

# Spatial and transcriptional heterogeneity of $\beta$ -cell neogenesis revealed by a time-resolved reporter system

**Shugo Sasaki**

BC Children's Hospital Research Institute

**Michelle Lee**

BC Children's Hospital Research Institute

**Yuka Wakabayashi**

Juntendo University Graduate School of Medicine

**Luka Suzuki**

Juntendo University Graduate School of Medicine

**Helena Winata**

BC Children's Hospital Research Institute

**Miwa Himuro**

Juntendo University Graduate School of Medicine

**Takaaki Matsuoka**

Osaka University

**Ichiro Shimomura**

Osaka University Graduate School of Medicine

**Hiroataka Watada**

Juntendo University <https://orcid.org/0000-0001-5961-1816>

**Francis Lynn**

BC Children's Hospital Research Institute

**Takeshi Miyatsuka** (✉ [miyatsuka-takeshi@umin.net](mailto:miyatsuka-takeshi@umin.net))

Juntendo University Graduate School of Medicine

---

**Article**

**Keywords:**

**Posted Date:** August 7th, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-46466/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# **Spatial and transcriptional heterogeneity of $\beta$ -cell neogenesis revealed by a time-resolved reporter system**

Shugo Sasaki<sup>1, 2, 3</sup>, Michelle Y. Y. Lee<sup>2</sup>, Yuka Wakabayashi<sup>4</sup>, Luka Suzuki<sup>4</sup>, Helena Winata<sup>2</sup>,  
Miwa Himuro<sup>4</sup>, Taka-aki Matsuoka<sup>1</sup>, Ichiro Shimomura<sup>1</sup>, Hirotaka Watada<sup>4, 5</sup>,  
Francis C. Lynn<sup>1, 2, 3</sup>, Takeshi Miyatsuka<sup>4, 5</sup>

<sup>1</sup> Department of Metabolic Medicine, Osaka University Graduate School of Medicine, Osaka, Japan

<sup>2</sup> Diabetes Research Group, BC Children's Hospital Research Institute, Vancouver, BC, Canada

<sup>3</sup> Department of Surgery, School of Biomedical Engineering, University of British Columbia, Vancouver, BC, Canada

<sup>4</sup> Department of Metabolism and Endocrinology, Juntendo University Graduate School of Medicine, Tokyo, Japan

<sup>5</sup> Center for Identification of Diabetic Therapeutic Targets, Juntendo University Graduate School of Medicine, Tokyo, Japan

Corresponding authors:

Takeshi Miyatsuka and Francis C. Lynn

Department of Metabolism & Endocrinology, Juntendo University Graduate School of Medicine, 2-1-1 Hongo, Bunkyo-ku, Tokyo 113-8421, Japan

Tel: +81-3-5802-1579 / Fax: +81-3- 3813-5996

E-mail: miyatsuka-takeshi@umin.net (T.M.), francis.lynn@ubc.ca (F.C.L)

Main text (not including Abstract, Methods, References and figure legends): 3,830/5,000 words

Title: 13/15 words, Abstract: 150/150 words, Methods: 1,524/3000 words, Figures/tables: 8/10

1 **Abstract**

2           While pancreatic  $\beta$  cells have been shown to originate from endocrine progenitors in ductal  
3 regions, it remains unclear precisely where  $\beta$  cells emerge and which transcripts define newborn  $\beta$   
4 cells. Here, we used a mouse model “*Ins1-GFP;Timer*” that provides spatial information during  $\beta$ -  
5 cell neogenesis with high temporal resolution. Fluorescent imaging demonstrated that, as expected,  
6 some newborn  $\beta$  cells arise close to the ducts; unexpectedly, all the others arise away from the ducts  
7 and adjacent to blood vessels. Single-cell RNA-sequencing (scRNA-seq) demonstrated five distinct  
8 populations of newborn  $\beta$  cells, confirming the spatial heterogeneity of  $\beta$ -cell neogenesis, and  
9 integration analysis with scRNA-seq of hESC-derived  $\beta$ -like cells uncovered transcriptional  
10 similarity with the data in mouse  $\beta$  cells. Thus, the combination of time-resolved histological imaging  
11 with single-cell transcriptional mapping demonstrated novel features of spatial and transcriptional  
12 heterogeneity in  $\beta$ -cell neogenesis, which will lead to a better understanding of  $\beta$ -cell differentiation  
13 for future cell therapy.

14

15 **KEYWORDS**

16  $\beta$  cell; differentiation; neogenesis; pancreatic duct; blood vessel; Mafa; Mafb; Gcg; single-cell  
17 RNA-sequencing

## 18 **Introduction**

19       Diabetes results from an absolute or relative deficiency of insulin-producing pancreatic  $\beta$  cells,  
20 hence there is significant interest in developing safe and efficient protocols for the derivation of  
21 replacement  $\beta$  cells. Understanding  $\beta$ -cell development provides clues towards a regeneration therapy  
22 for diabetes. To date, *in vivo* mouse models have uncovered hierarchical regulation of endocrine  
23 differentiation, showing that  $\beta$  cells are differentiated from *Neurog3*-expressing endocrine  
24 progenitors that emerge from a progenitor compartment during embryogenesis<sup>1,2</sup>, and that endocrine  
25 neogenesis infrequently occurs in the adult pancreas under physiological conditions<sup>1-6</sup>. In addition,  
26 live imaging using *Insulin 1 (Ins1)-GFP* transgenic mice, which express enhanced green fluorescent  
27 protein (GFP) under the control of the *Ins1* promoter<sup>7</sup>, demonstrated that individual  $\beta$  cells migrate  
28 significant distances, forming islet-like clusters over time<sup>8</sup>. However, there is little spatial  
29 information describing precisely where  $\beta$  cells arise from their endocrine progenitors, as there is no  
30 specific marker to distinguish newly generated  $\beta$  cells separately from more differentiated  $\beta$  cells.

31       Single-cell transcriptome analysis is a powerful new method to uncover cell characteristics in  
32 single cell resolution. In the field of pancreas development and diabetes, cell type characteristics and  
33 heterogeneity have been reported in healthy human pancreas<sup>9</sup>, the pancreas with type 2 diabetes<sup>10</sup>,  
34 human fetal pancreas<sup>11</sup>, pancreatic endocrine progenitors derived from human embryonic stem cells  
35 (hESCs)<sup>12</sup>, and mouse pancreatic endocrine cells<sup>13-15</sup>. However, single cell transcriptional profiling  
36 of  $\beta$ -cell neogenesis, downstream of endocrine differentiation, has not been reported.

37       We previously generated the “*Ins1-Timer*” reporter mouse model in which newly generated  $\beta$   
38 cells can be fluorescently separated from more differentiated  $\beta$  cells using fluorescence activated cell  
39 sorting (FACS). Using this model, we subsequently determined the expression profiles of these two  
40  $\beta$ -cell populations<sup>4</sup>. However, the fluorescence intensity of newborn  $\beta$  cells in the *Ins1-Timer* mice  
41 was below microscopic detection threshold (Figure 1A and Figure S1), and therefore the  
42 spatiotemporal characteristics of  $\beta$ -cell neogenesis could not be ascertained. To circumvent this

43 problem, we developed a new mouse model, “*Ins1-GFP; Timer*”, which provides the spatial  
44 information of the newly generated  $\beta$  cells with high intensity of green fluorescence, which can be  
45 distinguished from more differentiated green/red double-fluorescent  $\beta$  cells. Spatiotemporal imaging  
46 with *Ins1-GFP; Timer* mice demonstrated that  $\beta$  cells are generated not only in close proximity to the  
47 ductal lumen but also in an unexpected region that is away from the ducts and in close apposition to  
48 blood vessels and pre-existing islets. Moreover, single-cell transcriptome analyses demonstrated five  
49 distinct populations of newborn  $\beta$  cells, and further confirmed the spatial heterogeneity of  $\beta$ -cell  
50 neogenesis with unique distributions of MAFB- and GCG-expressing cells. Notably, single-cell RNA  
51 profiles of hES-derived insulin-producing cells confirmed transcriptional heterogeneity during  $\beta$ -cell  
52 neogenesis, which overlap with the cluster expression profiles of newborn mouse  $\beta$  cells. Thus, this  
53 new reporter system provides improved spatial and transcriptional information on newly generated  $\beta$   
54 cells which, in turn, sheds light on novel microenvironments that may orchestrate endocrine  
55 differentiation and  $\beta$ -cell neogenesis.

## 56 **Results**

### 57 **A novel reporter system for deciphering spatial heterogeneity during $\beta$ -cell neogenesis**

58 We previously developed reporter mouse models that express DsRed-E5 called “Fluorescent  
59 Timer”, which shifts its fluorescence from green to red in a time-dependent manner <sup>16</sup>, allowing  
60 temporal separation of pancreatic cell lineages <sup>4,17</sup>. Among these reporter models, the “*Insulin-Timer*”  
61 mouse (*Ins1-Timer*), which expresses DsRed-E5 under the control of the *Ins1* promoter, enabled us  
62 to separate newly generated  $\beta$  cells from more differentiated  $\beta$  cells by FACS. However, only  
63 green/red double-fluorescent cells were observed by microscopy, as the green fluorescence intensity  
64 of newly synthesized DsRed-E5 is below microscopic detection threshold in *Ins1-Timer* embryos  
65 (Figure S1A). To overcome this limitation, another fluorescent protein, GFP, was simply introduced  
66 together with the DsRed-E5 reporter protein, by crossing *Ins1-Timer* mice with *Insulin 1-GFP* (*Ins1-*  
67 *GFP*) mice, in which the insulin 1 promoter drives GFP <sup>7</sup>. Because of the higher fluorescence intensity  
68 and faster maturation of *Aequorea victoria* GFP compared with DsRed-E5 <sup>16,18</sup>, we hypothesized that  
69 newly-specified  $\beta$  cells would be bright green-dominant cells, which could be observed by  
70 fluorescence microscopy, in the *Ins1-GFP; Timer* double transgenic mice (Figure 1A and Figure S1A).  
71 As we had expected, green-fluorescent cells without red fluorescence were clearly observed by  
72 microscopy in the pancreas of *Ins1-GFP; Timer* double-transgenic mice, whereas only green/red  
73 double-fluorescent cells were observed in the pancreas of *Ins1-Timer* single-transgenic mice (Figure  
74 1B).

75 To verify whether green-fluorescent cells are the newly specified  $\beta$  cells that express red-  
76 fluorescence as they mature, the pancreata of *Ins1-GFP; Timer* embryos were dissected and cultured  
77 *ex vivo* for real-time imaging of fluorescent cells. Time-lapse imaging with the pancreatic explants  
78 revealed that the green-fluorescent cells exhibited red fluorescence over time (Figure 1C). The period  
79 of time required for the transition from green-fluorescent cells to green/red double-fluorescent cells  
80 was approximately 8 hours at the longest, which nicely agrees with the red maturation time of DsRed-

81 E5. As all fluorescent cells were stained with anti-insulin antibody (Figure S1B), these findings  
82 confirm that *Ins1-GFP; Timer* reporter mice are a useful tool for detecting and tracing newly  
83 generated  $\beta$  cells using microscopy.

84

## 85 **Characterization of newborn $\beta$ cells by microscopy**

86 Microscopic quantification of fluorescent  $\beta$  cells in *Ins1-GFP; Timer* embryos demonstrated that  
87 8.5% of whole fluorescent cells exhibited green fluorescence without red fluorescence at embryonic  
88 day 14.5 (E14.5), and the proportion of green-fluorescent cells gradually decreased at later stages of  
89 development (Figure 1D). There were few green-fluorescent cells at postnatal day 14, which was  
90 consistent with previous reports demonstrating that  $\beta$ -cell neogenesis rarely occurs after birth in mice  
91 <sup>3,4,6,19</sup>. In contrast with *Ins1-Timer* single-transgenic embryos, in which greater than 40% of  $\beta$  cells  
92 were labeled as newly generated  $\beta$  cells at E16.5 <sup>4</sup>, only 7.2% of  $\beta$  cells exhibited green fluorescence  
93 in *Ins1-GFP; Timer* double-transgenic embryos at the same stage (Figure 1D), suggesting that an  
94 earlier  $\beta$  cell population is labeled by *Ins1-GFP; Timer* mice than *Ins1-Timer* mice.

95 To investigate the characteristics of green-fluorescent cells, immunostaining for  $\beta$ -cell specific  
96 transcription factors was performed. All fluorescent cells were positive for pancreatic and duodenal  
97 homeobox 1 (PDX1) and NK6 homeobox1 (NKX6.1), which are known as essential transcription  
98 factors for  $\beta$  cell development (Figures 2A and 2B) <sup>20,21</sup>. In contrast, not all fluorescent cells were  
99 positive for musculoaponeurotic fibrosarcoma oncogene family A (MAFA), a transcription factor that  
100 was reported to be expressed exclusively in mature  $\beta$  cells (Figures 2C and 2D) <sup>22-24</sup>. The percentage  
101 of MAFA-positive green-fluorescent cells was significantly lower than that of MAFA-positive  
102 green/red double-fluorescent cells between E16.5 and postnatal day 0 (Figure 2D). As MAFA was  
103 shown to function downstream of NKX6.1 <sup>25</sup>, these results suggest that green-fluorescent cells are,  
104 not only temporally early but also, functionally immature  $\beta$  cells.

105

## 106 **Spatial heterogeneity of newborn $\beta$ cells**

107 To investigate where  $\beta$  cells are generated from their progenitors during development, the  
108 pancreata of *Ins1-GFP; Timer* mice were stained with DBA lectin, a pancreatic duct marker (Figures  
109 3A–3C), so that the relative position between green-fluorescent cells and ductal cells could be  
110 determined. We confirmed previous reports demonstrating that endocrine cells arise in close  
111 apposition with the ductal lumen<sup>1,17</sup> as green-fluorescent newborn  $\beta$  cells were observed adjacent to  
112 the pancreatic ducts (Figure 3A, arrows). Unexpectedly, a significant number of green-fluorescent  
113 cells were observed away from the ducts and were attached to green/red pre-existing  $\beta$  cells (Figure  
114 3A, arrowheads). To exclude the possibility that green-fluorescent cells, which were observed away  
115 from ductal cells on a thin section (5  $\mu\text{m}$ ), are actually observed close to ductal cells on neighboring  
116 sections, 3D images of 30–50  $\mu\text{m}$ -thick pancreata were acquired, and confirmed that there was some  
117 distance between green-fluorescent  $\beta$  cells and the ductal structures (Figure 3B).

118 To further determine the positional relationship between green-fluorescent cells and other cell  
119 types, immunostaining against PECAM1 was performed to label the endothelial cells within vessels,  
120 which demonstrated that many green-fluorescent cells located away from the ducts were in close  
121 contact with blood vessels (Figures 3C and S2). It was noted that none of the newborn  $\beta$  cells were  
122 detected inside the islet-like clusters. Thus, spatiotemporal imaging with *Ins1-GFP; Timer* uncovered  
123 two different populations of newborn  $\beta$  cells: 1) newborn “ $\beta^{\text{duct}}$  cells” near the ductal regions, and 2)  
124 newborn “ $\beta^{\text{vessel}}$  cells” adjacent to both blood vessels and pre-existing islet-like clusters, suggesting  
125 two distinct locations of  $\beta$ -cell neogenesis. Quantitative analysis demonstrated that all the newborn  $\beta$   
126 cells consist of  $\beta^{\text{duct}}$  and  $\beta^{\text{vessel}}$  cells at all developmental stages after E14.5 (Figure 3D).

127

## 128 **Single-cell mRNA profiles uncovered temporal heterogeneity of newborn $\beta$ cells**

129 To further investigate transcriptional dynamics in early stage  $\beta$  cells, single-cell RNA-  
130 sequencing (scRNA-seq) was performed using E16.5 *Ins1-GFP; Timer* mouse pancreas. Newborn  $\beta$

131 cells and more differentiated  $\beta$  cells were sorted as green-fluorescent cells and green/red double-  
132 fluorescent cells, respectively (Figure 4A). Libraries of two different sets of green-fluorescent cells  
133 (G1 and G2, 455 and 253 cells, respectively) and green/red double-fluorescent cells (GR, 1348 cells)  
134 were sequenced at a depth of 191,716, 144,968 and 43,668 mean reads per cell, respectively. G1, G2  
135 and GR cells were clustered, visualized and annotated using Seurat<sup>26,27</sup> and t-distributed stochastic  
136 neighbor embedding (t-SNE)<sup>28</sup> as 12 individual clusters (Figure S3A). Out of 12 clusters, four  
137 clusters were annotated as acinar, duct, endothelial and stellate clusters (375 cells, 18.2%), based on  
138 their expression profiles (Table S1) and eight other clusters were found to robustly express *Ins1* and  
139 *Ins2* (1,681 cells, 81.8%) (Figure S3B and S3C). Notably, unbiased clustering resulted in clear  
140 difference in distribution between green-fluorescent cells and green/red double-fluorescent cells  
141 (Figure 4B). Pseudotime trajectory analysis, generated with Monocle<sup>29,30</sup>, demonstrated that green-  
142 fluorescent cells (G1 and G2) were enriched at the beginning of pseudotime as an earlier cluster on  
143 pseudo-temporal axis (Figure 4C, 4D, and S3D). In addition, single-cell expression levels of *Ins1* and  
144 *Ins2* over pseudotime clearly showed sequential increase in *Ins1* and *Ins2* mRNAs as  $\beta$  cells  
145 differentiate (Figure S3E and S3F). These findings confirmed high consistency between  
146 computational pseudotime trajectory and the temporal shift of fluorescent spectrum designed by two  
147 reporter transgenes (Figure 1A).

148 To decipher detailed transcriptional dynamics in  $\beta$ -cell neogenesis, we focussed on green-  
149 fluorescent newborn  $\beta$  cells (G1 and G2), which were sequenced more deeply than green/red double-  
150 fluorescent cells (GR). After exclusion of non-endocrine lineage cells (Figure S4A and S4B), green-  
151 fluorescent newborn  $\beta$  cells were subclustered into five groups, which were named as  $G^{Neurog3}$ ,  $G^{Sst}$ ,  
152  $G^{Gcg}$ ,  $G^{Pdx1}$ , and  $G^{Ins2}$ , based on the most differentially expressed gene in each cluster (Fig 4E, 4F, and  
153 S5A). To reconstruct an unbiased lineage trajectory for these five subpopulations during  $\beta$ -cell  
154 neogenesis, pseudotime analysis was performed and demonstrated that  $G^{Sst}$  cells were located around  
155 bifurcation from  $G^{Neurog3}$  cells into  $G^{Pdx1}$  and  $G^{Ins2}$  cells or  $G^{Gcg}$  cells (Figure 5A, 5B). Furthermore,

156 branched expression analysis modelling (BEAM) clarified transcriptional dynamics along pseudo-  
157 temporal trajectories (Figure 5C and S5B–S5D). Among top 100 differentially expressed genes in  
158 BEAM, some transcription factors, which have been shown to play pivotal roles in endocrine  
159 differentiation, were extracted. For example, *Mafb* and *Isl1* were highly expressed at the intermediate  
160 regions of  $G^{Pdx1}/G^{Ins2}$  and  $G^{Gcg}$  branch, respectively. In contrast, *Arx* and *Irx2*, which are known to be  
161 expressed in  $\alpha$  cells, showed the peak expression at the end of  $G^{Gcg}$  branch (Figure 5C, and S5B).  
162 Taken together with temporal dynamics of *Ins2* and *Gcg* mRNAs showing the relative low expression  
163 of *Ins2* and the robust increase of *Gcg* at the end of  $G^{Gcg}$  branch, these findings indicate that at least  
164 part of  $G^{Gcg}$  cells may give rise to  $\alpha$ -cell lineage.

165 We further performed gene ontology analysis and found that UPR/ER stress-related genes, such  
166 as *Calr*, *Hspa5* and *Sdf2l1*, were upregulated in  $G^{Ins2}$  cells (Figure 5D, and S5C), implying that these  
167 genes may be associated with the response to robust increase in insulin proteins as well as *Ins1* and  
168 *Ins2* mRNAs (Figure S3F). Intriguingly, genes related to fatty acid oxidation are highly expressed in  
169  $G^{Pdx1}$  cells (Figure S5E). Thus, time-resolved sorting of newborn  $\beta$  cells and subsequent single-cell  
170 transcriptome analysis enabled us to probe transcriptional dynamics during  $\beta$ -cell neogenesis with  
171 high temporal resolution.

172

### 173 **Transcriptional heterogeneity during $\beta$ -cell maturation**

174 *Ins1-GFP; Timer* mouse model allows us to label not only newborn  $\beta$  cells but also more  
175 differentiated  $\beta$  cells (Figure 4A). To investigate gene sets important for  $\beta$ -cell maturation, green/red  
176 double fluorescent cells were analysed by scRNA-seq. Unbiased annotation revealed six distinct  
177 clusters in more differentiated  $\beta$  cells (Figure 6A, 6B). The trajectory given by Monocle showed one  
178 directional development toward  $GR^{Scg2}$  cells expressing *Scg2*, a member of the chromogranin family,  
179 and carboxypeptidase E (*Cpe*), which is a key enzyme for proinsulin processing (Figure 6C, 6D)<sup>31</sup>.  
180 Notably, the expression profiles of  $GR^{SpC25}$  cluster shared key characteristics of proliferating cells and

181 lacked markers of cells in G1 stage following cell cycle analysis (Figure 6B, 6E). The pseudotime  
182 analysis showed a unique branch consisting of  $GR^{Sp25}$  cells adjacent to  $GR^{Npy}$  cluster, which is located  
183 prior to most mature  $GR^{Scg2}$  cluster (Figure 6C). In contrast, unbiased annotation with green-  
184 fluorescent newborn  $\beta$  cells did not annotate proliferating insulin-positive population (Figure 4E, 4F,  
185 5A, S4A and S4B). These findings suggest that  $\beta$ -cell proliferation is tightly regulated during a  
186 specific time window during  $\beta$ -cell differentiation, which is consistent with our previous finding with  
187 *Ins1-Timer* embryos <sup>4</sup>.

188

### 189 **Orchestration between spatial and transcriptional heterogeneity during $\beta$ -cell neogenesis**

190 Histological imaging in *Ins1-GFP; Timer* embryos resulted in two distinct populations (i.e.  $\beta^{\text{duct}}$   
191 and  $\beta^{\text{vessel}}$  cells) during  $\beta$ -cell neogenesis, whereas unbiased clustering with single-cell mRNA profiles  
192 demonstrated five clusters ( $G^{Neurog3}$ ,  $G^{Sst}$ ,  $G^{Gcg}$ ,  $G^{Pdx1}$ , and  $G^{Ins2}$ ). To investigate whether  
193 transcriptional heterogeneity was indeed observed between  $\beta^{\text{duct}}$  and  $\beta^{\text{vessel}}$  cells, immunostaining was  
194 performed with antibodies against NEUROG3, SST, GCG, and MAFB, mRNAs of which were highly  
195 expressed in  $G^{Neurog3}$ ,  $G^{Sst}$ ,  $G^{Gcg}$ , and  $G^{Pdx1}$ , respectively (Figure 4F). NEUROG3 and SST proteins  
196 did not overlap with either green or red fluorescent cells (data not shown). Immunostaining against  
197 GCG demonstrated that 56% and 28% of newborn  $\beta^{\text{duct}}$  cells in *Ins1-GFP; Timer* embryos were  
198 positive for GCG at E14.5 and E16.5, respectively (Figure 7A and 7B), whereas none of  $\beta^{\text{vessel}}$  cells  
199 were positive for GCG at either stage, suggesting that  $G^{Gcg}$  cells comprise part of  $\beta^{\text{duct}}$  cells.  
200 Furthermore, immunostaining against MAFB demonstrated that more than 90% of  $\beta^{\text{duct}}$  cells were  
201 positive for MAFB, while only 20% of  $\beta^{\text{vessel}}$  cells were labeled (Figure 7C and 7D). These findings  
202 suggest that most of  $G^{Pdx1}$  cells, which highly express *Maifb*, belong to  $\beta^{\text{duct}}$  cells, while only part of  
203  $G^{Pdx1}$  cells compose  $\beta^{\text{vessel}}$  cells. Thus, since the expression pattern of MAFB uncovered unique spatial  
204 heterogeneity of newborn  $\beta$  cells, we next examined the immunostaining for MAFA, another member  
205 of MAF family transcription factor, which was detected in part of green-fluorescent cells (Figure 2C

206 and 2D), although *Mafa* transcripts were not sequenced in our scRNA-seq data. Intriguingly, the  
207 expression patterns of MAFA were opposite to those of MAFB (Figure 7E, 7F); more than 80% of  
208  $\beta^{\text{vessel}}$  cells were positive for MAFA, significantly higher than that of  $\beta^{\text{duct}}$  cells, in which less than  
209 30% of them expressed MAFA. The expression patterns of these three key proteins imply that  $\beta^{\text{vessel}}$   
210 cells are functionally more mature than  $\beta^{\text{duct}}$  cells despite having an identical time window after the  
211 *Ins1* promoter is activated. Considering that MAFA is thought to function as  $\beta$  cells mature, and the  
212 pseudotime trajectory showed that  $G^{\text{Ins2}}$  cells are more differentiated than  $G^{\text{Pdx1}}$  cells, these findings  
213 suggest that  $G^{\text{Ins2}}$  cells may comprise MAFA-expressing  $\beta^{\text{vessel}}$  cells. Thus, the histological analysis  
214 of newborn  $\beta$  cells accompanied by single-cells mRNA profiles demonstrated a unique orchestration  
215 between spatial and transcriptional heterogeneity during  $\beta$ -cell neogenesis.

216

## 217 **Human embryonic stem cell-derived $\beta$ -like cells share mouse newborn $\beta$ -cell features**

218 To date, *in vitro* differentiation protocols of human embryonic stem cells (hESCs) toward  $\beta$ -like  
219 cells have been developed, although the differentiated cells induced by most of the protocols are a  
220 mixed population of ~50% insulin-producing cells and ~50% of other cell types<sup>32–34</sup>. Additionally,  
221 these  $\beta$ -like cells are still immature because they need *in vivo* transplantation to give rise to mature  $\beta$   
222 cells that have the same glucose-responsive insulin secretion capacity as that of *bona fide*  $\beta$  cells. To  
223 investigate similarity between hESC-derived immature  $\beta$ -like cells and mouse newborn  $\beta$  cells,  
224 transcriptional profiles were compared using scRNA-seq data. Cyt49 hESCs were differentiated for  
225 27 days *in vitro* by a protocol reported by Nair et al.<sup>32</sup>, and 5,644 cells were analyzed after sequencing  
226 and filtering. After running Seurat, the clusters which contained 1,107 *INS*- highly expressing cells  
227 (20% of total cells) were selected and re-clustered to 7 clusters visualized on t-SNE (Figure 8A–8C).  
228 *MGST1* and *SYT13* are enriched in  $h\beta^{\text{MGST1}}$  and  $h\beta^{\text{SYT13}}$   $\beta$ -like populations, respectively, consistent to  
229 that mouse *Mgst1* and *Syt13* have the peak expression in very early  $\beta$  cells (Figure S5D). Furthermore,  
230 a trajectory generated by Monocle showed that *SST*-expressing  $h\beta^{\text{SST}}$  cluster was located prior to

231  $h\beta^{NEFM}$ ,  $h\beta^{IAPP}$ ,  $h\beta^{GAST}$ , and  $h\beta^{GCG}$  clusters (Figure 8D, 8E), as *Sst*-expressing  $G^{Sst}$  cluster was located  
232 at an earlier stage than  $G^{Pdx1}$ ,  $G^{Ins2}$ , and  $G^{Gcg}$  clusters (Figure 5A).

233 To further investigate transcriptional relevance between mouse  $\beta$  cells and hESC-derived  $\beta$ -like  
234 cells, the integration analysis was performed with data sets from mouse  $\beta$  cells from *Ins1-GFP; Timer*  
235 embryos and hESC-derived  $\beta$ -like cells (Figure S6A–S6C). Although the integrated t-SNE clustering  
236 showed few overlaps between mouse and human cells, *Neurog3*-, *Gcg*- and *Sst*-expressing  
237 populations in mouse embryos appeared in distinct areas that were overlapped with human cells by  
238 more than 15% distribution in each cluster (Figure S6D, S6E), suggesting that only  $\beta^{Neurog3}$ ,  $\beta^{Gcg}$  and  
239  $\beta^{Sst}$  share transcriptional profiles with *NEUROG3*-, *GCG*- and *SST*-expressing human cells. Thus,  
240 comparative analysis of hESC-derived  $\beta$ -like cells with mouse newborn  $\beta$  cells indicates potential  
241 shared transcriptional dynamics during early  $\beta$ -cell specification that remain to be further elucidated.

## 242 **Discussion**

243 Advances in imaging technology using fluorescent proteins both *in vivo* and *ex vivo* have enabled  
244 us to capture cellular dynamics that occur during pancreas development and  $\beta$ -cell differentiation<sup>7,35</sup>.  
245 While reporter mouse models that express fluorescent proteins have helped to trace the differentiation  
246 of specific cell types in endocrine lineages, it remains to be established as to how the newly specified  
247  $\beta$  cells can be detected by microscopy separately from more differentiated  $\beta$  cells, due to the lack of  
248 a specific marker for early  $\beta$ -cell neogenesis. In the present study, we demonstrated that a combination  
249 of two fluorescent proteins under the same *Ins1* promoter provides a discrete time window to address  
250 this difficulty, and we successfully demonstrated the two distinct pancreatic regions from which  $\beta$   
251 cells arise.

252 Previously a Cre/loxP-mediated reporter system combined with microscopy was used to show  
253 no evidence of  $\beta$ -cell neogenesis in adult pancreata<sup>19</sup>, which agrees with our findings with *Ins1-Timer*  
254<sup>4</sup> and *Ins1-GFP; Timer* mice (Figures 1D). On the other hand, at E14.5, Cre/loxP-mediated reporter  
255 mice labeled all the insulin-producing cells as a single population, whereas *Ins1-GFP; Timer* embryos  
256 clearly demonstrated temporal heterogeneity of  $\beta$ -cell differentiation; that is, around 8.5% of insulin-  
257 producing cells were labeled as green-fluorescent, newly differentiated  $\beta$  cells, which were clearly  
258 distinguished from green/red double-fluorescent, more differentiated  $\beta$  cells (Figures 1D). This  
259 difference is probably because the temporal resolution in the Cre/loxP reporter mice depended on the  
260 degradation of red fluorescent protein, which has a longer half-life, whereas our mouse model takes  
261 advantage of the shorter time window caused by the higher intensity and faster maturation of GFP  
262 than DsRed-E5 (Figures 1A and S1A). Thus, the short time window in *Ins1-GFP; Timer* embryos  
263 could provide novel insights into spatio-temporal heterogeneity during  $\beta$ -cell neogenesis.

264 Fluorescence imaging with *Ins1-GFP; Timer* embryos demonstrated the existence of two types  
265 of newborn  $\beta$  cells: “ $\beta^{\text{duct}}$  cells” near the pancreatic ducts and “ $\beta^{\text{vessel}}$  cells” near the blood vessels  
266 (Figures 3A–3C). As *Neurog3*-expressing endocrine progenitors emerge from the ductal region and

267 give rise to hormone-expressing endocrine cells <sup>1,2</sup>, we were not surprised to see *Neurog3*-expressing  
268 cells differentiate into  $\beta^{\text{duct}}$  cells near the ductal compartment without further migration. In contrast,  
269 the existence of  $\beta^{\text{vessel}}$  cells and the lack of green-fluorescent cells in every other region between  
270 pancreatic ducts and blood vessels supports the possibility that endocrine progenitors migrate away  
271 from the ductal area toward the region near the blood vessels prior to differentiation (Figure S7). On  
272 the other hand, it has been reported that *Neurog3*-expressing cells were observed in epithelial tip  
273 domains as well as in trunk domains <sup>36</sup>, implying that  $\beta^{\text{vessel}}$  cells might emerge from *Neurog3*-  
274 expressing progenitors in the tip domains. While several mouse models have been developed to study  
275 the regulatory mechanisms of islet cell migration <sup>37,38</sup>, it remains unclear when and how  $\beta$  cells are  
276 specified during migration. It would be of great interest to trace the emergence of  $\beta^{\text{duct}}$  and  $\beta^{\text{vessel}}$  cells  
277 under the circumstances where islet cell migration is disrupted in these mouse models.

278 Blood vessels and their surrounding environments have been demonstrated to play essential  
279 roles in various stages of pancreas development. For example, aortic endothelial cells are required  
280 for the maintenance of Pdx1 expression and for pancreas budding <sup>39</sup>, and ectopic vascularization leads  
281 to ectopic insulin expression <sup>40</sup>. The presence of  $\beta^{\text{vessel}}$  cells suggests that vascular endothelial cells  
282 may endow endocrine progenitors with a  $\beta$ -cell destiny. Further studies would be needed to address  
283 the microenvironment around  $\beta^{\text{duct}}$  and  $\beta^{\text{vessel}}$  cells, which may lead to the discovery of new substances  
284 critical for  $\beta$ -cell specification.

285 We performed single-cell transcriptome analysis with *Ins1-GFP; Timer* embryos and found that  
286 unbiased clustering resulted in five unique populations among green-fluorescent newborn  $\beta$  cells.  
287 Since *Mafb* mRNAs were differentially expressed in  $G^{\text{Pdx1}}$  cells (Figure 4F) and MAFB proteins were  
288 detected in a larger number of  $\beta^{\text{duct}}$  cells than in  $\beta^{\text{vessel}}$  cells, it is possible that *Mafb*-expressing  $G^{\text{Pdx1}}$   
289 cells may compose MAFB-expressing  $\beta^{\text{duct}}$  cells. However, there are some difficulties to validate the  
290 correlations between five populations annotated by scRNA-seq and two populations observed on  
291 microscopy. For example, neither SST nor NEUROG3 overlapped with fluorescent proteins probably

292 because the detection threshold of immunostaining was not sufficiently high. On the other hand, some  
293 transcripts like *Mafa* were not sequenced in scRNA-seq, although MAFA proteins were clearly  
294 detected (Figure 2C and 7E). Technical improvement in scRNA-seq methodologies, including  
295 improved sequencing depths, would advance our understanding concerning transcriptional  
296 heterogeneity during  $\beta$ -cell neogenesis.

297         Single-cell transcriptome analysis in hESC-derived  $\beta$ -like cells identified *SST*-expressing  $h\beta^{SST}$   
298 cells as an earlier cluster (Figure 8A–8E), showing similarity with mouse *Sst*-expressing  $G^{Sst}$  cluster  
299 in newborn  $\beta$  cells which was located at an earlier stage than some other clusters. In addition, the  
300 integration analysis between *Ins1-GFP; Timer* embryos and hESC-derived  $\beta$ -like cells confirmed  
301 shared transcriptional profiles among *Sst/SST*-expressing subsets (Figure S6D, S6E). These findings  
302 suggest that *Sst*-expressing cells are in an intermediate state during  $\beta$ -cell maturation in both mouse  
303 and human, which is consistent with recent findings that hESC-derived  $\beta$ -like cells co-expressed *SST*  
304 transcripts during differentiation <sup>12</sup>.

305         In summary, this time-resolved reporter system in mice revealed spatial and transcriptional  
306 heterogeneity during  $\beta$ -cell neogenesis, showing similarity and difference between mouse  $\beta$  cells and  
307 hESC-derived  $\beta$ -like cells. Although these novel findings lead to new additional questions that should  
308 be addressed, further time-resolved comprehensive analyses at the single-cell level will provide us  
309 with novel insights, which could be applied to regeneration therapy for diabetes.

310 **Methods**

311 **Animals**

312 *Ins1-Timer* and *Ins1-GFP* reporter mice were generated as previously described <sup>4,7</sup>. *Ins1-*  
313 *GFP; Timer* double-transgenic mice were obtained by crossing *Ins1-Timer* mice with *Ins1-GFP* mice.  
314 The mice were genotyped by PCR, using the forward primer 5'-AGTTCCAGTACGGCTCCAAG-3'  
315 and the reverse primer 5'-CAGCCCATGGTCTTCTTCTG-3' in the coding region of DsRed-E5 for  
316 *Ins1-Timer* and the forward primer 5'-GAACTTCAAGATCCGCCACA-3' and the reverse primer 5'-  
317 ACTCCAGCAGGACCATGTGA-3' in the coding region of GFP for *Ins1-GFP*, respectively. Timed  
318 mating was carried out with E0.5 being set as midday of the day of discovery of a vaginal plug. Mice  
319 were housed on a 12-h light-dark cycle in a controlled climate. All animal procedures were approved  
320 by the Ethics Review Committee for Animal Experimentation of Osaka University Graduate School  
321 of Medicine, and Juntendo University Graduate School of Medicine.

322

323 **Histological analyses**

324 Tissues were fixed in 4% paraformaldehyde in phosphate buffered saline (PBS) at 4 °C, washed  
325 in PBS, immersed in sucrose solution, and embedded in Tissue-Tek OCT Compound (Sakura, Japan).  
326 Frozen blocks were sectioned at 5- $\mu$ m thickness for regular observation and at 30- $\mu$ m thickness for  
327 3D analysis. Sections were blocked with 5% horse serum, incubated with primary antibodies  
328 overnight at 4 °C, and then visualized by incubation with secondary antibodies for 1 hr at room  
329 temperature. The primary antibodies used were the following: chicken anti-insulin (1:1,000; Abcam,  
330 Cambridge, MA, USA), rabbit anti-glucagon (1:1,000; Dako, Carpinteria, CA, USA), guinea pig anti-  
331 PDX1 (1:500; a kind gift from Dr. M. German) rabbit anti-NKX6.1 (1:200; Sigma-Aldrich, St Louis,  
332 MO, USA), rabbit anti-MAFA (1:500; Bethyl Laboratories, Montgomery, TX, USA), rat anti-  
333 PECAM1 (1:250; BD Biosciences, San Jose, CA, USA) and biotinylated lectin *Dolichos biflorus*  
334 agglutinin (DBA; 1:200; Vector Laboratories, Burlingame, CA, USA). The secondary antibodies used

335 were Alexa Fluor 647-conjugated anti-rabbit IgG, Alexa Fluor 647-conjugated anti-guinea pig IgG,  
336 Alexa Fluor 647-conjugated anti-rat IgG, Alexa Fluor 647 and 350-conjugated streptavidin, Alexa  
337 Fluor 350-conjugated anti-rabbit IgG, and DyLight 350-conjugated anti-chicken IgY (all at 1:200;  
338 Invitrogen, Carlsbad, CA, USA). After washing in PBS, sections were mounted in Vectashield  
339 mounting medium (Vector Laboratories). Images were acquired using an Olympus FV1000D  
340 confocal laser scanning microscope and a Keyence BZ-9000 all-in-one fluorescence microscope. For  
341 morphometric analysis, the number of fluorescent or immunostained cells was quantified (more than  
342 three sections for each mouse) with Image J software (NIH, Bethesda, MD, USA).

343

#### 344 ***Ex vivo* real-time imaging of the developing pancreas**

345 Developing pancreata were dissected from *Ins1-GFP; Timer* embryos at E14.5, and were stored  
346 in sterile PBS on ice until all dissections were completed. For growth, organ explants were transferred  
347 to coverglass-bottomed dishes (AGC Asahi Glass, Tokyo, Japan) coated with 30  $\mu$ L Matrigel™ (BD  
348 Biosciences, San Jose, CA, USA), and then incubated with growth medium (DMEM/F-12 with 10%  
349 FBS, antibiotics, and insulin–transferrin–selenium). Live imaging of the explants was carried out  
350 using an Olympus FV1000D confocal laser-scanning microscope, with an automated motorized stage,  
351 in a humidified heated chamber with CO<sub>2</sub>. For acquisition, digital images of 12 sections per z-stack  
352 were captured through multiple z-steps (5- $\mu$ m apart) every 30 min. The total time of imaging was 10  
353 hr. After imaging, z-stacks were collapsed into maximum intensity projections, and all time points  
354 were compiled and exported as a QuickTime (avi) file using Image J software (NIH, Bethesda, MD,  
355 USA).

356

#### 357 **Sorting cells for single-cell RNA-sequencing**

358 *Ins1-GFP; Timer* embryos were collected on E16.5 and dissected on ice. To generate single cells,  
359 embryonic pancreases were incubated in 2 mL of pre-warmed 37 °C 0.25% Trypsin with mild

360 agitation for 5 minutes. To stop digestion, 1 mL of cold FBS and 2 mL of cold PBS were added and  
361 mixed by inversion to stop digestion, followed by pipette filtering with a 40 mm nylon filter. Cells  
362 were then centrifuged for 5 minutes at 200g. After aspirating the supernatant, cells were resuspended  
363 in cold 2% FBS in PBS, placed on ice, and immediately sorted into green-fluorescent, and green/red  
364 double-fluorescent fractions using a Beckman Coulter MoFlo Astrios (Mississauga, ON, Canada).

365

### 366 **Preparation of human $\beta$ -like cells for single-cell RNA-sequencing**

367 CyT49 parental hESC line was obtained from ViaCyte, Inc. (San Diego, CA, USA). Studies with  
368 hESC were approved by the BC Children's and Women's Hospital Human Research Ethics Board  
369 (Approval #H09-00676). Undifferentiated cells were maintained as previously described<sup>15</sup>. CyT49  
370 hESCs were differentiated using the protocol published by Nair et al.<sup>32</sup> Briefly, dissociated  
371 undifferentiated CyT49 hESCs were plated on six-well plates at a density of 5.5 million cells in 5.5  
372 mL media per well. The plates were incubated at 37 °C and 5% CO<sub>2</sub> on an orbital shaker at 100 rpm  
373 to induce spheroid formation. After 24 h, seven-step differentiation was induced. Stage 7 cells were  
374 differentiated longer than the original protocol, and differentiation was stopped at day 27. ~100  
375 spheroids were washed once with PBS before adding 500  $\mu$ L of Accutase. Following 10 min at 37°C  
376 for dissociation, 500  $\mu$ L of 2% BSA CMRL medium was added, centrifuged for 5 min at 200  $\times$  g,  
377 washed once with PBS, and resuspended in 350  $\mu$ L of ice-cold PBS.

378

### 379 **Library preparations and single-cell RNA-sequencing**

380 scRNA-seq libraries were generated with 10 $\times$  Genomics Chromium<sup>TM</sup> pipeline (Pleasanton, CA,  
381 USA) as previously described<sup>15</sup>. The Single Cell 3' Reagent Kits v2 were used to generate single cell  
382 libraries. Briefly, cells were counted following FACS and cell suspensions were loaded for a targeted  
383 cell recovery of 1000-5000 cells per channel. The microfluidics platform was used to barcode single  
384 cells using Gel Bead-In-Emulsions (GEMs). RT is performed within GEMs, resulting in barcoded

385 cDNA from single cells. The full length, barcoded cDNA is PCR amplified followed by enzymatic  
386 fragmentation and SPRI double sided size selection for optimal cDNA size. End repair, A-tailing,  
387 Adaptor Ligation, and PCR are performed to generate the final libraries that have P5 and P7 primers  
388 compatible with Illumina sequencing. The libraries were pooled and sequenced using an Illumina  
389 NextSeq500 platform with a 150 cycle High Output v2 kit in paired-end format with 26 bp Read 1, 8  
390 bp I5 Index, and 85 bp Read 2.

391

## 392 **Data Analysis**

393 Following sequencing, data were analysed using publicly available software programs and R  
394 pipelines. First, cellranger mkfastq (10x Genomics) generates FASTQ files from the raw sequencing  
395 data. Next, cellranger count uses the FASTQ file to perform sequence alignment (mouse: GRCm38,  
396 human: GRCh38), filter sequences based on quality score, and generate single cell gene counts. For  
397 count alignment of the lineage-tracing transgenes, the mouse reference genome was annotated to  
398 include the GFP and Timer sequences. Cellranger aggr was used to combine data from multiple  
399 samples (G1, G2 and GR) and ensure all libraries have the same sequencing depth.

400 As minimal filtering is performed in cellranger, two additional R pipelines were used to filter  
401 out cells that did not meet the quality control standard. The first pipeline is Scater  
402 (<https://bioconductor.org/packages/release/bioc/html/scater.html>) and is a single cell analysis pipeline  
403 that places a great emphasis on quality control <sup>41</sup>. For this analysis, cells were discarded based on  
404 counts (transcripts/cell), genes (genes/cell) and percentage mitochondrial DNA expression, greater  
405 than 3 standard deviation away from the mean. Low-abundance genes, which were expressed by less  
406 than 3 cells, were removed. This QC dataset was then analyzed using the Seurat V2.0 and V3.0  
407 pipeline (<http://satijalab.org/seurat/>), another R toolkit for single cell genomics <sup>26,27</sup>. Seurat posed an  
408 additional filtering step that removed out cells with less than 500 genes. Seurat was then used for  
409 normalization, correcting cells by their library size, and linear dimensional reduction, removing

410 uninteresting source of variations. Finally, unsupervised shared nearest neighbor modularity  
411 optimization based clustering was performed using Seurat to group cells based on gene expression  
412 and to identify unique cell types within the populations. For cell cycle analysis, cell-cycle scoring  
413 and phase determination pipelines provided with Seurat was assigned to the GR library. Cross-species  
414 integration analysis was conducted using Seurat V3.0.

415

### 416 **Pseudotime analysis**

417 Pseudotime analysis was performed using Monocle v2.6.1 ([http://cole-trapnell-lab.github.io/](http://cole-trapnell-lab.github.io/monocle-release/docs/#constructing-single-cell-trajectories)  
418 [monocle-release/docs/#constructing-single-cell-trajectories](http://cole-trapnell-lab.github.io/monocle-release/docs/#constructing-single-cell-trajectories)). Quality controlled transcript data by  
419 Scater and Seurat were loaded into Monocle as a CellDataSet object. Same filtering threshold was  
420 used as Seurat, in while cells with less than 500 genes, and genes expressed by less than 3 cells were  
421 removed. Variable expressed genes were defined as a gene that was expressed in >50 cells.  
422 Unsupervised clustering was performed using genes that have a mean expression of > 0.1 and  
423 dimensional reduction was done using t-SNE. Next, differential gene expression analysis was done  
424 between clusters of interest and the top 1000 variable genes with  $qval < 0.01$  were used to learn a tree-  
425 like trajectory using DDRTree method and lastly order cells in the pseudotime. Branched expression  
426 analysis modelling (BEAM) provided with Monocle was conducted to determine gene expression  
427 patterns over pseudotime.

428

### 429 **Statistical analyses**

430 Measurements were performed on discrete samples unless otherwise stated. Statistical analyses  
431 were performed using the GraphPad Prism 8.0 and R software. Comparisons of two samples were  
432 performed by the unpaired two-tailed *t*-tests. Multiple groups were analysed by one-way ANOVA  
433 with a multiple comparison test and the Tukey-Kramer's post-hoc test was used to compare different

434 groups. A value of  $p < 0.05$  was considered to indicate a statistically significant difference between  
435 two groups. Data are presented as the mean  $\pm$  SEM.

436 **References**

- 437 1. Schwitzgebel, V. M. *et al.* Expression of neurogenin3 reveals an islet cell precursor population in  
438 the pancreas. *Development* **127**, 3533–42 (2000).
- 439 2. Gu, G., Dubauskaite, J. & Melton, D. A. Direct evidence for the pancreatic lineage: NGN3+ cells  
440 are islet progenitors and are distinct from duct progenitors. *Development* **129**, 2447–57 (2002).
- 441 3. Dor, Y., Brown, J., Martinez, O. I. & Melton, D. A. Adult pancreatic beta-cells are formed by self-  
442 duplication rather than stem-cell differentiation. *Nature* **429**, 41–6 (2004).
- 443 4. Miyatsuka, T. *et al.* Chronological analysis with fluorescent timer reveals unique features of  
444 newly generated  $\beta$ -cells. *Diabetes* **63**, 3388–93 (2014).
- 445 5. Solar, M. *et al.* Pancreatic exocrine duct cells give rise to insulin-producing beta cells during  
446 embryogenesis but not after birth. *Dev Cell* **17**, 849–60 (2009).
- 447 6. Teta, M., Rankin, M. M., Long, S. Y., Stein, G. M. & Kushner, J. A. Growth and regeneration of  
448 adult beta cells does not involve specialized progenitors. *Dev Cell* **12**, 817–26 (2007).
- 449 7. Hara, M. *et al.* Transgenic mice with green fluorescent protein-labeled pancreatic beta -cells. *Am*  
450 *J Physiol Endocrinol Metab* **284**, E177-83 (2003).
- 451 8. Puri, S. & Hebrok, M. Dynamics of embryonic pancreas development using real-time imaging.  
452 *Dev Biol* **306**, 82–93 (2007).
- 453 9. Enge, M. *et al.* Single-Cell Analysis of Human Pancreas Reveals Transcriptional Signatures of  
454 Aging and Somatic Mutation Patterns. *Cell* **171**, 321-330 e14 (2017).

- 455 10. Segerstolpe, A. *et al.* Single-Cell Transcriptome Profiling of Human Pancreatic Islets in Health  
456 and Type 2 Diabetes. *Cell Metab* **24**, 593–607 (2016).
- 457 11. Ramond, C. *et al.* Understanding human fetal pancreas development using subpopulation sorting,  
458 RNA sequencing and single-cell profiling. *Development* **145**, (2018).
- 459 12. Petersen, M. B. K. *et al.* Single-Cell Gene Expression Analysis of a Human ESC Model of  
460 Pancreatic Endocrine Development Reveals Different Paths to beta-Cell Differentiation. *Stem*  
461 *Cell Rep.* **9**, 1246–1261 (2017).
- 462 13. Qiu, W. L. *et al.* Deciphering Pancreatic Islet beta Cell and alpha Cell Maturation Pathways and  
463 Characteristic Features at the Single-Cell Level. *Cell Metab* **25**, 1194-1205 e4 (2017).
- 464 14. Zeng, C. *et al.* Pseudotemporal Ordering of Single Cells Reveals Metabolic Control of Postnatal  
465 beta Cell Proliferation. *Cell Metab* **25**, 1160-1175 e11 (2017).
- 466 15. Krentz, N. A. J. *et al.* Single-Cell Transcriptome Profiling of Mouse and hESC-Derived  
467 Pancreatic Progenitors. *Stem Cell Rep.* **11**, 1551–1564 (2018).
- 468 16. Terskikh, A. *et al.* ‘Fluorescent timer’: protein that changes color with time. *Science* **290**, 1585–  
469 8 (2000).
- 470 17. Miyatsuka, T., Li, Z. & German, M. S. Chronology of islet differentiation revealed by temporal  
471 cell labeling. *Diabetes* **58**, 1863–8 (2009).
- 472 18. Cormack, B. P., Valdivia, R. H. & Falkow, S. FACS-optimized mutants of the green fluorescent  
473 protein (GFP). *Gene* **173**, 33–8 (1996).

- 474 19. Xiao, X. *et al.* No evidence for beta cell neogenesis in murine adult pancreas. *J Clin Invest* **123**,  
475 2207–17 (2013).
- 476 20. Miyatsuka, T., Matsuoka, T. A. & Kaneto, H. Transcription factors as therapeutic targets for  
477 diabetes. *Expert Opin Ther Targets* **12**, 1431–42 (2008).
- 478 21. Miyatsuka, T. Uncovering the mechanisms of beta-cell neogenesis and maturation toward  
479 development of a regenerative therapy for diabetes. *Diabetol. Int.* **6**, 261–267 (2015).
- 480 22. Kataoka, K. *et al.* MafA is a glucose-regulated and pancreatic beta-cell-specific transcriptional  
481 activator for the insulin gene. *J Biol Chem* **277**, 49903–10 (2002).
- 482 23. Matsuoka, T. A. *et al.* Members of the large Maf transcription family regulate insulin gene  
483 transcription in islet beta cells. *Mol Cell Biol* **23**, 6049–62 (2003).
- 484 24. Olbrot, M., Rud, J., Moss, L. G. & Sharma, A. Identification of beta-cell-specific insulin gene  
485 transcription factor RIPE3b1 as mammalian MafA. *Proc Natl Acad Sci U A* **99**, 6737–42 (2002).
- 486 25. Matsuoka, T. A. *et al.* The MafA transcription factor appears to be responsible for tissue-specific  
487 expression of insulin. *Proc Natl Acad Sci U A* **101**, 2930–3 (2004).
- 488 26. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell  
489 transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**,  
490 411–420 (2018).
- 491 27. Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888-1902.e21 (2019).

- 492 28. van der Maaten, L. & Hinton, G. Visualizing data using t-SNE. *J Mach Learn Res* **9**, 2579–2605  
493 (2008).
- 494 29. Qiu, X. *et al.* Single-cell mRNA quantification and differential analysis with Census. *Nat.*  
495 *Methods* **14**, 309–315 (2017).
- 496 30. Trapnell, C. *et al.* The dynamics and regulators of cell fate decisions are revealed by  
497 pseudotemporal ordering of single cells. *Nat. Biotechnol.* **32**, 381–386 (2014).
- 498 31. Chen, Y.-C., Taylor, A. J. & Verchere, C. B. Islet prohormone processing in health and disease.  
499 *Diabetes Obes. Metab.* **20 Suppl 2**, 64–76 (2018).
- 500 32. Nair, G. G. *et al.* Recapitulating endocrine cell clustering in culture promotes maturation of  
501 human stem-cell-derived  $\beta$  cells. *Nat. Cell Biol.* **21**, 263–274 (2019).
- 502 33. Rezanian, A. *et al.* Reversal of diabetes with insulin-producing cells derived in vitro from human  
503 pluripotent stem cells. *Nat Biotechnol* **32**, 1121–33 (2014).
- 504 34. Pagliuca, F. W. *et al.* Generation of functional human pancreatic beta cells in vitro. *Cell* **159**, 428–  
505 39 (2014).
- 506 35. Gu, G. *et al.* Global expression analysis of gene regulatory pathways during endocrine pancreatic  
507 development. *Development* **131**, 165–79 (2004).
- 508 36. Scavuzzo, M. A. *et al.* Endocrine lineage biases arise in temporally distinct endocrine progenitors  
509 during pancreatic morphogenesis. *Nat. Commun.* **9**, 3356 (2018).

- 510 37. Miettinen, P. J. *et al.* Impaired migration and delayed differentiation of pancreatic islet cells in  
511 mice lacking EGF-receptors. *Dev. Camb. Engl.* **127**, 2617–2627 (2000).
- 512 38. Greiner, T. U., Kesavan, G., Stahlberg, A. & Semb, H. Rac1 regulates pancreatic islet  
513 morphogenesis. *BMC Dev Biol* **9**, 2 (2009).
- 514 39. Yoshitomi, H. & Zaret, K. S. Endothelial cell interactions initiate dorsal pancreas development  
515 by selectively inducing the transcription factor Ptf1a. *Development* **131**, 807–17 (2004).
- 516 40. Lammert, E., Cleaver, O. & Melton, D. Induction of pancreatic differentiation by signals from  
517 blood vessels. *Science* **294**, 564–7 (2001).
- 518 41. McCarthy, D. J., Campbell, K. R., Lun, A. T. L. & Wills, Q. F. Scater: pre-processing, quality  
519 control, normalization and visualization of single-cell RNA-seq data in R. *Bioinforma. Oxf. Engl.*  
520 **33**, 1179–1186 (2017).

521 **Acknowledgements**

522 We thank H. Kim (Korea Advanced Institute of Science and Technology, Daejeon, South Korea)  
523 and M.S. German (University of California San Francisco, CA, USA) for their helpful advice and  
524 criticism, and S. Takebe, H. Hibino and S. Ishikawa for their assistance with the experiments. We also  
525 acknowledge the support of the Mouse Facilities at Osaka University and Juntendo University. This  
526 work was supported by JSPS KAKENHI (No. 16K09766, 19H04060), the Takeda Science  
527 Foundation, the Suzuken Memorial Foundation, the Astellas Foundation for Research on Metabolic  
528 Disorders, and Eli Lilly Research Foundation (all to T.Mi.). Fellowship support was provided by the  
529 Manpei Suzuki Diabetes Foundation, the Michael Smith Foundation for Health Research, and the  
530 JDRF Postdoctoral Fellowship award (to S.S.).

531

532 **Author contributions**

533 T.Mi. and S.S. designed the whole project and wrote the manuscript. S.S., M.L., Y. W., L.S.,  
534 M.H., and H.Wi. contributed to the acquisition and analysis of the data. T.Ma., I.S., H.Wa., and F.C.L.  
535 contributed to the interpretation of the data. All authors revised the manuscript critically for important  
536 intellectual content and approved the final version of the manuscript.

537

538 **Competing interests**

539 The authors declare no competing interests.

540 **Figure legends**

541 **Figure 1. Improved “Timer” fluorescent system for labeling newly generated  $\beta$  cells**

542 (A) Schematic diagram of the concept for improved temporal labeling in *Ins1-GFP; Timer* (DsRed-  
543 E5) double-transgenic mice. Whereas green fluorescence in newly generated  $\beta$  cells of *Ins1-Timer*  
544 embryos is below the microscopic detection threshold, green fluorescence from *Ins1-GFP* mice is  
545 high enough to be detected by microscopy in newly generated  $\beta$  cells as well as in mature  $\beta$  cells.  
546 Therefore, *Ins1-GFP; Timer* double-transgenic mice enable the labeling of newly generated  $\beta$  cells  
547 as green-fluorescent cells by microscopy, which can be distinguished from mature  $\beta$  cells with green  
548 and red fluorescence. (B) Fluorescence images of the pancreas of *Ins1-Timer* and *Ins1-GFP; Timer*  
549 embryos at E16.5. Whereas all green-fluorescent cells exhibited red fluorescence in *Ins1-Timer*  
550 embryos, green fluorescent cells without red fluorescence (arrows) were observed in *Ins1-GFP;*  
551 *Timer* double-transgenic embryos. Scale bar, 50  $\mu$ m. (C) Time-lapse imaging of fluorescent cells  
552 within an E14.5 pancreatic bud of *Ins1-GFP; Timer* embryos. Enlarged images of a green-fluorescent  
553 cell at the starting point (white arrow) are shown in the right panels. Red fluorescence started to  
554 appear 6 hours after the detection of green fluorescence. Scale bar, 50  $\mu$ m. (D) Percentage of green-  
555 dominant cells among all fluorescent cells in *Ins1-GFP; Timer* embryos and neonates. Data represent  
556 the mean  $\pm$  SEM. n = 3–4 for each group.

557

558 **Figure 2. Expression pattern of key transcription factors in fluorescent cells of *Ins1-GFP; Timer***  
559 **mice**

560 (A-C) Immunofluorescence staining for PDX1, NKX6.1 and MAFA was performed in the pancreas  
561 of E16.5 *Ins1-GFP; Timer* embryos. Arrows indicate PDX1 (A), NKX6.1 (B) and MAFA (C)-positive  
562 green fluorescent cells. Scale bar, 20  $\mu$ m. (D) Percentage of MAFA-positive cells among green-  
563 fluorescent newborn  $\beta$  cells (green bars) and green/red double-fluorescent differentiated  $\beta$  cells

564 (yellow bars). Data represent the mean  $\pm$  SEM. \*  $p < 0.05$  versus newborn  $\beta$  cells;  $n = 3-6$  for each  
565 group.

566

### 567 **Figure 3. Spatial heterogeneity of newly generated $\beta$ cells**

568 (A) The pancreas of an *Ins1-GFP; Timer* embryo was dissected at E16.5 and stained with DBA lectin,  
569 a marker of pancreatic ducts (white). Whereas some green-fluorescent cells are attached to the duct  
570 cells (arrows;  $\beta^{\text{duct}}$  cells), others exist apart from the duct cells (arrowheads), adjacent to pre-existing  
571  $\beta$  cells. Scale bar, 20  $\mu\text{m}$ . (B) Three-dimensional images of  $\beta$  cells in *Ins1-GFP; Timer* embryos at  
572 E18.5. Green-dominant cells without red fluorescence are observed (arrows), located away from the  
573 ductal structures (white). (C) The pancreas of *Ins1-GFP; Timer* embryo was dissected at E16.5 and  
574 stained with DBA lectin (blue) and anti-PECAM1 antibody, a marker of blood vessels (white). There  
575 are green-fluorescent cells adjacent to blood vessels (arrows;  $\beta^{\text{vessel}}$  cells), but away from the duct  
576 cells. Scale bar, 20  $\mu\text{m}$ . (D) Percentage of  $\beta^{\text{duct}}$  cells and  $\beta^{\text{vessel}}$  cells among total fluorescent cells in  
577 the pancreata of *Ins1-GFP; Timer* embryos and neonates. Data represent the mean  $\pm$  SEM. \*  $p < 0.05$   
578 versus  $\beta^{\text{vessel}}$  cells;  $n = 3-5$  for each group.

579

### 580 **Figure 4. Single-cell transcriptome analysis for fluorescent cells of *Ins1-GFP; Timer* mice**

581 (A) FACS plot of E16.5 *Ins1-GFP; Timer* pancreatic cells used for library generation. (B) t-SNE plot  
582 of all three libraries; two green fluorescent (G1, G2) and one green/red fluorescent (GR) cell  
583 populations. (C-D) Pseudotime trajectory of G1, G2 and GR cells plotted by pseudotime (C) and  
584 library identities (D). (E) t-SNE plot of 5 cell clusters from E16.5 endocrine-lineage cells in green-  
585 fluorescent cells:  $G^{\text{Neurog3}}$  (purple),  $G^{\text{Sst}}$  (brown),  $G^{\text{Gcg}}$  (blue),  $G^{\text{Pdx1}}$  (pink) and  $G^{\text{Ins2}}$  (green) cells. (F)  
586 Top ten differentially expressed genes in the 5 endocrine-lineage clusters.

587

### 588 **Figure 5. Single-cell transcriptomic dynamics at branching point in early $\beta$ -cell specification**

589 (A) Minimal spanning tree of pseudotime analysis for  $G^{Neurog3}$  (purple),  $G^{Sst}$  (brown),  $G^{Gcg}$  (blue),  
590  $G^{Pdx1}$  (pink) and  $G^{Ins2}$  (green) cells generated by Monocle. Pseudotime from 0 to 15 orders  $G^{Neurog3}$   
591 cells first, following by  $G^{Sst}$ , and then  $G^{Gcg}$  and  $G^{Pdx1}/G^{Ins2}$  cells. (B) Single cell gene expression of  
592  $Neurog3$ ,  $Sst$ ,  $Gcg$ ,  $Pdx1$  and  $Ins2$  on the minimal spanning tree and mRNAs expression dynamics on  
593  $G^{Gcg}$  and  $G^{Pdx1}/G^{Ins2}$  branches over pseudotime. (C) Top one hundred differentially expressed genes  
594 on  $G^{Gcg}$  and  $G^{Pdx1}/G^{Ins2}$  branches generated by branched expression analysis modeling. (D) Gene  
595 ontology analysis for  $G^{Ins2}$  cells by Panther.

596

### 597 **Figure 6. Single-cell transcriptome analysis for late maturing $\beta$ cells**

598 (A) t-SNE plot of 6 cell clusters from E16.5 green/red-fluorescent cells:  $GR^{Spp1}$  (deep green),  $GR^{Pdx1}$   
599 (green),  $GR^{Syt4}$  (pink),  $GR^{Npy}$  (brown),  $GR^{Scg2}$  (blue) and  $GR^{Spc25}$  (purple) cells. (B) Top ten  
600 differentially expressed genes in the 6 green/red-fluorescent cell clusters. (C) pseudotime trajectory  
601 of  $GR^{Spp1}$ ,  $GR^{Pdx1}$ ,  $GR^{Syt4}$ ,  $GR^{Npy}$ ,  $GR^{Scg2}$  and  $GR^{Spc25}$  cells. Pseudotime from 0 to 30 orders  $GR^{Spp1}$ ,  
602  $GR^{Pdx1}$ ,  $GR^{Syt4}$ ,  $GR^{Npy}$ ,  $GR^{Scg2}$  cells and makes a branching point to replicating  $GR^{Spc25}$  cells before  
603 going to  $GR^{Npy}/GR^{Scg2}$  cells. (D) pseudotime trajectory plotted by pseudotime. (E) Percentage of cells  
604 on S, G2/M and G1 cell cycle stages in each cluster cells. Note no G1 cells detected in  $GR^{Spc25}$  cluster.

605

### 606 **Figure 7. Heterogeneous expression pattern of transcription factors by location of newly** 607 **generated $\beta^{duct}$ cells and $\beta^{vessel}$ cells**

608 (A-F) Immunofluorescence staining for Gcg, Mafb and Mafa was performed in the pancreas of E16.5  
609  $Ins1-GFP$ ;  $Timer$  embryos. (A) Arrows indicate Gcg-positive green fluorescent cells. Arrowheads  
610 indicate Gcg-positive green/red double fluorescent cells. (C) Arrows indicate Mafb-positive green  
611 fluorescent cells. (E) Arrows indicate Mafa-positive green fluorescent cells. (B, D, F) Percentage of  
612 Gcg (B), Mafb (D), Mafa (F)-positive cells among  $\beta^{duct}$  cells or  $\beta^{vessel}$  cells. Scale bar, 20  $\mu$ m. Data  
613 represent the mean  $\pm$  SEM. \*  $p < 0.05$ ; n = 3–6 for each group.

614

615 **Figure 8. Single-cell transcriptome analysis for human  $\beta$ -like cells**

616 (A) t-SNE plot of 7 cell clusters from day 27 human  $\beta$ -like cells:  $h\beta^{SST}$  (pink),  $h\beta^{GCG}$  (brown),  $h\beta^{GAST}$   
617 (purple),  $h\beta^{MGST1}$  (red),  $h\beta^{SYT13}$  (blue),  $h\beta^{IAPP}$  (deep green) and  $h\beta^{NEFM}$  (green) cells. (B) Top ten  
618 differentially expressed genes in the 7 human  $\beta$ -like cell clusters. (C) Single cell gene expression of  
619 *INS*, *GCG*, *SST*, *MAFB*, *MGST1* and *SYT13*. (D) pseudotime trajectory of  $h\beta^{SST}$ ,  $h\beta^{GCG}$ ,  $h\beta^{GAST}$ ,  
620  $h\beta^{MGST1}$ ,  $h\beta^{SYT13}$ ,  $h\beta^{IAPP}$  and  $h\beta^{NEFM}$  cells. Pseudotime from 0 to 15 orders  $h\beta^{SST}$  first, following  $h\beta^{GCG}$ ,  
621  $h\beta^{GAST}$ ,  $h\beta^{IAPP}$  and  $h\beta^{NEFM}$  cells.  $h\beta^{MGST1}$  and  $h\beta^{SYT13}$  cells are branched from  $h\beta^{GAST}$  and  $h\beta^{IAPP}$  cell,  
622 respectively. (E) Pseudotime trajectory plotted by pseudotime.

623

624 **SUPPLEMENTAL FIGURE LEGENDS**

625 **Figure S1. Improved detection of newly generated  $\beta$  cells in *Ins1-GFP; Timer* reporter mice**  
626 **through their increased fluorescence intensity, related to Fig. 1**

627 (A) Schematic diagram of the concept for improved temporal labeling in *Ins1-GFP; Timer* double-  
628 transgenic mice. As the intensity of green fluorescence is below the microscopic detection threshold  
629 in *Ins1-Timer* (DsRed-E5) embryos, green-fluorescent  $\beta$  cells can be detected only by flow cytometry.  
630 In contrast, the green fluorescence from *Ins1-GFP* mice is sufficiently high to be detected by  
631 microscopy, and hence *Ins1-GFP; Timer* double transgenic mice enable us to label newly generated  
632  $\beta$  cells as green-fluorescent cells, which can be detected by microscopy separately from mature  $\beta$   
633 cells with green and red fluorescence. (B) The pancreas was dissected from an *Ins1-GFP; Timer*  
634 embryo at E17.5 and stained with anti-insulin antibody. All fluorescent cells were positive for insulin.  
635 Scale bar, 50  $\mu$ m.

636

637 **Figure S2. Location analysis of newly generated  $\beta$  cells by three-dimensional images, related to**  
638 **Fig. 3**

639 Three-dimensional images of an *Ins1-GFP; Timer* embryo at E18.5, stained with DBA lectin (blue),  
640 and with anti-PECAM1 antibody (white). Green-fluorescent cells were observed adjacent to the blood  
641 vessels and away from the duct cells (arrows).

642

643 **Figure S3. Single-cell transcriptome analysis for green- and green/red-fluorescent cells, related**  
644 **to Fig. 4**

645 (A) t-SNE plot of 12 clusters from E16.5 green- and green/red-fluorescent cells: insulin-expressing  
646  $\beta$ -cell lineage clusters ( $\beta^{Spp1}$ ,  $\beta^{Neurog3}$ ,  $\beta^{Sst}$ ,  $\beta^{Gcg}$ ,  $\beta^{Pdx1}$ ,  $\beta^{Ins1}$ ,  $\beta^{Npy}$  and  $\beta^{Spc25}$ ), duct, acinar, stellate and  
647 endothelial cells. (B) Top ten differentially expressed genes in the 8  $\beta$ -cell lineage clusters. (C) Single  
648 cell gene expression of *Ins1* and *Ins2*. (D) Pseudotime trajectory of 8  $\beta$ -cell lineage clusters ( $\beta^{Spp1}$ ,  
649  $\beta^{Neurog3}$ ,  $\beta^{Sst}$ ,  $\beta^{Gcg}$ ,  $\beta^{Pdx1}$ ,  $\beta^{Ins1}$ ,  $\beta^{Npy}$  and  $\beta^{Spc25}$ ). Pseudotime from 0 to 40 orders  $\beta^{Spp1}$ ,  $\beta^{Neurog3}$ ,  $\beta^{Sst}$ ,  $\beta^{Gcg}$ ,  
650  $\beta^{Pdx1}$ ,  $\beta^{Ins1}$ ,  $\beta^{Npy}$  cells and makes a branching point to replicating  $\beta^{Spc25}$  cells before going to  $\beta^{Npy}$  cells.  
651 (E) Single cell gene expression of *Ins1* and *Ins2* on the minimal spanning tree. (F) *Ins1* and *Ins2* gene  
652 expression dynamics on the pseudotime trajectory over pseudotime.

653

654 **Figure S4. Single-cell transcriptome analysis for green-fluorescent cells, related to Fig. 4 and**  
655 **Fig. 5**

656 (A) t-SNE plot of 8 cell clusters from E16.5 green-fluorescent cells: endocrine (Endo1 and Endo2),  
657 trunk, duct, stellate, endothelial, replicating (Rep), and low number gene (Low nGene) cells. Only  
658 endocrine (Endo1 and Endo2) cluster cells are used for further analysis in Fig. 4E, 4F and Fig. 5A-E.  
659 (B) Top ten differentially expressed genes in the 8 green-fluorescent cell clusters.

660

661 **Figure S5. Single-cell transcriptome and pseudotime analysis for green-fluorescent cells, related**  
662 **to Fig. 4 and Fig. 5**

663 (A) Single cell gene expression of *Neurog3*, *Sst*, *Gcg*, *Pdx1* and *Ins2* on t-SNE plot of 5 cell clusters  
664 from E16.5 endocrine-lineage cells in green-fluorescent cells (related to Fig, 4E and 4F). (B-D) Single  
665 cell gene expression on the minimal spanning tree and gene expression dynamics on  $G^{Gcg}$  and  
666  $G^{Pdx1}/G^{Ins2}$  branches over pseudotime. Transcription factors important for pancreas development  
667 (*Maib*, *Isl1* and *Arx* in B), ER stress-associated highly expressed genes in  $G^{Ins2}$  (*Calr*, *Hspa5* and  
668 *Sdf2ll* in C) and highly expressed genes in  $G^{Pdx1}$  (*Mgst1* and *Syt13* in D) are shown, related to Fig.  
669 5A-E. (E) Gene ontology analysis for  $G^{Pdx1}$  cells by Panther.

670

671 **Figure S6. Integration analysis of mouse embryonic  $\beta$ -cell lineage cells and human  $\beta$ -like cells**

672 (A) t-SNE plot of 14 clusters from E16.5 green- and green/red-fluorescent cells in *Ins1-GFP; Timer*  
673 mice, integrated with day 27 human  $\beta$ -like cells: insulin-expressing  $\beta$ -cell lineage clusters ( $c\beta^{GCG}$ ,  
674  $c\beta^{SST}$ ,  $c\beta^{NEUROG3}$ ,  $c\beta^{INS}$ ,  $c\beta^{NEFM}$ ,  $c\beta^{PCSK1}$ ,  $c\beta^{GCG}$ , and  $c\beta^{SPP1}$ ), duct, acinar, stellate, endothelial,  
675 replicating (Rep), and miscellaneous (Misc) cells. (B) Top ten differentially expressed genes in the 8  
676  $\beta$ -cell lineage clusters. (C) Single cell gene expression of *NEUROG3*, *SST*, *GCG* and *INS*. (D) t-SNE  
677 plot grouped by species: human (pink) and mouse (blue) cells.  $c\beta^{GCG}$ ,  $c\beta^{SST}$  and  $c\beta^{NEUROG3}$  clusters  
678 are reflected by dotted lines. (E) Proportions of human and mouse cells in each cluster cells.  $c\beta^{GCG}$ ,  
679  $c\beta^{SST}$  and  $c\beta^{NEUROG3}$  clusters (surrounded by dotted line) contain more than 15% of human or mouse  
680 cells.

681

682 **Figure S7. Spatial heterogeneity of  $\beta$ -cell neogenesis revealed by *Ins1-GFP; Timer* embryos**

683 Proposed model for  $\beta$ -cell neogenesis. Endocrine progenitors differentiate into two types of  $\beta$  cells;  
684  $\beta^{duct}$  and  $\beta^{vessel}$  cells. The  $\beta^{duct}$  cells are observed in the ductal region where *Neurog3*-expressing  
685 endocrine progenitors emerge. In contrast, part of endocrine progenitors migrate away from the ductal  
686 area toward the region near the blood vessels and differentiate into  $\beta^{vessel}$  cells.

687

688

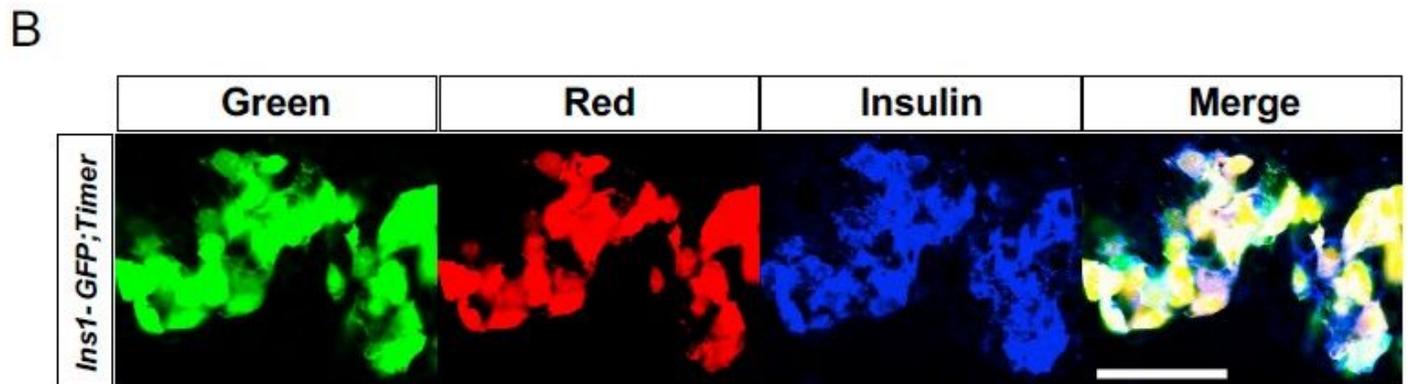
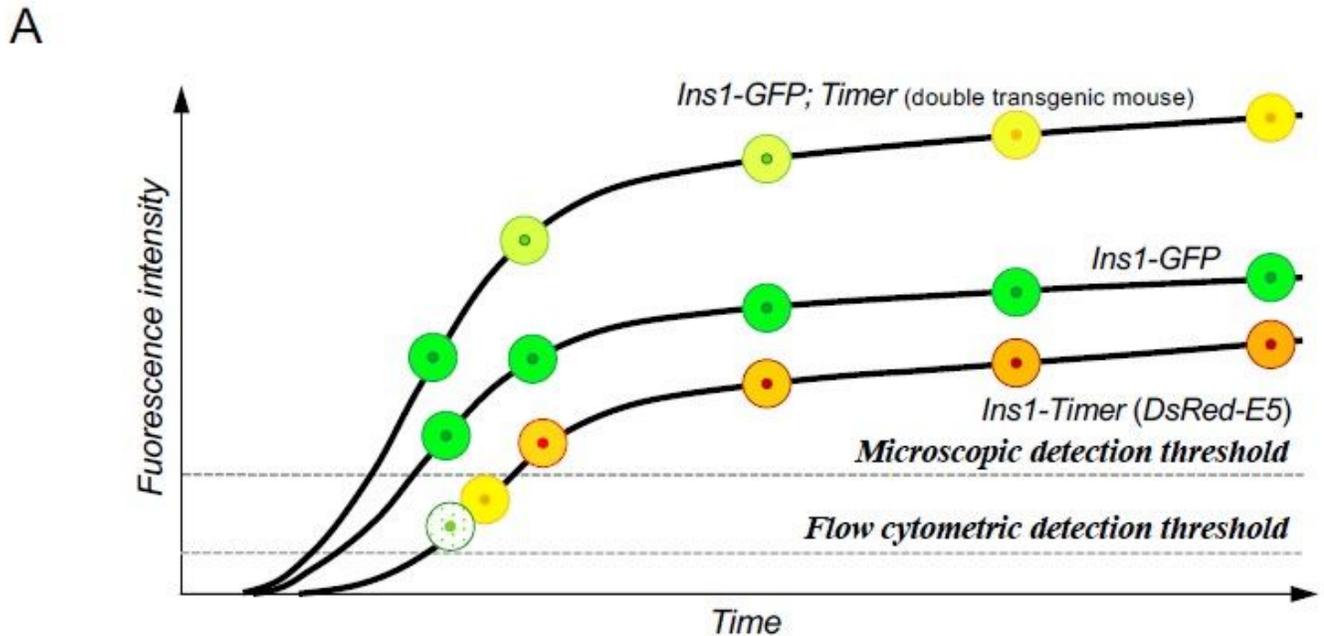
689 **Table S1. Top ten differentially expressed genes in the 4 non- $\beta$ -cell lineage clusters of E16.5**  
690 **green- and green/red-fluorescent cells, related to Figure S3**

691

692 **Movie S1. *Ex vivo* real-time imaging of the developing pancreas of *Ins1-GFP*; *Timer* mice**

693 The developing pancreata of *Ins1-GFP*; *Timer* embryos at E14.5 were dissected and cultured *ex vivo*  
694 to acquire real-time imaging of fluorescent cells. Images were acquired every 30 min for a total of 10  
695 hours.

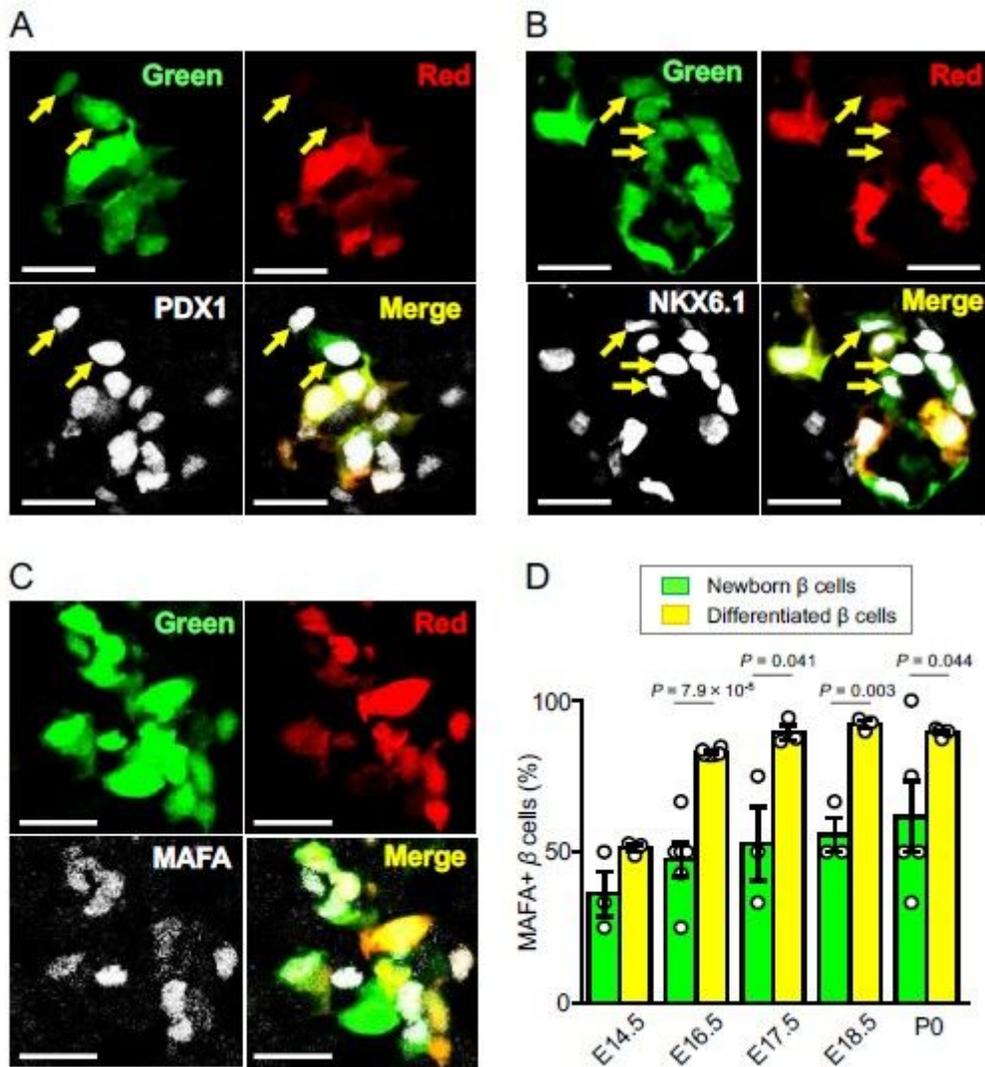
# Figures



**Figure 1**

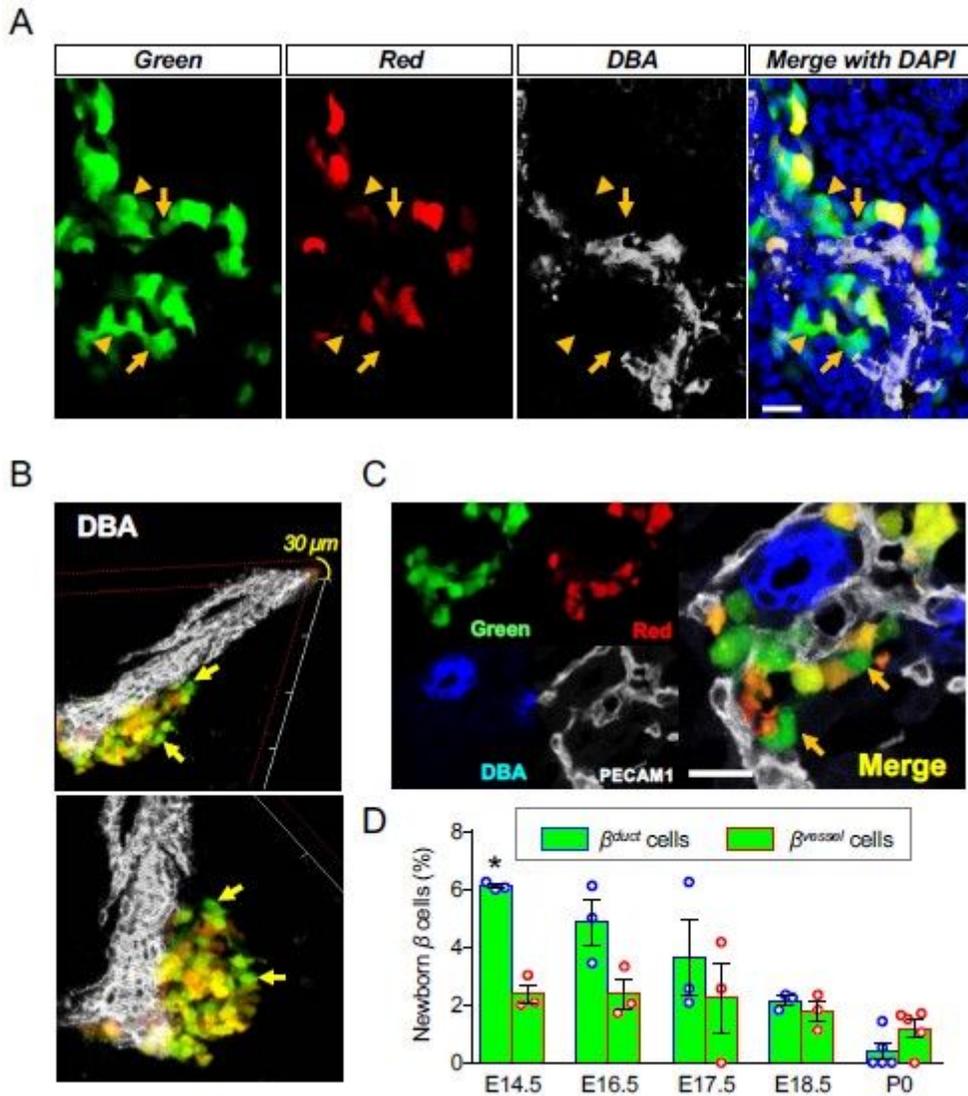
Improved “Timer” fluorescent system for labeling newly generated  $\beta$  cells (A) Schematic diagram of the concept for improved temporal labeling in *Ins1-GFP; Timer (DsRed-E5)* double-transgenic mice. Whereas green fluorescence in newly generated  $\beta$  cells of *Ins1-Timer* embryos is below the microscopic detection threshold, green fluorescence from *Ins1-GFP* mice is high enough to be detected by microscopy in newly generated  $\beta$  cells as well as in mature  $\beta$  cells. Therefore, *Ins1-GFP; Timer* double-transgenic mice enable the labeling of newly generated  $\beta$  cells as green-fluorescent cells by microscopy, which can be distinguished from mature  $\beta$  cells with green and red fluorescence. (B) Fluorescence images of the pancreas of *Ins1-Timer* and *Ins1-GFP; Timer* embryos at E16.5. Whereas all green-fluorescent cells exhibited red fluorescence in *Ins1-Timer* embryos, green fluorescent cells without red fluorescence (arrows) were observed in *Ins1-GFP; Timer* double-transgenic embryos. Scale bar, 50  $\mu$ m. (C) Time-lapse imaging of fluorescent cells within an E14.5 pancreatic bud of *Ins1-GFP; Timer* embryos. Enlarged

images of a green-fluorescent cell at the starting point (white arrow) are shown in the right panels. Red fluorescence started to appear 6 hours after the detection of green fluorescence. Scale bar, 50  $\mu\text{m}$ . (D) Percentage of green dominant cells among all fluorescent cells in Ins1-GFP; Timer embryos and neonates. Data represent the mean  $\pm$  SEM.  $n = 3-4$  for each group.



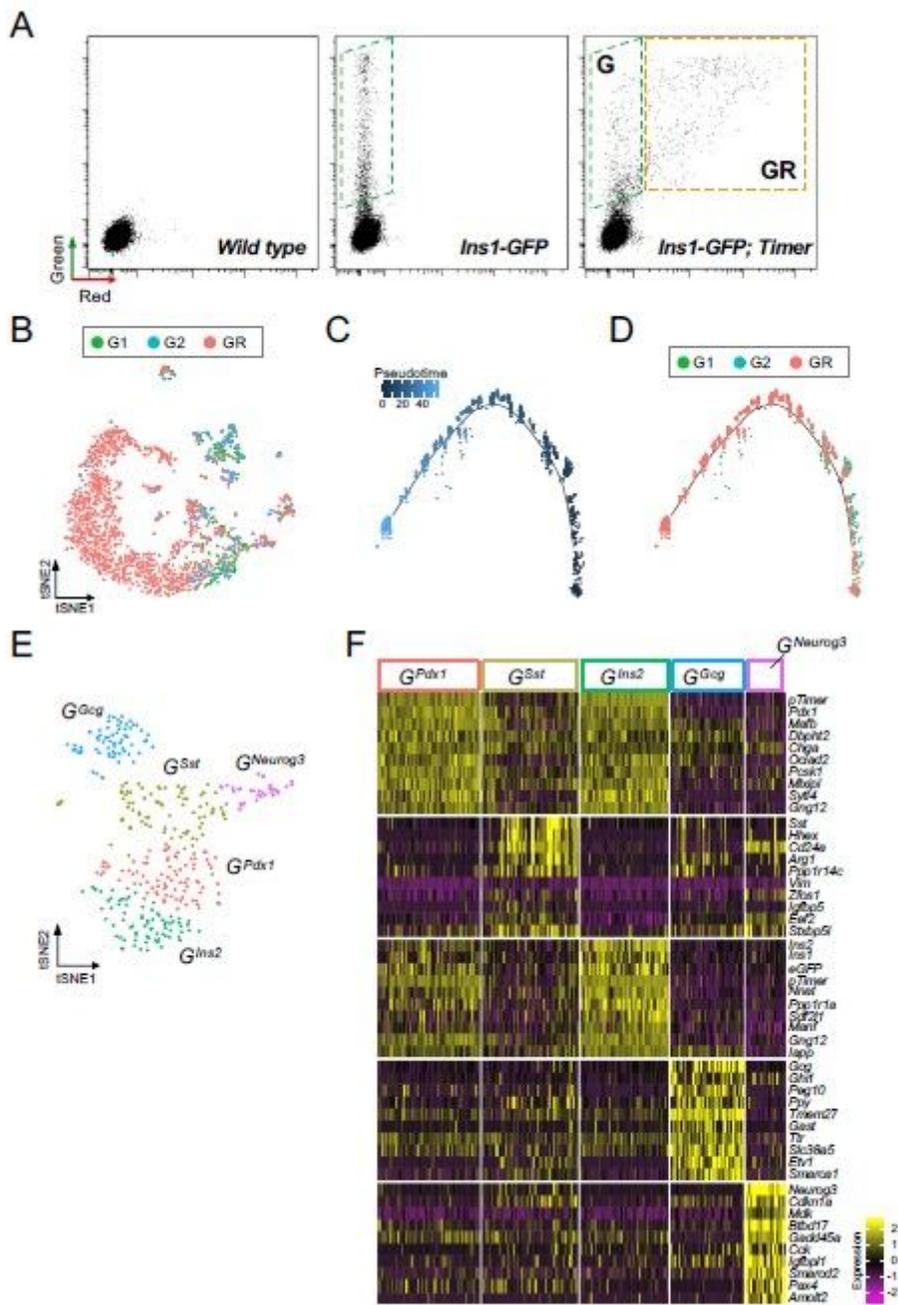
**Figure 2**

Expression pattern of key transcription factors in fluorescent cells of Ins1-GFP; Timer mice (A-C) Immunofluorescence staining for PDX1, NKX6.1 and MAFA was performed in the pancreas of E16.5 Ins1-GFP; Timer embryos. Arrows indicate PDX1 (A), NKX6.1 (B) and MAFA (C)-positive green fluorescent cells. Scale bar, 20  $\mu\text{m}$ . (D) Percentage of MAFA-positive cells among green fluorescent newborn  $\beta$  cells (green bars) and green/red double-fluorescent differentiated  $\beta$  cells (yellow bars). Data represent the mean  $\pm$  SEM. \*  $p < 0.05$  versus newborn  $\beta$  cells;  $n = 3-6$  for each group.



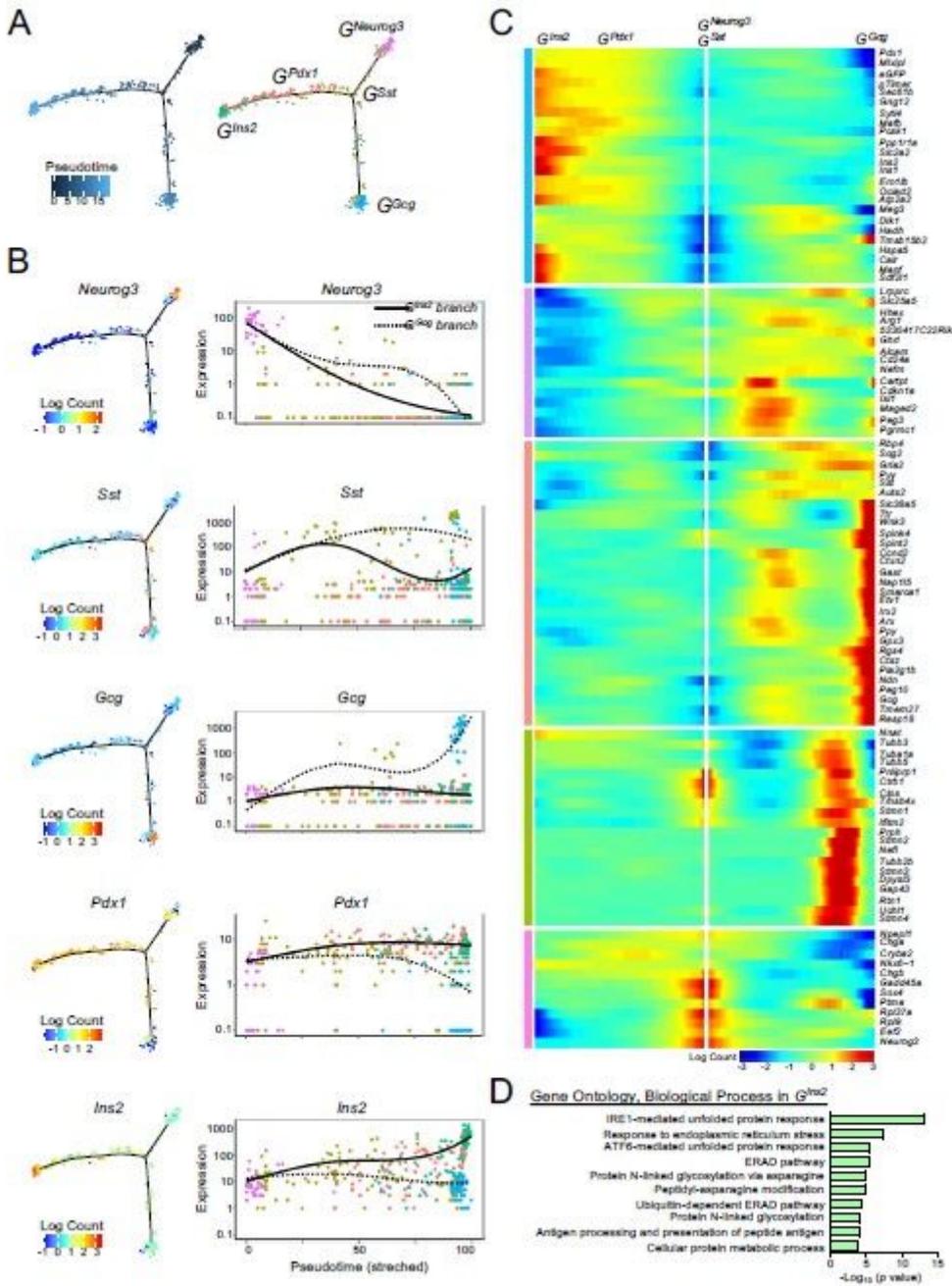
**Figure 3**

Spatial heterogeneity of newly generated  $\beta$  cells (A) The pancreas of an Ins1-GFP; Timer embryo was dissected at E16.5 and stained with DBA lectin, a marker of pancreatic ducts (white). Whereas some green-fluorescent cells are attached to the duct cells (arrows;  $\beta_{duct}$  cells), others exist apart from the duct cells (arrowheads), adjacent to pre-existing  $\beta$  cells. Scale bar, 20  $\mu$ m. (B) Three-dimensional images of  $\beta$  cells in Ins1-GFP; Timer embryos at E18.5. Green-dominant cells without red fluorescence are observed (arrows), located away from the ductal structures (white). (C) The pancreas of Ins1-GFP; Timer embryo was dissected at E16.5 and stained with DBA lectin (blue) and anti-PECAM1 antibody, a marker of blood vessels (white). There are green-fluorescent cells adjacent to blood vessels (arrows;  $\beta_{vessel}$  cells), but away from the duct cells. Scale bar, 20  $\mu$ m. (D) Percentage of  $\beta_{duct}$  cells and  $\beta_{vessel}$  cells among total fluorescent cells in the pancreata of Ins1-GFP; Timer embryos and neonates. Data represent the mean  $\pm$  SEM. \*  $p < 0.05$  versus  $\beta_{vessel}$  cells;  $n = 3-5$  for each group.



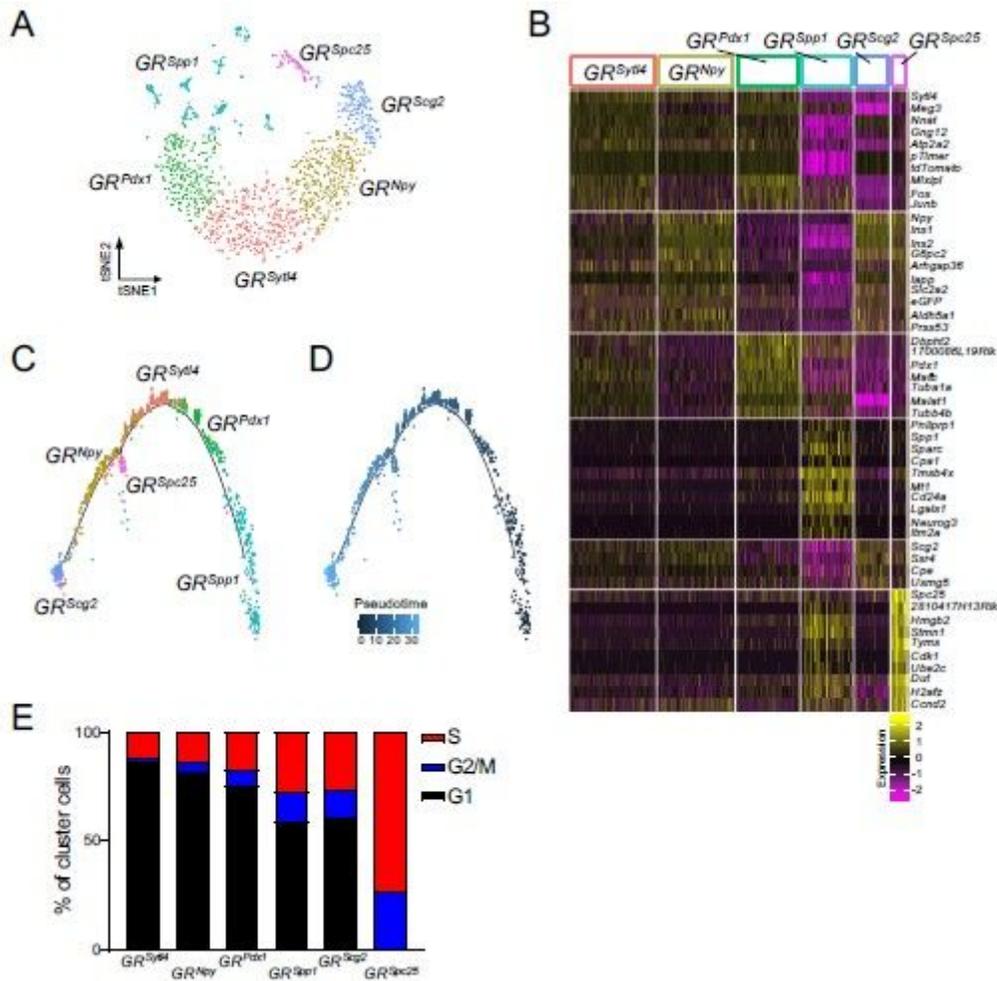
**Figure 4**

Single-cell transcriptome analysis for fluorescent cells of *Ins1-GFP; Timer* mice (A) FACS plot of E16.5 *Ins1-GFP; Timer* pancreatic cells used for library generation. (B) t-SNE plot of all three libraries; two green fluorescent (G1, G2) and one green/red fluorescent (GR) cell populations. (C-D) Pseudotime trajectory of G1, G2 and GR cells plotted by pseudotime (C) and library identities (D). (E) t-SNE plot of 5 cell clusters from E16.5 endocrine-lineage cells in green fluorescent cells: G<sub>Neurog3</sub> (purple), G<sub>Sst</sub> (brown), G<sub>Gcg</sub> (blue), G<sub>Pdx1</sub> (pink) and G<sub>Ins2</sub> (green) cells. (F) Top ten differentially expressed genes in the 5 endocrine-lineage cluster



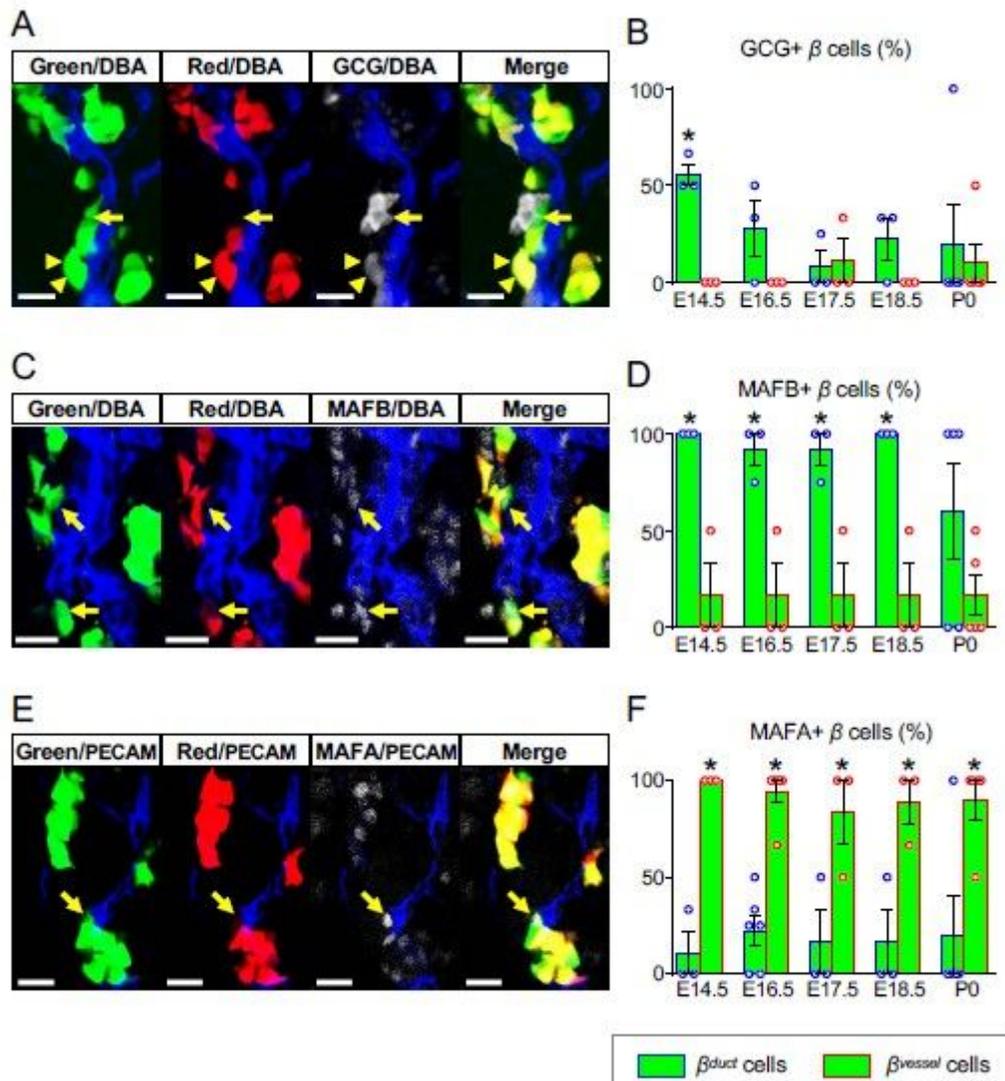
**Figure 5**

Single-cell transcriptomic dynamics at branching point in early  $\beta$ -cell specification (A) Minimal spanning tree of pseudotime analysis for GNeurog3 589 (purple), GSst (brown), GGcg (blue), GPdx1 (pink) and Glns2 (green) cells generated by Monocle. Pseudotime from 0 to 15 orders GNeurog3 cells first, following by GSst, and then GGcg and GPdx1/Glns2 cells. (B) Single cell gene expression of Neurog3, Sst, Gcg, Pdx1 and Ins2 on the minimal spanning tree and mRNAs expression dynamics on GGcg and GPdx1/Glns2 branches over pseudotime. (C) Top one hundred differentially expressed genes on GGcg and GPdx1/Glns2 branches generated by branched expression analysis modeling. (D) Gene ontology analysis for Glns2 cells by Panther.



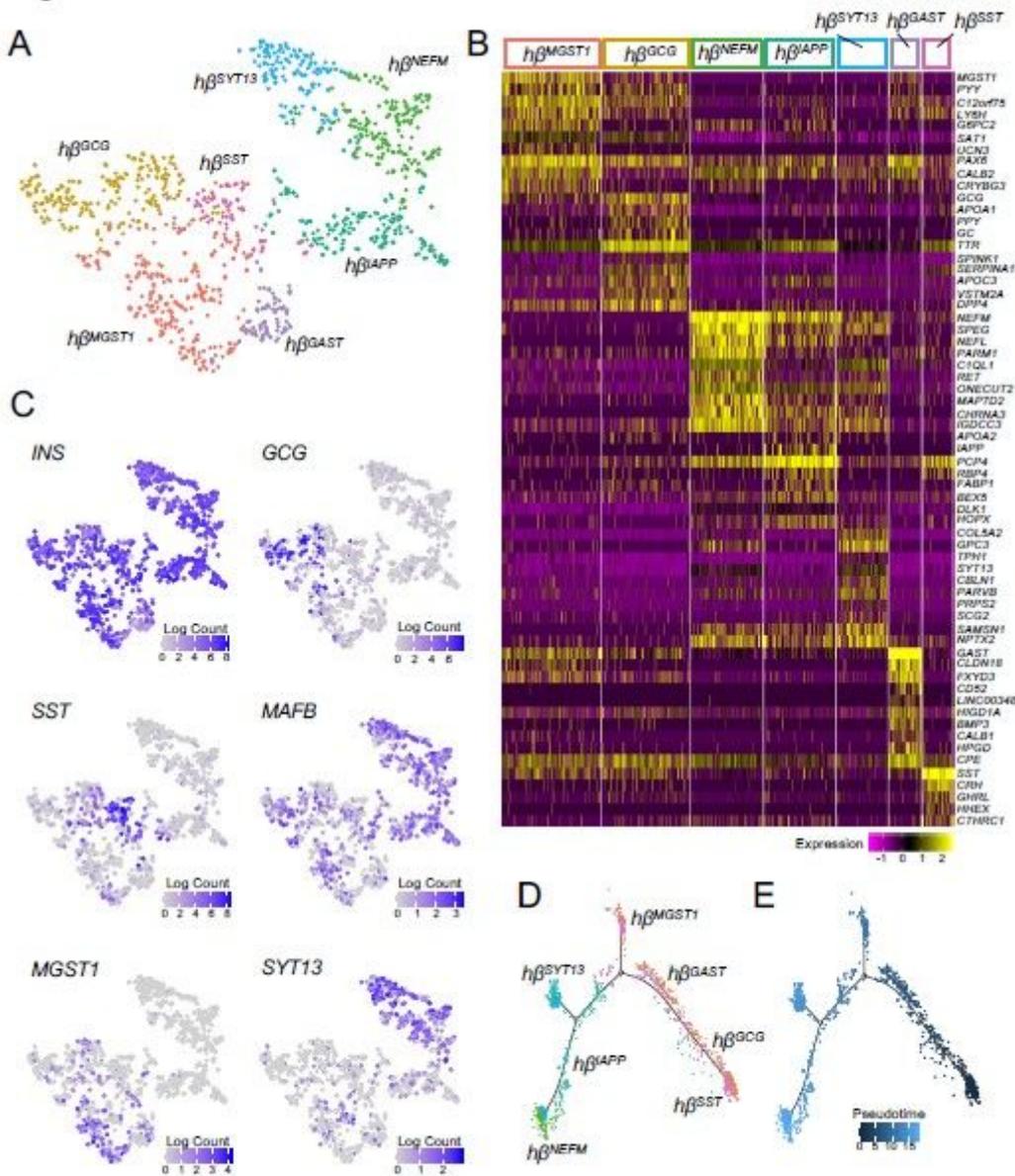
**Figure 6**

Single-cell transcriptome analysis for late maturing  $\beta$  cells (A) t-SNE plot of 6 cell clusters from E16.5 green/red-fluorescent cells: GR<sub>Spp1</sub> (deep green), GR<sub>Pdx1</sub> (green), GR<sub>Syt4</sub> (pink), GR<sub>Npy</sub> (brown), GR<sub>Scg2</sub> (blue) and GR<sub>Spc25</sub> (purple) cells. (B) Top ten differentially expressed genes in the 6 green/red-fluorescent cell clusters. (C) pseudotime trajectory of GR<sub>Spp1</sub>, GR<sub>Pdx1</sub>, GR<sub>Syt4</sub>, GR<sub>Npy</sub>, GR<sub>Scg2</sub> and GR<sub>Spc25</sub> cells. Pseudotime from 0 to 30 orders GR<sub>Spp1</sub>, GR<sub>Pdx1</sub>, GR<sub>Syt4</sub>, GR<sub>Npy</sub>, GR<sub>Scg2</sub> cells and makes a branching point to replicating GR<sub>Spc25</sub> cells before going to GR<sub>Npy</sub>/GR<sub>Scg2</sub> cells. (D) pseudotime trajectory plotted by pseudotime. (E) Percentage of cells on S, G2/M and G1 cell cycle stages in each cluster cells. Note no G1 cells detected in GR<sub>Spc25</sub> cluster.



**Figure 7**

Heterogeneous expression pattern of transcription factors by location of newly generated  $\beta$  duct cells and  $\beta$  vessel cells (A-F) Immunofluorescence staining for Gcg, Mafb and Mafa was performed in the pancreas of E16.5 *Ins1-GFP*; *Timer* embryos. (A) Arrows indicate Gcg-positive green fluorescent cells. Arrowheads indicate Gcg-positive green/red double fluorescent cells. (C) Arrows indicate Mafb-positive green fluorescent cells. (E) Arrows indicate Mafa-positive green fluorescent cells. (B, D, F) Percentage of Gcg (B), Mafb (D), Mafa (F)-positive cells among  $\beta$  duct cells or  $\beta$  vessel cells. Scale bar, 20  $\mu$ m. Data represent the mean  $\pm$  SEM. \*  $p < 0.05$ ;  $n = 3-6$  for each group.



**Figure 8**

Single-cell transcriptome analysis for human  $\beta$ -like cells (A) t-SNE plot of 7 cell clusters from day 27 human  $\beta$ -like cells:  $h\beta$ SST (pink),  $h\beta$ GCG (brown),  $h\beta$ GAST (purple),  $h\beta$ MGST1 (red),  $h\beta$ SYT13 (blue),  $h\beta$ IAPP (deep green) and  $h\beta$ NEFM (green) cells. (B) Top ten differentially expressed genes in the 7 human  $\beta$ -like cell clusters. (C) Single cell gene expression of *INS*, *GCG*, *SST*, *MAFB*, *MGST1* and *SYT13*. (D) pseudotime trajectory of  $h\beta$ SST,  $h\beta$ GCG,  $h\beta$ GAST,  $h\beta$ MGST1,  $h\beta$ SYT13,  $h\beta$ IAPP and  $h\beta$ NEFM cells. Pseudotime from 0 to 15 orders  $h\beta$ SST first, following  $h\beta$ GCG,  $h\beta$ GAST,  $h\beta$ IAPP and  $h\beta$ NEFM cells.  $h\beta$ MGST1 and  $h\beta$ SYT13 cells are branched from  $h\beta$ GAST and  $h\beta$ IAPP cell, respectively. (E) Pseudotime trajectory plotted by pseudotime.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [FigureS1.jpg](#)
- [FigureS2.jpg](#)
- [FigureS3.jpg](#)
- [FigureS4.jpg](#)
- [FigureS5.jpg](#)
- [FigureS6.jpg](#)
- [FigureS7.jpg](#)
- [TableS1.png](#)