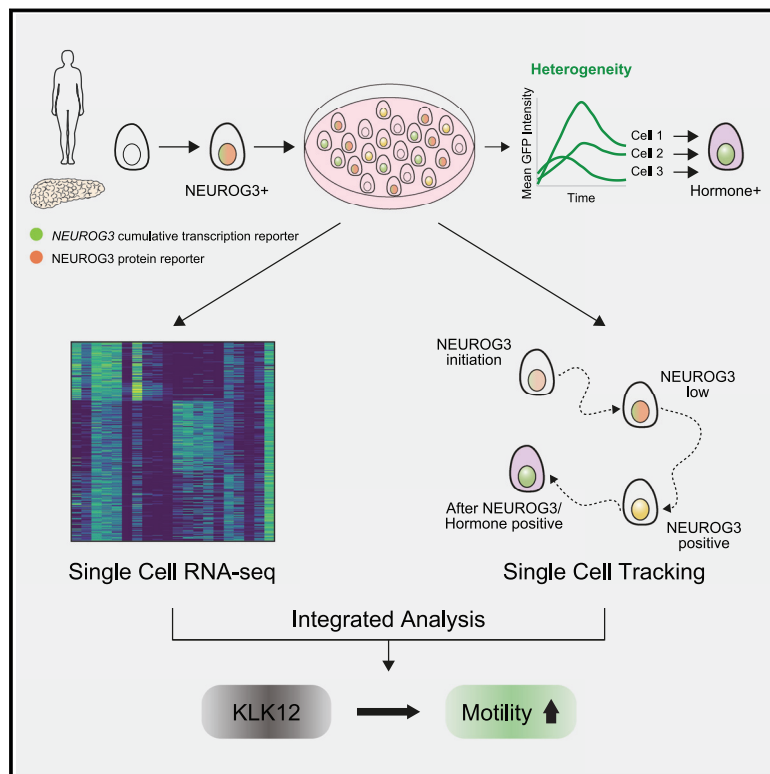


Developmental Cell

Integrating single-cell imaging and RNA sequencing datasets links differentiation and morphogenetic dynamics of human pancreatic endocrine progenitors

Graphical abstract



Authors

Belin Selcen Beydag-Tasöz,
 Joyson Verner D'Costa,
 Lena Hersemann, ..., Yung Hae Kim,
 Christoph Zechner,
 Anne Grapin-Botton

Correspondence

botton@mpi-cbg.de

In brief

Beydag-Tasöz et al. characterize the differentiation dynamics of single human pancreatic progenitors into endocrine cells in 2D and 3D models by RNA sequencing and live imaging. Statistical integration of these datasets identifies a role for *KLK12* in motility at *NEUROG3* expression onset, while also providing a methodology applicable to other contexts.

Highlights

- The pro-endocrine transcription factor *NEUROG3* has heterogeneous peak levels
- All *NEUROG3* peak levels can trigger differentiation into hormone-expressing cells
- Data-mapping methodology can link dynamic behaviors to transcriptomic signature
- *KLK12* promotes motility at the onset of *NEUROG3* expression in humans



Article

Integrating single-cell imaging and RNA sequencing datasets links differentiation and morphogenetic dynamics of human pancreatic endocrine progenitors

Belin Selcen Beydag-Tasöz,^{1,2} Joyson Verner D'Costa,² Lena Hersemann,² Byung Ho Lee,² Federica Luppino,^{2,3} Yung Hae Kim,^{1,2} Christoph Zechner,^{2,3,4} and Anne Grapin-Botton^{1,2,3,4,5,6,*}

¹The Novo Nordisk Foundation Center for Stem Cell Biology, Copenhagen 2200, Denmark

²Max Planck Institute of Molecular Cell Biology and Genetics, Dresden, Saxony 01307, Germany

³Center for Systems Biology Dresden Dresden 01307, Germany

⁴Cluster of Excellence Physics of Life, TU Dresden, Dresden 01062, Germany

⁵Paul Langerhans Institute Dresden of the Helmholtz Zentrum München at the University Clinic Carl Gustav Carus of Technische Universität Dresden, Helmholtz Zentrum München, Neuherberg, Germany

⁶Lead contact

*Correspondence: botton@mpi-cbg.de

<https://doi.org/10.1016/j.devcel.2023.07.019>

SUMMARY

Basic helix-loop-helix genes, particularly proneural genes, are well-described triggers of cell differentiation, yet information on their dynamics is limited, notably in human development. Here, we focus on Neurogenin 3 (*NEUROG3*), which is crucial for pancreatic endocrine lineage initiation. By monitoring both *NEUROG3* gene expression and protein in single cells using a knockin dual reporter in 2D and 3D models of human pancreas development, we show an approximately 2-fold slower expression of human *NEUROG3* than that of the mouse. We observe heterogeneous peak levels of *NEUROG3* expression and reveal through long-term live imaging that both low and high *NEUROG3* peak levels can trigger differentiation into hormone-expressing cells. Based on fluorescence intensity, we statistically integrate single-cell transcriptome with dynamic behaviors of live cells and propose a data-mapping methodology applicable to other contexts. Using this methodology, we identify a role for *KLK12* in motility at the onset of *NEUROG3* expression.

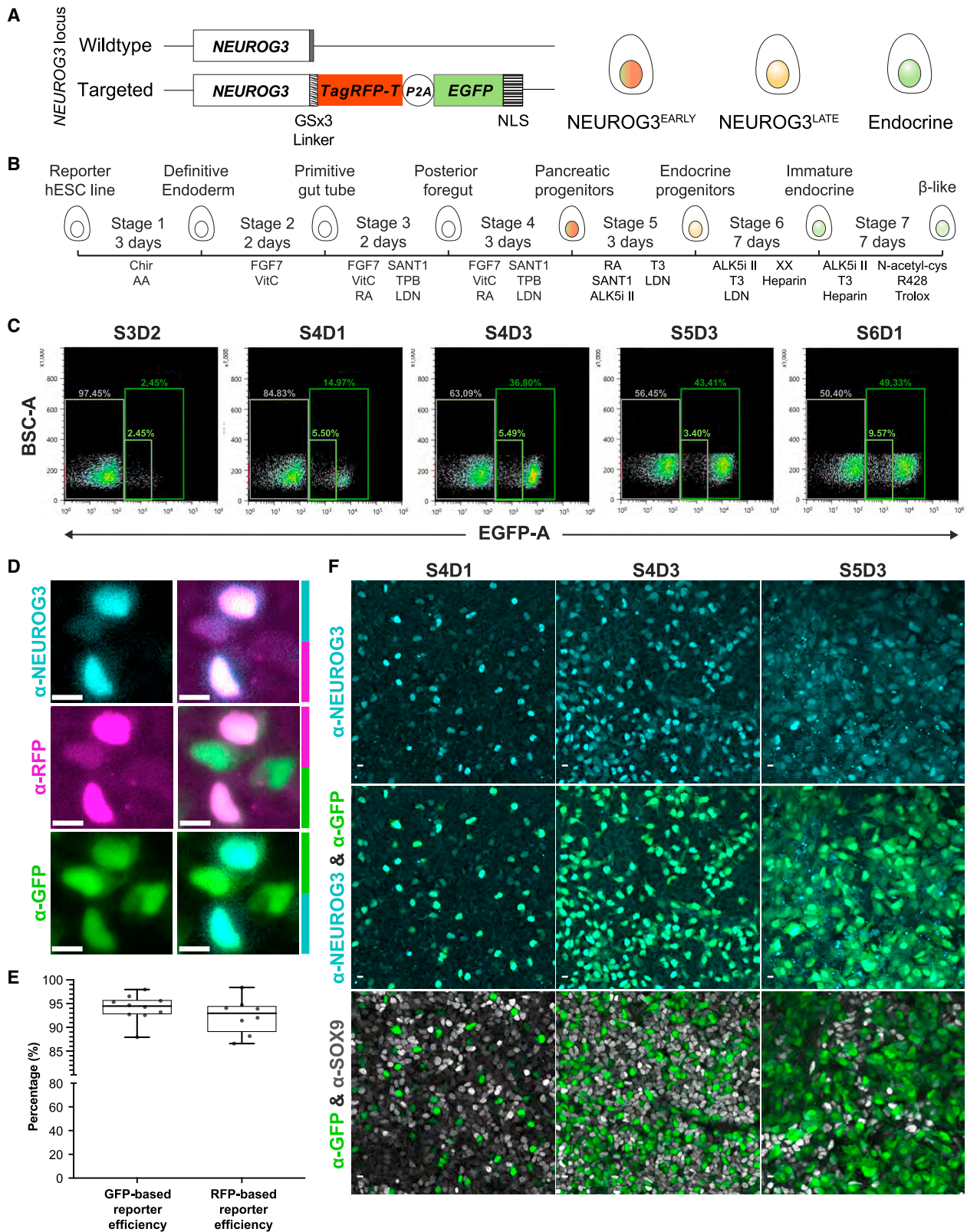
INTRODUCTION

Pancreatic organogenesis and lineage hierarchy are generally considered to be similar between humans and mice.^{1,2} In the context of endocrine differentiation, transient expression of the basic-helix-loop-helix (bHLH) transcription factor Neurogenin 3 (*NEUROG3*) initiates endocrine lineage formation in both species.^{3–11} *Neurog3* knockout mice fail to produce any pancreatic endocrine cells and die postnatally from diabetes mellitus.³ Similarly, patients with mutations severely impairing *NEUROG3* function in humans, develop diabetes mellitus early in life.^{9,10,12} However, some other aspects of endocrine pancreas development are not conserved between humans and mice. One example is the biphasic *Neurog3* expression observed in mice,¹³ which does not seem to exist in humans.^{1,14} Therefore, studying human pancreatic development is essential to uncover interspecies differences. We previously showed that *Neurog3* reaches its peak expression level in about 12 h in single cells in mouse embryonic pancreas explants.¹⁵ In comparison, the duration of *NEUROG3* transcript and protein expression in single cells has not yet been measured in a human system. With the increased levels of *NEUROG3*, the mouse endocrine progenitors start to exit

the cell cycle, as *NEUROG3* is known to induce the expression of cell-cycle inhibitors such as cyclin-dependent kinase inhibitor 1a (*Cdkn1a*) and p21 (RAC1) activated kinase 3 (*Pak3*).^{16,17} However, recent studies have shown that cells with low levels of *Neurog3* mRNA can proliferate, but the extent of this proliferation is not clear.¹⁸ Whether such proliferative, low *NEUROG3*-expressing cells exist in humans as well has not yet been addressed, but this is a key question as a proliferative population committed to the endocrine lineage would be interesting to propagate with the aim of producing β cells for cell therapy.

In this study, we set out to determine the kinetics of human endocrine progenitor differentiation and investigate their proliferation capacity using two-dimensional (2D) and three-dimensional (3D) *in vitro* human culture systems to model early steps of human pancreatic endocrine development.^{19–21} Moreover, we engineered a tagged human *NEUROG3* protein that enabled us to determine the duration of *NEUROG3* expression in human endocrine progenitors. Our construct also monitors *NEUROG3* gene expression, which can be compared with similar reporters in mice.¹⁵ Mapping live imaging to the deep single-cell RNA sequencing (scRNA-seq) when endocrine progenitors emerged allowed us to correlate cellular dynamics to transcriptomics





(legend on next page)

and identified that KLK12 promotes the motility of early *NEUROG3*⁺ endocrine progenitors. This work highlights the importance of using human systems to study pancreatic endocrine cell differentiation for a better control over *in vitro* endocrine cell production for diabetes mellitus therapy.

RESULTS

Generation of a dual reporter line monitoring *NEUROG3* expression and protein in a human system in real time

To visualize the first steps of endocrine pancreas specification in a human system, we inserted a dual reporter of *NEUROG3* in a human embryonic stem cell (hESC) line using CRISPR-Cas9 gene editing in one of the alleles at the endogenous *NEUROG3* locus (Figure 1A). This reporter line produces a fusion protein of *NEUROG3* with TagRFP-T (RFP) to live-monitor the human *NEUROG3* protein and nuclear EGFP (GFP) translated from the same mRNA as a proxy for cumulative expression. Considering that human *NEUROG3* protein can be degraded relatively quickly,^{22,23} the slow degradation of the GFP protein makes it possible to continuously follow the cells that turned off *NEUROG3* expression.¹⁹

Our *NEUROG3* dual reporter hESC line showed no genome abnormality (Figure S1A) and differentiated similarly to the mother H1 ESC line using a 7-stage *in vitro* human pancreatic endocrine differentiation protocol¹⁹ (Figures 1B and S1B). GFP⁺ and RFP⁺ endocrine progenitors started to emerge during stage (S) 4 of differentiation, with their percentage increasing throughout S4 and S5 (Figures 1C and S1C). Notably, the RFP signal was not detectable through flow cytometry, although it was detected by microscopy (Figure 1D). We found that 92% and 94% of *NEUROG3*⁺ cells expressed RFP and GFP, respectively, attesting to the reliability of the reporter (Figures 1D–1F). The reporters were also specific for *NEUROG3*-expressing cells as over 96% of RFP⁺ cells expressed *NEUROG3*.

NEUROG3 expression levels are heterogeneous among human pancreatic cells in 2D and 3D culture conditions

To visualize newly emerging *NEUROG3*⁺ endocrine progenitor cells, we used live imaging from the beginning of S4 (S3D2), at the onset of *NEUROG3* expression (Figures 1C and 2A; Video S1). We observed a high degree of heterogeneity in the *NEUROG3* expression peak level and dynamics (Figures 2B, S2A, and S2B).

Human *NEUROG3*-RFP fusion protein peaked around 11 h after induction but remained above base levels during 26 h of imaging (Figure 2C). Using the GFP expression proxy reporter, we compared the gene expression dynamics from the human *NEUROG3* locus in individual cells to the published work in which mouse *Neurog3* promoter was used to drive turboRFP expression.¹⁵ In human cells, the *NEUROG3* expression proxy reached its peak in around 22 h on average, whereas in the mouse cells, the transcription proxy reached it in 12 h accompanied by a steeper slope of reporter signal accumulation, suggesting a faster process in mice than in the human systems (Figure 2D).

As the reporters in the two species were not exactly the same and the experimental conditions were different, we further assessed the robustness of our findings, using the 3D human pancreatic progenitor spheroid system we previously developed.²¹ In this system, pancreas progenitors expand in 3D, and about 1%–2% of cells spontaneously differentiate into endocrine cells. We thus induced endocrine differentiation in pancreatic progenitor spheroids 48 h after seeding by adding a gamma-secretase inhibitor XX to the culture medium for downregulation of Hes family bHLH transcription factor 1 (*HES1*), an important Notch signaling effector that directly binds to *NEUROG3* promoter and suppresses its transcription.^{24,25} Starting light-sheet microscopy 48 h after treatment (Figure 3A; Video S2), we detected both GFP and RFP signals within the spheroids (Figure 3B; Video S2), indicating initiation of endocrine differentiation. Similar to the 2D conditions, there was a high degree of heterogeneity in the *NEUROG3* expression levels and dynamics in the 3D human pancreatic cells (Figures 3C, S3A, and S3B). The average GFP and RFP fluorescence accumulation rates and dynamics of the human *NEUROG3* fusion protein within single cells were similar in 3D and 2D. *NEUROG3* expression ramped up and culminated around 22 h, and the human *NEUROG3*-RFP fusion protein reached a peak at around 11 h and went back to baseline by 40 h (Figure 3D). This finding suggests that *NEUROG3* expression kinetics may be invariable between these two different differentiation systems, and as long as there is a sufficient trigger for the initiation of promoter activation, *NEUROG3* expression may occur with similar dynamics in individual cells.

Cells with low *NEUROG3* peak levels can differentiate

The observation of cells with heterogeneous *NEUROG3* expression peaks raised the question as to whether cells with low

Figure 1. Construction and characterization of a reliable hESC-based dual reporter line monitoring human *NEUROG3* protein and gene expression

(A) Schematic of dual fluorescent reporter construct introduced in the endogenous *NEUROG3* locus and the fluorescence progression of *NEUROG3*-expressing cells. The upper lane shows the wild-type *NEUROG3* allele, and the lower lane the targeted *NEUROG3* locus where the reporter sequence encoding *TagRFP-T* with a flexible GSx3 linker, a self-cleaving peptide (*P2A*), and *EGFP* tagged with a nuclear localization signal (NLS) was introduced before the stop codon of the *NEUROG3* coding sequence. With this setup, *NEUROG3*-expressing cells are nuclear-labeled with RFP and GFP. As *NEUROG3* is transcribed and translated, EGFP accumulates in the cell, and after *NEUROG3* is downregulated, longer-living EGFP allows the tracing of these cells using green fluorescence.

(B) Illustration of *in vitro* pancreatic differentiation protocol starting from hESCs. The individual stages are indicated along with the timescale and added components.

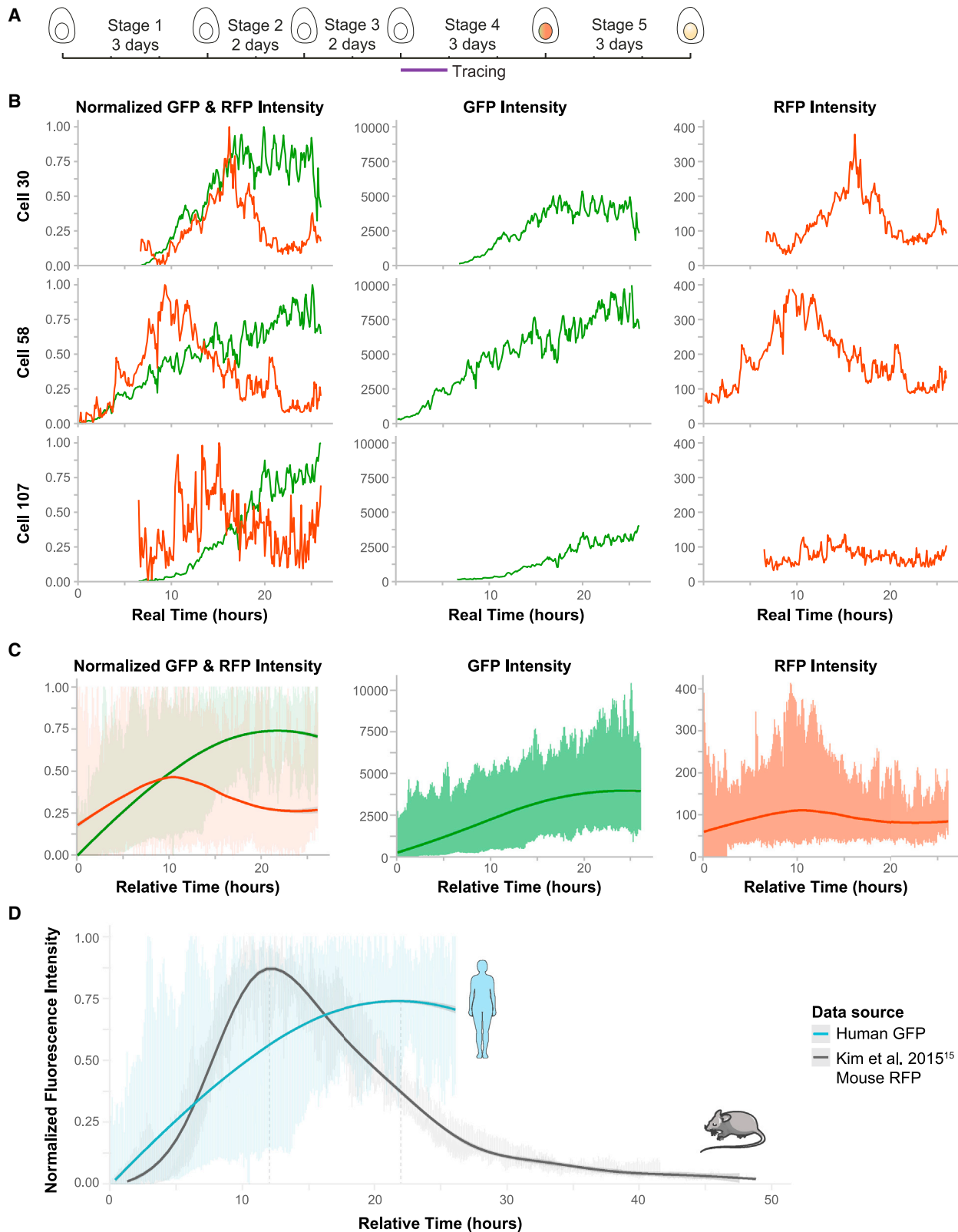
(C) Progression of the GFP expression from S3D2 to S6D1 of the *in vitro* pancreatic differentiation analyzed by flow cytometry. Backscatter (BSC) versus GFP profiles obtained from cells collected at different stages are shown. Light green gate marks GFP-low cells, while the dark green gate marks all GFP⁺ cells. Gray gate indicates GFP[−] cells.

(D) Representative immunofluorescence microscopy images showing *NEUROG3* (cyan), GFP (green), and RFP (magenta) expression in human pancreatic cells at S4D3. The left panels show single colors, while the right panels show the co-expression of pairs, as indicated by color bars. Scale bars, 10 μm.

(E) Boxplots showing the efficiency of the *NEUROG3* dual reporter line calculated by image-based quantification of the co-expression between *NEUROG3* and GFP or RFP within the *NEUROG3*⁺ compartment. Whiskers indicate minimum and maximum values. *NER* = 3 and *n* = 10 for GFP, and *NER* = 1 and *n* = 8 for RFP.

(F) Representative immunofluorescence images showing *NEUROG3* (cyan), GFP (green), and *SOX9* (gray). Scale bars, 10 μm.

S, stage; D, day.



(legend on next page)

expression could differentiate and whether they differentiated into the same endocrine subtypes. To address these questions, we imaged the differentiating cells in 2D for 6 days to selectively track hormone-producing cells (Figure 4A; Video S3). Identifying hormonal cells (C-peptide-positive, glucagon-positive, or C-peptide/glucagon double positive) at the last movie frame and focusing on cells in low-density areas to ensure reliable tracing, we back-traced 9 hormonal cells (Figures 4A–4C). We could detect the onset and their peaks of *NEUROG3* expression and these peaks were at heterogeneous amplitude, as seen before. It is thus clear that cells with low peaks can differentiate into hormone-expressing cells, as do the ones with higher amplitude peaks. Further, we observed that the 4 cells with the lowest amplitude became C-peptide-positive, whereas cells with high amplitude became glucagon positive or double positive. This experiment also gave us an idea of the timeline of human endocrine differentiation, showing that 64 h was sufficient for a cell to move from *NEUROG3* initiation to expressing hormones.

Human endocrine progenitors can divide after *NEUROG3* onset

Since mice studies have suggested that a proliferative population of endocrine progenitors may exist,¹⁸ we next addressed whether it was also true for human endocrine progenitors. Quantification from live imaging revealed that, on average, 4.8% of GFP+ cells divide in 26 h (Figures 5A–5C). We did not notice a pattern in timing (Figure 5C) or location (Figure S4A) of the division events that might indicate a hotspot for a time or place. Even when we normalized the timing of the division to the onset of GFP, cells divided rather asynchronously, relative to GFP onset (Figure 5C). In our 3D spheroid system, we also observed a number of GFP+ cells dividing (Figure 5D). In general, we noted that these proliferative endocrine progenitor cells tend to divide within 12 h after *NEUROG3* induction marked by low GFP levels, with a few exceptions (Figures 5C and 5E).

In 80% of the division events that we could record and categorize, both daughters of a GFP+ cell had similar trajectories of fluorescence accumulation, which we qualify as a symmetric division. This nomenclature does not refer to the uneven segregation of a marker during division but to the subsequent fluorescence trajectory of the daughter cells. For the remaining 20%, we observed that one of the daughter cells continued to upregulate GFP and RFP, whereas the other stayed at a low fluorescence intensity in both channels, which we termed as an asymmetric division (Figure 5E). This observation suggests that either *NEUROG3*+ cells act as an endocrine progenitor pool in which asymmetric cell division ensures the maintenance of a prolifera-

tive population with low *NEUROG3* levels,¹⁸ or that at this early point of *NEUROG3* transcription, the cells can still revert, gradually losing their fluorescent signal. However, no cells were seen to divide more than once over 5 days of imaging. Comparing the dividing and non-dividing *NEUROG3*+ cells revealed that the dividing *NEUROG3*+ cells accumulated GFP and RFP intensities with similar slope and profile on average to their non-dividing counterparts (Figures 5F, S4B, and S4C).

Our analysis so far indicated that although dividing *NEUROG3*+ cells exist throughout differentiation and in different culture systems, they may neither have unique characteristics nor be a part of a stable population. To rule out the possibility that the limited sample size did not allow us to uncover transcriptional differences of the dividing *NEUROG3*+ cells, we also performed bulk RNA-seq with four sorted cell populations: GFP+ in G0/G1, GFP+ in G2M, GFP– in G0/G1, and GFP– in G2M. Principal-component analysis (PCA) revealed that GFP+ cells in G2M cells represent a transitory state along the path of endocrine differentiation (Figures 5G and 5H). When cycling and non-cycling GFP+ cells were compared, the gene set enrichment analysis (GSEA) included genes related to G2M checkpoint and E2F targets, as expected from the cell-cycle category, as well as genes upregulated by Kirsten rat sarcoma virus (KRAS) activation, Notch signaling-related, and epithelial-mesenchymal transition-related genes being enriched in the cycling GFP+ cells (Figure S5A). Supporting the idea that the cycling GFP+ cells have newly activated endocrine differentiation, the non-cycling GFP+ cells expressed more β cell-related genes (Figure S5A).

Proliferating GFP+ cells were found throughout differentiation, from S3D2 to S6D1 using flow cytometry-based live cell-cycle analysis as well as immunostaining (Figures 5I, S4D, and S4E). Even though the percentage of GFP+ cells in G2M decreased over time due to the accumulation of GFP+ cells further along the differentiation path in culture, the percentage of cells in G2M was relatively stable over time within the GFP-low compartment, further demonstrating that the newly induced *NEUROG3*+ cells with low expression levels are more likely to divide (Figures 5I and S4D). Taken together, we show that human endocrine progenitors can divide after *NEUROG3* onset once, mainly in the early phase of *NEUROG3* expression.

Newly emerging human endocrine progenitors differentiate in a gradual and uniform trajectory

The increased propensity to proliferate at low *NEUROG3* levels suggested that there may be molecularly and functionally different populations among *NEUROG3*+ cells. To compare their molecular identity globally, we next performed deep scRNA-seq using the

Figure 2. Live imaging of *NEUROG3*+ human pancreatic cells reveals heterogeneity of *NEUROG3* expression and a slower accumulation of expression reporter signal compared with that of the mouse

(A) Schematic of live imaging and single-cell tracing over the *in vitro* pancreatic differentiation timeline.
 (B) Representative fluorescence tracing plots for three selected *NEUROG3*+ cells, showing GFP and RFP fluorescence intensities over time. On the left panel, fluorescence intensities were scaled between 0 and 1.
 (C) Plots showing fluorescent traces of all traced *NEUROG3*+ cells together. On the left panel, fluorescence intensities were scaled between 0 and 1. Green (GFP) and red (RFP) lines show smoothed mean intensities. The distribution represents individual values for each track per time point, and the gray area shows the confidence interval. NER = 3, n = 52.
 (D) Comparison of the dynamics of human *NEUROG3* expression with the mouse *Neurog3* transcription from Kim et al.¹⁵ The blue line shows smoothed mean GFP fluorescence intensities for the human *NEUROG3* reporter, and the gray line shows average mean RFP intensities for the mouse *Neurog3* reporter. The distribution represents individual values for each track per time point, and the gray area shows the confidence interval. NER = 3, n = 52 for human GFP, and n = 6 for mouse RFP.

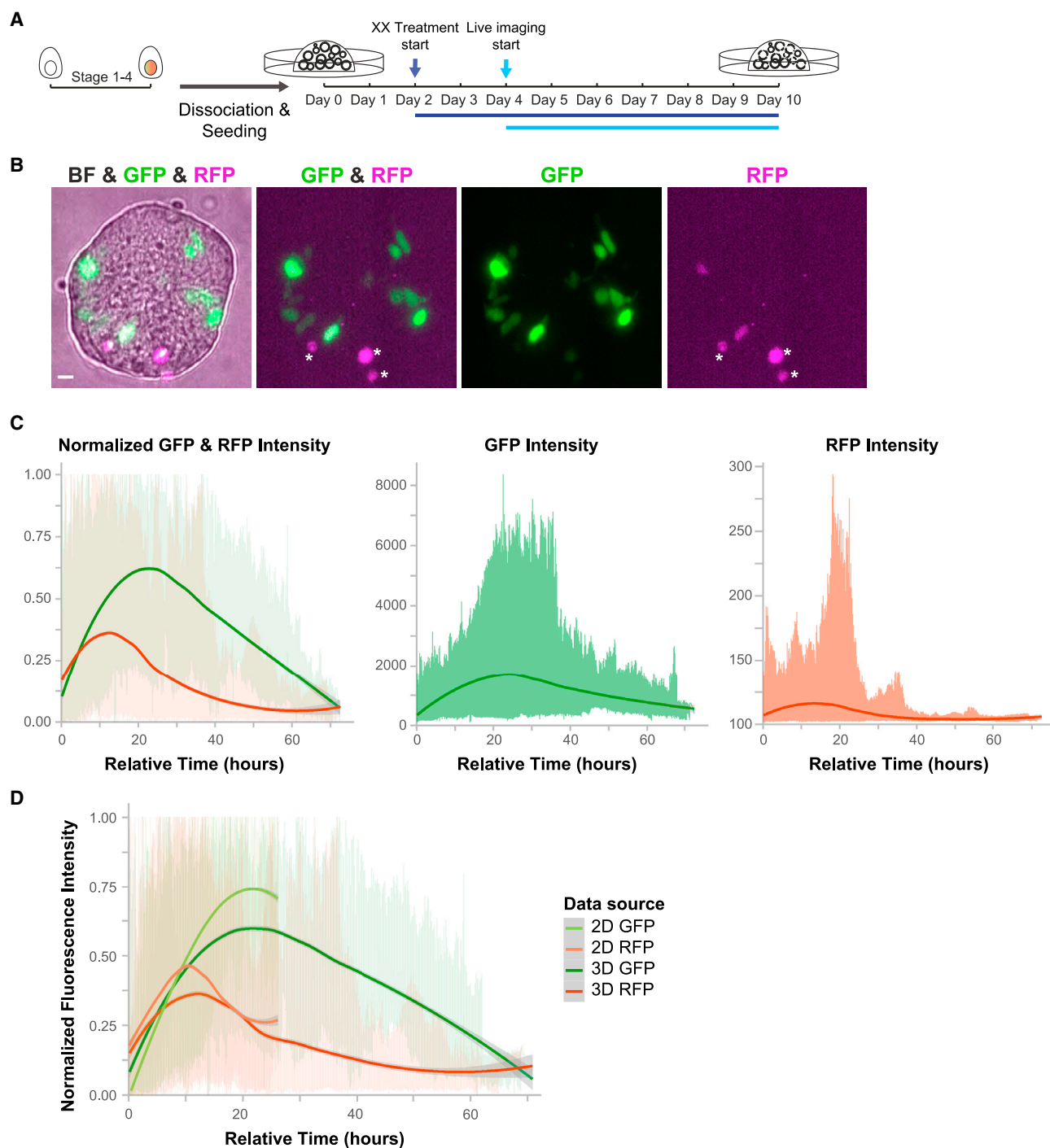


Figure 3. *NEUROG3* expression and protein dynamics are very similar in 2D and 3D culture systems

(A) Schematic of pancreatic progenitor spheroid culture and differentiation induction by a gamma-secretase inhibitor, XX, as well as live imaging timeline.
 (B) Sample fluorescence image from differentiated spheroids at day 6 of culture, showing brightfield (BF, gray), GFP (green), and RFP (magenta). Asterisks indicate debris. Scale bars, 10 μ m.
 (C) Plots showing fluorescent traces of all cells traced from 3D culture. On the left panel, fluorescence intensities were scaled between 0 and 1. Green and red lines indicate smoothed mean fluorescence intensities of GFP and RFP, respectively. The distribution represents individual values for each track per time point, and the gray area shows the confidence interval. NER = 1, n = 30.
 (D) Comparison of GFP and RFP intensities of traced cells from 2D and 3D cultures. Light green and red lines and dark green and red lines indicate smoothed mean fluorescence intensities from 2D and 3D, respectively. The distribution represents individual values for each track per time point, and the gray area shows the confidence interval.

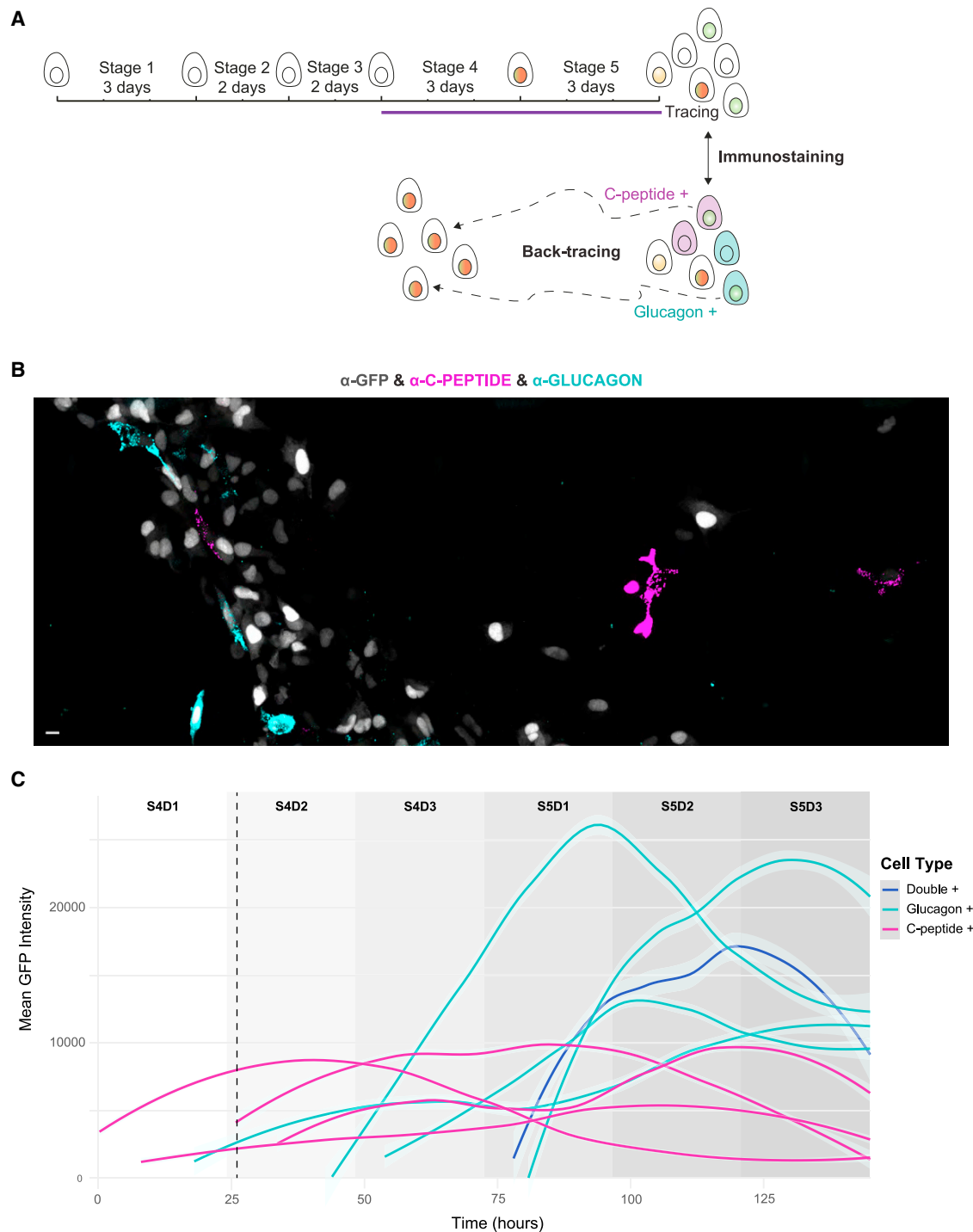


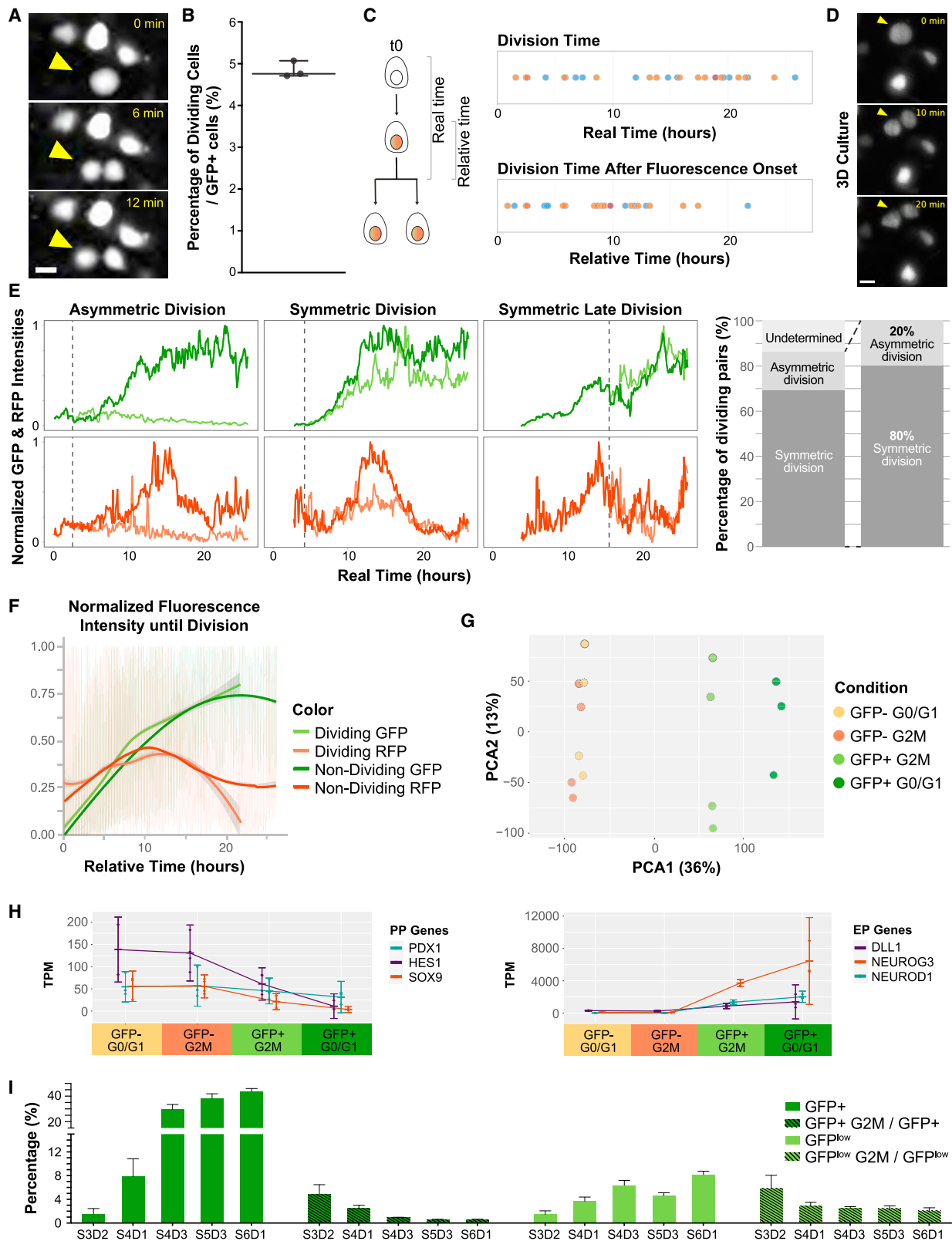
Figure 4. Cells with low and high *NEUROG3* peak levels can differentiate into hormonal cells

(A) Schematic of live imaging and single-cell back-tracing of immunostained hormonal cells over *in vitro* pancreatic differentiation timeline.

(B) Representative immunofluorescence images showing GFP (gray), C-peptide (magenta), and glucagon (cyan) expression in differentiated human pancreatic cells at S5D3. Scale bars, 10 μ m.

(C) Plot showing smoothed GFP mean intensities of C-peptide-positive cells (magenta), glucagon-positive cells (cyan), and a C-peptide and glucagon double-positive cell (blue) over live imaging time. The highlighted area along each line shows the confidence interval, and the dashed line indicates the endpoint of tracing in the earlier cell tracings in Figure 2A. NER = 3, n = 9.

S, stage; D, day.



(legend on next page)

plate-based Smart-seq2 platform,²⁶ prioritizing sequencing depth over the number of cells analyzed. We index-sorted pancreatic cells obtained in 2D at S4D1, corresponding to the end of the live imaging tracks, using fluorescence-activated cell sorting (FACS) sampling predetermined ratios of cells with no GFP, low GFP, and high GFP levels (Figures 1C and 6A). 97% of the *NEUROG3* mRNA-containing cells from the scRNA-seq also expressed the reporter (Figure S6A). Unsupervised cell embedding using uniform manifold approximation and projection (UMAP) revealed that cells align with their FACS-based GFP categories, reflecting their differentiation trajectory (Figures 6B, 6C, and S6B).

Though our scRNA-seq results indicated an overall differentiation continuum toward a pancreatic endocrine fate, clustering analysis enabled us to categorize the cells as pancreatic progenitors (PPs), proliferative cells, and cells in the early steps of the endocrine differentiation, named EP1, EP2, and EP3 (Figures 6D, S6C, and S6D). Finally, we identified a cluster of cells expressing early liver marker genes such as alpha fetoprotein (*AFP*), which were also observed in other pancreatic differentiation datasets^{21,28} and assumed to be an earlier progenitor population (Figures 6D, S6C, and S6D). Most cells in the G2M phase of the cell cycle formed a proliferative cluster, which included PPs as well as some *NEUROG3* mRNA-containing cells, further supporting that some of the *NEUROG3*⁺ cells do indeed express the transcriptional signature of dividing cells (Figures 6C and 6E). Cell-cycle classification for each identified cluster showed a decrease in the percentage of cycling cells from PPs to EP3 (Figure S6E). We also noted the expression of known cell-cycle exit-related genes, such as *Cyclin Dependent Kinase Inhibitor 1C* (*CDKN1C*), *BTG Anti-Proliferation Factor 2* (*BTG2*), and Growth Arrest And DNA Damage Inducible Alpha (*GADD45A*) in endocrine progenitor (EP) cells (Figure S6F). Using the data from Yu et al.,²⁷ where different populations from human fetal pancreata spanning from 9 to 19 weeks post-conception were sequenced using a deep scRNA-seq method,²⁷ we observed that some human fetal *NEUROG3*⁺ cells also expressed the transcriptional signature of cells that divide (Figures 6F–6H). Therefore, this characteristic of early *NEUROG3*⁺ cells is reflected *in vivo* as well.

RNA velocity analysis performed on the data after cell-cycle regression indicated a high degree of transcriptional changes,

with a faster change in transcription at the onset of *NEUROG3* expression (Figure 6I). In addition, proliferative *NEUROG3*⁺ cells without the cell-cycle signature were mostly scattered to EP1 and EP2 clusters, indicating their transitioning status with fast progression toward the endocrine lineage, further supporting our earlier conclusions (Figures 5H and 6I). At S4D1 of endocrine differentiation, we did not yet find many hormonal transcript-expressing cells or cells expressing marker genes for specific endocrine subtypes, such as aristaless-related homeobox (*ARX*), indicating their further endocrine fate (Figure S6B). *CHGA*⁺ cells were the most mature population in our dataset, suggesting the transition from *NEUROG3* onset to hormone transcript expression takes more than 26 h (Figure 6C). An exception is Ghrelin (*GHRL*) (Figure S6B), as previously been shown to be activated early in endocrine progenitors *in vitro* and *in vivo*.^{19,29}

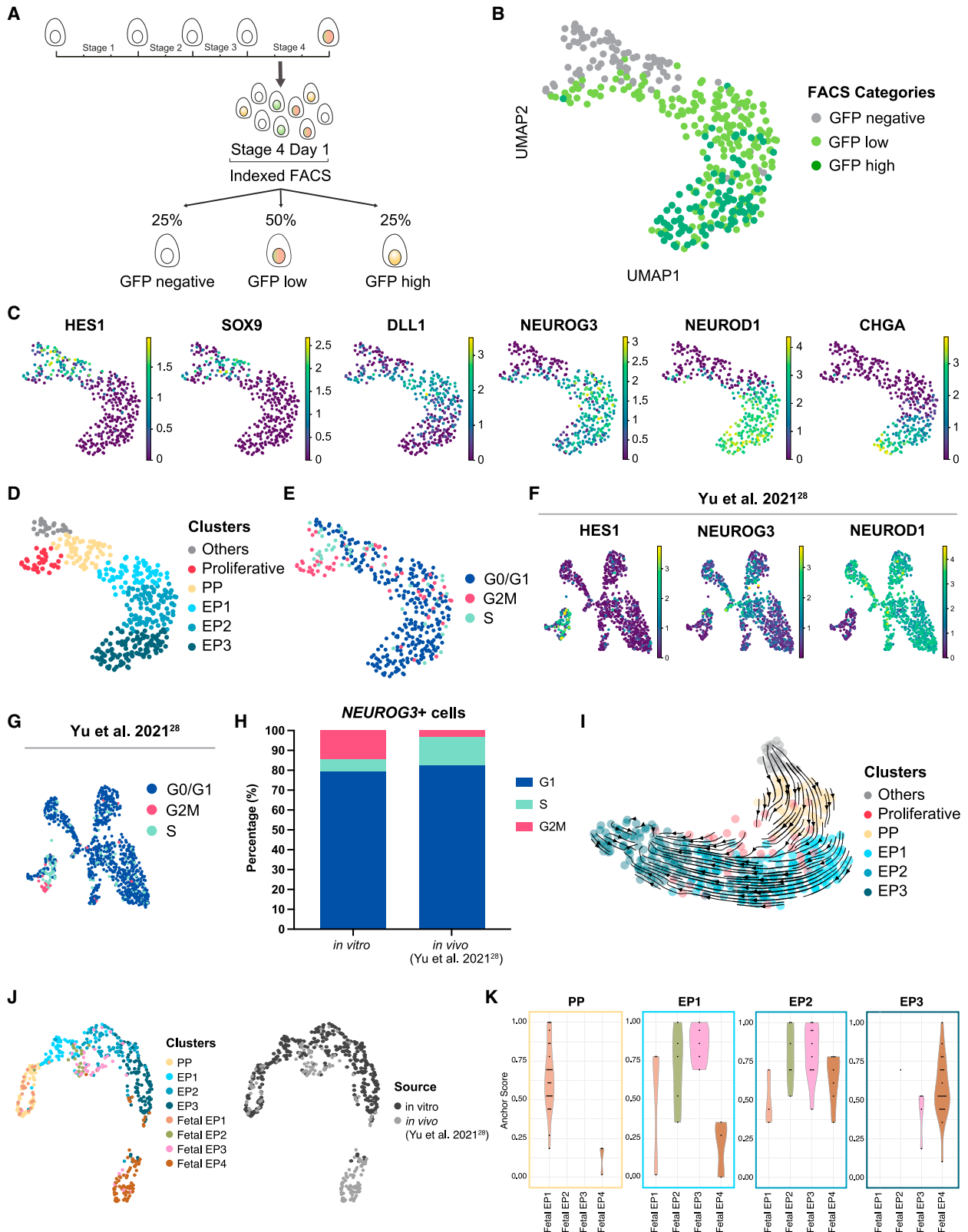
In order to deepen the comparison of our endocrine progenitor clusters with cells *in vivo*, we performed an integrated analysis of *in vitro* clusters PP and EP1–EP3 from our study with the clusters EP1–EP4 of human fetal endocrine progenitors from the study by Yu et al.²⁷ (Figures 6J and 6K). To check which stages of human pancreas development our identified cell clusters most closely resemble, we used highly variable genes shared by the two datasets for the integration, using the reciprocal PCA approach. As a result, we could identify the following defining characteristics: (1) the clustering of our PP cells close to fetal EP1, (2) a unique cluster of *in vitro* EP1 cells, which we most likely capture as our analysis has enriched for intermediates at low levels of GFP, (3) our EP2 cells being closest to fetal EP2–EP3, and (4) the fetal EP4 cells being somewhat close to our EP3 cells with a tendency to form a distinct cluster as expected, considering EP4 cells were more differentiated than our EP3 cells by their gene signature (Figure 6J). These observations were further supported by the scores of identified integration anchors as visualized in Figure 6K.

Integrating live imaging and scRNA-seq identifies *KLK12* as a protein promoting early endocrine progenitor movement

In order to identify gene signatures correlating with dynamic cell behaviors observed in movies, we integrated our live imaging

Figure 5. Dividing *NEUROG3*⁺ cells do not have a unique transcriptomic signature but are in transition to the endocrine progenitor lineage

- (A) Snapshot images showing division (arrowhead) of a GFP⁺ human endocrine progenitor cell in 2D at 0, 6, and 12 min (relative time). Scale bars, 10 μ m.
- (B) Boxplot showing the percentage of dividing cells within the GFP⁺ compartment from live imaging experiments. Whiskers indicate the highest and lowest observations. NER = 3.
- (C) Plots showing the division time of *NEUROG3*⁺ cells. The upper panel shows the division times with respect to the start of the movies (real time), and the lower panel shows division times normalized to the onset of GFP in each cell (relative time), as indicated in the schematic on the left. Colors indicate independent experiments. NER = 3, n = 27.
- (D) An example of a division event of a GFP⁺ cell during live imaging of 3D differentiating spheroids. Arrowheads indicate the dividing cell over time, with 3 panels showing 0, 10, and 20 min (relative time). Scale bars, 10 μ m.
- (E) Representative fluorescence tracing plots for three selected dividing *NEUROG3*⁺ cells in 2D, showing GFP and RFP fluorescence intensities over time for different division modes. Dark and light lines indicate daughter cells after the division event. Fluorescence intensities were scaled between 0 and 1. The right graph shows the quantification of different division modes. NER = 3, n = 29.
- (F) Plot showing fluorescent traces of all traced cells together for dividing (light colors, until division point) and non-dividing (dark colors) *NEUROG3*⁺ cells. Fluorescence intensities were scaled between 0 and 1. All lines show smoothed mean intensities. The distribution represents individual values for each track per time point, and the gray area shows the confidence interval. NER = 3, n = 52 for non-dividing cells, and NER = 3, n = 27 for dividing cells.
- (G) PCA plot showing the distance between transcriptomes of GFP⁺ and GFP[−] cells in G0/G1 and G2M. NER = 3 for GFP⁺ G0/G1 and NER = 4 for the other categories.
- (H) Mean TPMs (transcripts per million) with standard deviation error bars for selected pancreatic progenitor (PP) genes (*PDX1*, *HES1*, and *SOX9*) and endocrine progenitor (EP) genes (*DLL1*, *NEUROG3*, and *NEUROD1*) per sequencing category.
- (I) Flow cytometry-based live cell-cycle analysis of GFP-low and GFP⁺ cells from different stages of *in vitro* differentiation. Bars indicate mean percentage values \pm SEM. NER = 2 for S3D2 and NER = 4 for all the other groups (extended plot provided in Figure S4E).



(legend on next page)

and scRNA-seq dataset. Using quantile transformation, we aligned relative GFP intensity values from the index-sorted cells for scRNA-seq with those measured at the last time point of traced cells from live imaging (Figure 7A). Using Gaussian mixture modeling, we statistically predicted which sequencing cluster each non-dividing *NEUROG3*⁺ (GFP⁺) cell from our short-term and long-term live imaging dataset would most likely fall into (Figures S7A and S7E). Based on this prediction, it is possible to correlate the dynamic cellular properties of cells as observed by live image analysis to their transcriptomic profiles (Figures 7B, 7C, S7B, and S7C). Building upon prior literature on the motility of *NEUROG3*⁺ cells,³⁰ we compared the global displacement, as a readout for cell motility, for the cells in these three predicted clusters with different levels of *NEUROG3* expression. We observed that EP3 cells with decreasing *NEUROG3* expression were significantly slower than EP1 and EP2 cells (Figures 7B, 7C, S7B, and S7C). We did not find a positional bias for cells predicted to belong to different transcriptomic clusters (Figure S7D).

In our scRNA-seq analysis, we identified *KLK12* as a transcript that was associated with the migratory phase. Its transient expression started at the initiation of *NEUROG3* expression (EP1) and was downregulated in the *CHGA*⁺ cells (EP3) (Figures 7D and S6D). *KLK12* encodes kallikrein-related peptidase 12, a serine protease implicated in the sprouting and migration of endothelial cells through extracellular matrix remodeling. This protein was previously shown to cleave human fibronectin and tenascin.³¹ Its transient expression may enable the movement of cells aggregating into islets, possibly via the digestion of fibronectin produced by PPs at this stage (Figure S6D). We first tested if the finding had any *in vivo* relevance using the published dataset of human fetal pancreas atlas,²⁷ and we found that *KLK12* was expressed in the early phases of *NEUROG3* transcript accumulation *in vivo*, as it is *in vitro* (Figures 7D and 7E). We then used a previously identified and commercially available blocking antibody as well as Serpin Family F Member 2 (SERPINF2), a general serine protease inhibitor to test if interfering with *KLK12* function changed the motility of cells at EP1, EP2, and EP3 stages in a live imaging experiment. By quantifying the global displacement of GFP⁺ cells over time and comparing their behavior under perturbation, we found that the *KLK12* blocking antibody decreased cell motility at EP1 and EP2 stages

but not at the EP3 stage (Figure 7F). SERPINF2 also had a similar effect on the cells, where it decreased cell motility of EP1 cells and to a lesser extent (not significant) of EP2 cells but not EP3 cells. Interestingly, adding activated *KLK12* protein in the culture medium did not increase motility further and, on the contrary, reduced in EP1 and EP2 cells (Figure 7F), possibly because the matrix was too disrupted at the dose used or because the matrix disruption should be local and/or directional. We indeed see that PPs express inhibitors of *KLK* proteins and serine proteases that may counterbalance *KLK12*, such as serine protease inhibitor Kazal-type 1 (*SPINK1*) and Serpin Family A Member 1 (*SERPINA1*) (Figure S7F). Taken together, our integrated analysis method uncovered increased motility in the *NEUROG3*-high phase of human pancreatic endocrine progenitors and evidence that *KLK12* promotes their mobility, as a proof-of-concept of our methodology.

DISCUSSION

NEUROG3 dynamics are slower in humans compared with mice

The human pancreas develops over a longer period but is also larger than in mice, with its ratio of endocrine cells being similar. Using the GFP signal we show that in the human pancreatic cells, *NEUROG3* expression reached its peak in around 22 h, whereas in a mouse line where *Neurog3* promoter was used to drive turboRFP expression, it peaked at around 12 h.¹⁵ Though the constructs and the differentiation conditions used in mice and human cells were not identical, the two different conditions we tested in the human system gave very similar outcomes, strengthening our conclusions. Since EGFP used in human cells matures 6-fold faster than turboRFP in mice,³² the approximately 2-fold longer timing in human *NEUROG3*⁺ cells is likely due to longer transcription, splicing, and/or translation than in mice. This notion is supported by recent species comparison studies in presomitic mesoderm and motor neurons showing that differences in global biochemical rates play a substantial role in creating the divergence in the timespan of development and that human cells display twice slower kinetics of protein expression and decay than mouse cells.^{33,34} Live imaging using a mouse *Neurog3* fusion protein can more accurately determine the species differences, which has not yet been done, although

Figure 6. Deep single-cell sequencing of *NEUROG3*⁺ cells during human pancreatic endocrine differentiation indicates a gradual differentiation trajectory prior to hormonal transcription

- (A) Schematic depicting the indexed FACS strategy for the scRNA-seq experiment, indicating the percentage of cells in different GFP categories sequenced.
 (B) UMAP colored by FACS-based GFP categories.
 (C) Selected normalized marker gene expression levels projected on the UMAP, highlighting pancreatic differentiation trajectory.
 (D) UMAP colored by detected clusters.
 (E) UMAP colored by sequencing-based cell-cycle categories.
 (F) UMAP plots generated from *NEUROG3*⁺ cells from Yu et al.²⁷ showing selected normalized gene expression levels projected.
 (G) UMAP plots generated from *NEUROG3*⁺ cells from Yu et al.²⁷ showing sequencing-based cell-cycle categories projected.
 (H) Quantification of *NEUROG3*⁺ cells from scRNA-seq (*in vitro*) and Yu et al.²⁷ (*in vivo*) according to their sequencing-based cell-cycle categorization. 193 cells are represented for *in vitro*, and 969 cells are represented for *in vivo*.
 (I) RNA velocity projected onto the UMAP produced after cell-cycle regression, colored by detected clusters as in Figure 6H. Arrows indicate the extrapolated state.
 (J) UMAP plots generated from the integrated dataset of pancreatic progenitors (PP) and clusters EP1–EP3 identified in our study with sequential developmental stages EP1–EP4 of human fetal endocrine progenitor cells from the study by Yu et al.²⁷ The left panel indicates the clusters from both datasets, and the right panel indicates their sources.
 (K) Violin plots showing anchor scores of the anchors used for the pairwise data integration of clusters PP, EP1, EP2, and EP3 from this study, and EP1–EP4 from the *in vivo* dataset from Yu et al.²⁷

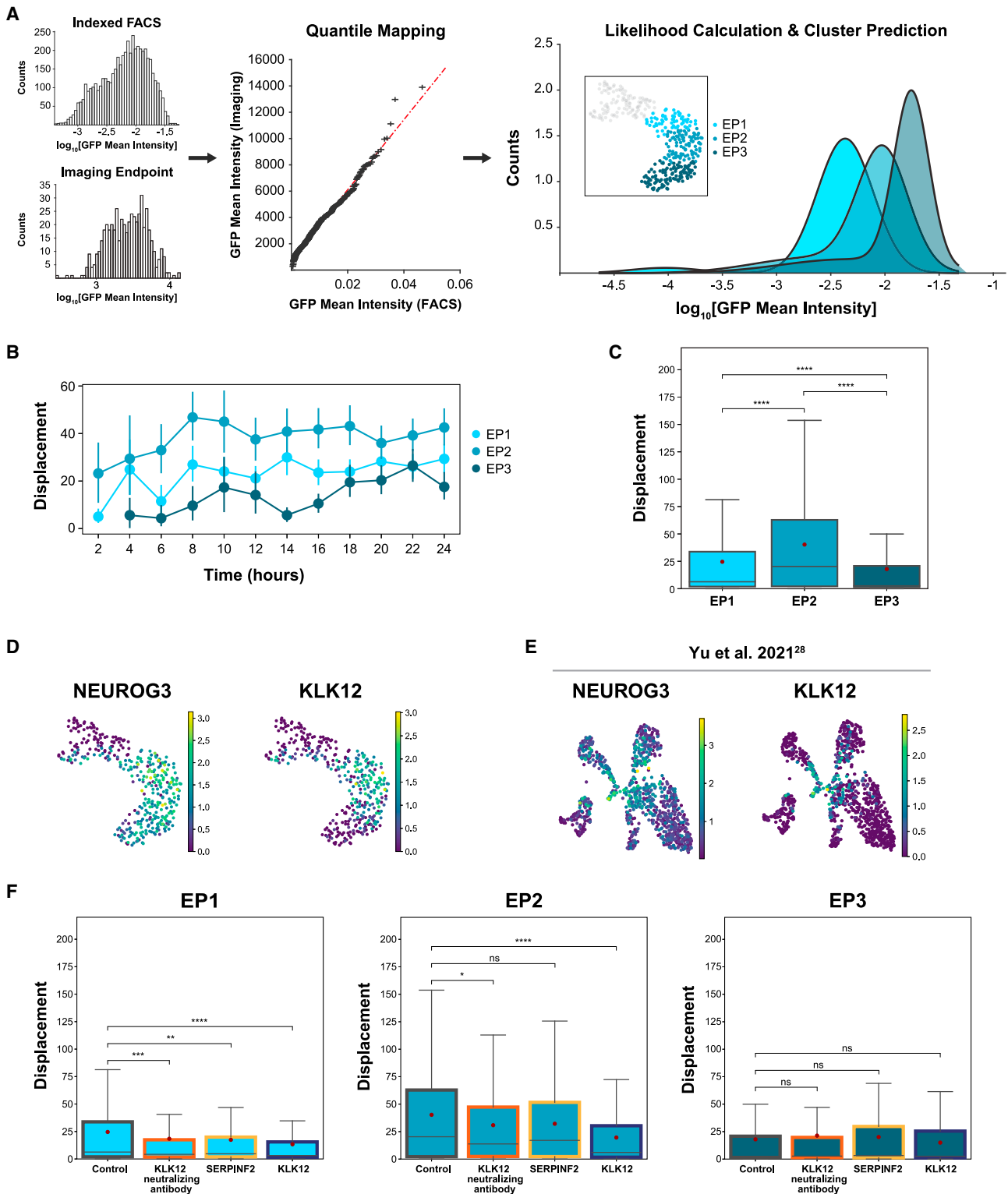


Figure 7. Integrating live imaging and single-cell sequencing datasets identifies KLK12 as a gene promoting early endocrine progenitor motility

(A) Schematic outline of the analysis pipeline. GFP measurements from live imaging were quantile-transformed to match the distribution of GFP values from flow cytometry. The GFP values of sequenced cells obtained from FACS were grouped according to three identified clusters. For each cluster, the corresponding conditional distribution over GFP levels was estimated using Gaussian mixture modeling.

(legend continued on next page)

a mouse model is available.³⁵ In our system, human *NEUROG3* protein peaks 11 h after its expression onset and remains for about 40 h. The reduction of *NEUROG3*-RFP levels by half, combining the protein decay and the progressively slower rate of new transcription, takes about 12 h after the peak. The protein degradation measurements indicate that *NEUROG3* half-life is shorter in mice, ranging from 12 to 15 min^{36,37} than in humans, 30 to 67 min,²³ although caution is needed as it was measured in less physiological systems and the half-life was cell type-dependent.²³

Even though it is risky to claim the relevance of *in vitro* observations to the development in the womb, we show robustness in the differentiation timing across culture systems. *In vitro* human *NEUROG3* expression and protein dynamics are very similar between different culture platforms (2D vs. 3D), media with different growth factors (high fibroblast growth factor (FGF) 2 in 3D), and even different induction methods (Notch inhibition in 3D and FGF7, retinoic acid, SANT1, LDN, and TPB in 2D), suggesting that these measurements may be relevant to an *in vivo* setting.

An important new observation we made was the high degree of variability in the *NEUROG3* peak expression among cells, both in the 2D and 3D conditions. Notably, we could show that cells with both low and high peak levels could further differentiate into endocrine cells by back-tracing. In mice, previous experiments using a bacterial artificial chromosome (BAC) reporter for *Neurog3* traced some progeny to the exocrine lineage.^{38,39} Since we were not able to forward trace thousands of cells, we cannot conclude that all cells expressing *NEUROG3* become endocrine. The limited number of tracings showed that cells with low *NEUROG3* peak expression were biased toward the insulin lineage, but tracing more cells at different time points would be necessary to be fully conclusive.

***NEUROG3*+ human pancreatic cells can divide but do not act as a specialized population**

In this work, we showed that *NEUROG3*+ human pancreatic cells could divide in different culture systems based on live imaging, flow cytometry-based cell-cycle analysis, as well as RNA-seq-based cell-cycle categorization. Dividing *NEUROG3*+ cells exist throughout the *in vitro* pancreatic differentiation, and the cells are more likely to divide when they have low *NEUROG3* levels. Following the daughters after division through live imaging, we saw that different division modes exist. In most cases, both daughters upregulated *NEUROG3* (80%), and in others, only one daughter upregulated *NEUROG3* while the other stayed low (20%). It is currently unclear whether the asymmetric division results in one daughter turning *NEUROG3* expression off and reverting to a PP fate or sustaining a low level of *NEUROG3* expression.

Even though human pancreatic cells can divide after *NEUROG3* initiation, we could neither detect any cell dividing more than once, even in 5-day movies, nor find any dynamic or transcriptomic signature associated with these cells to indicate they are a specialized population. Prior to division, these cells exhibited the same *NEUROG3* expression and protein dynamics as their non-dividing counterparts in live imaging. Bulk RNA-seq of *NEUROG3*+ endocrine progenitors in G2/M only uncovered cell division signatures and a signature of a transitional state between pancreas progenitors and endocrine progenitors. Together with the expression of cell-cycle exit-related genes such as *CDKN1C*,²⁵ *GADD45A*, and *BTG2*^{40,41} in *NEUROG3*+ EP1 population, we conclude that *NEUROG3*+ human pancreatic cells do not act as a stable endocrine progenitor pool as proposed in mice.^{18,42} Rather, these cells most likely finish their last cell-cycle round before further upregulating *NEUROG3* and becoming post-mitotic. Recent mouse studies found that G1 lengthening in bipotent progenitors drives endocrine differentiation by delaying *NEUROG3* phosphorylation, thereby stabilizing *NEUROG3* protein.^{22,36,37} Human *NEUROG3* phosphorylation was also shown,²² but its phosphorylation sites and mechanisms may differ from mouse *NEUROG3*.²³ If phosphorylation in humans acts mechanistically similarly as proposed in mice, our observations suggest that delayed phosphorylation may participate in a feed-forward loop enabling the ramping up of *NEUROG3*, rather than keeping cells in a cycling *NEUROG3*-low state with recurrent degradation over multiple cell cycles.

The initial steps of endocrine differentiation are similar *in vitro* and *in vivo*

A limitation of work on *in vitro* systems addressing human development is the difficulty of assessing their relevance to the *in vivo* context.²¹ Our scRNA-seq dataset indicated that human pancreatic cells undergo a gradual endocrine differentiation process after *NEUROG3* initiation. The similarity between *in vitro* and *in vivo* gene expression profiles indicates that the downstream transcriptional program after *NEUROG3* expression is robust and can be captured faithfully in *in vitro* pancreatic differentiation systems.²⁷ This finding extends previous observations as the convergence of the transcriptional program during endocrine induction was also reported in less extensive quantitative real-time PCR (qPCR)-based single-cell profiling of *in vivo*- and *in vitro*-derived pancreatic cells.²⁹

Even though the dataset from Yu et al.²⁷ predominantly included more mature endocrine populations than described here, finding *NEUROG3*+ human fetal pancreatic cells with a proliferation signature *in vivo* further validates the relevance of this study to the human fetal endocrine pancreas development. We could further draw comparisons between *in vitro* and

(B) Point plot showing the global displacement of cells over time during 26-h imaging throughout S4D1, binned every 2 h (time data points). Lines indicate mean with a confidence interval of 95%. NER = 2.

(C) Boxplot comparing distributions of global displacement of cells during 26-h imaging throughout S4D1 according to their predicted EP clusters. Whiskers indicate 25% and 75% quantiles with median (line in the box) and mean (red dot). p values were determined by the Kruskal-Wallis test. NER = 2.

(D) Normalized gene expression levels of *NEUROG3* and *KLK12* projected on the UMAP from our scRNA-seq.

(E) UMAP plots generated from Yu et al.²⁷ data showing normalized gene expression levels of *NEUROG3* and *KLK12* projected.

(F) Boxplots comparing distributions of global displacement of cells according to their predicted EP clusters, under different treatments (blocking antibody, SERPINF2, and activated *KLK12*) during stage 4 day 1. Whiskers indicate 25% and 75% quantiles with median (line in the box) and mean (red dot). p values were determined by the Kruskal-Wallis test. NER = 2.

fetal endocrine cells at different maturation stages within endocrine progenitor clusters, suggesting a similar differentiation progression.

Integrating live imaging and scRNA-seq connects dynamic behaviors to the transcriptomic signature and identifies KLK12 as a migration potentiator

Linking scRNA-seq profiles to functional readouts is a new emerging challenge in cell omics. In this study, by using a plate-based scRNA-seq platform in combination with indexed FACS, we were able to recall fluorescence intensities of sequenced cells in different transcriptomic clusters and compare them with the intensities of the live-imaged *NEUROG3*⁺ cells. With this mapping strategy, it is possible to correlate dynamic cell properties such as motility exemplified here, to their transcriptomic profiles. Using this approach, we found that endocrine progenitors with the highest *NEUROG3* levels, EP2 cells, exhibit increased motility, and the most mature endocrine progenitors, EP3 cells, start to slow down. This finding is in line with our previous studies showing that *NEUROG3* triggers motility and partial epithelial delamination,³⁰ while the *Neurog3*⁺ cells and their endocrine progeny remain connected to the epithelium.⁴³ Our observations extend these findings by identifying the serine protease KLK12 as a protein associated with the peak migration phase and promoting migration. Conceptually, our study reveals that differentiation dynamics during the initiation of endocrine commitment by *NEUROG3* in human pancreas development is slower than in mice and seemingly stochastic with regards to expression levels but very reproducible in its timing across cells and conditions.

The methodology used here could be transposed to other developmental processes to link changes in genome-wide gene expression to observable events in live imaging, which enables a comprehensive overview of a system in dynamically changing landscapes.

Limitations of the study

Our study focuses primarily on the *NEUROG3*⁺ human pancreatic endocrine progenitors at a specific stage of a commonly used *in vitro* human pancreatic differentiation protocol. Even though we validated some of our claims in a 3D pancreatic culture system and showed similar markers being expressed in *in vivo* datasets, this study could be extended to the progenitors produced later in the protocol to strengthen these claims. Furthermore, in the nervous system, some bHLH family members including *Neurog2* are known to oscillate,^{44,45} and recently, in the mouse fetal pancreas, *DLL1* and *HES1* were also shown to oscillate.⁴⁶ There is currently no data available on the potential oscillation of *NEUROG3* in any organ systems studied, and we did not observe oscillations in our data. However, it is possible that our reporter line and imaging conditions could not detect such a phenomenon, notably if it concerns the low expression phase, due to the inherent noise in this system.

STAR★METHODS

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

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
 - Lead contact
 - Materials availability
 - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
 - Culturing of human embryonic stem cells
- METHOD DETAILS
 - Construction of *NEUROG3* dual reporter hESC line
 - *In vitro* differentiation into the pancreatic lineage
 - Live imaging and cell tracking
 - Flow Cytometric Analyses
 - Immunostaining
 - Image analysis
 - FACS, scRNA-seq, and downstream analysis
 - RNA extraction and sequencing of human pancreatic cell populations
 - Integration of scRNA-seq datasets
 - Integration of live-imaging and sequencing datasets
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.devcel.2023.07.019>.

ACKNOWLEDGMENTS

The authors would like to thank DanStem Platforms of Genomics, Stem Cell Culture, Flow Cytometry (especially Paul van Dieken), and Imaging for their training, technical expertise, support, and the use of instruments. Additionally, we thank the Scientific Computing Facility (especially Noreen Walker, Gayathri Nadar, and Andre Gohr), Organoid and Stem Cell Facility, FACS Facility, and Light Microscopy Facility at MPI-CBG for their support and technical help. The Deep Sequencing Facility of the CMCB Technology Platform (Center for Regenerative Therapy in Dresden) provided invaluable service with library preparation and sequencing for the RNA sequencing experiments. Special thanks to Jifeng Liu and Allison Lewis for their support. The Novo Nordisk Foundation Center for Stem Cell Biology was supported by a Novo Nordisk Foundation grant number NNF17CC0027852. B.S.B.-T. is supported by the Copenhagen Bioscience PhD program financed by the Novo Nordisk Fonden (NNF16CC0020994).

AUTHOR CONTRIBUTIONS

Conceptualization, B.S.B.-T., Y.H.K., and A.G.-B.; methodology, B.S.B.-T., Y.H.K., L.H., B.H.L., F.L., C.Z., and A.G.-B.; investigation, B.S.B.-T. and J.V.D.; resources, A.G.-B.; writing – original draft, B.S.B.-T., Y.H.K., and A.G.-B.; writing – review & editing, B.S.B.-T., L.H., F.L., C.Z., Y.H.K., and A.G.-B.; visualization, B.S.B.-T.; supervision, Y.H.K. and A.G.-B.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: November 29, 2021

Revised: May 20, 2023

Accepted: July 27, 2023

Published: August 16, 2023

REFERENCES

1. Jennings, R.E., Berry, A.A., Kirkwood-Wilson, R., Roberts, N.A., Hearn, T., Salisbury, R.J., Blaylock, J., Piper Hanley, K.P., and Hanley, N.A. (2013).

- Development of the human pancreas from foregut to endocrine commitment. *Diabetes* 62, 3514–3522. <https://doi.org/10.2337/db12-1479>.
2. Petersen, M.B.K., Gonçalves, C.A.C., Kim, Y.-H., and Grapin-Botton, A. (2018). Recapitulating and deciphering human pancreas development from human pluripotent stem cells in a dish. *Curr. Top. Dev. Biol.* 129, 143–190. <https://doi.org/10.1016/bs.ctdb.2018.02.009>.
 3. Gradwohl, G., Dierich, A., LeMeur, M., and Guillemot, F. (2000). *neurogenin3* is required for the development of the four endocrine cell lineages of the pancreas. *Proc. Natl. Acad. Sci. USA* 97, 1607–1611. <https://doi.org/10.1073/pnas.97.4.1607>.
 4. Schwitzgebel, V.M., Scheel, D.W., Conners, J.R., Kalamaras, J., Lee, J.E., Anderson, D.J., Sussel, L., Johnson, J.D., and German, M.S. (2000). Expression of neurogenin3 reveals an islet cell precursor population in the pancreas. *Development* 127, 3533–3542. <https://doi.org/10.1242/dev.127.16.3533>.
 5. Gu, G., Dubauskaite, J., and Melton, D.A. (2002). Direct evidence for the pancreatic lineage: NGN3+ cells are islet progenitors and are distinct from duct progenitors. *Development* 129, 2447–2457. <https://doi.org/10.1242/dev.129.10.2447>.
 6. Heller, R.S., Jenny, M., Collombat, P., Mansouri, A., Tomasetto, C., Madsen, O.D., Mellitzer, G., Gradwohl, G., and Serup, P. (2005). Genetic determinants of pancreatic ϵ -cell development. *Dev. Biol.* 286, 217–224. <https://doi.org/10.1016/j.ydbio.2005.06.041>.
 7. Johansson, K.A., Dursun, U., Jordan, N., Gu, G., Beermann, F., Gradwohl, G., and Grapin-Botton, A. (2007). Temporal control of Neurogenin3 activity in pancreas progenitors reveals competence windows for the generation of different endocrine cell types. *Dev. Cell* 12, 457–465. <https://doi.org/10.1016/j.devcel.2007.02.010>.
 8. Ohsie, S., Gerney, G., Gui, D., Kahana, D., Martin, M.G., and Cortina, G. (2009). A paucity of colonic enteroendocrine and/or enterochromaffin cells characterizes a subset of patients with chronic unexplained diarrhea/malabsorption. *Hum. Pathol.* 40, 1006–1014. <https://doi.org/10.1016/j.humpath.2008.12.016>.
 9. Pinney, S.E., Oliver-Krasinski, J., Ernst, L., Hughes, N., Patel, P., Stoffers, D.A., Russo, P., and De León, D.D. (2011). Neonatal diabetes and congenital malabsorptive diarrhea attributable to a novel mutation in the human neurogenin-3 gene coding sequence. *J. Clin. Endocrinol. Metab.* 96, 1960–1965. <https://doi.org/10.1210/jc.2011-0029>.
 10. Rubio-Cabezas, O., Jensen, J.N., Hodgson, M.I., Codner, E., Ellard, S., Serup, P., and Hattersley, A.T. (2011). Permanent neonatal diabetes and enteric anendocrinosis associated with biallelic mutations in *NEUROG3*. *Diabetes* 60, 1349–1353. <https://doi.org/10.2337/db10-1008>.
 11. McGrath, P.S., Watson, C.L., Ingram, C., Helmuth, M.A., and Wells, J.M. (2015). The basic helix-loop-helix transcription factor *neurog3* is required for development of the human endocrine pancreas. *Diabetes* 64, 2497–2505. <https://doi.org/10.2337/db14-1412>.
 12. Solorzano-Vargas, R.S., Bjerknes, M., Wang, J., Wu, S.V., Garcia-Careaga, M.G., Pitukcheewanont, P., Cheng, H., German, M.S., Georgia, S., and Martin, M.G. (2020). Null mutations of *NEUROG3* are associated with delayed-onset diabetes mellitus. *JCI Insight* 5, e127657. <https://doi.org/10.1172/jci.insight.127657>.
 13. Villasenor, A., Chong, D.C., and Cleaver, O. (2008). Biphasic *Ngn3* expression in the developing pancreas. *Dev. Dyn.* 237, 3270–3279. <https://doi.org/10.1002/dvdy.21740>.
 14. Salisbury, R.J., Blaylock, J., Berry, A.A., Jennings, R.E., De Krijger, R., Piper Hanley, K.P., and Hanley, N.A. (2014). The window period of neurogenin3 during human gestation. *Islets* 6, e954436. <https://doi.org/10.4161/19382014.2014.954436>.
 15. Kim, Y.H., Larsen, H.L., Rué, P., Lemaire, L.A., Ferrer, J., and Grapin-Botton, A. (2015). Cell cycle-dependent differentiation dynamics balances growth and endocrine differentiation in the pancreas. *PLoS Biol.* 13, e1002111. <https://doi.org/10.1371/journal.pbio.1002111>.
 16. Miyatsuka, T., Kosaka, Y., Kim, H., and German, M.S. (2011). Neurogenin3 inhibits proliferation in endocrine progenitors by inducing *Cdkn1a*. *Proc. Natl. Acad. Sci. USA* 108, 185–190. <https://doi.org/10.1073/pnas.1004842108>.
 17. Piccand, J., Meunier, A., Merle, C., Jia, Z., Barnier, J.-V., and Gradwohl, G. (2014). Pak3 promotes cell cycle exit and differentiation of β -cells in the embryonic pancreas and is necessary to maintain glucose homeostasis in adult mice. *Diabetes* 63, 203–215. <https://doi.org/10.2337/db13-0384>.
 18. Bechard, M.E., Bankaitis, E.D., Hipkens, S.B., Ustione, A., Piston, D.W., Yang, Y.-P., Magnuson, M.A., and Wright, C.V.E. (2016). Precommitment low-level *neurog3* expression defines a long-lived mitotic endocrine-biased progenitor pool that drives production of endocrine-committed cells. *Genes Dev.* 30, 1852–1865. <https://doi.org/10.1101/gad.284729.116>.
 19. Petersen, M.B.K., Azad, A., Ingvorsen, C., Hess, K., Hansson, M., Grapin-Botton, A., and Honoré, C. (2017). Single-cell gene expression analysis of a human ESC model of pancreatic endocrine development reveals different paths to β -cell differentiation. *Stem Cell Rep.* 9, 1246–1261. <https://doi.org/10.1016/j.stemcr.2017.08.009>.
 20. Rezaia, A., Bruin, J.E., Arora, P., Rubin, A., Batushansky, I., Asadi, A., O'Dwyer, S., Quiskamp, N., Mojibian, M., Albrecht, T., et al. (2014). Reversal of diabetes with insulin-producing cells derived in vitro from human pluripotent stem cells. *Nat. Biotechnol.* 32, 1121–1133. <https://doi.org/10.1038/nbt.3033>.
 21. Gonçalves, C.A., Larsen, M., Jung, S., Stratmann, J., Nakamura, A., Leuschner, M., Hersemann, L., Keshara, R., Perlman, S., Lundvall, L., et al. (2021). A 3D system to model human pancreas development and its reference single-cell transcriptome atlas identify signaling pathways required for progenitor expansion. *Nat. Commun.* 12, 3144. <https://doi.org/10.1038/s41467-021-23295-6>.
 22. Krentz, N.A.J., van Hoof, D., Li, Z., Watanabe, A., Tang, M., Nian, C., German, M.S., and Lynn, F.C. (2017). Phosphorylation of *NEUROG3* links endocrine differentiation to the cell cycle in pancreatic progenitors. *Dev. Cell* 41, 129–142.e6. <https://doi.org/10.1016/j.devcel.2017.02.006>.
 23. Zhang, X., McGrath, P.S., Salomone, J., Rahal, M., McCauley, H.A., Schweitzer, J., Kovall, R., Gebelein, B., and Wells, J.M. (2019). A comprehensive structure-function study of Neurogenin3 disease-causing alleles during human pancreas and intestinal organoid development. *Dev. Cell* 50, 367–380.e7. <https://doi.org/10.1016/j.devcel.2019.05.017>.
 24. Jensen, J., Pedersen, E.E., Galante, P., Hald, J., Heller, R.S., Ishibashi, M., Kageyama, R., Guillemot, F., Serup, P., and Madsen, O.D. (2000). Control of endodermal endocrine development by *Hes-1*. *Nat. Genet.* 24, 36–44. <https://doi.org/10.1038/71657>.
 25. de Lichtenberg, K.H., Funari, N., Nakic, N., Ferrer, J., Zhu, Z., Huangfu, D., and Serup, P. (2018). Genome-wide identification of *HES1* target genes uncover novel roles for *HES1* in pancreatic development. Preprint at bioRxiv. <https://doi.org/10.1101/335869>.
 26. Picelli, S., Björklund, Å.K., Faridani, O.R., Sagasser, S., Winberg, G., and Sandberg, R. (2013). Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat. Methods* 10, 1096–1098. <https://doi.org/10.1038/nmeth.2639>.
 27. Yu, X.-X., Qiu, W.-L., Yang, L., Wang, Y.-C., He, M.-Y., Wang, D., Zhang, Y., Li, L.-C., Zhang, J., Wang, Y., and Xu, C.-R. (2021). Sequential progenitor states mark the generation of pancreatic endocrine lineages in mice and humans. *Cell Res.* 31, 886–903. <https://doi.org/10.1038/s41422-021-00486-w>.
 28. Veres, A., Faust, A.L., Bushnell, H.L., Engquist, E.N., Kenty, J.H.-R., Harb, G., Poh, Y.-C., Sintov, E., Gürtler, M., Pagliuca, F.W., et al. (2019). Charting cellular identity during human in vitro β -cell differentiation. *Nature* 569, 368–373. <https://doi.org/10.1038/s41586-019-1168-5>.
 29. Ramond, C., Beydag-Tasöz, B.S., Azad, A., van de Bunt, M., Petersen, M.B.K., Beer, N.L., Glaser, N., Berthault, C., Gloyn, A.L., Hansson, M., et al. (2018). Understanding human fetal pancreas development using subpopulation sorting, RNA sequencing and single-cell profiling. *Development* 145, dev165480. <https://doi.org/10.1242/dev.165480>.
 30. Gouzi, M., Kim, Y.H., Katsumoto, K., Johansson, K., and Grapin-Botton, A. (2011). Neurogenin3 initiates stepwise delamination of differentiating

- endocrine cells during pancreas development. *Dev. Dyn.* 240, 589–604. <https://doi.org/10.1002/dvdy.22544>.
31. Kryza, T., Parent, C., Pardessus, J., Petit, A., Burlaud-Gaillard, J., Reverdiou, P., Lochmann, S., Labas, V., Courty, Y., and Heuzé-Vourc'h, N. (2018). Human kallikrein-related peptidase 12 stimulates endothelial cell migration by remodeling the fibronectin matrix. *Sci. Rep.* 8, 6331. <https://doi.org/10.1038/s41598-018-24576-9>.
 32. Balleza, E., Kim, J.M., and Cluzel, P. (2018). Systematic characterization of maturation time of fluorescent proteins in living cells. *Nat. Methods* 15, 47–51. <https://doi.org/10.1038/nmeth.4509>.
 33. Matsuda, M., Hayashi, H., Garcia-Ojalvo, J., Yoshioka-Kobayashi, K., Kageyama, R., Yamanaka, Y., Ikeya, M., Toguchida, J., Alev, C., and Ebisuya, M. (2020). Species-specific segmentation clock periods are due to differential biochemical reaction speeds. *Science* 369, 1450–1455. <https://doi.org/10.1126/science.aba7668>.
 34. Rayon, T., Stamatakis, D., Perez-Carrasco, R., Garcia-Perez, L., Barrington, C., Melchionda, M., Exelby, K., Lazaro, J., Tybulewicz, V.L.J., Fisher, E.M.C., and Briscoe, J. (2020). Species-specific pace of development is associated with differences in protein stability. *Science* 369, eaba7667. <https://doi.org/10.1126/science.aba7667>.
 35. Bastidas-Ponce, A., Tritschler, S., Dony, L., Scheibner, K., Tarquis-Medina, M., Salinno, C., Schirge, S., Burtscher, I., Böttcher, A., Theis, F.J., et al. (2019). Comprehensive single cell mRNA profiling reveals a detailed roadmap for pancreatic endocrinogenesis. *Development* 146, dev173849. <https://doi.org/10.1242/dev.173849>.
 36. Roark, R., Itzhaki, L., and Philpott, A. (2012). Complex regulation controls Neurogenin3 proteolysis. *Biol. Open* 1, 1264–1272. <https://doi.org/10.1242/bio.20121750>.
 37. Azzarelli, R., Hurley, C., Sznurkowska, M.K., Rulands, S., Hardwick, L., Gamper, I., Ali, F., McCracken, L., Hindley, C., McDuff, F., et al. (2017). Multi-site Neurogenin3 phosphorylation controls pancreatic endocrine differentiation. *Dev. Cell* 41, 274–286.e5. <https://doi.org/10.1016/j.devcel.2017.04.004>.
 38. 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. <https://doi.org/10.1016/j.ydbio.2004.03.013>.
 39. Cortijo, C., Gouzi, M., Tissir, F., and Grapin-Botton, A. (2012). Planar cell polarity controls pancreatic beta cell differentiation and glucose homeostasis. *Cell Rep.* 2, 1593–1606. <https://doi.org/10.1016/j.celrep.2012.10.016>.
 40. Krentz, N.A.J., Lee, M.Y.Y., Xu, E.E., Sproul, S.L.J., Maslova, A., Sasaki, S., and Lynn, F.C. (2018). Single-cell transcriptome profiling of mouse and hESC-derived pancreatic progenitors. *Stem Cell Rep.* 11, 1551–1564. <https://doi.org/10.1016/j.stemcr.2018.11.008>.
 41. Schreiber, V., Mercier, R., Jiménez, S., Ye, T., García-Sánchez, E., Klein, A., Meunier, A., Ghimire, S., Birc, C., Jost, B., et al. (2021). Extensive NEUROG3 occupancy in the human pancreatic endocrine gene regulatory network. *Mol. Metab.* 53, 101313. <https://doi.org/10.1016/j.molmet.2021.101313>.
 42. Larsen, H.L., Martín-Coll, L., Nielsen, A.V., Wright, C.V.E., Trusina, A., Kim, Y.H., and Grapin-Botton, A. (2017). Stochastic priming and spatial cues orchestrate heterogeneous clonal contribution to mouse pancreas organogenesis. *Nat. Commun.* 8, 605. <https://doi.org/10.1038/s41467-017-00258-4>.
 43. Sharon, N., Chawla, R., Mueller, J., Vanderhooft, J., Whitehorn, L.J., Rosenthal, B., Gürtler, M., Estanboulieh, R.R., Shvartsman, D., Gifford, D.K., et al. (2019). A peninsular structure coordinates asynchronous differentiation with morphogenesis to generate pancreatic islets. *Cell* 176, 790–804.e13. <https://doi.org/10.1016/j.cell.2018.12.003>.
 44. Shimojo, H., Ohtsuka, T., and Kageyama, R. (2008). Oscillations in Notch signaling regulate maintenance of neural progenitors. *Neuron* 58, 52–64. <https://doi.org/10.1016/j.neuron.2008.02.014>.
 45. Imayoshi, I., Isomura, A., Harima, Y., Kawaguchi, K., Kori, H., Miyachi, H., Fujiwara, T., Ishidate, F., and Kageyama, R. (2013). Oscillatory control of factors determining multipotency and fate in mouse neural progenitors. *Science* 342, 1203–1208. <https://doi.org/10.1126/science.1242366>.
 46. Seymour, P.A., Collin, C.A., Egeskov-Madsen, A.R., Jørgensen, M.C., Shimojo, H., Imayoshi, I., de Lichtenberg, K.H., Kopan, R., Kageyama, R., and Serup, P. (2020). Jag1 modulates an oscillatory Dll1-Notch-Hes1 signaling module to coordinate growth and fate of pancreatic progenitors. *Dev. Cell* 52, 731–747.e8. <https://doi.org/10.1016/j.devcel.2020.01.015>.
 47. Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., and Zhang, F. (2013). Multiplex genome engineering using CRISPR/Cas systems. *Science* 339, 819–823. <https://doi.org/10.1126/science.1231143>.
 48. Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. *Nat. Methods* 9, 676–682. <https://doi.org/10.1038/nmeth.2019>.
 49. Schmidt, U., Weigert, M., Broaddus, C., and Myers, G. (2018). Cell detection with star-convex polygons. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018. Lecture Notes in Computer Science*, A. Frangi, J. Schnabel, C. Davatzikos, C. Alberola-López, and G. Fichtinger, eds. (Springer), pp. 265–273. https://doi.org/10.1007/978-3-030-00934-2_30.
 50. Haase, R., Rajasekhar, P., Lambert, T., grahamross123, Nunez-Iglesias, J., Lachie, Avenel, C., and Sandaltzopoulou, E. (2022). cEsperanto/pyclesperanto_prototype: 0.20.0. Zenodo. <https://doi.org/10.5281/zenodo.7389742>.
 51. Kechin, A., Boyarskikh, U., Kel, A., and Filipenko, M. (2017). cutPrimers: a new tool for accurate cutting of primers from reads of targeted next generation sequencing. *J. Comput. Biol.* 24, 1138–1143. <https://doi.org/10.1089/cmb.2017.0096>.
 52. Wu, T.D., and Nacu, S. (2010). Fast and SNP-tolerant detection of complex variants and splicing in short reads. *Bioinformatics* 26, 873–881. <https://doi.org/10.1093/bioinformatics/btq057>.
 53. Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30, 923–930. <https://doi.org/10.1093/bioinformatics/btt656>.
 54. McCarthy, D.J., Campbell, K.R., Lun, A.T.L., and Wills, Q.F. (2017). Scater: pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in R. *Bioinformatics* 33, 1179–1186. <https://doi.org/10.1093/bioinformatics/btw777>.
 55. Wolf, F.A., Angerer, P., and Theis, F.J. (2018). SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* 19, 15. <https://doi.org/10.1186/s13059-017-1382-0>.
 56. Traag, V.A., Waltman, L., and van Eck, N.J. (2019). From Louvain to Leiden: guaranteeing well-connected communities. *Sci. Rep.* 9, 5233. <https://doi.org/10.1038/s41598-019-41695-z>.
 57. La Manno, G., Soldatov, R., Zeisel, A., Braun, E., Hochgerner, H., Petukhov, V., Lidschreiber, K., Kastrioti, M.E., Lönnerberg, P., Furlan, A., et al. (2018). RNA velocity of single cells. *Nature* 560, 494–498. <https://doi.org/10.1038/s41586-018-0414-6>.
 58. Bergen, V., Lange, M., Peidli, S., Wolf, F.A., and Theis, F.J. (2020). Generalizing RNA velocity to transient cell states through dynamical modeling. *Nat. Biotechnol.* 38, 1408–1414. <https://doi.org/10.1038/s41587-020-0591-3>.
 59. 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. <https://doi.org/10.1186/s13059-014-0550-8>.
 60. Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., and Mesirov, J.P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA* 102, 15545–15550. <https://doi.org/10.1073/pnas.0506580102>.

61. Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J.P., and Tamayo, P. (2015). The molecular signatures database hallmark gene set collection. *Cell Syst.* *1*, 417–425. <https://doi.org/10.1016/j.cels.2015.12.004>.
62. Korotkevich, G., Sukhov, V., Budin, N., Shpak, B., Artyomov, M.N., and Sergushichev, A. (2021). Fast gene set enrichment analysis. Preprint at bioRxiv. <https://doi.org/10.1101/060012>.
63. Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W.M., III, Zheng, S., Butler, A., Lee, M.J., Wilk, A.J., Darby, C., Zager, M., et al. (2021). Integrated analysis of multimodal single-cell data. *Cell* *184*, 3573–3587.e29. <https://doi.org/10.1016/j.cell.2021.04.048>.
64. Weigert, M., Schmidt, U., Haase, R., Sugawara, K., and Myers, G. (2020). Star-convex polyhedra for 3D object detection and segmentation in microscopy. In 2020 IEEE Winter Conference on Applications of Computer Vision (WACV), pp. 3655–3662. <https://doi.org/10.1109/WACV45572.2020.9093435>.
65. Tirosh, I., Izar, B., Prakadan, S.M., Wadsworth, M.H., III, Treacy, D., Trombetta, J.J., Rotem, A., Rodman, C., Lian, C., Murphy, G., et al. (2016). Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* *352*, 189–196. <https://doi.org/10.1126/science.aad0501>.
66. Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., III, Hao, Y., Stoeckius, M., Smibert, P., and Satija, R. (2019). Comprehensive integration of single-cell data. *Cell* *177*, 1888–1902. <https://doi.org/10.1016/J.CELL.2019.05.031>.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Rat anti-Cpeptide-Alexa Fluor 647	BD Pharmigen	Cat#565831; RRID: AB_2739371
Mouse anti-Glucagon-BV421	BD Pharmigen	Cat#565891; RRID: AB_2739385
Sheep anti-NEUROG3	R&D Systems	Cat#AF3444; RRID: AB_2149527
Chicken anti-GFP	Abcam	Cat#ab13970; RRID: AB_300798
Rat anti-RFP	Chromotek	Cat#5F8; RRID: AB_2336064
Rabbit anti-MKI67	Abcam	Cat#ab16667; RRID: AB_302459
Rabbit anti-SOX9	Millipore	Cat#AB5535; RRID: AB_2239761
Bacterial and Virus Strains		
One Shot TOP10 E. coli	Life Technologies	Cat#C4040
Chemicals, Peptides, and Recombinant Proteins		
hESC-Qualified Matrigel	Corning	Cat#354277
mTESR1 medium	Stem Cell Technologies	Cat#85850
ROCK inhibitor Y-27632	Sigma	Cat#SCM075
Growth factor reduced Matrigel, phenol red-free	Corning	Cat#356231
MCDDB131 medium	Life Technologies	Cat#10372019
CHIR99021	Axon Medchem	Cat#MC-1279
Activin A	Peprotech	Cat#120-14
FGF7	Peprotech	Cat#100-19
Retinoic acid	Sigma	Cat#R2625
Sant-1	Sigma	Cat#S4572
LDN	Sigma	Cat#SML0559
TPB	Millipore	Cat#565740
Alk5i II	Santa Cruz	Cat#sc-221234B
T3	Sigma	Cat#T6397
Heparin	Sigma	Cat#H3149
Gamma-secretase inhibitor XX	EMD Millipore	Cat#565789
Ascorbic acid	Sigma	Cat#A4544
Zinc sulfate	Sigma	Cat#Z0251
FGF2	Peprotech	Cat#100-18B
DMEM/F-12, GlutaMAX	Thermo Fisher	Cat#10565018
B-27 supplement	Thermo Fisher	Cat#A1486701
Glutamax	Life Technologies	Cat#35050-038
ITS-X	Life Technologies	Cat#51500056
Sodium bicarbonate	Life Technologies	Cat#25080060
Ghost dye red 780	Cell Signaling Technology	Cat#18452S
Hoechst 33342 Ready Flow Reagent	Thermo Fisher	Cat#R37165
Critical Commercial Assays		
GeneArt Seamless PLUS Cloning and Assembly Kit	Thermo Fisher	Cat#A14603
RNeasy Micro Kit	QIAGEN	Cat#74004

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited Data		
Single-cell RNA sequencing of human embryonic stem cell-derived pancreatic cells enriched for early endocrine progenitors using a novel dual NEUROG3-reporter	This paper	ArrayExpress Accession number: E-MTAB-11208
Bulk RNA sequencing of cycling and non-cycling NEUROG3-positive and negative human embryonic stem cell-derived pancreatic cells using a dual NEUROG3-reporter	This paper	ArrayExpress Accession number: E-MTAB-11210
Human fetal pancreas mSTRTSeq dataset	Yu et al. ²⁷	BioProject: PRJCA004336
Experimental Models: Cell Lines		
H1 Human ESC line	WiCell	Cat#WA01
Oligonucleotides		
NEUROG3 guide RNA CTTTTGAGGGCCGCCCGT	IDT DNA	N/A
Recombinant DNA		
pX330-U6-Chimeric_BB-CBh-hSpCas9	Cong et al. ⁴⁷	Addgene #42230
pBluescript SK (-)	In house	N/A
Software and Algorithms		
Imaris	Bitplane	N/A
Prism	GraphPad Software	N/A
BD FACSDiva	BD Biosciences	N/A
FCS Express	De Novo Software	N/A
Python	Python Software Foundation	N/A
R	R Core Team	N/A
MATLAB	MathWorks	N/A
Fiji v1.53c	Schindelin et al. ⁴⁸	N/A
StarDist	Schmidt et al. ⁴⁹	N/A
py-clesperanto V0.20.0	Haase et al. ⁵⁰	N/A
cutadapt v2.6	Kechin et al. ⁵¹	N/A
gsnap v2018-07-04	Wu and Nacu ⁵²	N/A
featureCounts v1.6.3	Liao et al. ⁵³	N/A
scater v1.16.2	McCarthy et al. ⁵⁴	N/A
Scanpy v1.8.1	Wolf et al. ⁵⁵	N/A
Leiden	Traag et al. ⁵⁶	N/A
velocyto v0.17.17	La Manno et al. ⁵⁷	N/A
scVelo v0.2.3	Bergen et al. ⁵⁸	N/A
DESeq2 v1.22.1	Love et al. ⁵⁹	N/A
msigdb v7.2.1	Subramanian et al. ⁶⁰ Liberzon et al. ⁶¹	N/A
fgsea v1.8.0	Korotkevich et al. ⁶²	N/A
Seurat v4.1.3	Hao et al. ⁶³	N/A
ggpubr v0.4.0	https://github.com/kassambara/ggpubr	N/A

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact Anne Grapin-Botton (botton@mpi-cbg.de).

Materials availability

The *NEUROG3* dual reporter hESC line generated during this study will be provided upon reasonable request.

Data and code availability

The raw single-cell and bulk RNA sequencing datasets have been deposited to ArrayExpress ([ArrayExpress < EMBL-EBI](https://www.ebi.ac.uk/arrayexpress/)) with the identification numbers E-MTAB-11208 and E-MTAB-11210 respectively. Codes related to this paper can be found at <https://doi.org/10.21101/0000-0007-F075-6>. Any additional information required to reanalyze the data reported in this work paper is available from the [lead contact](#) upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Culturing of human embryonic stem cells

The H1 hESC line was purchased from WiCell. Approval to work on hESC was obtained from De Videnskabetiske Komiteer, Region Hovedstaden under number H-4-2013-057 and Central Ethics Commission for Stem Cell Research (ZES) and the Robert Koch Institute (RKI) (AZ: 3.04.02/0148). The hESCs were cultured on hESC-Qualified Matrigel (Corning) and maintained using mTESR1 medium (Stem Cell Technologies). The media was changed daily, and the cells were kept at 37°C and 5% CO₂. Approximately every 3-4 days, after the cell confluence had passed 70-80%, the cells were passaged by dissociating with TrypLE (Thermo Fisher) and replating at a density of 40,000 cells/cm². During the first 24 hours after passaging, hESCs were supplemented with 10 μM ROCK inhibitor Y-27632 (Sigma). The cells were periodically tested and cleared for mycoplasma and karyotypic abnormalities. The hESC work was conducted according to the guidelines.

METHOD DETAILS

Construction of *NEUROG3* dual reporter hESC line

For the construction of the *NEUROG3* dual reporter hESC line, insertion of GSx3 linker-TagRFP-T-P2A-EGFP-NLS-STOP coding sequence immediately before the stop codon of the *NEUROG3* coding sequence in H1 hESCs was achieved by homology-directed repair of a CRISPR/Cas9-induced DNA break.

A guide RNA targeting the *NEUROG3* locus (CTTTTGAGGGCCGCCCGCT) was designed with CRISPR design (<http://crispr.mit.edu>), using the pX330 vector that encodes for human codon-optimized SpCas9.⁴⁷ The DNA plasmid included around 1000 base pairs-long homology arms from *NEUROG3* genomic DNA with a point mutation of protospacer adjacent motif (PAM) sequence in a vector backbone (pBluescriptSK), with the desired construct and an Flippase (FLP)-based selection cassette. The fragments were assembled using the Seamless cloning kit (Thermo Fisher), according to the manufacturer's instructions. The two plasmids were co-electroporated into H1 hESCs, and antibiotic selection was performed. After the initial screening, 30 of the surviving colonies were analyzed for correct integration. Following validation, the selection cassette was excised by FLP recombinase. From the nine reporter colonies that were validated, the choice was made on *in vitro* pancreatic differentiation efficiency.

The karyogram of the *NEUROG3* dual reporter was analyzed on metaphase spreads at P14 (Cell Guidance Systems, Cambridge, UK) and P17 (Organisation Genetische Diagnostik Institut für Klinische Genetik Medizinische Fakultät Carl Gustav Carus TU Dresden).

In vitro differentiation into the pancreatic lineage

In vitro differentiation of hESCs into pancreatic endocrine cells was performed as previously described,^{19,29} based on the protocol published by Reznia et al.,²⁰ with minor adjustments. In summary, hESCs were dissociated into single cells using TrypLE and resuspended in mTeSR1 with 10 μM ROCK inhibitor Y-27632. The cells were seeded at a density of 350,000 cells/cm² onto 1:30 diluted growth-factor reduced Matrigel (Corning). The cells were incubated for 24 hours in mTeSR1 medium before transitioning to the differentiation media. Modifications from the original protocol²⁰ include the usage of MCDB131 medium (Life Technologies) as the basal medium throughout differentiation instead of BLAR medium, activin A (PeproTech) at 100 ng/ml during S1 instead of GDF8, CHIR99201 (Axon Medchem) at 3 μM and 0.3 μM for the first two days of S1 instead of MCX-928, and keeping the cells in 2D culture throughout the protocol instead of transitioning to air-liquid interface at S5. The modified protocol's differentiation efficiency was reported by Petersen et al.¹⁹ and is further documented for the reporter line in this paper (Figures 1 and S1). For the live imaging experiments, reporter hESCs and H1 hESCs were mixed in a 1:1 ratio and then differentiated for the ease of tracking fluorescent cells.

The generation of human pancreatic progenitor spheroids was performed as previously described in Gonçalves et al.²¹ Pancreatic progenitors differentiated from hESCs at the end of S4 using the described pancreatic differentiation protocol were dissociated into single cells with TrypLE and resuspended in a 3:1 mixture of growth-factor reduced Matrigel and expansion medium (DMEM/F12 Glutamax (Thermo Fisher), 10 μM ROCK inhibitor Y-27632, 64 ng/ml FGF2 (Peprotech), and B-27 supplement (Thermo Fisher)). The cells were seeded at a density of 1000 cells/μl of Matrigel mixture onto pre-warmed Nunc cell culture-treated 4-well dishes (Thermo Fisher) to form a 3D dome. The expansion medium was added after Matrigel polymerization and replenished every 3 days. The spheroids were passaged every 10 days. For the induction of endocrine differentiation in pancreatic progenitor spheroids, 48 hours after seeding, 100 nM gamma-secretase inhibitor XX (EMD Millipore) was added to the expansion medium and was replenished every other day.

Live imaging and cell tracking

Short-term live imaging experiments with the 2D differentiation system were carried out using a Deltavision microscope (PerkinElmer) with the 20X objective in a humidified, heated, and CO₂-controlled chamber. Tiled positions (4x4 tiles, 16 tiles per acquisition) were scanned every 6 minutes. Long-term live imaging experiments with the 2D differentiation system were carried out using the Andor Revolution WD Borealis Mosaic spinning disc microscope (Andor) with the 40X silicone objective in a humidified, heated, and CO₂-controlled chamber. Tiled positions (5x5 tiles, 25 tiles per acquisition) were scanned every 12 minutes with the z-step of 4 μm covering 60 μm in total. At the end of the imaging sessions, the cultures were immediately fixed with 4% PFA (Thermo Fisher) for immunocytochemistry. Imaris (Bitplane, Switzerland) software was used to perform manual tracking of NEUROG3+ cells over time, using spots of 5 μm diameter.

For the 3D cultures, a Viventis light-sheet microscope was used with a Nikon 25x 1.1 NA water immersion objective in a humidified, heated, and CO₂-controlled chamber. Set positions were images every 10 minutes. Manual tracing of NEUROG3+ cells was performed with Imaris (Bitplane, Switzerland) software, using spots of 5 μm diameter.

Flow Cytometric Analyses

For flow cytometry-based live cell cycle analysis, at different time points during human *in vitro* pancreatic differentiation, cells were dissociated and collected as described above and reconstituted in FACS buffer (1% BSA (Sigma) in 1X PBS (Thermo Fisher)). Ghost dye red 780 (Cell Signaling Technology) for live/dead stain and Hoechst 33342 Ready Flow Reagent (Thermo Fisher) were added with the concentration of 1 μl/1,000,000 cells and 2 drops/1,000,000 cells in 1 mL, respectively. The cells were incubated at 37°C for 1 hour, followed by filtering with a 35 μm cell strainer and flow analysis on a BD FACSAria (BD Biosciences).

For fixed flow cytometric analyses, after dissociating cells to a single cell suspension, the cells were washed with 1X PBS and incubated on ice in the dark with 1 μl of live/dead stain (Ghost dye red 780) in 1X PBS for 10 minutes. Afterward, the cells were washed with 1x PBS and fixed with 4% PFA on ice for 20 minutes. After spinning down, the cells were resuspended in a permeabilization buffer (0.2% Triton X-100 (Sigma) and 5% Donkey serum (Sigma) in 1X PBS) and incubated on ice for 30 minutes. After pelleting, primary antibodies conjugated with a fluorophore in a blocking buffer (0.1% Triton X-100 and 5% Donkey serum in 1X PBS) were added, and the samples were incubated at 4°C overnight. The next day, the cells were washed 3 times with an IC buffer (1% BSA in 1X PBS) and passed through a 35 μm cell strainer. Flow analysis was performed on a FACSAria or a Sony MA900 cell sorter (Sony). The primary antibodies for flow cytometry were: rat anti-C-peptide-Alexa Fluor 647 (1:200, 565831, BD Pharmingen) and mouse anti-Glucagon-BV421 (1:100, 565891, BD Pharmingen).

Immunostaining

For the staining of hESC-derived pancreatic cells, culture wells were washed with 1X PBS and fixed with 4% PFA at room temperature for 30 minutes. Afterward, the cells were washed with 1X PBS and permeabilized with 0.5% Triton X-100 in 1X PBS for 10 minutes. The cells were incubated in blocking buffer (0.1 M Tris-HCl (pH 7.5), 0.15 M NaCl, 0.5% TSA Blocking Reagent (PerkinElmer)) for 30 minutes, followed by overnight incubation at 4°C with primary antibodies diluted in 0.1% Triton X-100 in 1X PBS. The following day, cells were washed 3 times with 1X PBS for 5 minutes each and incubated with secondary antibodies in 0.1% Triton X-100 in 1X PBS for an hour at room temperature and imaged with a Leica Thunder microscope. The primary antibodies were: sheep anti-NEUROG3 (1:1000, AF3444, R&D Systems), chicken anti-GFP (1:1000, ab13970, Abcam), rat anti-RFP (1:1000, 5F8, Chromotek), rabbit anti-MKI67 (1:500, ab16667, Abcam), and rabbit anti-SOX9 (1:2000, AB5535, Millipore).

Image analysis

Image processing prior to manual cell tracing in 2D/3D and quantification of the overlap of NEUROG3 immunostaining and the fluorescence from the reporter line was performed using the ImageJ distribution Fiji (v1.53c) (<https://fiji.sc/>).⁴⁸ For live imaging processing prior to manual 2D cell tracing, an automated script was used to perform the following steps: Stitching tile positions, drift correction, background subtraction, and downscaling. For live imaging processing prior to manual 3D cell tracing, cropping, and bit conversions were applied.

For quantification of the overlap of NEUROG3 immunostaining and the fluorescence from the reporter line, nuclei were segmented using StarDist⁴⁹ (<https://imagej.net/plugins/stardist>) from merged channel images. An 8 μm diameter size filter and nuclei signal filter (> background mean intensity) were applied to the segmented nuclei to select positive cells, and overlapping cells between channels were identified.

The segmentation of GFP-positive nuclei in 3D-imaged 2D cultures was performed using StarDist.^{49,64} First, random images from live imaging were manually annotated using Napari as training data to create a StarDist model. After, the trained model was applied to predict and segment the nuclei. The nuclear segmented output was used to quantify the nuclear GFP intensities that were normalized by mean background subtraction using py-clesperanto V0.20.0.⁵⁰

In order to understand the correlation of motility of the endocrine progenitor cells in different transcriptomic clusters with the intensity levels of GFP, the global displacement of cells was quantified. This was achieved by extracting the centroids of the segmented nuclei using py-clesperanto and these centroids were then treated as a point-cloud in the corresponding timepoint. The distance between the median position of the point-clouds between successive timepoints was used as a measure to quantify the global displacement of EP cells over time.

FACS, scRNA-seq, and downstream analysis

At S4D1 of differentiation, the cells were dissociated into single cells with TrypLE and resuspended into a FACS buffer containing 0.5% BSA and 2 mM EDTA (Thermo Fisher) in 1X PBS. The cells were index-sorted according to their GFP expression (categorized into GFP-low and GFP-high based on the distribution of global GFP values) into 384-well plates on a BD FACSAria and snap-frozen. 397 sorted cells were processed using a modified version of the Smart-seq2 protocol.²⁶

FastQC (<http://www.bioinformatics.babraham.ac.uk>) was used to perform basic quality control of resulting sequencing data. Trimming of low quality bases, Smartseq2 adapter sequences, and poly(A)/poly(T) stretches was done with cutadapt (v2.6)⁵¹ and the following parameters: “-q 5 -m 20 -a smartseq1=AAGCAGTGGTATCA -a smartseq2=AACGCAGAGTGCAGTGC -b smartseq3=AACGCAGAGTGCAGTGC -a smartseq4=CTGTCTCTTATA -a polyA=AAAAAAAA -a polyT=TTTTTTTTTTT -b smartseq5=GGTATCAACGCAGA -times 5”. Trimmed fragments were aligned to the human reference genome hg38 with support of the Ensembl 92 splice sites using the aligner gsnap (v2018-07-04).⁵² The sequence of the *NEUROG3* fusion reporter was included in the reference without the *NEUROG3* sequence to allow for mapping to the endogenous *NEUROG3* gene. Counts per gene and sample were obtained based on the overlap of uniquely mapped fragments with the same Ensembl annotation and the included fusion reporter sequence using featureCounts (v1.6.3).⁵³ After removing three sequencing controls (positive, negative, and bulk controls) from the count matrix, the remaining 397 cells were checked for overall quality using the scater package v1.16.2.⁵⁴ Detected outlier cells were removed from the data set, and the remaining 380 cells were further analyzed with Scanpy v1.8.1⁵⁵ to perform clustering with the Leiden algorithm⁵⁶ and marker gene detection using the Wilcoxon rank-sum test.

For RNA velocity analysis, spliced and unspliced transcripts were quantified with smartseq2 command of velocity v0.17.17⁵⁷ and served as input for further analysis with scVelo v0.2.3⁵⁸ using dynamical modeling. RNA velocity was projected upon a UMAP obtained with Scanpy after additional regression of cell cycle scores assigned based on cell cycle genes defined in Tirosh et al.⁶⁵

RNA extraction and sequencing of human pancreatic cell populations

Bulk RNA sequencing was performed on cells from 4 independent *in vitro* pancreatic differentiation experiments. The cells were FACS-sorted into RLT buffer (QIAGEN) with 1% β -mercaptoethanol (Sigma Aldrich) according to their GFP fluorescence intensity and cell cycle category with live cell cycle staining described above on a Sony MA900 sorter. RNA was extracted from each sample using the RNeasy Micro Kit (QIAGEN), according to the manufacturer's guidelines. Subsequent RNA quality assessment was done with Bioanalyzer (Agilent). Samples of extracted RNA were sequenced with the Smart-Seq2 protocol for the low RNA abundance, providing datasets of unstranded, paired-end (2x101) reads of sizes between 54 Mio and 102 Mio. Smart-Seq2 adapters and poly-A tails were aggressively cut from both ends of the paired-end reads with Cutadapt v1.16, discarding reads shorter than 15 nucleotides, which left datasets of sizes between 37 Mio and 80 Mio. Reads were mapped to the human genome version GRCh38.p13 with splice-aware mapper STAR v2.7.3. with high alignment rates (93.3% to 94.9%), and the number of uniquely mapping, unstranded fragments per gene were determined with STAR option `-quantMode GeneCounts`.

These fragment counts per gene (sum over all genes 28 Mio to 63 Mio) were analyzed for differential gene expression with R v3.5.1 and DESeq2 v1.22.1.⁵⁹ Normalizing with DESeq2 sizeFactors for different dataset sizes and applying PCA to regularized-log-transformed counts, we identified one outlier sample `gfpp_g1_2` whose location in the PC1-PC2 plot is far from its biological replicates. Thus, we discarded this outlier sample from any further analysis. To be considered as differentially expressed genes comparing any two groups of samples, we required that these genes had at least 10 counts in at least one sample, an absolute \log_2 -fold change larger than 1, and an adjusted p-value of max 0.05. Gene set enrichment analyses were run against MSigDB hallmark genes set with more than 15 and less than 500 genes (`msigdb v7.2.1`)^{60,61} using `fgsea v1.8.0`⁶² by applying an adjusted p-value cutoff of 0.05. For this analysis, the ranking of input genes was done using the negative \log_{10} of obtained p-values which were then additionally multiplied with -1 in case of a negative \log_2 fold-change.

Integration of scRNA-seq datasets

Data integration of our *in vitro* clusters PP and EP1-EP3 and *in vivo* clusters EP1-EP4 of the published human fetal pancreas mSTRTSeq dataset (BioProject: PRJCA004336²⁷) was done with Seurat v4.1.3.⁶³ The count matrices were filtered for features expressed in at least 10 cells and cells with at least 2500 detected features, resulting in 323 cells and 19130 features for our *in vitro* data and 163 cells and 11376 features for the *in vivo* dataset. Based on 4000 highly variable genes detected per dataset, 1000 integration features were selected and used for the identification of anchors using reciprocal PCA.⁶⁶ The first two principal components of the PCA space were used for the analysis. In total, 95 anchors were identified with a medium score of 0.69 and used for pairwise data integration.

Integration of live-imaging and sequencing datasets

Intensity measurements were used to link transcriptome and live-cell data statistically. Due to potential differences between FACS and live imaging in terms of sensitivity and dynamic range, the raw GFP measurements of both methods are not directly comparable. To account for this, GFP measurements obtained from live imaging were quantile-transformed to match the distribution of GFP values from flow cytometry. The latter was normalized to forward scatter (FSC) as a proxy for intensity per volume for the entire population of S4D1 cells. This strategy assumes that the ordering of GFP values revealed by both techniques is approximately equal. The GFP values of sequenced cells obtained from FACS were then grouped according to the three transcriptional clusters identified (EP1-3) using the Leiden algorithm. For each cluster, we estimated the corresponding conditional distribution over GFP levels using

Gaussian mixture modeling. Each cell tracked from imaging was then classified into different clusters by picking the cluster for which it had the highest likelihood. Different features revealed from imaging were then analyzed according to the predicted clusters. Analyses were performed in Matlab using a custom code.

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical significance tests were performed using the `stat_compare_means` function in the `ggpubr` R package v0.4.0 (<https://github.com/kassambara/ggpubr>). Significance was defined as * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$. P values were determined by Wilcoxon and Kruskal-Wallis tests. The number of replicates is provided in the figure legends. N denotes the number of independent experiments, and n denotes the total number of measurements.