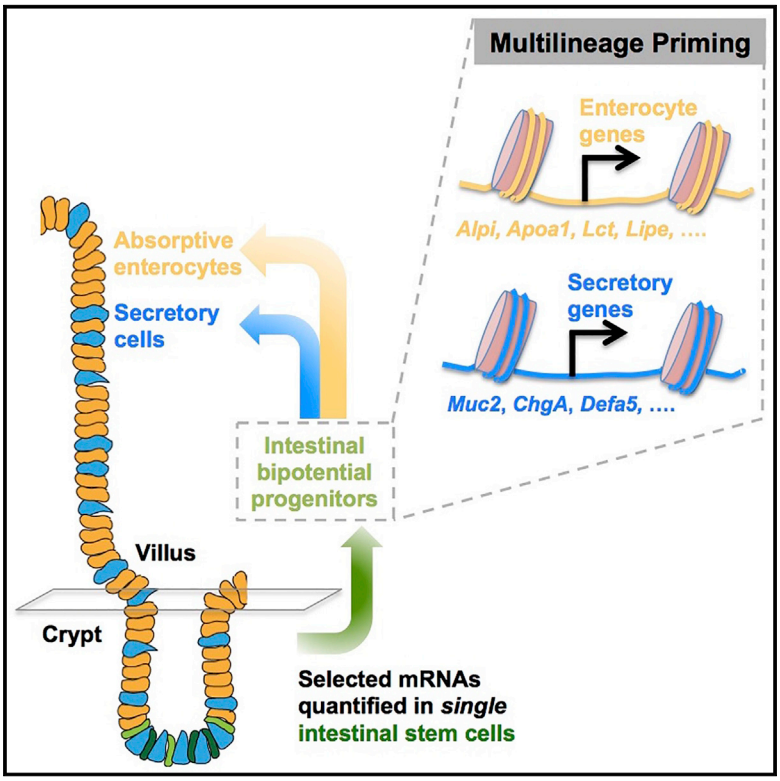
Cell Reports

Single-Cell Transcript Profiles Reveal Multilineage Priming in Early Progenitors Derived from Lgr5$^+$ Intestinal Stem Cells

Graphical Abstract



Highlights

- Single-cell analyses of intestinal Lgr5$^+$ cells identify two discrete populations
- Both pools express stem-cell genes, but only one activates terminal cell markers
- Multilineage-primed Lgr5$^+$ cells express features of bipotential progenitors
- A suite of informatics tools reveals that these progenitors originate in Lgr5$^+$ ISCs

Authors

Tae-Hee Kim, Assieh Saadatpour, Guoji Guo, ..., Stuart H. Orkin, Guo-Cheng Yuan, Ramesh A. Shivdasani

Correspondence

gcyuan@jimmy.harvard.edu (G.-C.Y.), ramesh_shivdasani@dfci.harvard.edu (R.A.S.)

In Brief

Characterizing the earliest cells to exit in vivo stem-cell compartments is a challenge. Kim et al. demonstrate multilineage priming—co-expression of markers for both the absorptive and secretory daughter lineages—in the earliest progeny of Lgr5$^+$ intestinal crypt stem cells.

Kim et al., 2016, Cell Reports *16*, 1–8
August 23, 2016
http://dx.doi.org/10.1016/j.celrep.2016.07.056

CellPress

Please cite this article in press as: Kim et al., Single-Cell Transcript Profiles Reveal Multilineage Priming in Early Progenitors Derived from Lgr5+ Intestinal Stem Cells, Cell Reports (2016), http://dx.doi.org/10.1016/j.celrep.2016.07.056

Cell Reports
Report

Single-Cell Transcript Profiles Reveal Multilineage Priming in Early Progenitors Derived from Lgr5+ Intestinal Stem Cells

Tae-Hee Kim,[1,2,8,9,10] Assieh Saadatpour,[3,8] Guoji Guo,[4,11] Madhurima Saxena,[1,2] Alessia Cavazza,[1,2] Niyati Desai,[5] Unmesh Jadhav,[1,2] Lan Jiang,[3] Miguel N. Rivera,[5] Stuart H. Orkin,[4,6,7] Guo-Cheng Yuan,[3,6,*] and Ramesh A. Shivdasani[1,2,6,12,*]

[1]Department of Medical Oncology and Center for Functional Cancer Epigenetics, Dana-Farber Cancer Institute, Boston, MA 02215, USA
[2]Department of Medicine, Harvard Medical School, Boston, MA 02215, USA
[3]Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute and Harvard T.H. Chan School of Public Health, Boston, MA 02215, USA
[4]Department of Pediatric Oncology, Dana-Farber Cancer Institute, Boston Children's Hospital and Harvard Medical School, Boston, MA 02215, USA
[5]Department of Pathology and Center for Cancer Research, Massachusetts General Hospital and Harvard Medical School, Boston, MA 02114, USA
[6]Harvard Stem Cell Institute, Cambridge, MA 02138, USA
[7]Howard Hughes Medical Institute, Boston, MA 02115, USA
[8]Co-first author
[9]Present address: Developmental and Stem Cell Biology Program, The Hospital for Sick Children, Toronto, ON M5G 0A4, Canada
[10]Present address: Department of Molecular Genetics, University of Toronto, Toronto, ON M5S 1A8, Canada
[11]Present address: Center of Stem Cell and Regenerative Medicine, Zhejiang University, Hangzhou, Zhejiang 310058, China
[12]Lead Contact
*Correspondence: gcyuan@jimmy.harvard.edu (G.-C.Y.), ramesh_shivdasani@dfci.harvard.edu (R.A.S.)
http://dx.doi.org/10.1016/j.celrep.2016.07.056

SUMMARY

Lgr5+ intestinal stem cells (ISCs) drive epithelial self-renewal, and their immediate progeny—intestinal bipotential progenitors—produce absorptive and secretory lineages via lateral inhibition. To define features of early transit from the ISC compartment, we used a microfluidics approach to measure selected stem- and lineage-specific transcripts in single Lgr5+ cells. We identified two distinct cell populations, one that expresses known ISC markers and a second, abundant population that simultaneously expresses markers of stem and mature absorptive and secretory cells. Single-molecule mRNA in situ hybridization and immunofluorescence verified expression of lineage-restricted genes in a subset of Lgr5+ cells in vivo. Transcriptional network analysis revealed that one group of Lgr5+ cells arises from the other and displays characteristics expected of bipotential progenitors, including activation of Notch ligand and cell-cycle-inhibitor genes. These findings define the earliest steps in ISC differentiation and reveal multilineage gene priming as a fundamental property of the process.

INTRODUCTION

Cell turnover in the small bowel relies on pools of 12–15 Wnt-responsive Lgr5+ intestinal stem cells (ISCs) that lie at the base of each intestinal crypt and replicate daily to produce new ISCs and transit-amplifying (TA) progenitors (Barker et al., 2007). Other cells present near crypt tier 4 express a combination of *Bmi1*, *mTert*, and *Hopx1* (Barker et al., 2012) and may represent Paneth cell precursors that are recruited into the stem-cell pool upon epithelial injury (Buczacki et al., 2013). Both Lgr5+ ISCs and TA cells replicate briskly, albeit at different rates, and TA cells quickly adopt a single fate—absorptive or secretory—whereas ISCs stay multipotent; the basis for these cardinal differences is unknown. In another self-renewing tissue, blood cell progenitors simultaneously activate genes specific to each daughter lineage before distinct cell types are specified, a phenomenon known as multilineage priming (Hu et al., 1997; Miyamoto et al., 2002). Because absorptive and secretory fates are determined by lateral inhibition, a means for reciprocal cell specification (Pellegrinet et al., 2011; Stamataki et al., 2011), it is unclear whether the progeny of Lgr5+ ISCs traverse a similar phase. Lateral inhibition likely occurs in intestinal bipotential progenitors (IBPs), which have never been captured and may represent the earliest, albeit transient, progeny of Lgr5+ ISCs.

Lgr5+ cells show a range of GFP signals in *Lgr5^{Gfp}* mice (Barker et al., 2007), and cells at the center of the crypt base produce larger clones than cells located at the periphery (Ritsma et al., 2014). Not all Lgr5+ cells spawn functional clones in vivo (Kozar et al., 2013), and some of them correspond to non-cycling Paneth-cell precursors (Buczacki et al., 2013). Although these observations suggest that early progenitors might arise among Lgr5+ cells, a recent single-cell mRNA study (Grün et al., 2015) reported that Lgr5^{hi} cells are homogeneous, possibly because the method has low sensitivity for transcripts expressed at low

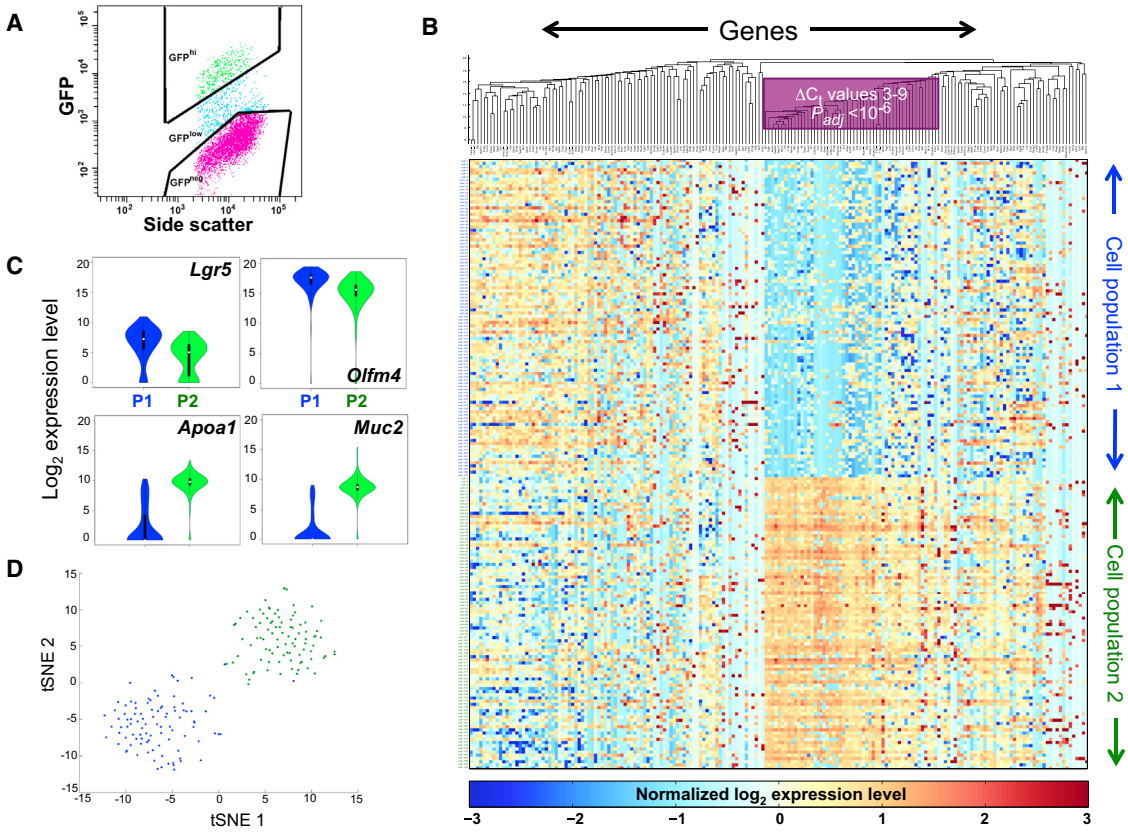


Figure 1. Targeted mRNA Profiles Identify Two Populations of Lgr5$^+$ Intestinal Crypt Cells

(A) Flow cytometry plot, showing the gates applied to isolate Lgr5^{hi} (green) cells.

(B) Heatmap display of *k*-means (*k* = 2) clustering of Ct values from 183 mRNAs (x axis, five genes are represented by two primer sets each) in 192 single Lgr5$^+$ intestinal crypt cells (y axis). Blue represents absent to low, and yellow to amber represent increasing, transcript levels. Genes are ordered by hierarchical clustering with the average linkage method and Euclidean distance. A block of genes that best distinguishes the two cell populations, including most mature villus markers, is boxed.

(C) Violin plots showing differential expression of representative stem (*Lgr5* and *Olfm4*) and differentiated (*Apoa1* and *Muc2*) cell markers in all cells in populations 1 (P1; blue) and 2 (P2; green).

(D) t-SNE analysis of the qRT-PCR data, demonstrating discrete Lgr5$^+$ cell populations (blue and green); overlaid colors are from the adjoining *k*-means clusters. See also Figure S1 and Tables S1 and S2.

abundance. To overcome this limitation, we measured 185 transcripts for selected stem cell and lineage-specific markers in single GFP$^+$ (Lgr5$^+$) intestinal crypt cells isolated from the same *Lgr5^{GFP}* mice (Barker et al., 2007). We identified a distinct population that expresses slightly reduced levels of known ISC transcripts and co-expresses markers of mature secretory cells and enterocytes. Immunofluorescence and single-molecule mRNA in situ hybridization (ISH) confirmed the presence of these cells in vivo, and analysis of transcript networks indicates that they represent early ISC-derived bipotential progenitors.

RESULTS AND DISCUSSION

We used microfluidic qRT-PCR following targeted pre-amplification of 185 genes from defined categories (Table S1), including genes previously identified as Lgr5$^+$ cell specific (Kim et al., 2014; Muñoz et al., 2012); targets of various signaling pathways; markers specific to mature enterocytes or secretory cells (Kim et al., 2014); and tissue-restricted transcription factors. To ensure reproducibility and RNA quality, we assessed three housekeeping genes (*Actb*, *Gapdh*, and *Hprt*) and used two separate primer pairs to measure five genes. From *Lgr5^{Gfp}* mice (Barker et al., 2007), we captured crypt epithelial cells that showed strong GFP fluorescence in flow cytometry (Figure 1A) but might, nevertheless, include LGR5$^+$ cells on the verge of ISC exit. Fluorescence microscopy and direct visualization verified the recovery of dilute, viable GFP$^+$ singlets (Figure S1A). Following reverse transcription with primers specific to the selected genes and PCR amplification of cDNA, we excluded wells that gave cycle threshold (Ct) values <13 in qRT-PCR for *Actb*, further eliminating possible rare doublets. Different primers for each of five selected genes gave concordant results (Table S1), indicating a robust protocol.

We measured the levels of all 185 genes in 192 cells captured on 2 separate days and pooled the data for subsequent analyses (Table S2); two genes, *Zg16* and *Ido1*, gave no signal in any cell

and were excluded from the analysis. *k*-means clustering of the RNA data, using the Silhouette measure (Kaufman and Rousseeuw, 1990) to identify the best *k* (Figure S1B), revealed two distinct cell populations that were roughly equal in size (Figure 1B) and expressed similar levels of markers historically assigned to quiescent ISCs (Figure S1C). The salient differences between these two populations were a modestly higher (2- to 8-fold) expression of ISC markers, such as *Lgr5* and *Olfm4*, in one pool and an 8- to >100-fold higher expression of many genes in the other (Figures 1B and 1C); adjusted p (p_{adj}), $<10^{-7}$ to $<10^{-5}$. After confirming efficient qPCR by selected primer pairs, we estimated copy numbers of some of the latter mRNAs at 3% to 8% of *Hprt* copies (Figure S1D). Cells isolated on different days were similarly distributed in the two pools, and, to verify the results from *k*-means clustering, we used t-distributed Stochastic Neighbor Embedding (t-SNE) (van der Maaten and Hinton, 2008). The two cell populations identified by *k*-means clustering remained distinct on a t-SNE map (blue and green dots in Figure 1D), and the high concordance of RNA profiles in each group (Figure 1B), together with the absence of outliers in t-SNE, strongly supports the absence of cell doublets.

Among the 185 genes we interrogated, 35 genes discriminated the two cell populations without ambiguity ($\Delta Ct > 3$, $p_{adj} < 10^{-6}$; Figure 1B; shaded in Table S1), and 31 of these transcripts were higher in population 2. Weighted gene co-expression network analysis (WGCNA) (Zhang and Horvath, 2005) revealed two specific, highly coordinated gene modules in this population (Figure 2A), compared to the modest connectivity of expressed genes in population 1 (Figure S2A), and the transcripts elevated in population 2 overlapped significantly with these modules (Figure 2B). Eighteen of the 27 common genes represented secretory or enterocyte-specific markers (Figure 2C) that were not mutually exclusive but appeared at similar levels in nearly every cell in population 2 and were virtually absent in the other cells (Figures 1B, 2C, and 2D). The simultaneous expression of different lineage programs is reminiscent of multilineage priming in blood progenitors (Hu et al., 1997; Miyamoto et al., 2002), and the lack of any instance of unilineage expression suggests that population 2 may represent IBPs. Single-cell latent variable modeling (scLVM) (Buettner et al., 2015) attributed only 12.2% of the variation to cell replication, and transcript profiles were very similar before and after correcting for cell-cycle effects (Figure S2B). Cell-cycle-related transcripts that were increased in IBPs included both positive and negative regulators of the cell cycle, and *Pcna*, *Mki67*, and targets of Wnt signaling were expressed at comparable levels (Figure S2C). Thus, the distinct mRNA profiles do not trivially reflect differential mitotic activity, and both populations seem to include cycling cells.

Superficially, the presence of numerous candidate IBPs among $Lgr5^+$ cells contrasts with recent evidence of population homogeneity by single-cell mRNA sequencing (mRNA-seq) (Grün et al., 2015). One explanation is that Grün et al. examined cells with higher GFP levels than we did. Thus, our population 1 might represent homogeneous GFP^{hi} ISCs, whereas population 2 may contain cells with modestly lower *Lgr5* mRNA (Figure 1C) and protein levels, i.e., cells leaving the ISC compartment. Another explanation is the low sensitivity of single-cell RNA sequencing (RNA-seq) for low-abundance transcripts, and, indeed, the method did not reliably capture genes that distinguish ISCs from IBPs in our qRT-PCR study. Although a few lineage markers—such as *Defa5*, *Muc2*, and *Ang4*—were detected in some cells, most markers were not (Figure S2D). Nevertheless, to exclude the possibility that our qRT-PCR signals are spurious, we performed bulk (ensemble) RNA-seq analysis on triplicate samples of $Lgr5^{hi}$ cells, sorted using the same parameters as in our single-cell analysis, and also queried bulk RNA data from $Lgr5^{hi}$ cells profiled on microarrays (Muñoz et al., 2012). Every lineage marker we detected in single cells was represented among the >11,000 genes identified in these ensemble studies (Figure S3A), compared to <4,000 genes in the single-cell mRNA-seq study (Grün et al., 2015).

In light of the multilineage profiles of putative IBPs, transcripts specific to enterocytes or secretory cells might persist in specified progenitors of the other type. This was, indeed, evident in ensemble analysis of the respective purified progenitors (Figure S3A); e.g., whereas high *Alpi* levels are restricted to enterocytes in vivo (Tetteh et al., 2016), levels ~10-fold lower than those found in bulk villus cells are equally abundant in both enterocyte and secretory progenitors. Conversely, we detected many secretory genes in enterocyte progenitors. Because this *Atoh1* null population categorically lacks secretory cells (Kim et al., 2014; Yang et al., 2001), genes from this lineage were likely activated in a preceding cell generation, IBP. Together, these observations imply that the earliest cells to leave the ISC compartment activate genes of both intestinal lineages, at levels that elude detection at the current resolution of single-cell RNA-seq.

To confirm our findings by independent methods, first, we used single-molecule mRNA ISH with branched DNA (bDNA) signal amplification (Player et al., 2001). Probes for the villus cell markers *Alpi*, *Chga*, *Neurog3*, and *Cck* gave the expected signals in most (enterocyte) or few (enteroendocrine) wild-type mouse villus cells, respectively, with weaker signals in crypt epithelium and virtually none in the lamina propria; conversely, *Lgr5* probes carrying a different chromophore stained only crypt base columnar cells (Figure S3B). We detected low levels of mature villus cell marker mRNAs in up to 24.7% of *Lgr5*-expressing cells (Figures 3A and 3B; Figure S3C), greatly exceeding the background of red signals and compatible with the different sensitivities of single-cell qRT-PCR and single-mRNA ISH to detect transcripts of low abundance. Second, we used $Atoh1^{Gfp}$ knockin mice (Rose et al., 2009) to examine protein levels of ATOH1, a transcription factor whose RNA is restricted to the pool of putative IBPs (Figure S3D). After verifying ATOH1/GFP expression in $lysozyme^+$ Paneth cells and occasional secretory progenitors positioned higher than crypt tier 5 (red arrow in Figure 3C), we restricted attention to $Lgr5^+$ cells in the crypt base (open arrows, Figure 3C; n = 454), which showed distinct populations of $ATOH1^+$ and $ATOH1^-$ nuclei (filled or open arrows, respectively, in Figures 3D, 3E, and S3E). As protein expression must trail new transcripts, the fraction of $ATOH1^+$ cells (23.7%; Figure 3F) is compatible with that detected by qRT-PCR (47.9%). mRNA ISH and ATOH1/GFP stains did not localize lineage-marker-expressing $Lgr5^+$ cells to high crypt tiers, which suggests that cell heterogeneity may originate—perhaps stochastically—among ISCs at the crypt bottom and that cells with this feature preferentially exit the ISC compartment. The cells we regard as IBPs may,

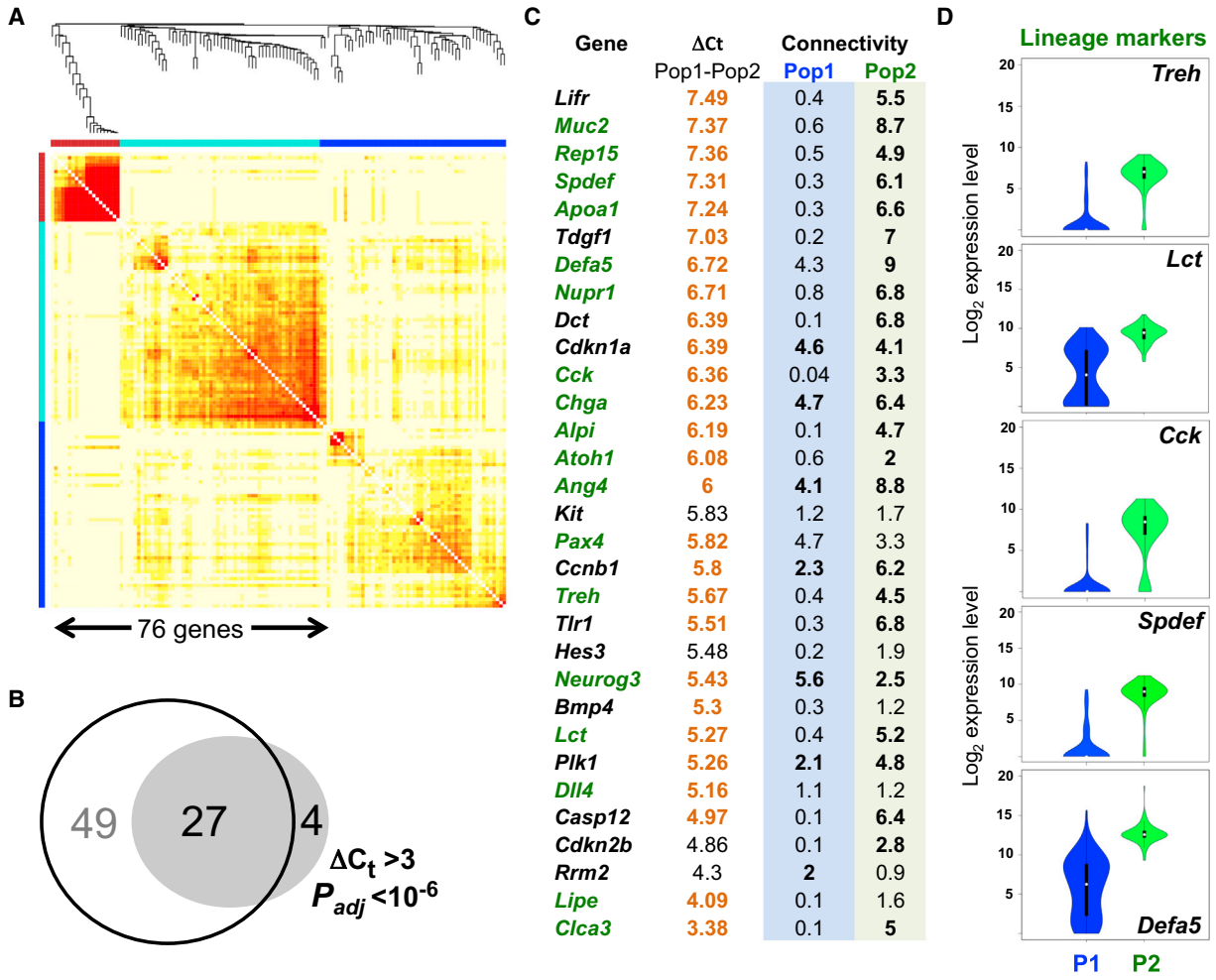


Figure 2. Candidate IBPs Show Multilineage Priming

(A) Results of WGCNA, showing network modules of genes that are strongly co-expressed across the cells in population 2. In contrast, population 1 showed limited connectivity (Figure S2A).

(B) Overlap of 31 genes showing differentially high expression in IBP with 76 genes showing high network connectivity (see Supplemental Experimental Procedures).

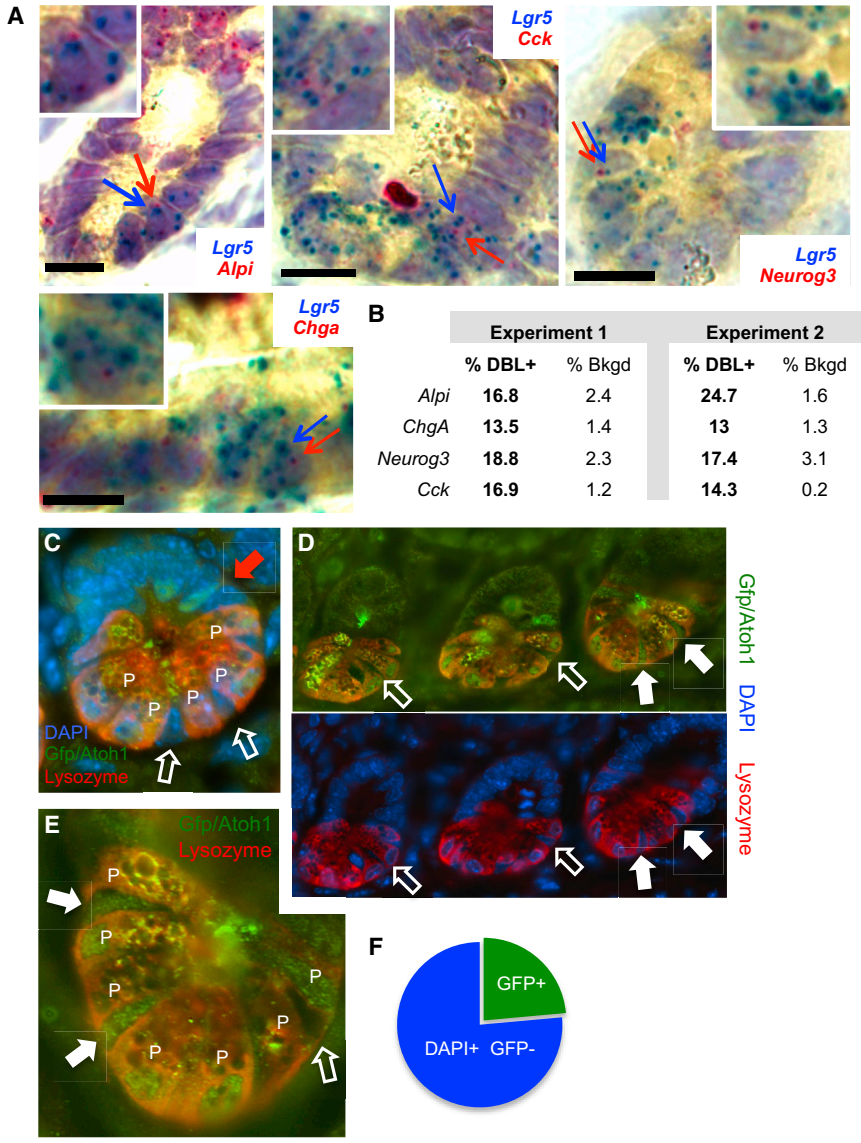
(C) ΔCt and connectivity values for the 31 genes that best distinguish population 2; lineage-specific markers are labeled green.

(D) Violin plots showing highly differential expression of markers of each major terminal intestinal cell type in all cells in populations 1 (P1; blue) and 2 (P2; green): *Lct* and *Treh* (enterocytes), *Cck* (endocrine), *Spdef* (goblet and other secretory cells), and *Defa5* (Paneth cells).

See also Supplemental Experimental Procedures and Figure S2.

however, correspond to a GFP^{low} population in vivo (Basak et al., 2014), and their property of multilineage priming is significant, regardless of the precise crypt location.

To examine further the relationship between populations 1 and 2, we considered that any transition among them is likely not abrupt; rather, transcripts from one cell state might decline, while those from the other begin to accumulate. The foregoing cluster analysis (Figure 1B), which is discrete, would fail to detect such a transition, but the non-branching structure of the t-SNE map (Figure 1D) permits the use of principal curves to infer cell trajectories (Hastie and Stuetzle, 1989). We derived such a principal curve, then divided all the cells into ten groups according to the inferred pseudo-time (Marco et al., 2014), and identified 28 cells at the boundary between the two major populations (Figure 4A). Average expression of each of the 183 genes in the ten groups of cells revealed 66 genes that discriminate between ISCs and IBPs (denoted by a box on the cluster dendrogram and heatmap in Figure 4B) and, as expected, include nearly every gene that had shown high ΔCt values (Figure 4C). Whereas ISCs and IBPs expressed uniformly higher levels of different subsets in this gene group, the 28 boundary cells varied in expression (Figure 4C), with declining average levels of stem cell markers, such as *Lgr5*, and concomitant increase of mature markers (Figure 4D). Average expression values were similar for different numbers of bins. For example, using eight bins instead of ten, the histogram of cell numbers identified 12 boundary cells, and mean expression over these 12 cells was highly correlated (R^2 = 0.95) with that in the 28 cells identified



	Experiment 1		Experiment 2	
	% DBL+	% Bkgd	% DBL+	% Bkgd
Alpi	**16.8**	2.4	**24.7**	1.6
ChgA	**13.5**	1.4	**13**	1.3
Neurog3	**18.8**	2.3	**17.4**	3.1
Cck	**16.9**	1.2	**14.3**	0.2

Figure 3. Expression of Lineage Markers in Lgr5$^+$ Crypt Base Cells In Vivo

(A) Representative images of single-molecule mRNA ISH for *Alpi*, *ChgA*, *Cck*, *Neurog3* (red), and *Lgr5* (blue), showing red and blue signals in the same crypt base cells. Cells with arrows pointing to co-expressed red and blue dots are magnified in the respective insets. Scale bars, 15 μm.

(B) Fraction of double-positive (DBL+, red and blue) cells and background (Bkgd) of extra-epithelial cells with red dots in intestines from four mice in two experiments.

(C) Immunostaining of *Atoh1*$^{Gfp/Gfp}$ crypts with lysozyme (red) and GFP (green) antibody (Ab) and DAPI nuclear stain (blue). GFP (ATOH1) was present in lysozyme$^+$ Paneth cells (P) at the crypt base and in occasional TA cells (red arrow); only slim columnar cells wedged between Paneth cells (white arrows) were assessed further.

(D) Absence (open arrows) or presence (filled arrows) of ATOH1 in a representative z-section of three consecutive crypts, with fluorescence channels separated for clarity.

(E) Magnified view of a single crypt, showing that ATOH1 signals in some putative IBP are similar to those in neighboring Paneth (P) cells. Open arrows, absence of ATOH1; filled arrows, presence of ATOH1.

(F) Fraction of ATOH1/GFP$^+$ cells among 454 columnar DAPI$^+$ nuclei in tiers 0–3 of *Atoh1*$^{Gfp/Gfp}$ mouse crypts.

See also Figure S3.

using ten bins. Gradual accumulation of terminal cell markers, as revealed in boundary cells, strongly suggests a cell transition from ISCs to putative IBPs.

The high census of IBPs suggests that they are distinct from the small, label-retaining fraction of Lgr5$^+$ cells. Transcripts recently assigned to the latter—*Nfatc3*, *Nfat5*, and *Cd82* (Buczacki et al., 2013)—were essentially similar in ISCs and IBPs (Figure S3F) and may increase only in Paneth-cell precursors. Notably, and in line with recent evidence for extreme plasticity in crypts (Kim et al., 2014; Tian et al., 2011; van Es et al., 2012), IBPs may be unstable cells that revert to ISCs as readily as they differentiate into absorptive or secretory cells. The latter event occurs as some cells use DLL1 or DLL4 to signal to Notch receptors on their neighbors (Pellegrinet et al., 2011; Stamataki et al., 2011). Because lateral inhibition requires equipotent cells to deliver or respond to Notch signals, increased expression of these ligands is one feature expected in IBPs. Indeed, average *Dll1* mRNA is higher in IBPs, and *Dll4* increases substantially in most of these cells (Figure 4E).

In summary, microfluidic qRT-PCR reveals a distinct cell population that seems to represent the earliest progeny of Lgr5$^+$ ISCs: putative IBPs with multilineage priming and modestly reduced Lgr5/GFP expression. Although multilineage priming was originally inferred from bulk cell populations (Hu et al., 1997; Miyamoto et al., 2002), recent studies suggest that single blood progenitors express genes exclusive to one lineage or another (Paul et al., 2015; Perié et al., 2015). In contrast, our analysis revealed no cell expressing genes specific to just one intestinal lineage (Figure 1), and enterocyte progenitors continue to express secretory genes (Figure S3A); these findings likely reflect features particular to lineage specification by lateral inhibition. Levels of certain TF mRNAs—*Atoh1*, *Spdef*, *Pax4*, and *Tbx3*—first rise in IBPs, where they may initiate the lineage-affiliated programs. Although equal expression of *Mki67* and *Pcna* in ISCs and IBPs supports the idea that all crypt cells other than Paneth cells and their precursors replicate, high mRNA levels of cell-cycle inhibitors *Cdkn1a*, *1b*, *2a*, and *2b* in IBPs (Figure S2C) suggest that they, or their immediate progeny, may replicate more slowly than ISCs or TA cells.

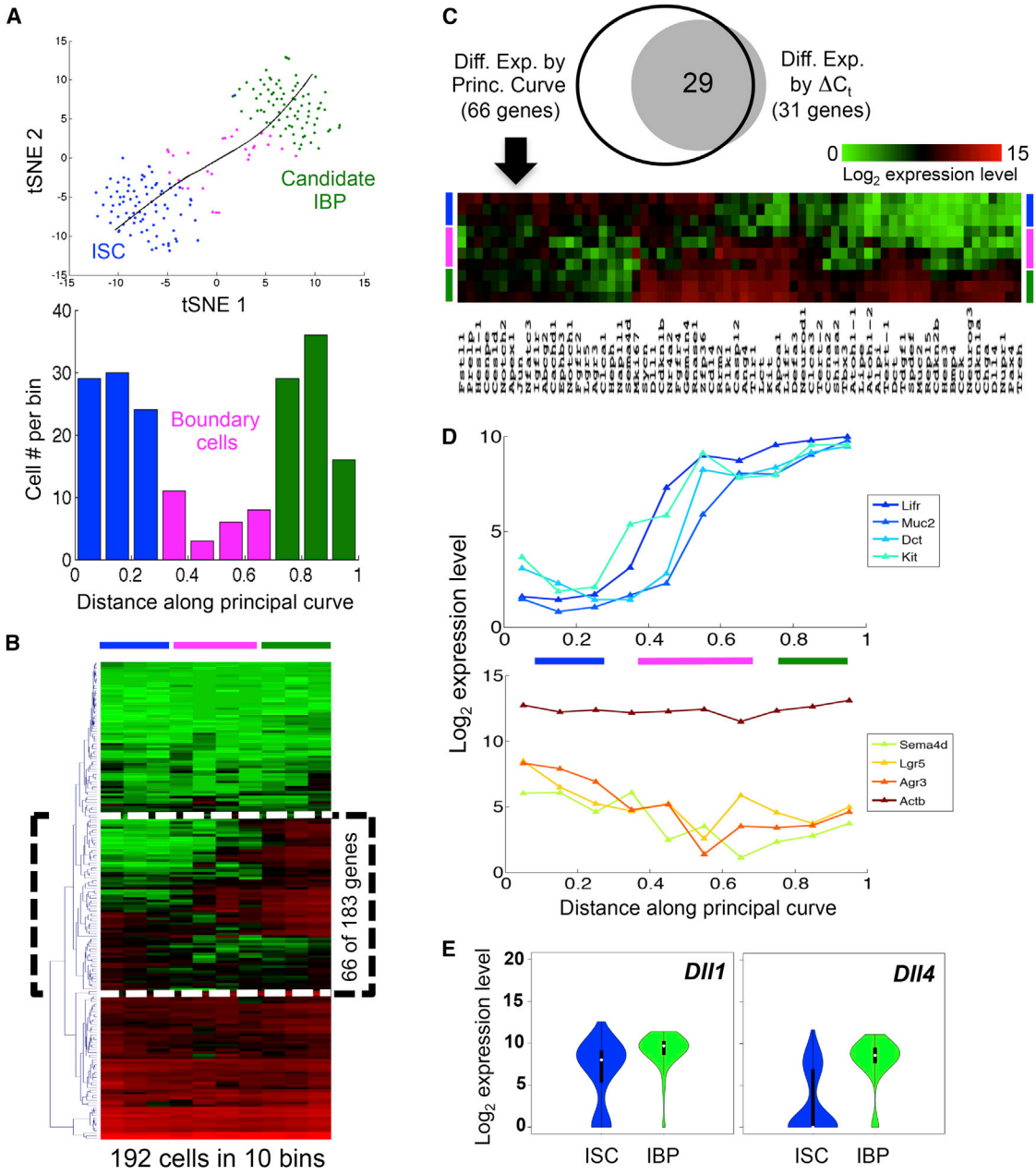


Figure 4. Evidence that Lgr5+ ISCs Transition into the IBP Population

(A) Principal curve analysis (black curve) projected on the t-SNE map from Figure 1D reveals the relationship of the two populations, based on the proximity of gene expression, as a non-branching curve. The 28 boundary cells—determined by partitioning of the principal curve into ten bins of equal distance—are now represented in pink. The graph indicates cell numbers in each bin; blue and green denote ISCs and IBPs, respectively.

(B) Heatmap of the global analysis (183 genes × 192 single cells; red indicates high expression, and green indicates low expression) partitioned in ten bins according to the aforementioned principal curve analysis. 66 transcripts denoted by a dotted box provide discrimination.

(C) The latter transcripts include nearly every gene that distinguished populations 1 and 2 by ΔCt (Figure 1B), and the dotted box in (B) is here expanded and rotated 90° to show the trajectory of expression in ISCs (blue), boundary cells (pink), and IBPs (green). Diff. Exp., different expression; Princ., principal.

(D) Average levels of representative IBP-enriched (*Lifr*, *Muc2*, *Dct*, and *Kit*), ISC-enriched (*Lgr5*, *Agr3*, and *Sema4d*), and *Actb* mRNAs in cell groups defined by distance along the principal curve.

(E) Violin plots for expression of Notch ligand genes *Dll1* and *Dll4* in all ISCs and IBPs.

Despite clear differences in gene activity, IBPs are unlikely to show different behaviors than ISCs by lineage tracing or in organoids, where even ISCs and specified progenitors are difficult to distinguish (Buczacki et al., 2013; Tetteh et al., 2016; van Es et al., 2012). Moreover, no *Cre* driver or surface marker is likely expressed exclusively in IBPs, i.e., not also in ISCs or specified

progenitors. Thus, our targeted single-cell analysis, reinforced by localization of transcripts in vivo, reveals features of a crucial and transient cell population that is likely difficult to isolate or to characterize by other means.

EXPERIMENTAL PROCEDURES

Isolation of Single Lgr5+ ISCs

Intestines harvested from *Lgr5*^{GFP} mice (Barker et al., 2007) were washed with PBS. Villi were scraped away using coverslips, and the crypt epithelium was collected by shaking in 5 mM EDTA for 1 hr at 4°C (Kim et al., 2014). Single cells were obtained on 2 separate days by digestion in 5× TrypLE (Invitrogen) for 1 hr at 37°C and verified by fluorescence microscopy. GFP^{hi} cells were sorted into individual wells in 96-well plates using a BD FACSAria II sorter (Becton Dickinson). Cells from one of the two isolations were also examined visually in microfluidic channels. Animals were handled according to protocols approved and monitored by the Animal Care and Use Committee of the Dana-Farber Cancer Institute.

Single-Cell Gene Expression Analysis by Microfluidic qRT-PCR

The pre-amplification solution in 96 wells included 5 μl of a master mix containing 2.5 μl CellsDirect reaction mix (Invitrogen), 0.5 μl primer pool (0.1 μM [Table S1], synthesized at Bioneer), 0.1 μl reverse transcriptase (RT)/Taq polymerase (Invitrogen), and 1.9 μl nuclease-free water. Lysed cells were treated with this mix at 50°C for 1 hr, followed by inactivation of RT, activation of Taq at 95°C for 3 min, and 20 cycles of sequence-specific cDNA amplification (15 s denaturation at 95°C, 15 min annealing and elongation at 60°C). Amplified single-cell cDNAs were first tested in control qRT-PCR reactions for *Actb*, and samples giving Ct values between 13 and 17 were selected for subsequent analysis with the full primer pools, Universal PCR Master Mix (Applied Biosystems), and EvaGreen Binding Dye (Biotium), using the 96 × 96 Dynamic Array on the BioMark System (Fluidigm). Table S2 lists the Ct values for each gene in each cell, calculated using BioMark Real-Time PCR Analysis software (Fluidigm).

Computational Analyses

mRNA levels were estimated by subtracting the Ct values from the background level of 28 (start of the tail of the distribution in the histogram of Ct values), which approximates $\log_2$ gene expression levels. We conducted k-means clustering in MATLAB using the squared Euclidean distance of normalized data (z scores). To determine the optimal k, we applied every value from 2 to 20, assessed the average Silhouette value (Kaufman and Rousseeuw, 1990) for each clustering result (Figure S1B), and selected k = 2, which gave the largest mean Silhouette value. Differentially expressed genes were identified using a two-sided Wilcoxon-Mann-Whitney rank-sum test implemented in the "coin" package in R. Differences between populations were determined by subtracting mean Ct values (equivalent to $\log_2$ expression levels). The p values were adjusted for multiple testing (Benjamini and Hochberg, 1995). Violin plots were generated in R using the package "vioplot." For t-SNE analysis (van der Maaten and Hinton, 2008), we used the MATLAB toolbox for dimensionality reduction (http://homepage.tudelft.nl/19j49/t-SNE.html). The pseudotime of individual cells was estimated as previously described (Marco et al., 2014), fitting a principal curve (Hastie and Stuetzle, 1989) to the single-cell expression data. We used the R package "princurve," with the options "smoother = lowess" and "maxit = 200." Heatmaps (Figure 2) were prepared with the MultiExperiment Viewer (http://www.tm4.org/mev.html), using the Euclidean distance and average linkage as parameters for unsupervised hierarchical clustering of genes. Latent variable modeling and analysis of co-expression gene networks are described in the Supplemental Experimental Procedures.

Analysis of Public mRNA-Seq Data

Processed mRNA-seq data on 192 isolated Lgr5+ mouse intestinal cells (Grün et al., 2015) were obtained from GEO: GSE62270 (accession file GSE62270_data_counts_Lgr5SC.txt.gz). Violin plots for genes relevant to our study were generated using the Vioplot2 function in R. The accession number for the ensemble RNA-seq is GEO: GSE71713.

Single-mRNA ISH with bDNA Amplification

Intestines from C57BL/6J mice were fixed overnight in 4% paraformaldehyde, embedded in paraffin, and cut in 5-μm sections. ISH was performed twice on two intestines each, using Quantigene ViewRNA probes (Affymetrix) for two-color ISH, as described in the Supplemental Experimental Procedures. Between 320 and 460 Lgr5+ crypt base cells were counted in at least 50 crypts from each mouse (n = 4). Cells were scored as double positive (DBL+) when at least one dot for a mature-cell marker mRNA (red) was present in a cell expressing *Lgr5* mRNA (blue dots). Background signals were estimated from counts of red dots in 370 to 440 nucleated sub-epithelial cells for each mature-cell marker in each sample.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, three figures, and two tables and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2016.07.056.

AUTHOR CONTRIBUTIONS

T.-H.K. and R.A.S. conceived the study; T.-H.K., G.G., M.S., N.D., and U.J. performed experiments; A.S. performed computational analyses; A.S., T.-H.K., M.S., A.C., L.J., G.-C.Y., and R.A.S. analyzed data; S.H.O. and M.N.R. supervised portions of the study; G.-C.Y. supervised computational analyses; and R.A.S. provided overall supervision. R.A.S., A.S., T.-H.K., and G.-C.Y. drafted the manuscript, with input from all authors.

ACKNOWLEDGMENTS

This work was supported by grants R01DK081113 (including a supplement from the Office of the Director [NIH Common Fund]), U01DK103152 (the Intestinal Stem Cell Consortium of the NIDDK and NIAID), K99DK095983 (to T-H.K.), F32DK103453 (to U.J.), and P50CA127003; an American-Italian Cancer Foundation fellowship (to A.C.); and funds from the Harvard Stem Cell Institute (to G-C.Y. and S.H.O.), Affymetrix (to N.D. and M.N.R.), and the Lind family (to R.A.S.). We thank L. Deary and D. Breault for help with microscopy.

Received: February 12, 2016
Revised: April 18, 2016
Accepted: July 20, 2016
Published: August 11, 2016

REFERENCES

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

Barker, N., van Oudenaarden, A., and Clevers, H. (2012). Identifying the stem cell of the intestinal crypt: strategies and pitfalls. Cell Stem Cell *11*, 452–460.

Basak, O., van de Born, M., Korving, J., Beumer, J., van der Elst, S., van Es, J.H., and Clevers, H. (2014). Mapping early fate determination in Lgr5+ crypt stem cells using a novel Ki67-RFP allele. EMBO J. *33*, 2057–2068.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. Roy. Statist. Soc. B *57*, 289–300.

Buczacki, S.J., Zecchini, H.I., Nicholson, A.M., Russell, R., Vermeulen, L., Kemp, R., and Winton, D.J. (2013). Intestinal label-retaining cells are secretory precursors expressing Lgr5. Nature *495*, 65–69.

Buettner, F., Natarajan, K.N., Casale, F.P., Proserpio, V., Scialdone, A., Theis, F.J., Teichmann, S.A., Marioni, J.C., and Stegle, O. (2015). Computational analysis of cell-to-cell heterogeneity in single-cell RNA-sequencing data reveals hidden subpopulations of cells. Nat. Biotechnol. *33*, 155–160.

Please cite this article in press as: Kim et al., Single-Cell Transcript Profiles Reveal Multilineage Priming in Early Progenitors Derived from Lgr5$^+$ Intestinal Stem Cells, Cell Reports (2016), http://dx.doi.org/10.1016/j.celrep.2016.07.056

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

Hastie, T., and Stuetzle, W. (1989). Principal curves. J. Am. Stat. Assoc. *84*, 502–516.

Hu, M., Krause, D., Greaves, M., Sharkis, S., Dexter, M., Heyworth, C., and Enver, T. (1997). Multilineage gene expression precedes commitment in the hemopoietic system. Genes Dev. *11*, 774–785.

Kaufman, L., and Rousseeuw, P.J. (1990). Finding Groups in Data: An Introduction to Cluster Analysis (John Wiley & Sons).

Kim, T.H., Li, F., Ferreiro-Neira, I., Ho, L.L., Luyten, A., Nalapareddy, K., Long, H., Verzi, M., and Shivdasani, R.A. (2014). Broadly permissive intestinal chromatin underlies lateral inhibition and cell plasticity. Nature *506*, 511–515.

Kozar, S., Morrissey, E., Nicholson, A.M., van der Heijden, M., Zecchini, H.I., Kemp, R., Tavaré, S., Vermeulen, L., and Winton, D.J. (2013). Continuous clonal labeling reveals small numbers of functional stem cells in intestinal crypts and adenomas. Cell Stem Cell *13*, 626–633.

Marco, E., Karp, R.L., Guo, G., Robson, P., Hart, A.H., Trippa, L., and Yuan, G.C. (2014). Bifurcation analysis of single-cell gene expression data reveals epigenetic landscape. Proc. Natl. Acad. Sci. USA *111*, E5643–E5650.

Miyamoto, T., Iwasaki, H., Reizis, B., Ye, M., Graf, T., Weissman, I.L., and Akashi, K. (2002). Myeloid or lymphoid promiscuity as a critical step in hematopoietic lineage commitment. Dev. Cell *3*, 137–147.

Muñoz, J., Stange, D.E., Schepers, A.G., van de Wetering, M., Koo, B.K., Itzkovitz, S., Volckmann, R., Kung, K.S., Koster, J., Radulescu, S., et al. (2012). The Lgr5 intestinal stem cell signature: robust expression of proposed quiescent '+4' cell markers. EMBO J. *31*, 3079–3091.

Paul, F., Arkin, Y., Giladi, A., Jaitin, D.A., Kenigsberg, E., Keren-Shaul, H., Winter, D., Lara-Astiaso, D., Gury, M., Weiner, A., et al. (2015). Transcriptional heterogeneity and lineage commitment in myeloid progenitors. Cell *163*, 1663–1677.

Pellegrinet, L., Rodilla, V., Liu, Z., Chen, S., Koch, U., Espinosa, L., Kaestner, K.H., Kopan, R., Lewis, J., and Radtke, F. (2011). Dll1- and dll4-mediated notch signaling are required for homeostasis of intestinal stem cells. Gastroenterology *140*, 1230–1240.e1, 7.

Perié, L., Duffy, K.R., Kok, L., de Boer, R.J., and Schumacher, T.N. (2015). The branching point in erythro-myeloid differentiation. Cell *163*, 1655–1662.

Player, A.N., Shen, L.P., Kenny, D., Antao, V.P., and Kolberg, J.A. (2001). Single-copy gene detection using branched DNA (bDNA) in situ hybridization. J. Histochem. Cytochem. *49*, 603–612.

Ritsma, L., Ellenbroek, S.I., Zomer, A., Snippert, H.J., de Sauvage, F.J., Simons, B.D., Clevers, H., and van Rheenen, J. (2014). Intestinal crypt homeostasis revealed at single-stem-cell level by in vivo live imaging. Nature *507*, 362–365.

Rose, M.F., Ren, J., Ahmad, K.A., Chao, H.T., Klisch, T.J., Flora, A., Greer, J.J., and Zoghbi, H.Y. (2009). Math1 is essential for the development of hindbrain neurons critical for perinatal breathing. Neuron *64*, 341–354.

Stamataki, D., Holder, M., Hodgetts, C., Jeffery, R., Nye, E., Spencer-Dene, B., Winton, D.J., and Lewis, J. (2011). Delta1 expression, cell cycle exit, and commitment to a specific secretory fate coincide within a few hours in the mouse intestinal stem cell system. PLoS ONE *6*, e24484.

Tetteh, P.W., Basak, O., Farin, H.F., Wiebrands, K., Kretzschmar, K., Begthel, H., van den Born, M., Korving, J., de Sauvage, F., van Es, J.H., et al. (2016). Replacement of lost Lgr5-positive stem cells through plasticity of their enterocyte-lineage daughters. Cell Stem Cell *18*, 203–213.

Tian, H., Biehs, B., Warming, S., Leong, K.G., Rangell, L., Klein, O.D., and de Sauvage, F.J. (2011). A reserve stem cell population in small intestine renders Lgr5-positive cells dispensable. Nature *478*, 255–259.

van der Maaten, L.J.P., and Hinton, G.E. (2008). Visualizing high-dimensional data using t-SNE. J. Mach. Learn. Res. *9*, 2579–2605.

van Es, J.H., Sato, T., van de Wetering, M., Lyubimova, A., Nee, A.N., Gregorieff, A., Sasaki, N., Zeinstra, L., van den Born, M., Korving, J., et al. (2012). Dll1+ secretory progenitor cells revert to stem cells upon crypt damage. Nat. Cell Biol. *14*, 1099–1104.

Yang, Q., Bermingham, N.A., Finegold, M.J., and Zoghbi, H.Y. (2001). Requirement of Math1 for secretory cell lineage commitment in the mouse intestine. Science *294*, 2155–2158.

Zhang, B., and Horvath, S. (2005). A general framework for weighted gene co-expression network analysis. Stat. Appl. Genet. Mol. Biol. *4*, e17.

Cell Reports, Volume *16*

Supplemental Information

Single-Cell Transcript Profiles Reveal Multilineage Priming in Early Progenitors Derived from Lgr5^{+} Intestinal Stem Cells

Tae-Hee Kim, Assieh Saadatpour, Guoji Guo, Madhurima Saxena, Alessia Cavazza, Niyati Desai, Unmesh Jadhav, Lan Jiang, Miguel N. Rivera, Stuart H. Orkin, Guo-Cheng Yuan, and Ramesh A. Shivdasani

Single-cell transcript profiles reveal multilineage priming in early progenitors derived from Lgr5^{+} intestinal stem cells

Tae-Hee Kim, Assieh Saadatpour, Guoji Guo, Madhurima Saxena,
Alessia Cavazza, Niyati Desai, Unmesh Jadhav, Lan Jiang,
Miguel N. Rivera, Stuart H. Orkin, Guo-Cheng Yuan, Ramesh A. Shivdasani

SUPPLEMENTAL INFORMATION

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

Estimation of mRNA copy number in single-cell analysis. cDNA was synthesized from total RNA extracted from bulk Lgr5$^+$ ISC, using SuperScript III Reverse Transcriptase system (Thermo Fisher Scientific), and a serial dilution of this cDNA (1 ng, 100 pg, 10 pg, 1 pg) used to determine PCR efficiency for specific primer pairs: *Gapdh* and *Hprt* as references and *Alpi*, *Spdef*, *Lgr5*, *Olfm4* and *Casp12* for interrogation. We then assumed that controls lacking template cDNA contain zero transcripts at a background C_t value x and that one lower C_t value (x-1) signifies 1 mRNA copy. Based on the exponential increase in transcripts expected for every 3.32 unit drop in C_t, values of (x-1)-3.32 would represent 10 copies of a transcript and so on; the values obtained in this manner from bulk Lgr5$^+$ ISC confirmed efficient qPCR in each case and were considered the "calibrator". To estimate mRNA copy numbers from single-cell qPCR data averaged from ISC and IBP, we applied background-subtracted C_t values (Suppl. Table 1) to the formula $2^{(\text{target gene Ct} - \text{reference gene Ct in calibrator}) - (\text{target gene Ct} - \text{reference gene Ct in sample})}$ and expressed the results with respect to the average values for *Gapdh* and *Hprt* transcript copies.

Single-mRNA in situ hybridization (ISH) with branched DNA amplification. Tissue sections were deparaffinized, boiled for 10 minutes, and digested with protease for 20 min. After hybridization with Quantigene ViewRNA probes for 2 h at 40°C, branched DNA was amplified following the manufacturer's protocol (Affymetrix). Signals were detected with Fast Red and Fast Blue substrates, and the tissue was counterstained with hematoxylin. Representative images were captured on a Nikon Eclipse E800 microscope using SPOT 5.0 software and processed in Photoshop CS5 (Adobe). Between 320 and 460 Lgr5$^+$ crypt base cells were counted in at least 50 crypts from each mouse (N=4). Cells were scored as DBL+ when at least one dot for a mature-cell marker mRNA (red) was present in a cell expressing *Lgr5* mRNA (blue dots). DBL+ cells are represented as the fraction of all Lgr5$^+$ cells examined for each marker. To estimate background signals, we counted the red dots in 370 to 440 nucleated sub-epithelial cells for each mature-cell marker in each sample.

Immunostaining. Frozen 5-µm sections of intestines from *Atoh1*^{Gfp} mice (Rose et al., 2009) (Jackson Laboratories, B6.129S-Atoh1 tm4.1Hzo/J) were stained simultaneously with chicken

GFP (Abcam ab13970) and rabbit lysozyme (Dako A0099) Ab (1:1000 each), followed by goat anti-chicken and anti-rabbit Alexafluors AF488 and AF568 (Life Technologies, 1:500), and counterstained with DAPI (Vector Labs). Images were captured on a Nikon Eclipse 90i epifluorescence microscope using NIS-Elements Advanced Research 3.2 software (Nikon) and processed using Photoshop CS5 (Adobe). More than 450 GFP^+ and GFP^- crypt base columnar cells within the Paneth-cell zone were examined in 230 crypts.

Ensemble mRNA-seq. $Lgr5^+$ ISC were isolated by GFP flow cytometry using the same gates as for single-cell analysis (Suppl. Fig. S1A). Bulk populations of secretory (Sec, from intestinal crypts of mice treated with dibenzazepine) and enterocyte (Ent, from crypts of *Atoh1*$^{-/-}$ small intestine) progenitors and mature villus cells were isolated as described previously (Kim et al., 2014). Total RNA was extracted using Trizol reagent (Life Technologies). RNA quality (RNA Integrity Number >8) was verified using Bioanalyzer 2100 (Agilent Technologies) and 1 μg was used to prepare libraries with TruSeq RNA Sample Preparation Kit v2 (Illumina). Single-end sequences were obtained on Illumina HiSeq 2000 (50-bp reads) or NextSeq 500 (75-bp reads) instruments and aligned to the mouse genome (Mm9, NCBI build 37) using TopHat version 2.0.6. mRNA levels of genes in triplicate ($Lgr5^+$ ISC and villus cells) or duplicate (Sec and Ent progenitors) were calculated as reads per kb of transcript per 1M mapped reads (RPKM) using Cufflinks version 2.0.2 (Trapnell et al., 2012).

Co-expression gene network analysis. Co-expression gene networks were analyzed using the WGCNA package implemented in R (Langfelder and Horvath, 2008; Zhang and Horvath, 2005). Anti-$\log_2$ transformation was applied to convert $\log_2$ expression levels to a normal scale. Unsigned weighted networks were then constructed using the power adjacency function $a_{ij}=|\mathrm{cor}(x_i,x_j)|^{\beta}$, which defines the connection strength between genes x_i and x_j using a soft power threshold β. For both ISC and IBP, β=4 was used to approximate a scale-free topology to the network with the highest fitting index (R^2 ~0.8 is the correlation between $\log_{10}(p(k))$ and $\log_{10}(k)$, wherein $p(k)$ denotes the frequency distribution of the network connectivity k). Modules were detected with hierarchical clustering, using the average linkage method, a dissimilarity measure based on the topological overlap matrix (TOM) and with a dynamic tree-cut algorithm. The TOM dissimilarity matrix is calculated according to $d_{ij}^{\omega}=1-\omega_{ij}$, where

$\omega_{ij} = \frac{(\sum_u a_{iu} a_{uj}) + a_{ij}}{\min\{k_i, k_j\} + 1 - a_{ij}}$ is the topological overlap, and $k_i = \sum_u a_{iu}$ is the node connectivity (Zhang and Horvath, 2005). Any two genes have high topological overlap if they connect to roughly the same group of genes in the network. All 185 genes were considered in constructing networks, but only genes assigned to any co-expressed module are shown. The diagonal of the dissimilarity matrix was set to NA and the matrix was raised to the 7th power to reveal the module structure. Connectivity values for each node were calculated as the sum of adjacencies to the other nodes in the network.

Single-cell latent variable modeling (scLVM). We adapted the scLVM approach originally developed for single-cell RNA-seq data (Buettner et al., 2015) to single-cell qPCR data for the cell cycle genes listed in Suppl. Table S2. Technical noise was set to 0 because this variable is much lower in single-cell qPCR than in single-cell RNA-seq data.

SUPPLEMENTAL REFERENCES

Buettner, F., Natarajan, K.N., Casale, F.P., Proserpio, V., Scialdone, A., Theis, F.J., Teichmann, S.A., Marioni, J.C., and Stegle, O. (2015). Computational analysis of cell-to-cell heterogeneity in single-cell RNA-sequencing data reveals hidden subpopulations of cells. Nat Biotechnol *33*, 155-160.

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

Jadhav, U., Nalapareddy, K., Saxena, M., O'Neill, N.K., Pinello, L., Yuan, G-C, Orkin, S.H., Shivdasani, R.A. (2016). Acquired tissue-specific promoter bivalency underlies PRC2 necessity in adult cells. Cell *165*, 1389-1400.

Kim, T.H., Li, F., Ferreiro-Neira, I., Ho, L.L., Luyten, A., Nalapareddy, K., Long, H., Verzi, M., and Shivdasani, R.A. (2014). Broadly permissive intestinal chromatin underlies lateral inhibition and cell plasticity. Nature *506*, 511-515.

Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. BMC bioinformatics *9*, 559.

Munoz, J., Stange, D.E., Schepers, A.G., van de Wetering, M., Koo, B.K., Itzkovitz, S., Volckmann, R., Kung, K.S., Koster, J., Radulescu, S.*, et al.* (2012). The Lgr5 intestinal stem cell signature: robust expression of proposed quiescent '+4' cell markers. EMBO J *31*, 3079-3091.

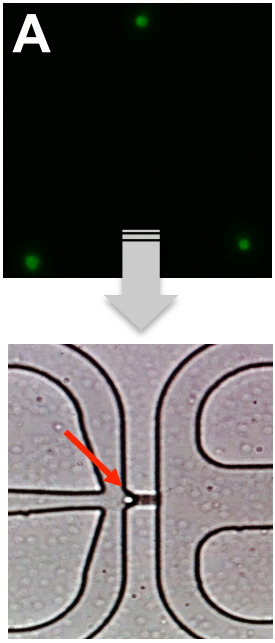
Rose, M.F., Ren, J., Ahmad, K.A., Chao, H.T., Klisch, T.J., Flora, A., Greer, J.J., and Zoghbi, H.Y. (2009). Math1 is essential for the development of hindbrain neurons critical for perinatal breathing. Neuron *64*, 341-354.

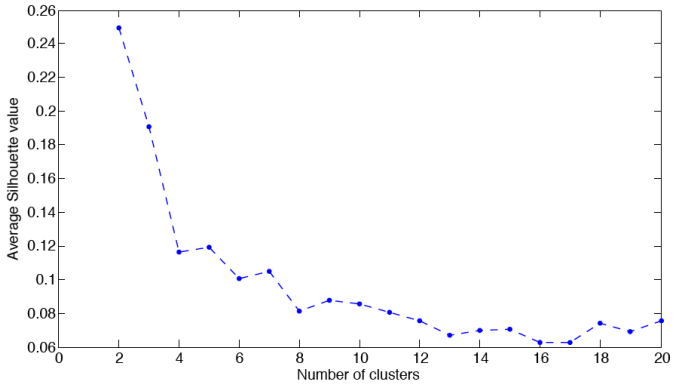
Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., Pimentel, H., Salzberg, S.L., Rinn, J.L., and Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nature protocols *7*, 562-578.

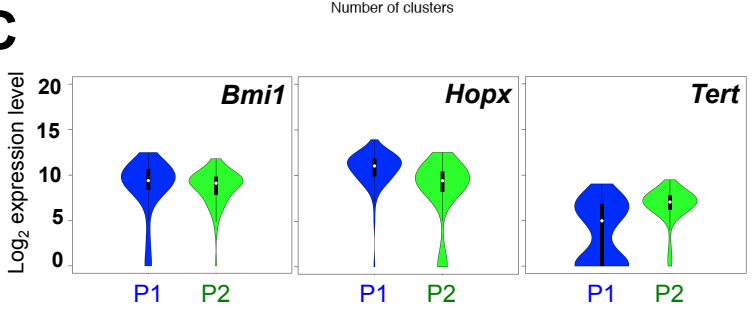
Zhang, B., and Horvath, S. (2005). A general framework for weighted gene co-expression network analysis. Stat Appl Genet Mol Biol *4*, Article 17.

Supplemental Figure S1 (related to Figure 1)**. Lgr5^{+} intestinal crypt cell subpopulations identified by targeted mRNA profiling. (A)** Cells with high LGR5/GFP signal were isolated by fluorescence-activated cell sorting (FACS) and verified as singlets by fluorescence microscopy as well as visualization in microfluidic channels. **(B)** Average Silhouette values for *k*-values between 2 and 20 in *k*-means clustering, indicating that *k*=2 is optimal. **(C)** Violin plots showing similar expression of markers historically attributed to quiescent +4 ISC (*Bmi1*, *Hopx*, *Tert*) in all 192 cells belonging to the two populations, P1 (blue) and P2 (green), identified in this study. Violin width (x-axis) represents the fraction of cells with the corresponding Log_2 expression level (y-axis). **(D)** mRNA copy number estimates for selected genes in ISC and putative IBP, expressed in relation to *Hprt* transcript numbers.

Kim et al., Suppl. Fig. S1





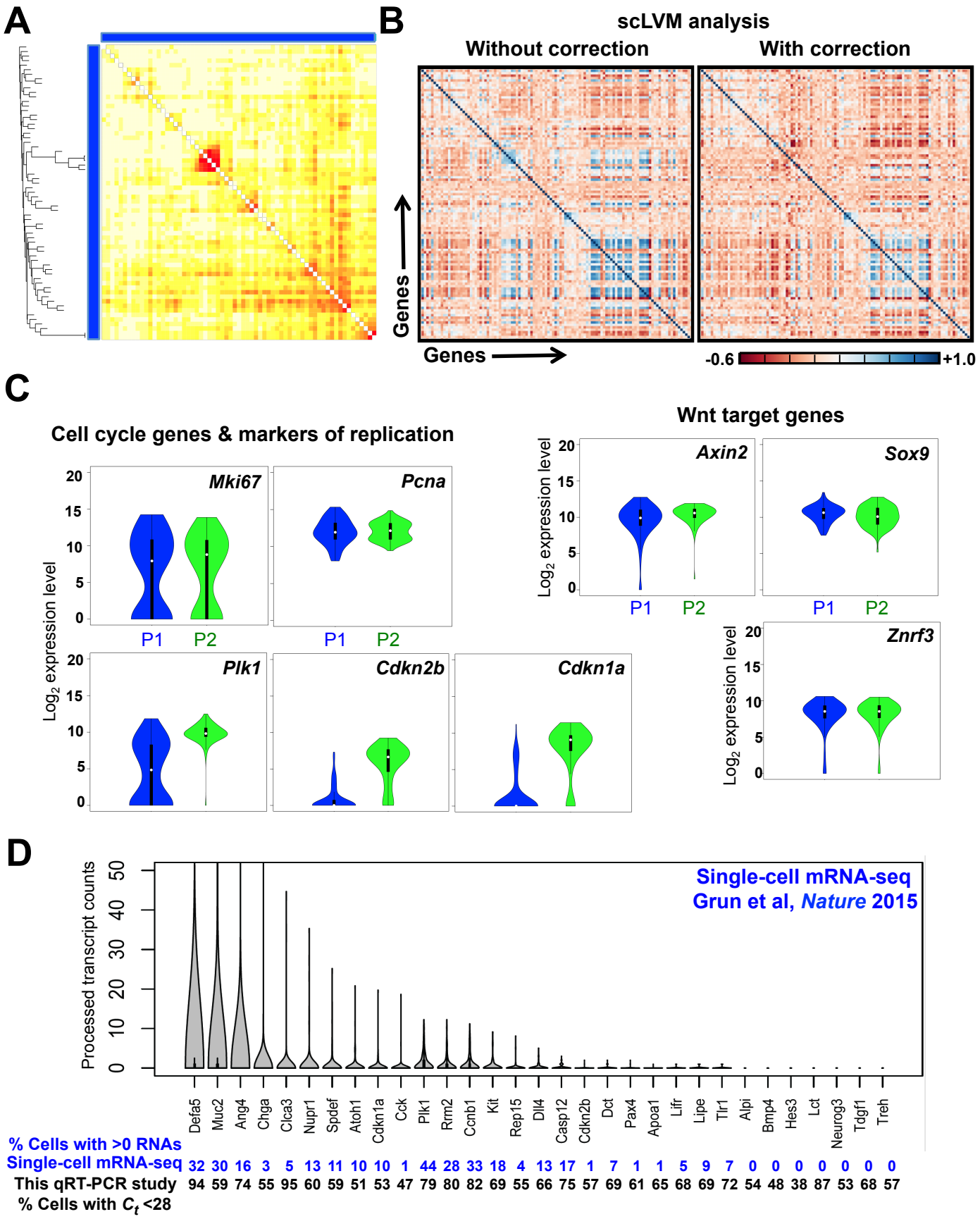


D

Gene	PCR efficiency	% *Hprt* copies (ave.) in ISC	% *Hprt* copies (ave.) in IBP
Lgr5	0.90	0.98	0.20
Olfm4	0.99	6014	1327
Casp12	0.86	0.10	3.26
Alpi	0.88	0.07	5.11
Spdef	0.97	0.05	7.94

Supplemental Figure S2 (related to Figure 2)**. Analyses of single-cell qRT-PCR data. (A)** Weighted Gene Co-expression Network Analysis (WGCNA), showing limited connectivity – concordant expression across many cells – among genes in Population 1. The corresponding map for Population 2 (Fig. 2A) shows high connectivity in two discrete gene modules. **(B)** Single-cell latent variable modeling (scLVM) reveals fundamentally similar gene modules before and after correction for cell-cycle effects. **(C)** Violin plots comparing expression of cell cycle-associated genes (*Plk1*, *Cdkn2b* and *Cdkn1a*), markers of cell replication (*Mki67* and *Pcna*) and targets of intestinal Wnt signaling (*Axin2*, *Sox9* and *Znrf3*) in all 192 cells from Populations (P) 1 (blue) and 2 (green). **(D)** Data from single-cell mRNA-seq analysis of 192 fresh $Lgr5^{+}$ mouse crypt ISC (Grun et al., 2015) on the 31 genes that robustly distinguished Population 2 (putative IBP) from Population 1 in our qRT-PCR study. Most of these genes were undetectable by RNA-seq, where the vast majority of cells showed 0 transcripts. The same genes were, however, readily detected in bulk (ensemble) RNA-seq analysis (Suppl. Fig. S3A).

Kim et al., Suppl. Fig. S2



Supplemental Figure S3 (related to Figure 3)**. Expression of lineage markers in bulk crypt cells and in Lgr5^{+} crypt base cells *in vivo*. (A)** Ensemble (bulk) mRNA-seq analysis of

- Lgr5^{+} ISC, isolated by flow cytometry as shown in Fig. S1A,
- Enterocyte (Ent-) and secretory (Sec-) progenitors (Pro) isolated as described before (Kim et al., 2014), and
- Unfractionated intestinal villus cells (Jadhav et al., 2016).

Data are shown as $\log_2$ of the number of sequence tags per kb of coding region per 1M reads (RPKM), together with published data from microarray profiles of Lgr5^{hi} ISC (Munoz et al., 2012). Every lineage marker detected by qRT-PCR in single cells was identified in the bulk populations. **(B)** mRNA ISH data for *Alpi*, *ChgA* (both red) and *Lgr5* (blue), showing red dots in nearly all (*Alpi*, enterocytes) or rare (*ChgA*, endocrine, black arrows) villus cells and blue signals restricted to the crypt base. High-magnification images are shown at the bottom right. **(C)** Additional high-magnification views of single-molecule mRNA ISH signals for lineage-specific genes *Alpi* and *Cck* (red dots and arrows), showing co-expression with *Lgr5* (blue dots and arrows). Scale bars (B-C), 15 μm. **(D)** Violin plot of differential *Atoh1* mRNA expression in all single ISC (blue) and putative IBP (green). **(E)** Additional examples of absence (open arrows) or presence (filled arrows) of ATOH1/GFP in cells at the crypt base in a representative z-section. Lysozyme marks Paneth cells and GFP signal is subtracted from the lower image to display the cells clearly. **(F)** Similar mRNA levels of label-retaining cells (LRC, Paneth-endocrine cell precursors) in single ISC and putative IBP.

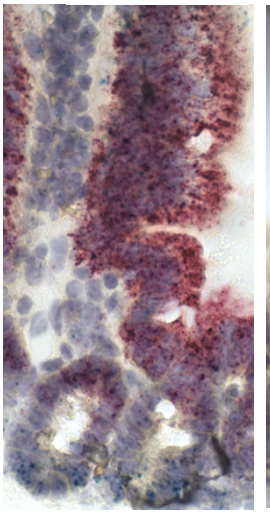
A

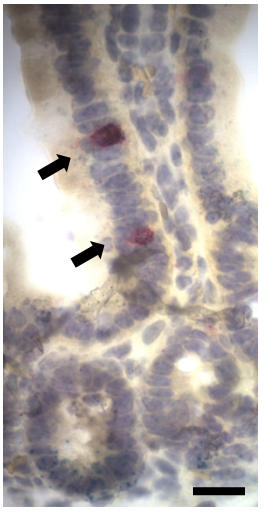
Cell-specific	marker	Bulk ISC by microarray analysis*	Bulk RNA-seq, Log2 (RPKM+1)**			
			ISC	Ent-Pro	Sec-Pro	Villus
Goblet	*Muc2*	8.47	6.94	1.59	7.35	6.24
	Gcnt3	3.5	3.48	4.45	4.55	4.35
Endocrine	*Chga*	5.31	5.78	2.69	5.74	4.8
	Pax4	3.07	1.74	0	1.75	0.2
Paneth	*Defa5*	11.5	11.38	0.27	8.76	5.13
	Ang4	9.28	11.2	2.84	9.66	9.8
	Nupr1	5.14	6.55	1.88	4.79	4.03
Tuft	*Gfi1b*	2.48	0.27	2.99	0.88	0.56
Enterocyte	*Alpi*	6.06	4.31	6.21	6.44	9.93
	Apoa1	6.77	6.68	10.33	10.58	12.95
	Lct	4.7	2.82	4.61	5.82	8.35
	Lipe	2.65	2.37	3.26	3.9	5.52
	Treh	2.93	2.49	4.2	5.06	8.21
Controls	*Actb*	11.54	9.16	10.22	10.6	10.14
	Hprt	9.53	6.26	6.37	6.67	5.51
	Vil1	10.01	7.18	8.54	8.85	10

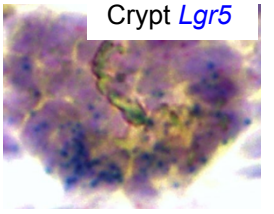
*From Munoz et al., *EMBO J* 2012; 31:3079-3091

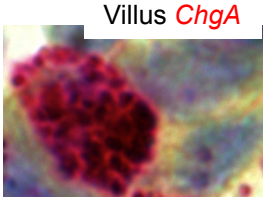
**From Jadhav et al., *Cell* 2016; 165:1389-1400

B

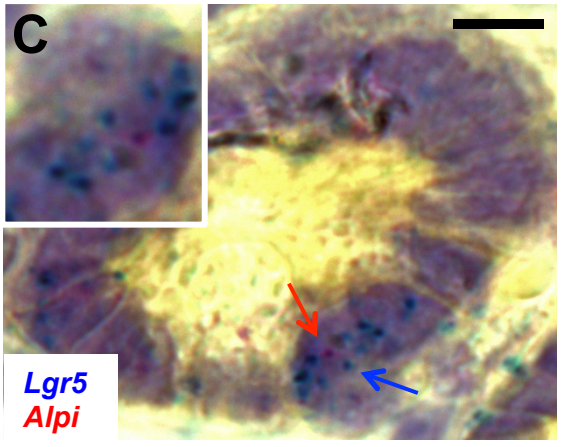


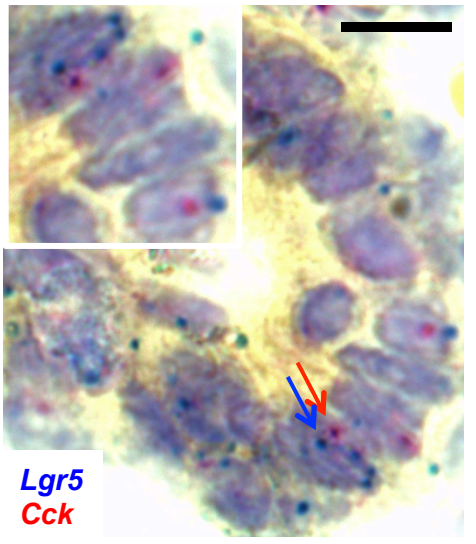




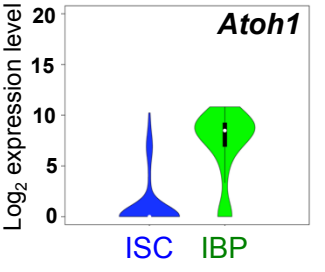


C

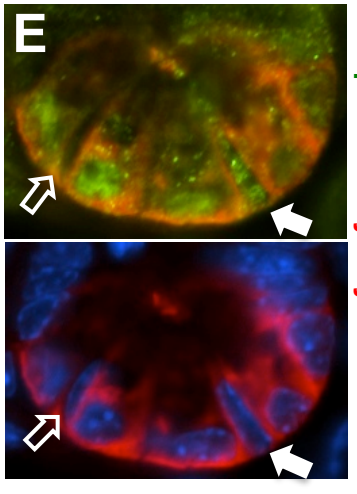




D



E



F