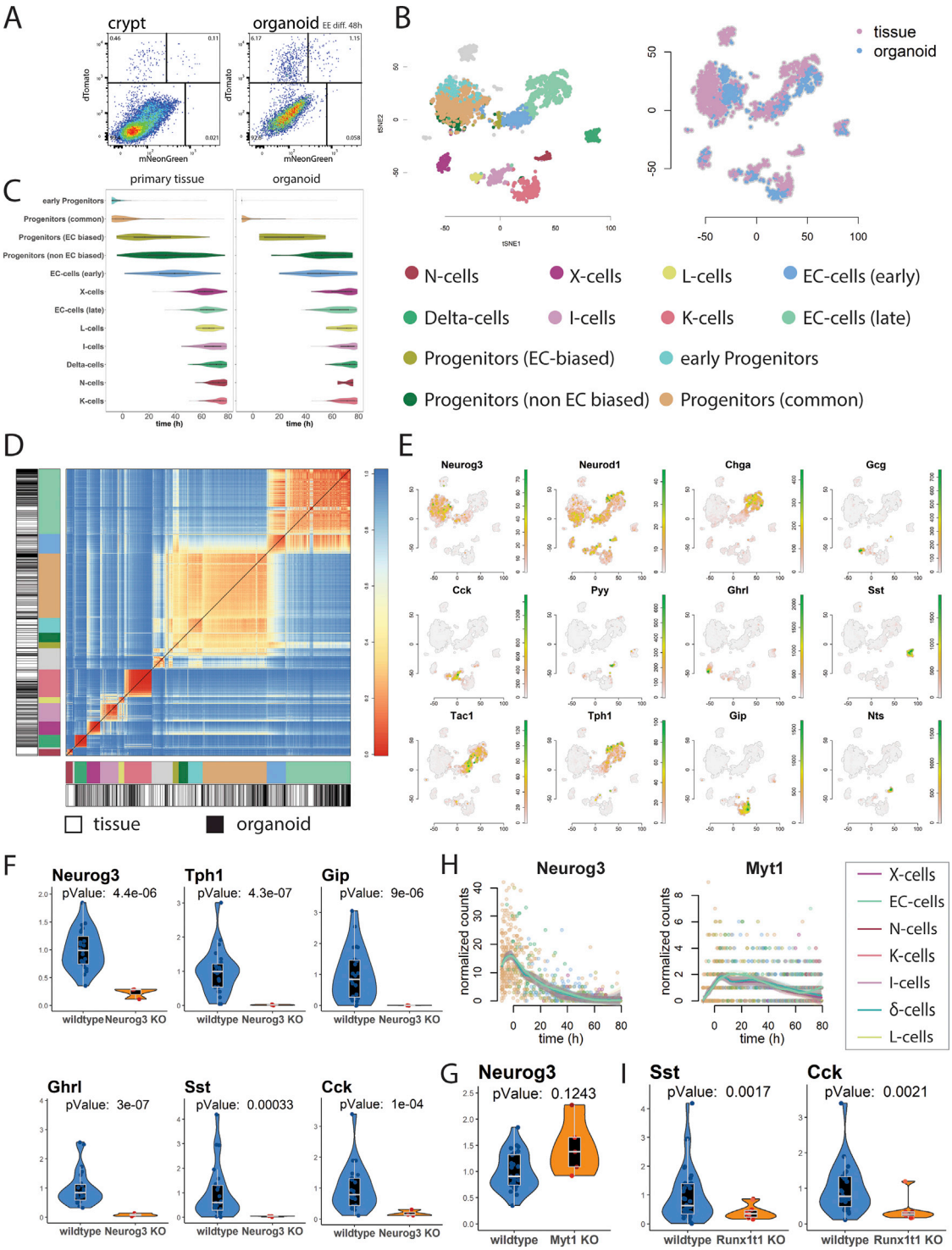


Figure S7. Confirmation of Candidate Genes in Intestinal Organoids, Related to Figure 7

(A) Comparative flow cytometry of isolated small intestinal crypt cells of homozygous Neurog3Chrono mice and cells from homozygous Neurog3Chrono organoids 48h after start of EE differentiation.

(B) tSNE map of a combined dataset of 1750 Neurog3Chrono positive cells and 950 Neurog3Chrono positive cells from intestinal organoids sorted 0h, 24h, 48h and 72h after induction of differentiation. (left) cluster identity (right) cell origin.

(C) Violin/Boxplots of time distribution within clusters in (B) separated by cell origin.

(D) Cluster-heatmap corresponding to (B) with annotated cluster identity and cell origin.

(legend continued on next page)

(E) tSNE maps indicating expression of EE marker genes. Numbers represent normalized unique transcript counts.
(F) mRNA expression of indicated genes in *Neurog3* knockout organoids measured by qPCR relative to mean expression of all wild-type clones. Points denote independent experiments.
(G) mRNA expression of *Neurog3* in 2 independent *Myt1* knockout clones (point color) measured by qPCR relative to mean expression of all wild-type clones.
(H) Lineage-resolved expression profiles of *Neurog3* and *Myt1*. Colored lines represent lineage-specific Loess-smoothed expression means. Numbers indicate normalized unique transcript counts.
(I) mRNA expression of indicated genes in 3 independent *Runx1t1* knockout clones (point color) measured by qPCR relative to mean expression of all wild-type clones.
Statistical significance for (F, G and I) was determined by unpaired Student's t-test.