# Mex3a marks a slowly dividing subpopulation of Lgr5+ intestinal stem cells

Francisco M. Barriga[1], Elisa Montagni[1], Miyeko Mana[4], Maria Mendez-Lago[5,6], Xavier Hernando-Momblona[1], Marta Sevillano[1], Amy Guillaumet-Adkins[5,6],Gustavo Rodriguez-Esteban[5,6], Simon J. A. Buczacki[2], Marta Gut[5,6], Holger Heyn[5,6], Douglas J. Winton[2], Omer H. Yilmaz[4], Camille Stephan-Otto Attolini[1], Ivo Gut[5,6] and Eduard Batlle[1,3]

1. Institute for Research in Biomedicine (IRB Barcelona), The Barcelona Institute of Science and Technology. Baldiri Reixac 10, 08028 Barcelona, Spain.
2. Cancer Research UK Cambridge Research Institute, Li Ka Shing Centre, Robinson Way, Cambridge CB2 0RE, UK.
3. Institució Catalana de Recerca i Estudis Avançats (ICREA), Pg. Lluís Companys 23, 08010 Barcelona, Spain
4. The David H. Koch Institute for Integrative Cancer Research at MIT, Cambridge, MA 02139, Department of Biology, MIT, Cambridge, MA 02139 USA
5. CNAG-CRG, Centre Nacional d´Anàlisi Genòmica (CNAG) - Centre for Genomic Regulation (CRG), Barcelona Institute of Science and Technology (BIST), Baldiri i Reixac 4, 08028 Barcelona, Spain
6. Universitat Pompeu Fabra (UPF), Barcelona, Spain

Correspondence should be addressed to:

Eduard Batlle (eduard.batlle@irbbarcelona.org)

Lead contact:

Eduard Batlle (eduard.batlle@irbbarcelona.org)

## HIGHLIGHTS

- Mex3a is expressed in a subset of Lgr5+ cells that proliferate slowly
- Mex3a+ cells are multipotent and can generate all intestinal lineages
- Mex3a+ cells can convert to fast-dividing ISCs to maintain homeostasis
- Mex3a+ cells regenerate the epithelium after chemotherapeutic insult

## ETOC:

Lgr5+ intestinal stem cells are considered to be a homogenous and rapidly proliferating population. Barriga et al. show that the RNA binding protein Mex3a defines a subset of slowly proliferating Lgr5+ cells that contribute to all intestinal lineages with slow kinetics, are resistant to chemotherapy, and support intestinal regeneration.

**SUMMARY**

Highly proliferative Lgr5+ stem cells maintain the intestinal epithelium and are thought to be largely homogenous. While quiescent intestinal stem cells (ISCs) populations have been described, the identity and features of such a population remain controversial. Here we report unanticipated heterogeneity within the Lgr5+ ISC pool. We found that expression of the RNA-binding protein Mex3a labels a slowly cycling subpopulation of Lgr5+ ISCs that contribute to all intestinal lineages with distinct kinetics. Single cell transcriptome profiling revealed that Lgr5+ cells adopt two discrete states, one of which is defined by a Mex3a expression program and relatively low levels of proliferation genes. During homeostasis, Mex3a+ cells continually shift into the rapidly dividing, self-renewing ISC pool. Chemotherapy and radiation preferentially target rapidly dividing Lgr5+ cells but spare the Mex3a-high/Lgr5+ population, helping to promote regeneration of the intestinal epithelium following toxic insults. Thus, Mex3a defines a reserve-like ISC population within the Lgr5+ compartment.

## INTRODUCTION

The small intestine is the fastest self-renewing tissue in mammals (Clevers, 2013). Cells are generated within invaginations of the epithelium called crypts, migrate rapidly towards the surface and die at the tip of finger-like protrusions known as villus. This whole process takes less than a week. In homeostasis, cell loss in the villus is compensated by continuous cell production in crypts. The rapid cellular turnover of the intestine is powered by Lgr5+ intestinal stem cells (ISCs) that reside at the bottommost positions of the crypts (Barker et al., 2007). Lgr5+ cells actively proliferate and give rise to progenitors that differentiate as they reach the top of the crypts. Large numbers of absorptive cells (enterocytes) and mucosecreting cells (goblet cells) that populate the intestinal epithelium are generated by amplification of the progenitor pool through several rounds of cell division before differentiation. A subset of progenitor cells undergo differentiation to Paneth cells, which remain intermingled with Lgr5+ cells at the base of the crypt. In addition, the small intestine contains three low abundance cell types; Enteroendocrine cells which are hormone secreting cells scattered throughout the crypt and villus, Microfold (M) cells which line the Peyer patches and initiate mucosal immunity and Tuft cells which are dedicated to sense and trigger responses to helminth parasites. Enteroendocrine, M- and Tuft cells in the intestine represent less than 1 epithelial cell in every 100.

Lgr5+ ISCs proliferate as a homogenous cell pool (Snippert et al., 2010), with a rate of about 1 division per day (Barker et al., 2007). Such high proliferation rate, renders Lgr5+ ISCs sensitive to DNA and cytostatic damage (Tao et al., 2015). Despite single cell profiling supported the absence of heterogeneity in Lgr5+ cell population (Grün et al., 2015), the intestine displays a remarkable capacity to recover from these insults, suggesting the existence of reserve or facultative stem cells in a relatively quiescent state. Numerous studies have proposed that quiescent ISCs occupy the +4 crypt position and express markers such as Bmi1, mTert, Lrig1 or Hopx (Montgomery et al., 2010; Powell et al., 2012; Takeda et al., 2011; Yan et al., 2012). Yet, the existence and identity of such quiescent ISCs has been largely controversial (Muñoz et al., 2012) and remains a matter of debate. It was also shown that upon damage of the Lgr5+ pool, committed progenitor cells undergo dedifferentiation and act as facultative stem cells by regenerating the ISC compartment. For example, Alpi+ enterocytes act as facultative stem cells upon genetic ablation of Lgr5+ cells (Tetteh et al., 2016). Likewise, crypt progenitors that express the Notch ligand Dll1 give rise to secretory cells in homeostatic conditions yet they produce Lgr5+ ISCs upon radiation-induced damage of the intestine (van Es et al., 2012). Secretory cell progenitors are slow

proliferating, retain DNA labels and are relatively resistant to chemotherapeutic drugs (Buczacki et al., 2013).

Mex3a belongs to the Mex3 family which in mammals contains four members encoded by different genes: Mex3a, Mex3b, Mex3c and Mex3d. Mex3 proteins have highly conserved RNA binding domains and a C-terminal RING finger domain with E3 ubiquitin ligase activity (Buchet-Poyau et al., 2007). The role of Mex genes in mammals is largely unknown, yet their C. elegans homologue – mex3 -, is required for germline stem cell identity and maintenance (Ciosk et al., 2006) whereas human MEX3A has been correlated to stemness in colon cancer cell lines (Pereira et al., 2013). Here we report that Mex3a labels a subpopulation of slow proliferating progenitor cells located around +3/+4 crypt position. In homeostatic conditions, Mex3a-high cells give rise to cells that differentiate to all intestinal lineages with low output. A substantial proportion of Mex3a-high cells also produce rapidly proliferating ISCs in homeostasis. Upon damage of the intestinal epithelium with chemotherapy or irradiation, Mex3a-high cells are spared and a larger proportion of this population contributes to regenerate the rapidly proliferating ISC pool.

## RESULTS

### Characterization of Mex3a-expressing intestinal cells

We compared the specific transcriptional programs of ISCs of the mouse small intestine (Muñoz et al., 2012), mouse large intestine and human colon (Jung et al., 2011) (**Fig S1A**. and **Table S1**). These three sets of ISCs were characterized by expression of canonical intestinal stem cell markers such as Lgr5 (Barker et al., 2007), Ascl2 (van der Flier et al., 2009), EphB3 (Batlle et al., 2002) and Smoc2 (Muñoz et al., 2012). We focused our attention on the RNA binding protein Mex3a, which was enriched in the three types of ISCs yet it had not been previously studied in the intestine. We confirmed by RT-qPCR that Mex3a was upregulated in human and mouse ISCs (**Figure S1B**). By RNAscope in situ hybridization (ISH) on small intestine tissue we found that Mex3a mRNA was restricted to crypt base columnar cells (CBCs) (**Figure 1A**). Most CBCs displayed low levels yet about half of the crypts in each histological section contained 1 to 3 CBCs with relative higher Mex3a mRNA abundance (**Figure 1B**). Quantification revealed that Mex3a-high cells were preferentially located at the crypt position +3 near the boundary of the Paneth cell compartment (**Figure 1C**). To analyze functionally the population of cells that express Mex3a, we generated mice bearing a transcriptional reporter cassette knocked-in at the start codon of the Mex3a locus. The reporter cassette consisted in tdTomato and CreERT2 cDNAs separated by self-cleavage T2A peptide (**Figure 1D**, **Figure S1C**). We termed this mouse strain Mex3a$^{Tom\text{-}CreERt2/+}$ (from now on referred to as Mex3a$^{Tom/+}$).

Mex3a$^{Tom/+}$ were born at mendelian ratios, had normal lifespans, were fertile, and showed no overt phenotype. Analysis of reporter expression in the small intestine revealed Tomato positive (Tom+) cells positioned at the crypt base. The expression of the reporter was low, which is in agreement with the low endogenous levels of Mex3a mRNA detected by RT-qPCR and ISH. We observed cells with higher Tomato levels located most frequently near to uppermost Paneth cells, i.e. the +3/+4 crypt region (**Figure 1D**). In flow cytometry analysis of dissociated crypt-enriched epithelial cells, approximately 4% were Tom+. We defined the Tom-high cell population as the top 20% brightest cells of the Tom+ fraction, which correspond to about 1% of all crypt cells (**Figure 1E**). Given that each crypt is formed by 200-300 epithelial cells, this threshold coincides with 2-3 Tom-high cells per crypt, which is roughly the frequency of Mex3a-high cells observed by ISH. Indeed, Tom-high cells expressed highest levels of Mex3a mRNA (**Figure 1F**).

Genes driven by WNT signaling in ISCs such as Lgr5, Ascl2, Axin2, Fzd2 or Smoc2 were expressed at similar levels in Mex3a-high compared to Mex3a-low cells whereas Mex3a-neg cells displayed several fold lower expression (**Figure 1F**). In contrast, expression of markers of mucosecreting and enterocyte differentiation was highest in the Mex3a-negative cell population (**Figure 1G**). We also assessed the expression of putative +4 ISC markers. We observed a small yet consistent enrichment in Bmi1 and Tert in Mex3a-high cells (**Figure 1H**). To further define the phenotype of Mex3a-expressing cell populations, we performed transcriptomics of Mex3a-high, -low and –neg cells (**Table S2**). Gene ontology analysis showed that Mex3a-high cells were enriched in genes encoding secretory cell functions, hormone metabolic processes and also in genes that regulate negatively the cell cycle (**Figure 1I**, **Table S3**). This profile differed substantially from that of Lgr5-high cells, which were characterized by expression of genes that promote mitosis, Wnt signaling and biosynthetic processes as previously reported (**Figure 1I**, **Table S3**)(Muñoz et al., 2012).

**Mex3a-high cells are slow proliferating**

In agreement with the expression of a gene program of negative regulation of cell proliferation in Mex3a-high cells, we confirmed high levels of the cyclin-dependent kinase inhibitors p57 (Cdkn1c) and p21 (Cdkn1a) in Mex3a-high cells (**Figure 2A**). To study the proliferative status of Mex3a-expressing cells in vivo, we interrogated Cyp1a1::H2B-YFP mice which express H2B-YFP under the control of the beta-naphthoflavone-inducible Cyp1a1 promoter in all epithelial intestinal cells except for Paneth cells. In pulse-chase experiments, only long-lived slow proliferating cells retain the H2B-YFP label after 10 days of induction with beta-napthoflavone (Buczacki et al., 2013). Label-retaining cells (LRCs) were characterized by elevated levels of both Mex3a and Lgr5 (**Figure 2B**). Analysis of the transcriptomes of LRCs revealed 68 genes specifically upregulated in LRCs (H2B-YFP+ compared to H2B-YFP- cells and to Paneth cells, **Table S4**). This LRC signature was very significantly enriched in Mex3a-high cells (**Figure 2C**). We confirmed expression of LRC-specific genes such as Rfx6 and Peg3 by RT-qPCR in Mex3a-high cells (**Figure 2D**).

To functionally validate the association of Mex3a-high cells with LRCs, we crossed Cyp1a1::H2B-YFP mice with Mex3a$^{Tom/+}$ mice and analyzed the distribution of the two reporters in compound mice. We had previously demonstrated that high surface abundance of the receptor tyrosine kinase EphB2 identify cells at the bottommost positions of the crypt and enable the isolation of ISCs from mouse and human intestine

(Jung et al., 2011; Merlos-Suárez et al., 2011). Pulse chase-experiments demonstrated that Mex3a-high/EphB2-high cells retained H2B-YFP compared to the Mex3a-neg/EphB2-high cell population (**Figure 2E, 2F**). Within H2B-YFP retaining cells, LRC gene expression was largely restricted to the Mex3a-high population (**Figure S2A**). We also performed classical DNA labeling of proliferative cells using the nucleotide analogue EdU (**Figure 2G**). These experiments confirmed that Mex3a-high cells retained about 15 fold more EdU than Mex3a-neg cells 10 days after labeling (**Figure 2H**). From these experiments, we concluded that Mex3a marks a population of slow proliferating cells within the crypt base.

**Lineage tracing from Mex3a-high cells**

We next sought to analyze the contribution of Mex3a-high cells to intestinal homeostasis (**Figure 3**). To this end, we crossed $Mex3a^{Tom/+}$ mice to Rosa26 mTmG mice (Muzumdar et al., 2007). In this mouse model, induction with tamoxifen switches on the activity of Mex3a-driven CreERT2, which in turn activates the expression of membrane bound GFP (GFP) in Mex3a-expressing cells and their progeny. In parallel, we analyzed the dynamics of Lgr5+ cells using $Lgr5^{GFP/+}$; $Rosa26R^{LacZ}$ mice (**Figure S3A**). These experiments were performed in conditions that trigger recombination in one cell per crypt. We focused our analysis in the small intestine and first established the initial position of Mex3a-derived clones. 36 hours after induction, around 80% of clones were located just above or below the uppermost Paneth cells, within the crypt +3/+4 region (**Figure 3A**). Lineage tracing from the $Lgr5^{GFP/+}$ mice showed a distribution of labeled cells complementary to that observed in $Mex3a^{Tom/+}$ mice at this early time point, i.e. 80% of LacZ+ cells were positioned at the crypt base whereas the rest marked +3/+4 crypt cells (**Figure 3A**). Thus, the initial position of Mex3a+ clones coincided approximately with that of cells expressing highest levels of Mex3a mRNA (**Figure 1A, 1D**). However, we could only find about 1 recombined (GFP+) cell every 700 crypts and this frequency remained in a similar range over the first month of tracing despite the progressive increase in clone size (**Figure 3B**). Thus, recombination of the Rosa26 mTmG reporter allele only occurred in a small fraction of Tomato-CreERT2 expressing cells, which probably reflects poor Cre recombinase activity as a result of low expression of the Mex3a locus. Nevertheless, we confirmed that creERT2 mRNA was largely restricted to Mex3a-high cells (**Figure S3B**) implying that most traces originate from this population. We also found that clones arising 3 days post tamoxifen, which were mostly composed of one or two cells, retained higher levels of Mex3a and

p57 mRNA compared to non-recombined cells (**Figure S3C**). To overcome the limitations imposed by the low frequency of recombination and capture the behavior of the whole Mex3a+ cell population, we assessed a large number of clones for each time point in subsequent experiments (n=56-331 clones; **Figure 3B**).

Clones generated by Mex3a+ cells increased progressively in size yet displayed slow growth kinetics compared to those produced by Lgr5+ cells (**Figure 3C-D**). At 3 days of tracing around 80% of GFP+ clones in Mex3a$^{Tom/+}$ mice appeared as one or two labeled cells around crypt +3/+4 region (**Figure 3C and examples in 3G**). At 7 days, approximately 50% of clones were still composed of 1 or 2 cells whereas only one third had expanded beyond 20 cells (**Figure 3C and examples in Figure 3H**). At this time point, most clones within crypts remained localized at +3/+4 position (**Figure 3E**) and included one or two CBCs (**Figure 3F**). In contrast, Lgr5+ cells generated clones that after 7 days appeared as ribbons of cells that expanded from the crypt base up to the villus tip (**Figure 3D and examples in Figure S3E**). This pattern is generated by the rapid and continuous proliferation of ISCs as described elsewhere (Barker et al., 2007; Lopez-Garcia et al., 2010). Importantly, we did not observe a single ribbon in Mex3a$^{Tom/+}$ mice at 7 days of tracing implying that we did not initially mark any fast dividing ISCs. However, ribbons equivalent to those observed in Lgr5$^{GFP/+}$ mice appeared after 14 days of tamoxifen induction and progressively became more abundant up to the point that they were the only clone type present at 28 days (**Figure 3C, example in S3E**). Quantification of clone number per crypt indicated that frequency decreased progressively until day 28 to then stabilize coinciding with the preponderance of ribbon clones (**Figure 3B**). Analysis of clone position in Mex3a$^{Tom/+}$ mice showed that the majority migrated into the villus over a period of two weeks yet some clones remained in the crypt even after 14 days (**Figure 3E**). These delayed clones may result from either slow cell migration or simply represent new clones generated by traced Mex3a+ cells. Altogether, these observations suggest that Mex3a labels a population of slow dividing progenitor cells. The emergence of ribbon clones 2 weeks after tamoxifen induction implies that a substantial fraction of Mex3a+ cells or of their progeny give rise to rapidly proliferating ISCs in homeostatic conditions.

We next studied the cell composition of the clones formed by Mex3a+ cells. At 3 days post-induction, the majority of single cells and small clones were located within the crypt. None expressed differentiation markers (n=59 clones; **Figure 4A**). At 1 week, 60 % of clones had migrated into the upper region of the crypt and the villus (**Figure 3E**). Villus clones formed by 1 or 2 cells expressed DCLK1, CHGA or ANPEP implying differentiation towards tuft, enteroendocrine or adsorptive lineages respectively

(**examples Figure 4B, S4A**). Multi-cell clones (3-20 cells) were composed of both absorptive (ANPEP+) and mucosecreting cells (MUC2+) (**examples Figure 4C**). 6 clones out of 241 analyzed contained Paneth cells located at the crypt base (examples in **Figure S4B**). Quantification revealed approximately a 3:1 ratio of absorptive versus secretory cell differentiation (**Figure 4D**). Those clones that after 7 days remained positioned near the crypt base did not express differentiation makers (examples in **Figure 4E, S4C, S4D**). From these results, we conclude that similarly to Lgr5+ ISCs, Mex3a-high cells generate multilineage progeny albeit with low output.

**Mex3a is expressed in a subset of Lgr5+ cells**

To explore the relationship between Mex3a+ and Lgr5+ cell populations, we crossed $Mex3a^{Tom/+}$ mice with the $Lgr5^{GFP/+}$ strain (Barker et al., 2007) and analyzed the distribution of the two fluorescent reporters in crypt cells of the small intestine. Previous studies had shown that Lgr5-high cells represent ISCs whereas Lgr5-low cells are early transient amplifying cells (van der Flier et al., 2009). GFP-negative cells are not informative of Lgr5 mRNA as a result of mosaicism in transgene expression in these mice (Schuijers et al., 2014). We thus limited our analyses to the distribution of Mex3a-Tomato within the Lgr5-GFP+ gate. Lgr5-high/Mex3a-high cells represented around 5% of total GFP cells and Lgr5-low/Mex3a-high cell represented around 4% of total GFP cells. (**Figure 5A, Figure S5A**). Thus, 1 in every 5 Lgr5-high and 1 in every 16 Lgr5-low cells are Mex3a-high.

We measured expression of marker genes by RT-qPCR across the 6 cell populations defined by Mex3a-Tomato and Lgr5-GFP levels (**Figure 5B-D** and **Table S5**). ISC-specific and WNT-driven genes were elevated in Lgr5-high cells regardless of Mex3a expression. Mex3a mRNAs were highest in Mex3a-high/Lgr5-high cells (**Figure 5B**). Paneth cell-specific genes such as Defa and Lyz1 were expressed in Mex3a-high/Lgr5-high cells (**Figure 5D**) yet levels of these genes were 80 fold lower in Mex3a-high cells than in mature Paneth cells (**Figure S5C**). Consistent with our previous data, Mex3a-high cell populations contained 2-3-fold more Bmi1 mRNA (**Figure S5B**).

The Lgr5-low cell population was characterized by reduced expression of ISC/WNT genes and elevated levels of markers of intestinal differentiation as previously reported (**Figure 5B-C**) (van der Flier et al., 2009). Mucosecreting (Muc2+/Tff3+) and absorptive progenitor cells (Alpi+) were captured in Lgr5-low/Mex3a-low and Lgr5-low/Mex3a-neg gates implying that these cells represent transient amplifying cells (**Figure 5C**). Dll1+ a

marker of secretory precursors, was in the limit of detection yet it did not show differential expression between subpopulations (**Figure S5B**). Interestingly, Lgr5-low/Mex3a-high cell population expressed genes characteristic of low abundant lineages; tuft cells (Dclk1), enteroendocrine cells (Chga) and M cells (Spib) (**Figure 5D**). Altogether, these expression patterns indicate that Mex3a-high/Lgr5-high cells resemble ISCs with an incipient expression of genes characteristic of both Paneth and rare secretory cells. Of note, most of the clones observed in lineage tracing experiments using the Mex3a$^{Tom}$ driver probably originated from Mex3a-high/Lgr5-high population as inferred from the observation that these cells expressed highest levels of creERT2 (**Figure 5B**).

We next assessed the clonogenic potential of Mex3a/Lgr5 subpopulations using as readout their ability to form *in vitro* organoids (Sato et al., 2009). As previously reported, Lgr5-high cells showed highest organoid forming capacity in media supplemented with RSPO1, EGF and NOGGIN (Sato et al., 2009). We found, however, that Mex3a-high/Lgr5-high formed 3-fold and 10-fold more organoids than Mex3a-low/Lgr5-high and Mex3a-neg/Lgr5-high cell populations respectively (**Figure 5E, 5F**). Organoids generated from Mex3a-high/Lgr5-high cells contained all intestinal lineages (**Figure 5G**) and could be maintained during multiple passages (n=8) implying self-renewal and multilineage differentiation potential. Finally, to test whether Mex3a-high/Lgr5-low cells could gain clonogenic potential under conditions that promote self-renewal, we enforced WNT and NOTCH signaling using CHIR99021 and valproic acid, two small molecules that maximize signaling from these pathways in the absence of ligands (Yin et al., 2014). In these conditions, cells remain blocked in an ISC-like state as indicated by the formation of spheroids yet Mex3a-high/Lgr5-low and Mex3a-low/Lgr5-low remained poorly clonogenic (**Figure S5D**).

**Single cell transcriptomics identifies a subpopulation of Lgr5-high cells enriched in the Mex3a signature.**

The finding that a subset of Lgr5-high cells expresses Mex3a and that Mex3a-high/Lgr5-high display slow proliferation kinetics implies heterogeneity of the ISC pool. Our data contrast with recent single cell transcriptomic analysis of Lgr5-high cells (Grün et al., 2015), which concluded that this population is largely homogenous. Indeed, we reanalyzed the Grün dataset and failed to identify an obvious Lgr5+ subpopulation expressing the Mex3a-specific signature. We, however, reasoned that this negative result might simply be due to underrepresentation of the Mex3a population as the Grün dataset contained sequences of only 53 individual Lgr5-high cells at a relatively low

coverage (mean of 0.4 million reads per cell). Therefore, we set out to generate transcriptomic data for a larger number of Lgr5-high cells at a higher sequencing depth. To this end, the top 25% of brightest Lgr5-GFP+ cells from the small intestine were isolated and loaded into a high-throughput microfluidic chip. We generated single cell transcriptomic profiles by RNA sequencing of 400 capture sites. Subsequently, we excluded both empty capture sites and those sites that contained visually identified cell doublets. Additionally, we applied stringent quality control filters to further discard cells showing either low quality sequencing data or transcript distributions that scored with elevated probability of belonging to cell aggregates (see methods for details). The final high-quality dataset contained 245 individual Lgr5-high cells sequenced at an average of 1.8 million reads per cell (**Figure 6A**). We identified unique transcripts (using unique molecular identifiers, UMIs) and assessed cell population distribution using principal component analysis (PCA). Unsupervised analysis grouped Lgr5-high cells into two well-defined clusters. Cluster 1 and cluster 2 contained 140 and 105 Lgr5-high cells respectively (**Figure 6B**). Optimal cluster number was not affected by coverage as shown by downsampling analysis (**Figure S6A**), yet a reduction of cell numbers below 50% precluded the identification of the two Lgr5-high cell populations (**Figure S6B**). We estimated that the classification (out-of-bag) error of cells into the two predefined cell clusters increased upon downsampling of the number reads per cell below 20% (**Figure S6C**). Analysis of the dataset based on t-distributed stochastic neighbor embedding (tSNE) also grouped Lgr5-high cells in two clusters (**Figure S6E-G**). Thus, a minimal number of cells with sufficient coverage are required to identify heterogeneity within the Lgr5-high cell population.

The ISC-specific genes Lgr5, Smoc2, Olfm4 and Axin2 were expressed at equal levels in cluster 1 and 2 (**Figure 6C**). Furthermore, the average level of the Lgr5-GFP-high specific gene expression signature was equivalent in the two clusters (**Figure 6D and Figure S6E**). Therefore, both cluster 1 and 2 contain bonafide Lgr5-high cells. Lowly expressed genes such as Mex3a, Tert or Dll1 were captured with very low coverage in our dataset (i.e. an average of 0.24, 0.06, and 0.47 UMIs per cell respectively), which precluded reaching reliable conclusions about their expression pattern. However, we found that the Mex3a-high signature was largely upregulated in cells belonging to cluster 2 (**Figure 6E and Figure S6F**). Genes upregulated in Mex3a-high/Lgr5-high cells compared to Mex3a-neg/Lgr5-high cells were also enriched in cluster 2 (**Table S6 and Figure S6H**). We found no correlation between expression levels of Lgr5-high and Mex3a-high signatures (**Figure 6F**).

The above data imply that cluster 2 contains Lgr5-high cells that express the gene program characteristics of Mex3a-high cells. Additionally, the PCA plots showed that the Mex3a gene program was expressed in a graded manner; this is, it increased as cells from cluster 2 separate from cluster 1 along principal component 1 (PC1). We found that the graded expression of the Mex3a program correlated inversely with a signature characteristic of Ki67+ crypt cells (corr: -0.38; p value: $7.8 \times 10^{-10}$) (**Figure 6G, 6H)** obtained from Ki67-RFP knock-in mice (**Table S6),** (Basak et al., 2014) which is enriched in genes required for mitosis and progression through the cell cycle. Indeed, Cluster 1 and 2 in the PCA and tSNE plots differed in the expression levels of the Ki67 signature (**Figure 6G and S6G**). Therefore, as Lgr5-high cells become progressively Mex3a positive they lower expression of the proliferation program.

Finally, a recent study has identified a subset of Lgr5-high cells characterized by incipient expression of makers of both of absorptive and secretory lineages. It was proposed that these "primed" Lgr5-high cells, termed intestinal bipotent progenitors (IBPs), may represent ISCs undertaking the first step toward differentiation before commitment to either lineage (Kim et al., 2016). We used genes enriched in IBPs to identify this population in our dataset and confirmed that a fraction of Lgr5-high cells expressed markers of both absorptive and secretory lineages. IBP-like cells were present in both cluster 1 and 2 and therefore they did not represent a main source of cell heterogeneity in the PCA (**Figure S6I**).

**Mex3a-expressing cells are resistant to both chemotherapy and γ-radiation**

Standard chemotherapy and γ-radiation treatments are aimed at killing rapidly dividing cells. A main example is 5-Fluorouracil (5FU), which blocks thymidine synthesis thus impeding DNA replication. γ-radiation induces double strand breaks in all cell types, albeit with higher toxicity in highly proliferative cells. Consistent with this notion, treatment of mice with both high dose 5FU and IR reduced significantly the Lgr5-high cell population (**Figure 7A**-**B**, **Figure S7A**). In contrast, the frequency of Mex3a-high cells increased upon treatment (**Figure 7A-B, Figure S7B**). Analysis of compound $Mex3a^{Tom/+}$; $Lgr5^{GFP/+}$ mice showed that Mex3a-high/Lgr5-high cells were more resistant to 5-FU and IR than Lgr5-high/Mex3a-low and Lgr5-high/Mex3a-neg cells (**Figure 7C, 7D**). Mex3a expression also segregated Lgr5-low cells according to their sensitivity to 5-FU and IR. These experiments also showed that Lgr5-low/Mex3a-neg cells, which represent the bulk of the transient amplifying compartment, were highly sensitive to these insults whereas the relative numbers of Lgr5-low/Mex3a-high and Lgr5-low/Mex3a-low cells increased upon treatment (**Figure 7C, 7E**). Therefore, the majority

of Lgr5-high and Lgr5-low cells that resisted 5FU expressed Mex3a. As we assessed cell numbers 48 hours after treatment, the observed frequencies likely reflect relative survival rates rather than proliferation and subsequent expansions of each subpopulation.

We next studied the contribution of Mex3a-high cells to tissue renewal upon chemotherapy by mapping their fate using lineage tracing analysis. We did not find significant differences in clone size or numbers at early time points of treatment (**Figure S7C, S7D**). Importantly, 1 week after 5FU treatment, Mex3a+ cells produced larger clones, including 10% of ribbons that were never present in untreated mice at this time point (**Figure 7F, 7G**). All clones produced in control or 5FU-treated $Mex3a^{Tom/+}$ mice arose from +3/+4 crypt region, except for ribbons that started below within the crypt base (**Figure S7E**). Two weeks after 5FU treatment, the percentage of crypt base-to-villus ribbons generated by Mex3a-high cells increased by over two fold compared to controls (**Figure 7G**). These patterns fit well with the notion that Mex3a-high cells are not only relatively resistant to chemotherapy but also contribute to regenerate the rapidly proliferating ISC pool after damage.

## DISCUSSION

The features and behavior of Mex3a+ cells are unique among other previously characterized crypt cell populations. Mex3a-high cells resemble Lgr5+ crypt base columnar cells in that they express high levels of WNT-driven ISC-specific genes. However, the progeny of Mex3a-high cells originates largely from +3/+4 crypt region and is substantially less abundant than that produced by bulk Lgr5+ cell population during equivalent periods. Mex3a+ cells generate multilineage progeny and therefore they do not represent progenitor committed to particular lineages. Dll1, Dclk1, Ngn3 or Alpi mark short-lived progenitor cells that differentiate towards particular lineages (Nakanishi et al., 2013; van Es et al., 2012; Wang et al., 2007). None of these populations contribute long-term to renewal of the epithelium in homeostatic conditions. In contrast, a large fraction of Mex3a population is recalled to the ISC pool and regenerates the epithelium over months.

Mex3a-high cells display features reminiscent of crypt Label Retaining Cells (LRCs) originally identified by Chris Potten (Potten et al., 1978) and later characterized by the Winton lab (Buczacki et al., 2013). The observation that the H2B-YFP mark that accumulates in LRCs is inherited by enteroendocrine and Paneth cells led to the

proposal that these cells are precursors of these secretory lineages (Buczacki et al., 2013). Experiments of lineage-tracing using a bipartite Cre recombinase in which the enzyme is reconstituted only in label retaining cells indicated that this cell population does not retain clonogenic capacity in homeostasis (Buczacki et al., 2013). Similarly to LRCs, Mex3a-high cells accumulate nucleotide analogues as a result of their low proliferation rates and display incipient expression of Paneth Cell markers. Yet, Mex3a-derived clones contain larger proportion of absorptive cells, which reflect the greater abundance of this cell type in the small intestine, and therefore Mex3a+ progeny is not biased towards the secretory lineage. More importantly, unlike LRCs described by Winton and colleagues (Buczacki et al., 2013), Mex3a-high cells or its progeny produce ribbon clones in homeostatic conditions implying long-term clonogenic potential. The reasons for the discrepancy between the behavior of H2B-YFP-retaining cells reported in the previous study (Buczacki et al., 2013) and that of the Mex3a-high cells described herein are unclear but indicate that these two populations are not completely overlapping. A potential confounding effect is the fact that Mex3a-high cells include both Lgr5-high and Lgr5-low populations. Mex3a-high/Lgr5-high cells express the ISC/WNT program and are highly clonogenic ex vivo. In contrast, Mex3a-high/Lgr5-low cells are characterized by higher levels of enteroendocrine and tuft cell genes and give rise organoids with very low efficiency. We thus speculate that lineage tracing from the bipartite cre/H2B present in the previous study (Buczacki et al., 2013) may largely reflect the behavior of the Mex3a-high/Lgr5-low population. Testing this hypothesis will require generation of genetic tools that enable fate mapping from each of these two populations.

Single cell transcriptomic analysis revealed previously unanticipated heterogeneity of Lgr5-high cell pool. Our data indicates that two classes of Lgr5-high cells co-exist within the crypt base. Both are marked by genes that define canonical ISCs whereas cluster 2 cells are enriched in the gene program specific of Mex3a+ cells. The average expression of the Mex3a signature is anti-correlated with that of Ki67+ progenitor cells, an observation that fits in well with the limited cellular output generated by Mex3a+ cells in lineage tracing experiments. Nevertheless, the Ki67 program is not fully silenced in cluster 2 cells, which is in agreement with Lgr5-high/Mex3a-high cells being slow proliferating rather than fully quiescent. Thus, single cell expression data reflects to a large extent rapid versus slow proliferative states adopted by ISCs. Based on the graded levels of Mex3a-specific gene signature across cluster 2 cells, we speculate that some cluster 1 cells may progressively acquire the slow diving Mex3a+ phenotype. As the Lgr5-driven creERT2 marks both rapid and slow proliferating ISCs, lineage

tracing alleles specific of cluster 1 cells will be required to track conversion between these two states.

It is important to consider that as a consequence of their lower division rates, Mex3a-high cells have a modest contribution to sustain renewal of the epithelium under homeostatic conditions. According to the lineage tracing experiments, individual cells and small clones derived from Mex3a-high cells last for about two to three weeks suggesting that this is the approximate life-span of the Mex3a-high cell population. This data fits in well with the possibility that slow proliferating Mex3a-high cells are continuously displaced by the rapidly dividing ISCs as a result of the neutral competition dynamics that drive renewal of the intestinal epithelium (Snippert et al., 2010). An alternative explanation for the long residence time of Mex3a-derived clones in crypts would be that over a period of two weeks, Mex3a+ cells keep producing new progeny, which is observed as individual cells or small cell clones in each experimental time point. Importantly, a fraction of Mex3a-derived cells are converted to rapidly proliferating ISCs as shown by the emergence of crypt base-to-villus clones ("ribbons") in two weeks tracing experiments. The fact that we never observed ribbon-like clones before this period implies that Mex3a-driven creERT2 does not initially mark fast dividing ISCs.

Finally, Mex3a cells are largely resistant to chemotherapy and radiotherapy when compared to bulk Lgr5+ ISCs (Tao et al., 2015). This property is likely the consequence of their relative slow proliferation rates. In addition, the observation that isolated Mex3a-high/Lgr5-high cells display enhanced organoid formation capacity may suggest that this cell population adapts particularly well to stress conditions such as those present in *ex vivo* cultures. We show that upon 5FU treatment, the rate of conversion of Mex3a-high cells to ISCs is exacerbated. Mounting evidence indicates that Dll1+ secretory progenitors (van Es et al., 2012) and even differentiated enterocytes (Tetteh et al., 2016) exhibit plasticity and regain stemness upon depletion of the Lgr5+ pool. Yet, we argue that reversion of differentiated cells to a stem cell state is probably a rare event that does not contribute substantially to regenerate the epithelium after chemotherapy or radiation. Indeed, De Sauvage and colleagues demonstrated that Lgr5+ cell population is necessary to regenerate the epithelium upon irradiation (Metcalfe et al., 2014). Our experiments show that chemotherapy and radiation eliminates preferentially Mex3a-neg/Lgr5-low cell population, which contains the bulk of transient amplifying cells in the crypt. In contrast, the Mex3a+ cell population continues to generate progeny immediately after 5FU treatment. In these conditions, about 40% of Mex3a-high cells produce ribbon-like clones implying that a

large proportion contribute to generate the compartment of fast dividing Lgr5+ ISCs. Given the fact that Mex3a-high/Lgr5-low cells show little clonogenic capacity in *ex vivo* assays, we favor the idea that resilient Mex3a-high/Lgr5-high cells contribute the most to regeneration after damage. Therefore Mex3a-high cells represent a reservoir of slow-dividing chemotherapy- and radiotherapy-resistant Lgr5+ cells.

## Author Contributions

F.M.B. and E.B conceived the project. F.M.B. designed and characterized the Mex3a$^{Tom\text{-}creERT2}$ allele, performed FACS, histology and expression analysis of the different mouse models and analyzed results. E.M. performed organoid growth assays and histology experiments. X.H.M performed mouse handling and in vivo experimental manipulations. C.S.A. developed the biostatistical pipeline and methods to analyze single cell data. M.S. performed histology experiments. H.H, I.G, M.G, A.G-A and G.R-E performed library preparation, sequencing and transcript mapping in single cell transcriptomic analysis. M.N. and O.Y performed RNAscope ISH. S.A.J. B. and D.J.W. provided the Cyp1a1::H2B-YFP mouse model. E.B. supervised the project, analyzed results, and wrote the manuscript with the assistance of F.M.B.

## Acknowledgements

We thank Elena Sancho, Anna Merlos-Suárez, Andreu Casali, Peter Jung, Travis H. Stracker and Alexandra Avgustinova for helpful discussions. We thank all members of the Batlle laboratory for support and discussions. We are also grateful for the assistance by the IRB Barcelona core facilities for Histopathology, Functional Genomics, Biostatistics, and Advanced Digital Microscopy, as well as the PCB Flow Cytometry platform. HH is a Miguel Servet (CP14/00229) researcher funded by the Spanish Institute of Health Carlos III (ISCIII) and supported by the Olga Torres Foundation. This work has been financed by the European Research Council (ERC advanced grant 340176) and by the Spanish Ministry of Science and Competitivity (SAF2011-27068). Work in the laboratory of Dr. Batlle is supported by Fundación Botín and Banco Santander, through Santander Universities.

## REFERENCES


Baker, S.C., Bauer, S.R., Beyer, R.P., Brenton, J.D., Bromley, B., Burrill, J., Causton, H., Conley, M.P., Elespuru, R., Fero, M.*, et al.* (2005). The External RNA Controls Consortium: a progress report. Nature methods *2*, 731-734.

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.*, et al.* (2007). Identification of stem cells in small intestine and colon by marker gene Lgr5. Nature *449*, 1003-1007.

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.

Batlle, E., Henderson, J.T., Beghtel, H., Van den Born, M., Sancho, E., Huls, G., Meeldijk, J., Robertson, J., Van de Wetering, M., Pawson, T.*, et al.* (2002). B-Catenin and TCF Mediate Cell Positioning in the Intestinal Epithelium by Controlling the Expression of EphB/EphrinB. Cell *111*, 251-263.

Buchet-Poyau, K., Courchet, J., Le Hir, H., Seraphin, B., Scoazec, J.Y., Duret, L., Domon-Dell, C., Freund, J.N., and Billaud, M. (2007). Identification and characterization of human Mex-3 proteins, a novel family of evolutionarily conserved RNA-binding proteins differentially localized to processing bodies. Nucleic acids research *35*, 1289-1300.

Buczacki, S.J.A., 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.

Ciosk, R., DePalma, M., and Priess, J.R. (2006). Translational regulators maintain totipotency in the Caenorhabditis elegans germline. Science *311*, 851-853.

Clevers, H. (2013). The intestinal crypt, a prototype stem cell compartment. Cell *154*, 274-284.

Cunningham, F., Amode, M.R., Barrell, D., Beal, K., Billis, K., Brent, S., Carvalho-Silva, D., Clapham, P., Coates, G., Fitzgerald, S.*, et al.* (2015). Ensembl 2015. Nucleic acids research *43*, D662-669.

Eklund, A.C., and Szallasi, Z. (2008). Correction of technical bias in clinical microarray data improves concordance with known biological information. Genome Biol *9*, R26.

Gonzalez-Roca, E., Garcia-Albéniz, X., Rodriguez-Mulero, S., Gomis, R.R., Kornacker, K., and Auer, H. (2010). Accurate Expression Profiling of Very Small Cell Populations. PloS one *5*, e14418.

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.

Huang da, W., Sherman, B.T., and Lempicki, R.A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nature protocols *4*, 44-57.

Irizarry, R.A., Hobbs, B., Collin, F., Beazer-Barclay, Y.D., Antonellis, K.J., Scherf, U., and Speed, T.P. (2003). Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics *4*, 249-264.

Islam, S., Zeisel, A., Joost, S., La Manno, G., Zajac, P., Kasper, M., Lonnerberg, P., and Linnarsson, S. (2014). Quantitative single-cell RNA-seq with unique molecular identifiers. Nature methods *11*, 163-166.

Jung, P., Sato, T., Merlos-Suárez, A., Barriga, F.M., Iglesias, M., Rossell, D., Auer, H., Gallardo, M., Blasco, M.A., Sancho, E.*, et al.* (2011). Isolation and in vitro expansion of human colonic stem cells. Nature medicine *17*, 1225-1227.

Kim, T.H., Saadatpour, A., Guo, G., Saxena, M., Cavazza, A., Desai, N., Jadhav, U., Jiang, L., Rivera, M.N., Orkin, S.H.*, et al.* (2016). Single-Cell Transcript Profiles Reveal Multilineage Priming in Early Progenitors Derived from Lgr5(+) Intestinal Stem Cells. Cell reports *16*, 2053-2060.

Kivioja, T., Vaharautio, A., Karlsson, K., Bonke, M., Enge, M., Linnarsson, S., and Taipale, J. (2012). Counting absolute numbers of molecules using unique molecular identifiers. Nature methods *9*, 72-74.
Lopez-Garcia, C., Klein, A.M., Simons, B.D., and Winton, D.J. (2010). Intestinal stem cell replacement follows a pattern of neutral drift. Science *330*, 822-825.
Marco-Sola, S., Sammeth, M., Guigo, R., and Ribeca, P. (2012). The GEM mapper: fast, accurate and versatile alignment by filtration. Nature methods *9*, 1185-1188.
Merlos-Suárez, A., Barriga, Francisco M., Jung, P., Iglesias, M., Céspedes, María V., Rossell, D., Sevillano, M., Hernando-Momblona, X., da Silva-Diz, V., Muñoz, P.*, et al.* (2011). The Intestinal Stem Cell Signature Identifies Colorectal Cancer Stem Cells and Predicts Disease Relapse. Cell stem cell *8*, 511-524.
Metcalfe, C., Kljavin, N.M., Ybarra, R., and de Sauvage, F.J. (2014). Lgr5+ stem cells are indispensable for radiation-induced intestinal regeneration. Cell stem cell *14*, 149-159.
Montgomery, R.K., Carlone, D.L., Richmond, C.A., Farilla, L., Kranendonk, M.E.G., Henderson, D.E., Baffour-Awuah, N.Y., Ambruzs, D.M., Fogli, L.K., Algra, S.*, et al.* (2010). Mouse telomerase reverse transcriptase (mTert) expression marks slowly cycling intestinal stem cells. Proceedings of the National Academy of Sciences *108*, 179-184.
Mudge, J.M., and Harrow, J. (2015). Creating reference gene annotation for the mouse C57BL6/J genome assembly. Mamm Genome *26*, 366-378.
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. The EMBO journal *31*, 3079-3091.
Muzumdar, M.D., Tasic, B., Miyamichi, K., Li, L., and Luo, L. (2007). A global double-fluorescent Cre reporter mouse. Genesis *45*, 593-605.
Nakanishi, Y., Seno, H., Fukuoka, A., Ueo, T., Yamaga, Y., Maruno, T., Nakanishi, N., Kanda, K., Komekado, H., Kawada, M.*, et al.* (2013). Dclk1 distinguishes between tumor and normal stem cells in the intestine. Nat Genet *45*, 98-103.
Pereira, B., Sousa, S., Barros, R., Carreto, L., Oliveira, P., Oliveira, C., Chartier, N.T., Plateroti, M., Rouault, J.P., Freund, J.N.*, et al.* (2013). CDX2 regulation by the RNA-binding protein MEX3A: impact on intestinal differentiation and stemness. Nucleic acids research *41*, 3986-3999.
Potten, C.S., Hume, W.J., Reid, P., and Cairns, J. (1978). The segregation of DNA in epithelial stem cells. Cell *15*, 899-906.
Powell, A.E., Wang, Y., Li, Y., Poulin, E.J., Means, A.L., Washington, M.K., Higginbotham, J.N., Juchheim, A., Prasad, N., Levy, S.E.*, et al.* (2012). The pan-ErbB negative regulator Lrig1 is an intestinal stem cell marker that functions as a tumor suppressor. Cell *149*, 146-158.
Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic acids research *43*, e47.
Satija, R., Farrell, J.A., Gennert, D., Schier, A.F., and Regev, A. (2015). Spatial reconstruction of single-cell gene expression data. Nature biotechnology *33*, 495-502.
Sato, T., Vries, R.G., Snippert, H.J., van de Wetering, M., Barker, N., Stange, D.E., van Es, J.H., Abo, A., Kujala, P., Peters, P.J.*, et al.* (2009). Single Lgr5 stem cells build crypt-villus structures in vitro without a mesenchymal niche. Nature *459*, 262-265.
Schuijers, J., van der Flier, Laurens G., van Es, J., and Clevers, H. (2014). Robust Cre-Mediated Recombination in Small Intestinal Stem Cells Utilizing the Olfm4 Locus. Stem Cell Reports *3*, 234-241.
Snippert, H.J., van der Flier, L.G., Sato, T., van Es, J.H., van den Born, M., Kroon-Veenboer, C., Barker, N., Klein, A.M., van Rheenen, J., Simons, B.D.*, et al.* (2010). Intestinal crypt homeostasis results from neutral competition between symmetrically dividing Lgr5 stem cells. Cell *143*, 134-144.

Soriano, P. (1999). Generalized lacZ expression with the ROSA26 Cre reporter strain. Nat Genet *21*, 70-71.
Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S.*, et al.* (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A *102*, 15545-15550.
Takeda, N., Jain, R., LeBoeuf, M.R., Wang, Q., Lu, M.M., and Epstein, J.A. (2011). Interconversion between intestinal stem cell populations in distinct niches. Science *334*, 1420-1424.
Tao, S., Tang, D., Morita, Y., Sperka, T., Omrani, O., Lechel, A., Sakk, V., Kraus, J., Kestler, H.A., Kuhl, M.*, et al.* (2015). Wnt activity and basal niche position sensitize intestinal stem and progenitor cells to DNA damage. The EMBO journal *34*, 624-640.
Tetteh, Paul W., Basak, O., Farin, Henner F., Wiebrands, K., Kretzschmar, K., Begthel, H., van den Born, M., Korving, J., de Sauvage, F., van Es, Johan H.*, et al.* (2016). Replacement of Lost Lgr5-Positive Stem Cells through Plasticity of Their Enterocyte-Lineage Daughters. Cell stem cell *18*, 203-213.
van der Flier, L.G., van Gijn, M.E., Hatzis, P., Kujala, P., Haegebarth, A., Stange, D.E., Begthel, H., van den Born, M., Guryev, V., Oving, I.*, et al.* (2009). Transcription factor achaete scute-like 2 controls intestinal stem cell fate. Cell *136*, 903-912.
van der Maaten, L., and Hinton, G. (2008). Visualizing Data using t-SNE. Journal of Machine Learning Research *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. Nature cell biology *14*, 1099-1104.
Wang, Y., Giel-Moloney, M., Rindi, G., and Leiter, A.B. (2007). Enteroendocrine precursors differentiate independently of Wnt and form serotonin expressing adenomas in response to active beta-catenin. Proceedings of the National Academy of Sciences of the United States of America *104*, 11328-11333.
Yan, K.S., Chia, L.A., Li, X., Ootani, A., Su, J., Lee, J.Y., Su, N., Luo, Y., Heilshorn, S.C., Amieva, M.R.*, et al.* (2012). The intestinal stem cell markers Bmi1 and Lgr5 identify two functionally distinct populations. Proceedings of the National Academy of Sciences of the United States of America *109*, 467-471.
Yin, X., Farin, H.F., van Es, J.H., Clevers, H., Langer, R., and Karp, J.M. (2014). Niche-independent high-purity cultures of Lgr5+ intestinal stem cells and their progeny. Nat Methods *11*, 106-112.

## FIGURE LEGENDS

**Figure 1: Characterization of Mex3a-expressing intestinal cells.**

**(A)** ISH of Mex3a expression in mouse small intestine. (Left panel) Representative staining of Mex3a in proximal mouse small intestine. Scale bar represents 50 µm. (Right panel) High magnification of small intestine crypt stained for Mex3a mRNA expression. The position of crypt-base columnar cells is shown. Red arrowheads point to Mex3a-high CBCs and black arrowheads point to Mex3a-low CBCs. Blue P refers to Paneth cells. Scale bar represents 20 µm.

**(B)** Frequency of Mex3a-high cells per crypt (n = 72 crypts).

**(C)** Frequency of the position of Mex3a-high cells in small intestine crypts (n = 51 Mex3a-high cells in 72 crypts).

**(D)** Design and expression of the Mex3a reporter allele. (Left) A tdTomato/T2A/Cre-ERT2/bGHpolyA cassette was inserted in the translation start site of Mex3a. This construct results in a transcriptional reporter for Mex3a expression driving a tdTomato protein and a tamoxifen inducible Cre recombinase. (Right) Immunofluorescence against tdTomato in the small intestine of a $Mex3a^{Tom/+}$ mouse. Scale bar represents 20 µm.

**(E)** Representative FACS profile of dissociated small intestine crypt cell preparations from $Mex3a^{Tom/+}$ mice. Mex3a-high cells are defined as the top 20% of the Tomato+ population. Frequency of populations is referred to the number of viable epithelial cells.

**(F – H)** Relative expression of known ISC genes (F), abundant lineage genes (G), and putative crypt +4 cell marker genes (H) in Mex3a populations. Bars depict the mean and upper/lower limits of relative expression determined by RT qPCR obtained from a representative sorting experiment.

**(I)** Selected GO biological processes enriched in Mex3a and/or Lgr5 signatures. GO category enrichment and statistical analysis was performed using the DAVID Analysis platform.

**Figure 2: Mex3a expression identifies intestinal label-retaining cells.**

**(A)** Mex3a identifies cells with elevated expression of p21 and p57. Expression of cell cycle inhibitors Cdkn1a (p21/WAF1) and Cdkn1c (p57/KIP2) in Mex3a populations. Bars depict the mean and upper/lower limits of relative expression obtained from a representative sorting experiment.

**(B)** Mex3a is enriched in crypt-base label retaining cells. Lgr5 and Mex3a expression data in label-retaining (YFP+) vs proliferative (YFP-) cells. Bars depict the mean and upper/lower limits from a representative RT qPCR.

**(C)** GSEA analysis of the LRC gene expression signature (**Table S4**) in Mex3a-high vs Mex3a-low cells.

**(D)** RT qPCR of LRC genes Rfx6 and Peg3 in Mex3a populations. Bars depict the mean and upper/lower limits of relative expression obtained from a representative sorting experiment.

**(E)** Experimental protocol. $Mex3a^{Tom/+}$/Cyp1a1::H2B-YFP compound reporters were generated and LRCs were followed by chasing for 10 days after induction. To ensure the analysis of ISC enriched cells, an EphB2 staining was incorporated. Representative FACS plots of LRCs (YFP+) in Mex3a-high and Mex3a-neg cells.

**(F)** Quantification of the distribution of LRCs in Mex3a-high and Mex3a-neg cells. ***, p value < 0.001 in a two-tailed t-test (n = 3 mice). Bars depict the mean ± SEM.

**(G)** Experimental design to follow EdU retention. EdU was injected into $Mex3a^{Tom/+}$ mice and 10 days later Mex3a-high and Mex3a-neg cells were sorted and stained for the presence of EdU by FACS. Representative FACS plots of EdU retention in Mex3a-high (Left panel) and Mex3a-neg cells (Right panel).

**(H)** Quantification of the retention of EdU in Mex3a-high and Mex3a-neg cells. **, p value < 0.01 in a two-tailed t-test (n = 3 mice). Bars depict the mean ± SEM.

**Figure 3: Lineage tracing from Mex3a-high cells shows a distinct behavior from Lgr5-high cells.**

**(A)** Initial clone position in Mex3a and Lgr5-driven lineage tracing. Representative image of a Mex3a GFP+ cell in position +3 (Far left panel), Mex3a GFP+ cell in position +4 (Left panel), and an Lgr5 LacZ+ cell in position +2 (Right panel) after 36

hours of induction. Scale bars represent 20 µm. (Far right) Position of Mex3a (n = 45) and Lgr5 (n = 55) early clones.

**(B)** Frequency of small intestine Mex3a-derived clone number / 1000 crypts over time. Numbers within bars are the total number of clones observed at each timepoint and used for subsequent analysis. Data was obtained from n>3 mice (with > 5 independently stained sections per mouse) per time point. Bars depict mean +/- SD.

**(C)** Clone size distribution over time in small intestine Mex3a-driven tracing. Data is presented for tracing after 3 (n = 146 clones), 7 (n = 302 clones), 14 (n = 331 clones), 21 (n = 65 clones), 28 (n = 61 clones) and >180 days (n = 56 clones).

**(D)** Comparison of clone size distribution after 7 days of tracing for Mex3a (n = 302) and Lgr5 (n = 72 clones).

**(E)** Distribution of Mex3a-derived clones in crypts or villi over time. At least 100 independent clones were scored per time analyzed.

**(F)** Number of labeled crypt base columnar cells (CBCs) in 7 day clones from Mex3a (n = 52 clones) and Lgr5 (n = 61 clones) derived tracing. Only clones observed within crypts were included in the analysis.

**(G-I)** (G) Representative images of Mex3a-derived clones after 3 days of tracing. Scale bar represents 20 µm. (H) Representative images of Mex3a derived clones after 7 days of tracing. Scale bar represents 20 µm.

**Figure 4: Mex3a-high cells are multipotent**

**(A)** Early Mex3a-derived clones do not express differentiation markers. Pictures show representative examples of 3 day GFP+ clones co-stained for markers of Paneth (LYZ1, far left), enteroendocrine (CHGA, left), tuft (DCLK1, right) or absorptive cells (ANPEP, far right). Arrowhead point towards differentiated cells. Scale bar represents 20 µm.

**(B-D)** 7 day Mex3a-derived clones present multilineage differentiation. (B) Examples of 1 – 2 cell clones positive for tuft (left) and enteroendocrine (middle) and enterocyte (right) markers. (C) Examples of multicellular clones with secretory (left) and absorptive (right) cells. Arrowheads point towards GFP+ differentiated cells of the specified lineage. Note that multicellular clones are composed of more than a single lineage

(Red arrowheads). Scale bar represents 20 µm. (D) Mex3a-derived clones are composed of both absorptive and secretory cells. A pan secretory antibody mix (i.e. combined anti-CHGA, -DLCK1, -MUC2 and -LYZ1 antibodies) and anti-ANPEP to label absorptive cells were used to assess the lineage composition of 7 day Mex3a-derived clones. Scale bar represents 20 µm.

**(E)** A fraction of Mex3a-derived clones remains undifferentiated after 7 days of tracing. (Left) A representative undifferentiated clone that is located within the crypt negative for the pan-secretory antibody mix. (Right) Representative small GFP clone negative for ANPEP. Scale bar represents 20 µm.

**Figure 5: Mex3a-high cells are a subpopulation of Lgr5+ cells**

**(A)** Representative FACS plot of small intestine Lgr5-GFP+ cells in $Lgr5^{GFP/+}$; $Mex3a^{Tom/+}$ mice.

**(B - D)** Characterization of the Lgr5 subpopulations defined in (A) by RT qPCR analysis of known intestinal cell markers. Relative expression of ISC genes (B), abundant-cell lineage genes (C), and rare-cell lineage genes (D). Values are normalized to the subpopulation with the highest expression of each gene. Data represent the mean of four independent sorting experiments. (Full expression data is presented in **Table S5**).

**(E-F)** Mex3a-high/Lgr5-high cells are the most efficient in organoid forming assays. (E) Representative images of organoid cultures derived from Mex3a/Lgr5 subpopulations. (F) Quantification of organoid forming efficiency. Bars depict mean ± SEM (n = 6). *, $p < 0.05$; ***, $p < 0.001$ in a two-way ANOVA followed by Tukey´s multiple comparison test.

**(G-H)** Mex3a-high/Lgr5-high cells self-renew and are multipotent in vitro. (G) Experimental approach to study self-renewal and multipotency in vitro. (H) Mex3a-high/Lgr5-high derived organoids were stained for stem/progenitor cells (EPHB2), enterocytes (ANPEP), enteroendocrine cells (CHGA), Paneth cells (LYZ1), tuft cells (DCLK1) and goblet cells (MUC2). Arrowheads point towards stained cells within organoids. Scale bar represents 20 µm.

**Figure 6: Single cell transcriptomics identifies a subpopulation of Lgr5-high cells enriched in the Mex3a signature.**

**(A)** (Left) FACS plot of Lgr5-GFP+ cells obtained from the small intestine of an $Lgr5^{GFP/+}$ mouse. (Right) Experimental pipeline to obtain single cell RNA transcriptomic profile of Lgr5 GFP-high cells.

**(B)** Principal component analysis of 245 Lgr5-high cells unbiasedly identifies two distinct clusters of cells. Each point represents a single Lgr5-high cell.

**(C)** Violin plots of relative expression of ISC genes Lgr5, Smoc2, Axin2, and Olfm4 between Cluster 1 and Cluster 2.

**(D)** The Lgr5-signature is equally expressed in Clusters 1 and 2. (Left) Mean expression of the Lgr5-signature plotted in each Lgr5-high cell. Lines are drawn on the edges of clusters to facilitate their visualization. (Right) Box-plots of relative expression of the Lgr5-signature in Cluster 1 and Cluster 2. n.s., $p > 0.05$ in Kruskal-Wallis test.

**(E)** The Mex3a-signature is enriched in Cluster 2. (Left) Mean expression of the Mex3a-signature plotted in each Lgr5-high cell. Lines are drawn on the edges of clusters to facilitate their visualization. (Right) Box-plots of relative expression of the Lgr5-signature in Cluster 1 and Cluster 2. ***, $p < 0.001$ in Kruskal-Wallis test with at least 10% fold change.

**(F)** There is no correlation between Mex3a and Lgr5 signatures. Pearson correlation of the mean expression of the Lgr5-signature with the Mex3a-signature.

**(G)** The Ki67 signature is enriched in Cluster 1. (Left) Mean expression of the Ki67-signature plotted in each Lgr5-high cell. Lines are drawn on the edges of clusters to facilitate their visualization. (Right) Box-plots of relative expression of the Ki67-signature in Cluster 1 and Cluster 2. ***, $p < 0.001$ in Kruskal-Wallis test with at least 10% fold change.

**(H)** The Mex3a-signature is inversely correlated with the Ki67-signature. Pearson correlation of the mean expression of the Ki67-signature with the Mex3a-signature.

## Figure 7: Mex3a-expressing cells are resistant to 5FU and radiation.

**(A)** Experimental setup to assess stress response. Mice were treated with either two consecutive doses of 5-FU (100 mg/kg per day) or ionizing radiation (12 Gy in one day). Mex3a and Lgr5 populations were analyzed by FACS 48h after the final dose.

**(B)** Mex3a-high cells are more resistant to stress than Lgr5-high cells. Quantification of Lgr5-high (Left) and Mex3a-high (right) cells in control (n = 4), 5-FU (n = 5) and IR (n = 4) treated mice. *, $p < 0.05$; ***, $p < 0.001$ in a one-way ANOVA followed by Tukey´s multiple comparison test.

**(C)** Representative FACS plots of Lgr5-GFP+ cells from untreated (left panel), 5-FU treated (middle panel), and IR treated (right panel) $Lgr5^{GFP/+}$; $Mex3a^{Tom/+}$ compound mice. Frequencies are referred to total Lgr5-GFP+ cells.

**(D)** Quantification of Mex3a populations within Lgr5-high cells in control (n = 4), 5-FU treated (n = 5) and IR treated (n = 4) mice. n.s., $p > 0.05$; ***, $p < 0.001$ in a two-way ANOVA followed by Sidak´s multiple comparison test to untreated mice.

**(E)** Quantification of Mex3a populations within Lgr5-low cells in control (n = 4), 5-FU treated (n = 5) and IR treated (n = 4) mice. ***, $p < 0.001$ in a two-way ANOVA followed by Sidak´s multiple comparison test to untreated mice.

**(F-G)** (F) Experimental setup. Tracing was induced with tamoxifen and followed by two doses of 5-FU. Clones were assessed at shown timepoints. (G) Clone-size distribution in control and 5-FU treated mice at 3 (Control n = 146 clones; 5-FU n = 134 clones), 7 (Ctrl n = 302 clones, 5-FU n = 170 clones) and 14 days (Ctrl n = 331 clones, 5-FU n = 113 clones) from the initial Tamoxifen treatment. Data was obtained from n>4 mice (> 5 independently stained sections per mouse) per time point.

## STAR METHODS

### CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for reagents may be directed to, and will be fulfilled by the corresponding author, Dr. Eduard Batlle (eduard.batlle@irbbarcelona.org).

### EXPERIMENTAL MODEL AND SUBJECT DETAILS

All experimental protocols involving mice were approved by institutional and governmental ethics board. Mex3a$^{tm1(tdTomato-T2A-CreERT2)EBa}$ were generated at the Institute for Research in Biomedicine (IRB) Barcelona. Lgr5$^{tm1(cre/ERT2)Cle}$ mice have been previously described (Barker et al., 2007) and were obtained from the laboratory of Dr. Hans Clevers. Tg(Cyp1a1-H2B-YFP)$^{Dwi}$ mice have been previously described (Buczacki et al., 2013) and were obtained from the laboratory of Dr. Douglas J Winton. Reporter strains Gt(ROSA)26Sor$^{tm4(ACTB-tdTomato,-EGFP)Luo}$ and Gt(ROSA)26Sor$^{tm1Sor}$ have been previously described (Muzumdar et al., 2007; Soriano, 1999), and were obtained from The Jackson Laboratory. Mex3a reporter mice were used to identify the population within the mouse small intestine that expresses this gene. To assess their proliferation rate, lineage tracing and overlap with the small intestine Lgr5 population, Mex3a mice were crossed with the different models described above. For these in vivo experiments, age and sex-matched littermates were always used, with n >3 mice in all experiments (detailed description and number of mice are detailed in the respective figure legends and method details).

### METHOD DETAILS

*Analysis of ISC signatures:*

To identify robust ISC genes we compared three lists of ISC expression signatures derived from mouse and small intestine populations. We overlapped the expression profile of three different ISC signatures: The first is the small intestine Lgr5-GFP-hi signature (GEO dataset: GSE36497) (Muñoz et al., 2012) , which is defined as those genes enriched in Lgr5-high compared to Lgr5-low cells in both Agilent and Affymetrix platforms (Fold change > 2, p value < 0.05). The second is the previously characterized human EPHB2 derived signature (Jung et al., 2011), which is defined by those genes enriched in EPHB2-high cells in a graded manner (i.e. EPHB2-high > EPHB2-med > EPHB2-low. Fold chage > 2, p value < 0.05). The third signature is obtained from

mouse colon Lgr5-GFP cells defined as follows those genes enriched over 2.5 fold in Lgr5-GFP-high compared to Lgr5-GFPlow cells (no p value was used as cut-off). Consistent ISC genes were those present in all signatures (**Table S1**).

*Generation of the Mex3a reporter allele:*

A cassette bearing a tdTomato/T2A/Cre-ERT2/bGH polyA was inserted in frame with the Mex3a start codon in exon 1. The targeting vector was generated by Genebridges. The vector was verified by restriction enzyme digestion as well as sequencing. The vector was electroporated into W4 mouse embryonic stem cells (mESCs) and stably transfected cells were selected with G418. After this, single colonies were picked and triplicate plates were generated of resistant clones. The clones were then screened by long range PCR by using the SequalPrep Kit (Invitrogen). Positive clones were then expanded and further analyzed by Southern Blot (as described in (Barker et al., 2007)) to assure correct and unique genomic integration. The clones verified by southern blot were then transfected with an FlpO-bearing plasmid to remove the Neomycin resistance. Single clones were picked and screened by PCR for recombination and presence of the reporter cassette. Positive clones were then expanded and analyzed by Southern once again to ensure the appropriate integration of the cassette, as well as to independently confirm the deletion of the Neomycin resistance cassette. Finally, positive clones were selected and blastocyst injection was done to generate chimeras. The chimeras with successful germline transmission were bred with C57Bl6 mice and the offspring was used to generate the final mouse colonies. mESC culture, blastocyst injection of selected clones and generation of chimeras were conducted in the Mouse Mutant Core Facility of IRB Barcelona following institutional guidelines. See **Table S7** for genotyping primers.

*Mouse models:*

Lgr5$^{tm1(cre/ERT2)Cle}$ mice have been previously described (Barker et al., 2007) and were obtained from the laboratory of Dr. Hans Clevers. Tg(Cyp1a1-H2B-YFP)Dwi mice have been previously described (Buczacki et al., 2013) and were obtained from the laboratory of Dr. Douglas J Winton. Reporter strains Gt(ROSA)26Sor$^{tm4(ACTB-tdTomato,-EGFP)Luo}$ and Gt(ROSA)26Sor$^{tm1Sor}$ have been previously described (Muzumdar et al., 2007; Soriano, 1999), and were obtained from The Jackson Laboratory. Mice were housed and bred according to IRB Barcelona and PCB institutional guidelines.

*Lineage tracing:*

*Rationale:* For lineage tracing in the Mex3a$^{Tom/+}$ we used CAG-driven mTmG cassette inserted in the ROSA26 locus. We used this model due to the high sensitivity of this allele, which allows the identification of very low number of recombination events. In the case of Lgr5$^{GFP/+}$ model, we turned to the ROSA26 lox-stop-lox LacZ model since the Lgr5-driven GFP would be confounded with the recombination of the ROSA26mTmG cassette.

*Dosage:* For lineage tracing experiments in the Mex3a$^{Tom/+}$ model, mice were injected with two consecutive intraperitoneal doses of tamoxifen (20 mg/kg, Sigma). For early time points (i.e. 36 hours), mice were treated with a single dose of 4-OH tamoxifen (5 mg/kg, Sigma). For lineage tracing from the Lgr5$^{GFP/+}$ model, 1 dose of diluted Tamoxifen (2 mg/kg) was used to ensure tracing from single cell conditions. The time of tracing is referred to the first dose of Tamoxifen.

*Clone identification:* In the case of Mex3a$^{Tom/+}$ mice, the presence of tdTomato in the mTmG cassette is confounded with the Mex3a-driven tdTomato. Thus for all lineage tracing experiments, intestines were stained to detect GFP+ cells by either IHC or immunofluorescence (IF) without considering the tdTomato signal. In the case of Lgr5-driven tracing, intestines were stained to detect β-galactosidase activity. (See below for the protocols used for staining).

*Timepoints for tracing:* Clone scoring was performed at 36 h (n = 11 mice), 3 days (n = 6 mice), 7 days (n = 12 mice), 14 days (n = 9 mice), 21 days (n = 4 mice), 28 days (n = 4 mice) and after 6 months (n = 3 mice) for Mex3a-driven lineage tracing. Lgr5-driven lineage tracing was assessed at 36 h (n = 3 mice) and 7 days (n = 3 mice). At least 5 independent sections were stained per mouse and clones were scored according to their size and position (see below). For tracing after 5-FU treatment, clone scoring was performed at 3 days (n = 4 mice), 7 days (n = 6 mice) and 14 days (n = 8 mice) after the final dose of Tam.

Clone scoring:

*Position*: GFP+ or LacZ+ clones were scored for their position in three categories: crypt base (crypt position +1/+2); suprabasal (crypt position +3/+4) and “crypt top or villus” (clone outside of the crypt base). Clone position was only assessed when clones were

unequivocally localized in each of these compartments (i.e. clones spanning crypts and villi were discarded from the analysis.)

*Size*: GFP+ or LacZ+ clone size was estimated by counting the number of cells. Clones were then classified as single cells, two-cell clones, small clones (3-20 cells), large clones (>20 cells) and ribbons. Clones were scored as ribbons only when GFP+ or LacZ+ cells were present both at the crypt base (position +1/+2) and up to the tip of the villus in the section analyzed.

*CBC clone number*: Number of crypt-base columnar cells were assessed in clones localized at the crypt base where the orientation of the crypt allowed assessment of CBC positions from +1 to +4. Since we focused on sections were we could observe the crypt-villus axis, the maximum number was 8 CBCs per crypt.

*Lineage bias:* In order to analyze potential differentiation bias of the Mex3a population, 7 day clones were co-stained for GFP and either absorptive (ANPEP) or a pan-secretory antibody mix (CHGA, DCLK1, LYZ1, MUC2). The 7 day tracing time was chosen because it has the highest diversity in terms of clone position and clone size.

*In vivo treatments:*

Cyp1a1::H2B-YFP pulse chase experiments: Transgenic mice bearing the Cyp1a1::H2B-YFP allele were given 3 intraperitoneal doses of β-napthoflavone (80 mg/kg, Sigma) over a 36 hour period to induce activity of the Cyp1a1 promoter and mark intestinal cells with H2B-YFP as previously described (Buczacki et al., 2013) . After 10 days, intestines were dissociated and stained for EPHB2 (see below for detailed staining protocol). EPHB2-high cells that retained H2B-YFP mark were quantified 10 days after the initial dose by flow cytometry (Buczacki et al., 2013).

Label retention assay: For EdU label-retention, Mex3a$^{Tom/+}$ mice were treated with 3 intraperitoneal doses (1 mg/25 g, Life technologies) in a 36 hour period and analyzed 10 days after the final dose.

5-FU and IR treatments: Damage to the intestinal epithelium was done by injecting two intraperitoneal doses of 5-fluorouracil (100 mg/kg, Sigma) over a 48h period or 12 Gy of γ-radiation from a Cs source. For FACS of single and compound reporters, intestinal cells were analyzed 48h after the last 5-FU dose. For FACS analysis 4-5 mice (age and

sex-matched littermates) were analyzed per group. Treatments were randomized but no blinding was incorporated. For lineage tracing experiments mice were injected with tamoxifen, treated with 5-FU (100 mg/kg, Sigma) and analyzed at 3, 7 or 14 days after the first dose of Tamoxifen (refer to lineage tracing section for exact number of clones and mice analyzed per timepoint).

*Crypt purification and staining:*

Intestinal crypt cells were purified as previously described (Merlos-Suárez et al., 2011). Proximal small intestines (duodenum and jejunum) were dissected from age and sex-matched mice and cut open longitudinally. They were later incubated with HBSS EDTA (8 mM) for 5 min at room temperature. After this incubation, intestines were shaken vigorously to remove villi and the first supernatant was discarded. The intestines were then incubated with HBSS EDTA on ice for 15 min, after which shaking was repeated. This yielded the first crypt fraction. This process was repeated twice and after microscopic inspection of the samples the fractions enriched in crypts were kept (usually fractions 2 and 3). These fractions were then pooled together for each mouse, centrifuged at 1200 rpm for 5 min (4 ºC) and filtered through a 70 µm mesh (BD Biosciences) to further remove villi. Filtered fractions were enzymatically disaggregated by adding dispase (0.4 mg/mL, Gibco) and incubated for 20 min at 37ºC. The media was then supplemented with 5% FBS (Gibco) to neutralize dispase and the samples were filtered through a 40 µm mesh. For intestinal LRC analysis in compound reporter mice, single cells were stained with an APC-conjugated antibody raised against EphB2 (Clone mAb 2H9, 6 µg/mL, BD). Cells were stained with DAPI (1 µg/mL, Sigma Aldrich) to exclude dead cells. The single-cell preparations of small intestinal crypts were analyzed by flow cytometry in a FACS Aria 2.0 (BD). Data was analyzed with FACSDiva software (BD) or FlowJo for EdU retention assays.

*Population gating rationale:* To define subpopulations with the reporter mice we used the following criteria: for the Mex3a reporter allele, Tomato-high cells were defined as the top 20% brightest cells within the Tomato+ population. For the Lgr5 reporter allele, GFP-high cells were defined as the top 25% of the GFP+ population. For compound reporter mice, only the GFP+ population was analyzed due to mosaicism of the Lgr5-GFP allele and Mex3a gates were defined within the Lgr5-GFP high and Lgr5-GFP low populations. For all experiments a Mex3a+/+ mouse was processed in parallel to set the sorting gates in both single and compound reporter FACS experiments.

*In vitro organoid growth:*

Compound Mex3a / Lgr5 reporter mice were used to isolate the populations expressing different levels of Mex3a and Lgr5. As described above, these experiments were done by gating the Lgr5::GFP+ cells and isolating single cells from the respective Mex3a-high, Mex3a-low and Mex3a-neg cells within both Lgr5-high and Lgr5-low populations. 4500 cells of each population were sorted into organoid growth media (Advanced DMEM/F12 supplemented with Glutamax, HEPES, B27, recombinant EGF (50 ng/mL), recombinant NOGGIN (100 ng/mL), recombinant RSPO1 (500 ng/mL) and Y27632 (10 μM). Sorted cells from each population were seeded in three independent Matrigel drops (1500 cells/drop). Media was changed every 48 hours and organoid formation was quantified 10 days after seeding. To assess organoid formation in stem cell conditions, CHIR99021 (3 μm) and Valproic acid (1 mM) were added to the cultures. To test the self-renewal of Mex3a-high Lgr5-high cells, established organoids were serially passaged in Matrigel every 7 days and kept in culture for over 8 passages. To test for multipotency, Mex3a-high Lgr5-high derived organoids were formalin fixed and paraffin embedded. Sections were then stained for EPHB2, LYZ1, CHGA, ANPEP, MUC2 and DCLK1 with standard IF protocols (see below for details and refer to Table S7 for antibodies and concentrations used).

*RT qPCR analysis:*

RNA was extracted from sorted cells by using the Trizol Reagent (Life technologies) following manufacturer´s instructions. The only modification was the addition of glycogen (40 ng/mL, Roche) to the aqueous phase to visualize the RNA pellet after precipitation. RNA was quantified using a Nanodrop and cDNA was produced with the High-Capacity cDNA Reverse Transcription kit (Applied Biosystems) following manufacturer´s instructions. For experiments with low cellular yield, cells were directly sorted on lysis buffer and subjected to cDNA amplification as previously described (Gonzalez-Roca et al., 2010). To assess changes in expression of selected genes RT-qPCR was done with their respective TaqMan probes or primers for SYBR Green. For TaqMan assays the TaqMan Universal PCR Master Mix (Applied Biosystems) was used. In the case of SYBR Green reactions the Power SYBR Master Mix (Applied Biosystems) was used. To calculate gene expression to a normalization gene the comparative threshold cycle method was used. See **Table S7** for a list of Taqman assays and SYBR primers used.

*GO Analysis:*

A signature for Mex3a-high cells was derived by filtering genes differentially expressed with a p value < 0.05 and a fold change > 2 fold compared to Mex3a-low cells. The previously defined Lgr5 mRNA cell signature (Muñoz et al., 2012) was used. These signatures were then analyzed with the DAVID analysis platform and the GO biological process enrichment was retrieved for each signature (Huang da et al., 2009).

*Immunohistochemistry:*

*Tissue fixation:* Intestines were processed and cut longitudinally to expose villi. They were fixed in formalin for 24h at room temperature and later processed for paraffin embedding as previously described (Merlos-Suárez et al., 2011).

*Tissue staining:* Paraffin sections were processed for standard histological staining as previously described (Merlos-Suárez et al., 2011). Antibody concentrations for IHC are shown in **Table S7**. For the detection of tdTomato by immunofluorescence (IF), primary anti tdTomato antibody was incubated overnight at room temperature followed by an incubation with HRP-coupled polymer (Envision+). After this incubation, a tyramide-biotin conjugation was performed and later stained with streptavidin-Alexa532 following manufacturer´s instructions (Perkin Elmer). β-galactosidase staining was performed as previously described (Barker et al., 2007).

*GSEA Analysis:*

We ranked all genes measured in Mex3a-high and Mex3a-low cells by microarrays according to differential expression between the two populations. We then assessed the enrichment of the LRC signature (LRC genes > 2 fold, p value < 0.05 when compared to both Paneth and crypt cycling cells. For details refer to Table S4) between phenotypes against the ranked file using GSEA. To validate the statistical relevance of the enrichment, the enrichment score of 1000 random signatures of matching size was computed and compared to the signature of interest. GSEA analysis was performed using the implementation from the Broad Institute (Merlos-Suárez et al., 2011; Subramanian et al., 2005).

*In situ hybridization*:

Single-molecule in situ hybridization was performed using Advanced Cell Diagnostics RNAscope 2.5 HD detection kit (REF 322360) with a probe against Mex3a RNA (REF

318531). Small intestines from 4-month old mice were formalin fixed for 24 hours, paraffin-embedded and cut into 5-micron thick slices. The single-molecule ISH protocol was followed according to manufacturer's instructions with the following modifications: 1) after deparafinization and rehydration, tissue was fixed in 10% formaldehyde/PBS for 16 hours prior to pretreatment steps, 2) protease treatment for 20 minutes, and 3) Amp5 incubation for 20 minutes. Images were acquired with an Olympus BX43 upright microscope and DP73 color camera.

*Transcriptomic analysis:*

25 ng of RNA from sorted cells were used to hybridize mouse M430 PM strip microarrays (Affymetrix). For Lgr5 colon cells, a pool of proximal colon epithelium from 5 different mice was used to obtain the necessary material. In the case of Mex3a and Mex3a/Lgr5 populations, proximal small intestine epithelium was used. All samples were age and sex matched. Affymetrix arrays were normalized using RMA background correction and summarization (Irizarry et al., 2003) as implemented in the "affyPLM" package from the R statistical framework . Annotations for the "HT-430_PM" array version na34 were downloaded from Affymetrix. Technical metrics capturing RNA degradation, array saturation, noise floor and quantity of starting material were computed for each sample as described in (Eklund and Szallasi, 2008). All samples passed quality controls. A linear model was fitted to find differentially expressed genes between conditions of interest with technical metrics and scan strip as covariates. The "lmFit" function from the "limma" package (Ritchie et al., 2015) was used for fitting the model.

*Single cell RNA sequencing:*

A single cell preparation from the small intestine of Lgr5 GFP/+ mice was obtained as described above. After stringent gating, the top 25% of GFP+ cells were collected. Isolation of single cells (400 capture sites), cDNA synthesis and amplification were performed on the C1 mRNA Seq HT IFCs medium chip (10-17µm) (Fluidigm) with the SMARTer Ultra Low RNA Kit (Clonetech). Spike-in artificial transcripts (ERCC RNA Spike-In Mix, Ambion) from The External RNA Controls Consortium were added during sample preparation as internal control to support subsequent quality control steps (Baker et al., 2005). Quantity and quality of cDNA were assessed with a High Sensitivity DNA Chip (Agilent Technologies). Libraries were generated by Nextera XT DNA Sample Prep Kit (Illumina) and sequenced on a HiSeq 2500 (Illumina) using 100-bp mode of Truseq Rapid SBS kit v2. Sequencing was carried out as paired-end reads. Read one (12 bp) contained cell and unique molecular barcodes (Islam et al., 2014;

Kivioja et al., 2012); read two (85 bp) contained the transcript sequence. Primary data analysis was carried out with the standard Illumina pipeline.

Fluidigm C1 HT-Seq technology includes a two-level indexing system, which allows the multiplexed sequencing of multiple single cells and library pools. Here, we processed 10 libraries (up to 40 cells per library) of Lgr5-high cells, with a total number of 546,853,233 reads. Quality check was performed with FastQC quality control suite. All samples reached the quality standards and were processed to deconvolute reads by demultiplexing library and cell barcodes. Prior to mapping, reads were filtered for poly-T sequences. Then, reads were aligned to the mouse reference genome GRCm38.p4 (Cunningham et al., 2015) using Gencode M8 annotations (Mudge and Harrow, 2015) with the GEMTools 1.7.0 RNASeq pipeline (Marco-Sola et al., 2012). In-house scripts were used for gene expression quantification based on molecular identifiers. ERCC spike-ins were mapped and quantified in parallel.

QUANTIFICATION AND STATISTICAL ANALYSIS

*General experimental design:*

For all experiments n represents the number of independent biological replicates. In mice experiments, all groups consisted of age and sex-matched littermates. No blinding nor sample-size estimations were incorporated to the experimental setup and all data points were included in the analyses (i.e. no exclusion criteria was used). For the experiments of 5-FU and IR treatments, mice were randomized.

General statistical analysis

Graphs and statistical analysis for figures 1, 2, 3, 4, 5 and 7 were done with GraphPad Prism 7.0. In these figures n indicates the number of biological replicates (populations obtained from independent mice). For panels 2F and 2H, significant differences were assessed by a t-test. For panel 7B, a one-way ANOVA was used followed by Sidak´s multiple comparison test. For panels 5F and 7D, a two-way ANOVA was used followed by Tukey´s multiple comparison test. Asterisks in figures indicate the following: *, p value < 0.05; **, p value < 0.01; ***, p value < 0.001. n.s., p value > 0.05. Each experiment

Single cell transcriptomic analysis

*Filtering and preprocessing of SCS data*

Cells with a percentage of reads aligning to spikes larger than 40% and exon percentage lower than 30% were removed from the dataset. 10 out of 400 cells were filtered by this criterion.

The gene expression *y* for a cell with vector of UMI counts *x* was computed as

$$y = \log 2\left(\frac{(x+1)}{sum(x+1)} * 1e^{6}\right)$$

and quantile normalization was applied to the resulting matrix. The expression matrix was further filtered by removing genes with expression in less than 4 cells and cells with less than 1000 genes expressed. The minimum expression was set at 4.5.

Doublets and empty wells were detected by visual inspection of microscopy images (81 doublets and 24 empty). In order to remove any undetected doublet we performed unsupervised clustering and looked for clusters enriched in these categories. First, tSNE coordinates (van der Maaten and Hinton, 2008) were computed without removing any artifact and clusters were found using Normal Mixture Modeling as implemented in the R package “Mclust” with default parameters and variable number of clusters. For each cluster we computed the median of the classification uncertainty for empty and non-empty cells. Clusters for which this value for empty cells was lower than for non-empty cells were classified as enriched (3 out of 7 clusters). All cells belonging to this cluster with certainty larger than 0.95 were marked as empty wells (65 of 390).

A similar approach was followed to remove doublets: starting with the dataset without low quality or empty cells, clusters were found and defined to be enriched in doublets with the same criterion as before except that clusters were only considered to be enriched if they contained more than 8% of cells marked as doublets. 2 out of 7 clusters were classified as enriched and 55 cells were removed from the dataset.

*Analysis of final dataset*

The final dataset consisted of 245 cells after removing visually detected and predicted artifacts. The mean number of UMIs per cell was 31630 (min=4459, median=25920 and max=70040). Count data was normalized as described before and principal components were computed using the most variable and expressed genes (top 11% of genes according to the median absolute deviation and mean expression). For plotting

of the dataset, dimension reduction was performed through tSNE using the “Seurat” R package (Satija et al., 2015). In order to detect possible heterogeneity in the data, clustering was computed on the first 3 principal components or in the tSNE coordinates accordingly. The optimal number of clusters was decided according to the Bayesian Information Criterion for expectation-maximization as implemented in the Mclust function. 2 clusters were found in both PCA and tSNE analysis. For PCA analysis, clusters 1 and 2 were of sizes 140 and 105 cells respectively. For tSNE analysis, clusters 1 and 2 were of 165 and 80 cells respectively. Gene set enrichment analysis (GSEA) as implemented in (Subramanian et al., 2005) was performed on ranked lists by fold change of expression between clusters. Fold changes and p-value for differential expression were computed with a bimodal model as implemented in the function “find.markers” from the R package Seurat.

*Robustness analysis of clusters*

Downsampling of the total number of UMIs per cell was performed in order to investigate the robustness of the resulting clusters. Datasets were simulated 20 times for total number of UMIs ranging from 10 to 90% of the original value. Unsupervised clustering was repeated as with the original data and the number of clusters and expression was stored. The number of clusters remained relatively constant for all simulations (**Figure S6A**). A random forest classifier was fitted to the simulated data and original cluster labels. The number of individuals may be important for cluster discovery therefore we downsampled our dataset choosing subsets of cells and repeating the process. We found that with fewer than 50% of the cells the second cluster was lost in half of the simulations and almost completely with 30% of the cells (**Figure S6B**). The out of bag error of classification is depicted (**Figure S6C,S6D**), showing that most cells are correctly classified even with 10% of reads or cells analyzed.

## DATA AND SOFTWARE AVAILABILITY

Microarray data from Mex3a/Lgr5 subpopulations has been deposited in the GEO database with the accession number GSE90629.

Single cell transcriptomics of Lgr5-high cells has been deposited in the GEO database with the accession number GSE90856.

**KEY RESOURCES TABLE**

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---|---|---|
| Antibodies | | |
| Rabbit anti tdTomato | Rockland | Cat# 600-401-379 |
| Rabbit anti GFP | Abcam | Cat# ab6556 |
| Goat anti GFP | Abcam | Cat# ab6673 |
| Mouse anti BrdU | BD | Cat# 347580 |
| Rabbit anti Lysozyme | Dako | Cat# A0099 |
| Goat anti Chromogranin A | Santa Cruz | Cat# sc-1488 |
| Rabbit anti Dclk1 | Abgent | Cat# AP7291b |
| Goat anti Anpep | RD systems | Cat# AF2335 |
| Rabbit anti Muc2 | Santa Cruz | Cat# sc-15334 |
| Goat anti EphB2 APC | BD | Cat# 564699 |
| Chemicals, Peptides, and Recombinant Proteins | | |
| Ulex Europaeus Agglutinin I, FITC labeled | Sigma | Cat# L9006 |
| 5-fluorouracil | Sigma | Cat# F6627 |
| Tamoxifen | Sigma | Cat# T5648 |
| 4-OH Tamoxifen | Sigma | Cat# H6278 |
| Beta-napthoflavone | Sigma | Cat# N3633 |
| 5-ethynyl-2´-deoxyuridine (EdU) | Life technologies | Cat# A10044 |
| 5-bromo-2´-deoxyuridine (BrdU) | Sigma | Cat# B5002 |
| Ethylenediaminetetracetic acid (EDTA) | Sigma | Cat# E5134 |
| Dispase | ThermoFisher | Cat# 17105-041 |
| Trizol | Life technologies | Cat# 15596018 |
| TaqMan Universal PCR Master Mix | Applied Biosystems | Cat# 4364341 |
| SYBR Green PCR Master Mix | Applied Biosystems | Cat# 4368706 |
| Matrigel | BD | Cat# 356231 |
| Recombinant human EGF | Peprotech | Cat# AF-100-15 |
| Recombinant human RSPO1 | In house | n/a |
| Recombinant human NOGGIN | In house | n/a |
| CHIR99021 | Stemgent | Cat# 04-0004-02 |
| Y27632 | Sigma | Cat# Y0503 |
| Valproic Acid | Sigma | Cat# P4543 |
| Critical Commercial Assays | | |
| RNAscope 2.5 HD detection kit | ACD | Cat# 322360 |
| High-Capacity cDNA Reverse Transcription kit | Applied Biosystems | Cat# 4368813 |
| C1 Single-Cell mRNA Seq HT IFC, 10–17 µm | Fluidigm | Cat# 101-0222 |
| SMART-Seq v4 Ultra® Low Input RNA Kit | Clontech | Cat# 635025 |
| Nextera XT DNA Sample Prep Kit | Illumina | Cat# FC-131-1096 |
| Deposited Data | | |
| Microarray analysis of Mex3a Lgr5 populations | This paper | GEO90629 |
| Single cell RNA Seq of Lgr5-high cells | This paper | GEO90856 |
| Experimental Models: Organisms/Strains | | |
| Mex3a$^{tm1(tdTomato-T2A-CreERT2)EBa}$ | This paper | n/a |
| Lgr5$^{tm1(cre/ERT2)Cle}$ | Barker et al, 2007 | JAX Stock nº: 008875 |
| Gt(ROSA)26Sor$^{tm1Sor}$ | Soriano et al, 1999 | JAX Stock nº: 003309 |
| Gt(ROSA)26Sor$^{tm4(ACTB-tdTomato,-EGFP)Luo}$ | Muzumdar et al, 2007 | JAX Stock nº: 007576 |
| Tg(Cyp1a1-H2B-YFP)Dwi | Buczacki et al, 2013 | n/a |
| Sequence-Based Reagents | | |
| ISH Mex3a probe | ACD | Cat# 318351 |
| Genotyping and RT qPCR assays | Table S7 | |

| *Software and Algorithms* | | |
|---|---|---|
| affyPLM package | | (Irizarry et al., 2003) |
| lmFit | limma package | (Ritchie et al., 2015) |
| GEMTools 1.7.0 RNASeq pipeline | | (Marco-Sola et al., 2012) |
| tSNE analysis | | (van der Maaten and Hinton, 2008) |
| Seurat | R package | (Satija et al., 2015) |

Figure 1

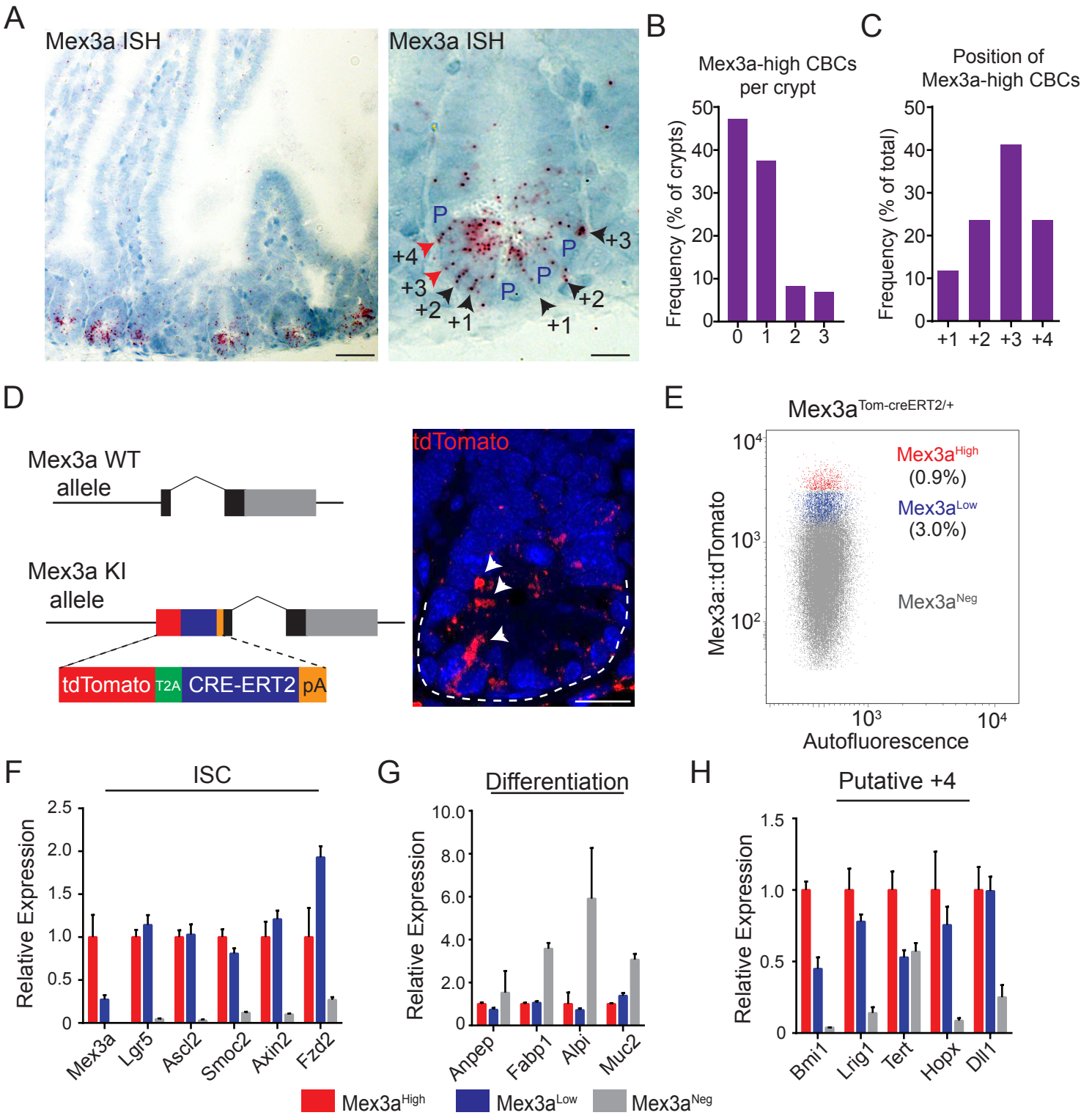

A
Mex3a ISH
Mex3a ISH
P
+4
+3
+3
+2
+1
B
Mex3a-high CBCs per crypt
Frequency (% of crypts)
C
Position of Mex3a-high CBCs
Frequency (% of total)
D
Mex3a WT allele
Mex3a KI allele
tdTomato
T2A
CRE-ERT2
pA
tdTomato
E
Mex3a Tom-creERT2/+
Mex3a::tdTomato
Mex3a High (0.9%)
Mex3a Low (3.0%)
Mex3a Neg
Autofluorescence
F
ISC
Relative Expression
Mex3a
Lgr5
Ascl2
Smoc2
Axin2
Fzd2
G
Differentiation
Anpep
Fabp1
Alpi
Muc2
H
Putative +4
Bmi1
Lrig1
Tert
Hopx
Dll1
Mex3a High
Mex3a Low
Mex3a Neg
I
Mex3a
Lgr5
Cellular component morphogenesis
Cell fate specification
Epithelium development
Intracellular signaling cascade
Mitotic cell cycle
Positive regulation of cell proliferation
Regulation of sodium ion transport
Wnt receptor signaling pathway
Positive regulation of cellular biosynthetic process
Negative regulation of cell proliferation
Response to wounding
Hormone metabolic process
Defense response
Regulation of secretion
Common
Lgr5-specific
Mex3a-specific
ns
>1.3
>2.0
>3.0
>4.0
log10(1/P)

Figure 2

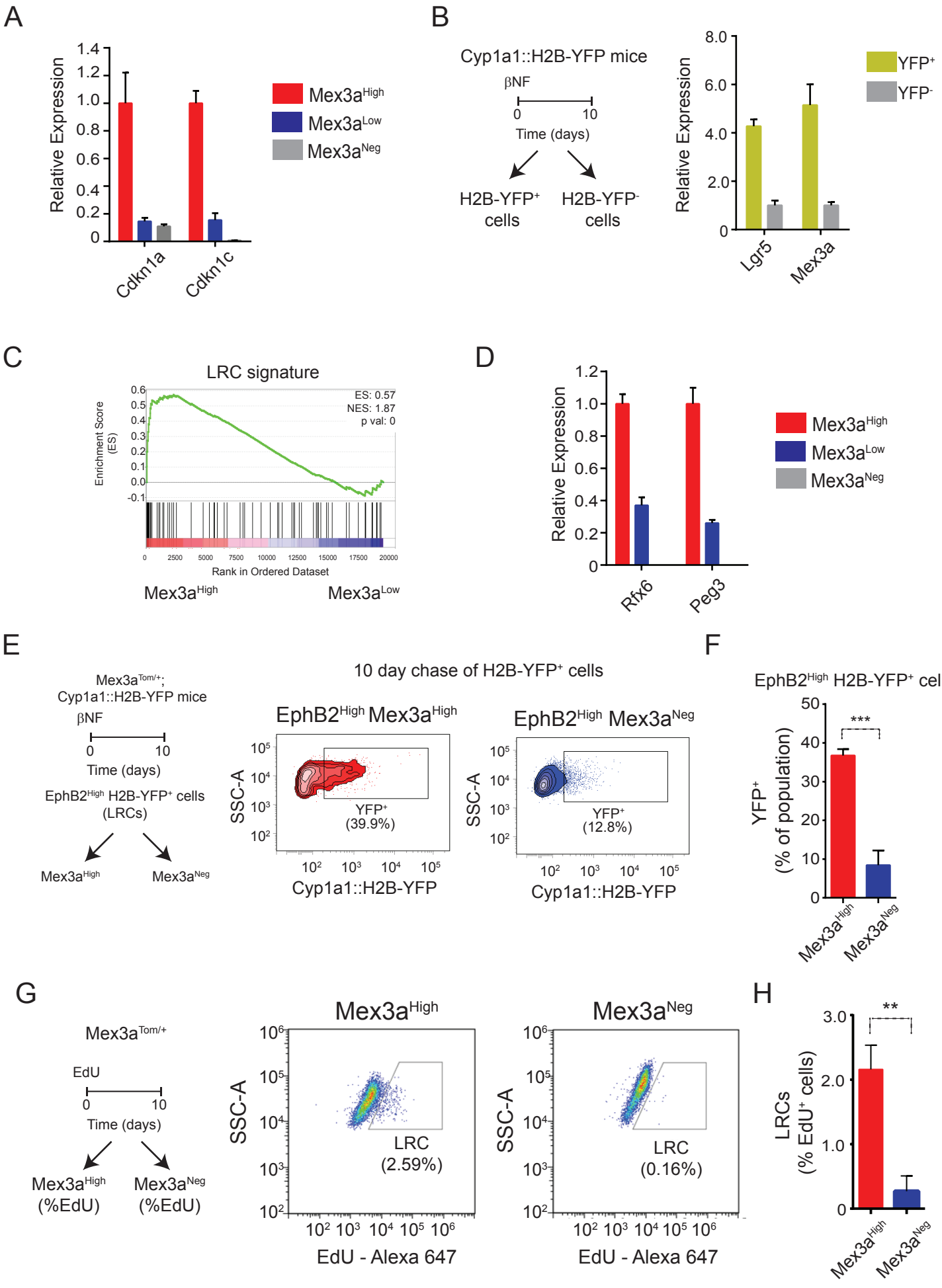

A
Relative Expression
Mex3a^High
Mex3a^Low
Mex3a^Neg
Cdkn1a
Cdkn1c
B
Cyp1a1::H2B-YFP mice
βNF
Time (days)
H2B-YFP+ cells
H2B-YFP- cells
YFP+
YFP-
Lgr5
Mex3a
C
LRC signature
ES: 0.57
NES: 1.87
p val: 0
Enrichment Score (ES)
Rank in Ordered Dataset
D
Rfx6
Peg3
E
Mex3a^Tom/+; Cyp1a1::H2B-YFP mice
EphB2^High H2B-YFP+ cells (LRCs)
10 day chase of H2B-YFP+ cells
EphB2^High Mex3a^High
EphB2^High Mex3a^Neg
YFP+ (39.9%)
YFP+ (12.8%)
SSC-A
Cyp1a1::H2B-YFP
F
EphB2^High H2B-YFP+ cells
YFP+ (% of population)
***
G
Mex3a^Tom/+
EdU
Mex3a^High (%EdU)
Mex3a^Neg (%EdU)
LRC (2.59%)
LRC (0.16%)
EdU - Alexa 647
H
LRCs (% EdU+ cells)
**

Figure 3

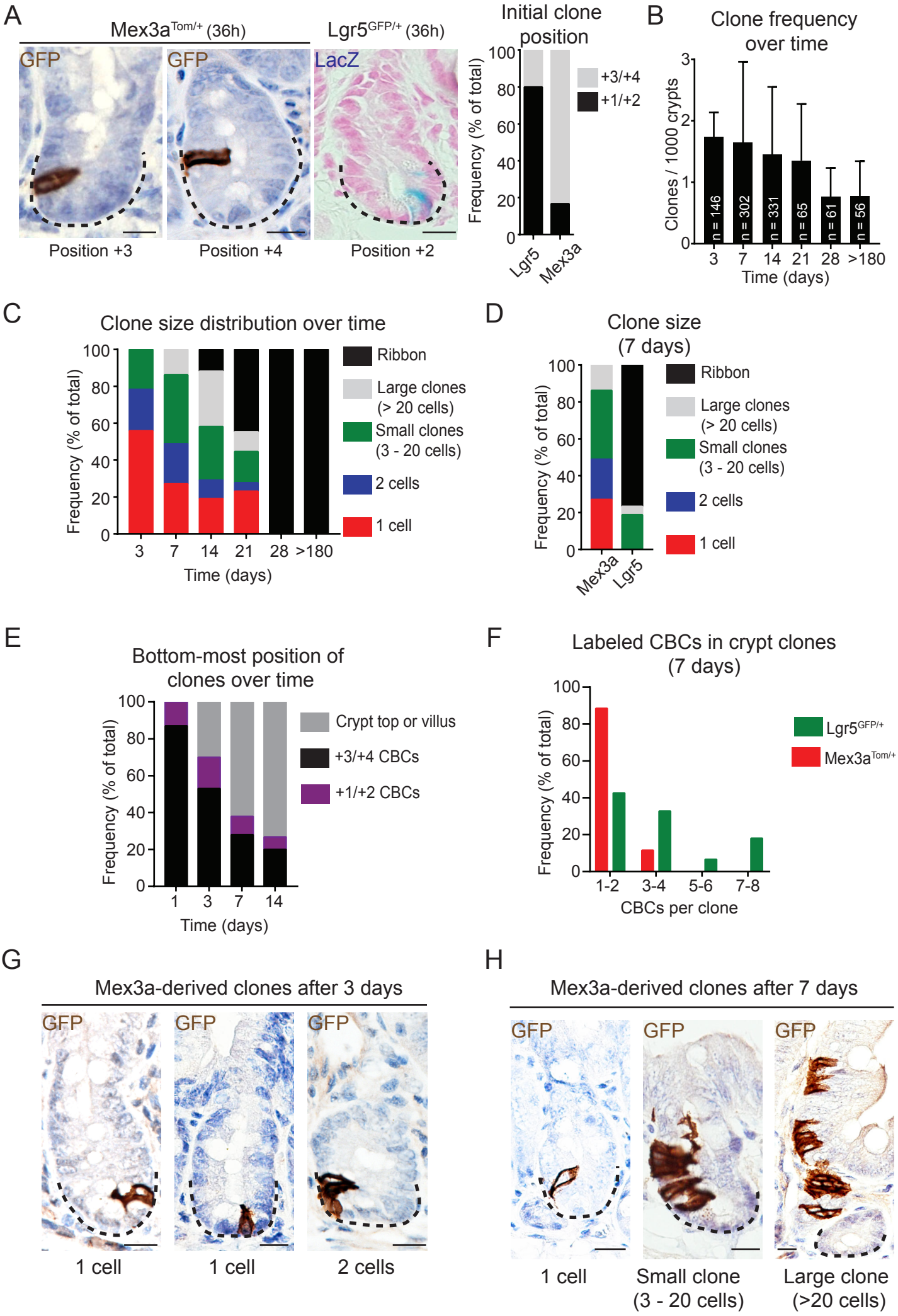

A
Mex3a^Tom/+ (36h)
Lgr5^GFP/+ (36h)
GFP
GFP
LacZ
Position +3
Position +4
Position +2
Initial clone position
Frequency (% of total)
+3/+4
+1/+2
Lgr5
Mex3a
B
Clone frequency over time
Clones / 1000 crypts
n = 146
n = 302
n = 331
n = 65
n = 61
n = 56
3 7 14 21 28 >180
Time (days)
C
Clone size distribution over time
Frequency (% of total)
Ribbon
Large clones (> 20 cells)
Small clones (3 - 20 cells)
2 cells
1 cell
3 7 14 21 28 >180
Time (days)
D
Clone size (7 days)
Frequency (% of total)
Ribbon
Large clones (> 20 cells)
Small clones (3 - 20 cells)
2 cells
1 cell
Mex3a
Lgr5
E
Bottom-most position of clones over time
Frequency (% of total)
Crypt top or villus
+3/+4 CBCs
+1/+2 CBCs
1 3 7 14
Time (days)
F
Labeled CBCs in crypt clones (7 days)
Frequency (% of total)
Lgr5^GFP/+
Mex3a^Tom/+
1-2 3-4 5-6 7-8
CBCs per clone
G
Mex3a-derived clones after 3 days
GFP
GFP
GFP
1 cell
1 cell
2 cells
H
Mex3a-derived clones after 7 days
GFP
GFP
GFP
1 cell
Small clone (3 - 20 cells)
Large clone (>20 cells)

Figure 4

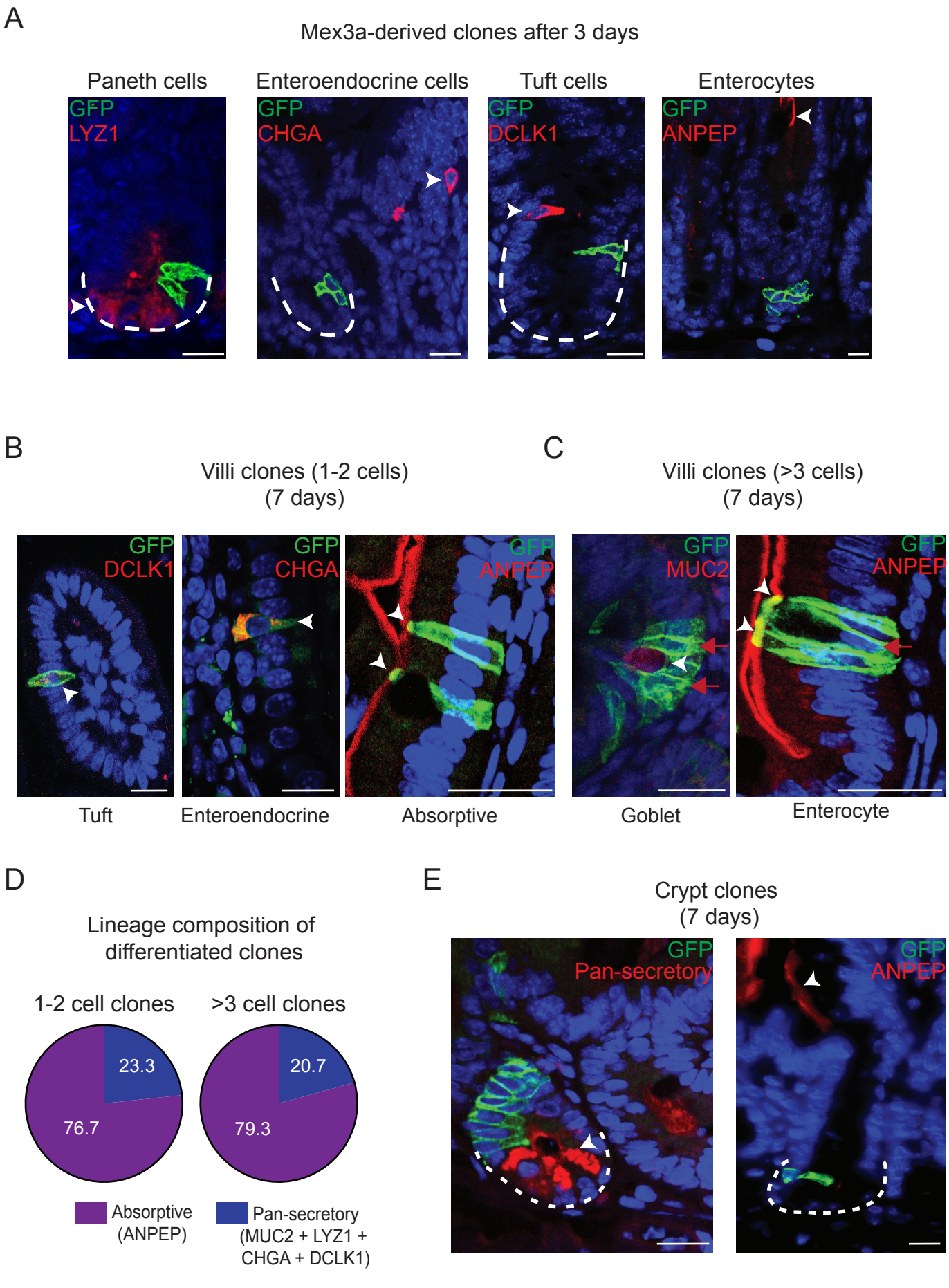

A
Mex3a-derived clones after 3 days
Paneth cells
Enteroendocrine cells
Tuft cells
Enterocytes
GFP
LYZ1
GFP
CHGA
GFP
DCLK1
GFP
ANPEP
B
Villi clones (1-2 cells)
(7 days)
C
Villi clones (>3 cells)
(7 days)
GFP
DCLK1
GFP
CHGA
GFP
ANPEP
GFP
MUC2
GFP
ANPEP
Tuft
Enteroendocrine
Absorptive
Goblet
Enterocyte
D
Lineage composition of
differentiated clones
1-2 cell clones
>3 cell clones
23.3
76.7
20.7
79.3
Absorptive
(ANPEP)
Pan-secretory
(MUC2 + LYZ1 +
CHGA + DCLK1)
E
Crypt clones
(7 days)
GFP
Pan-secretory
GFP
ANPEP

Figure 5

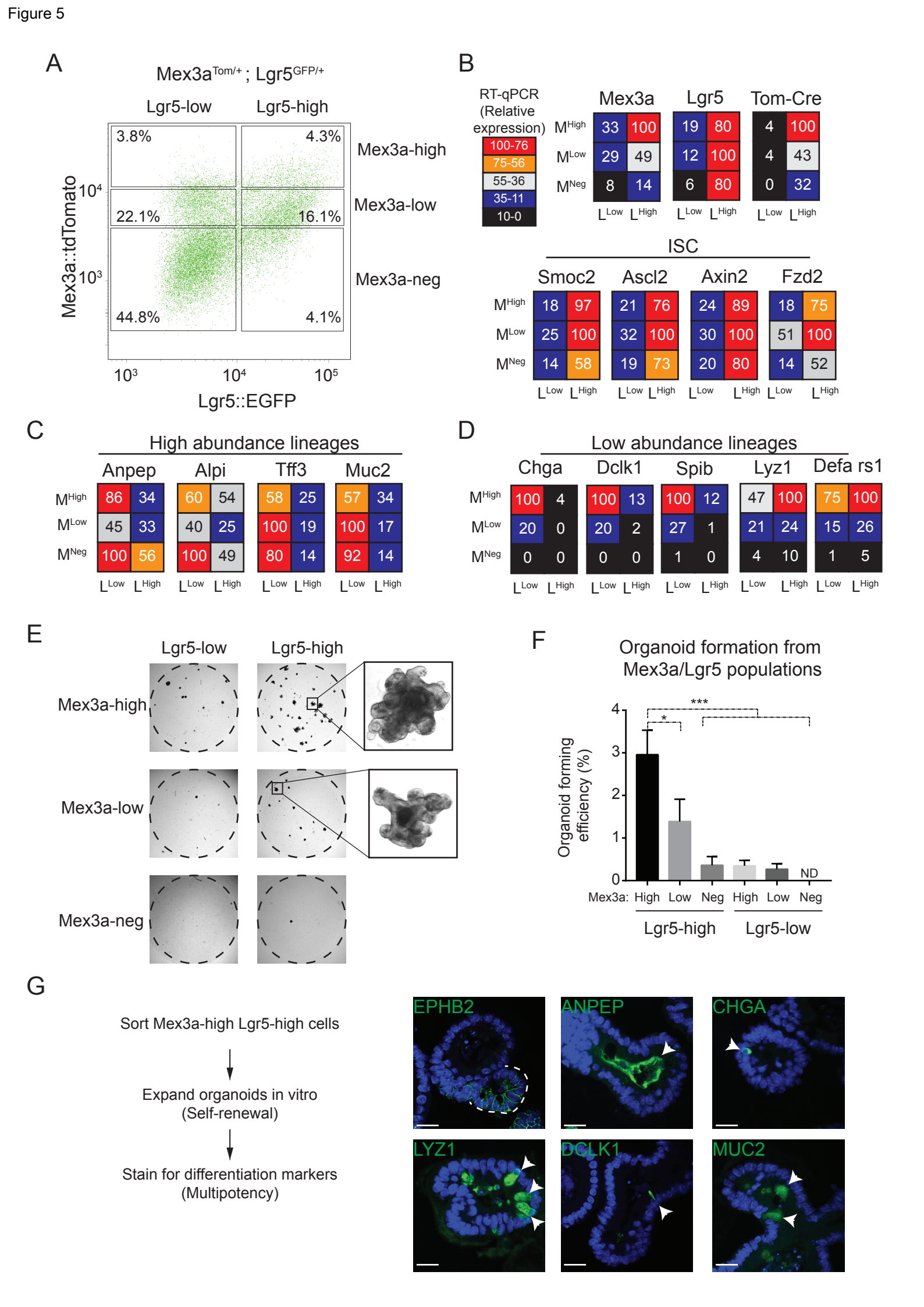

A
Mex3a^Tom/+ ; Lgr5^GFP/+
Lgr5-low
Lgr5-high
3.8%
4.3%
22.1%
16.1%
44.8%
4.1%
Mex3a-high
Mex3a-low
Mex3a-neg
Mex3a::tdTomato
Lgr5::EGFP
B
RT-qPCR (Relative expression)
100-76
75-56
55-36
35-11
10-0
Mex3a
Lgr5
Tom-Cre
ISC
Smoc2
Ascl2
Axin2
Fzd2
C
High abundance lineages
Anpep
Alpi
Tff3
Muc2
D
Low abundance lineages
Chga
Dclk1
Spib
Lyz1
Defa rs1
E
Lgr5-low
Lgr5-high
Mex3a-high
Mex3a-low
Mex3a-neg
F
Organoid formation from Mex3a/Lgr5 populations
Organoid forming efficiency (%)
ND
Mex3a: High Low Neg High Low Neg
Lgr5-high
Lgr5-low
G
Sort Mex3a-high Lgr5-high cells
Expand organoids in vitro (Self-renewal)
Stain for differentiation markers (Multipotency)
EPHB2
ANPEP
CHGA
LYZ1
DCLK1
MUC2

Figure 6

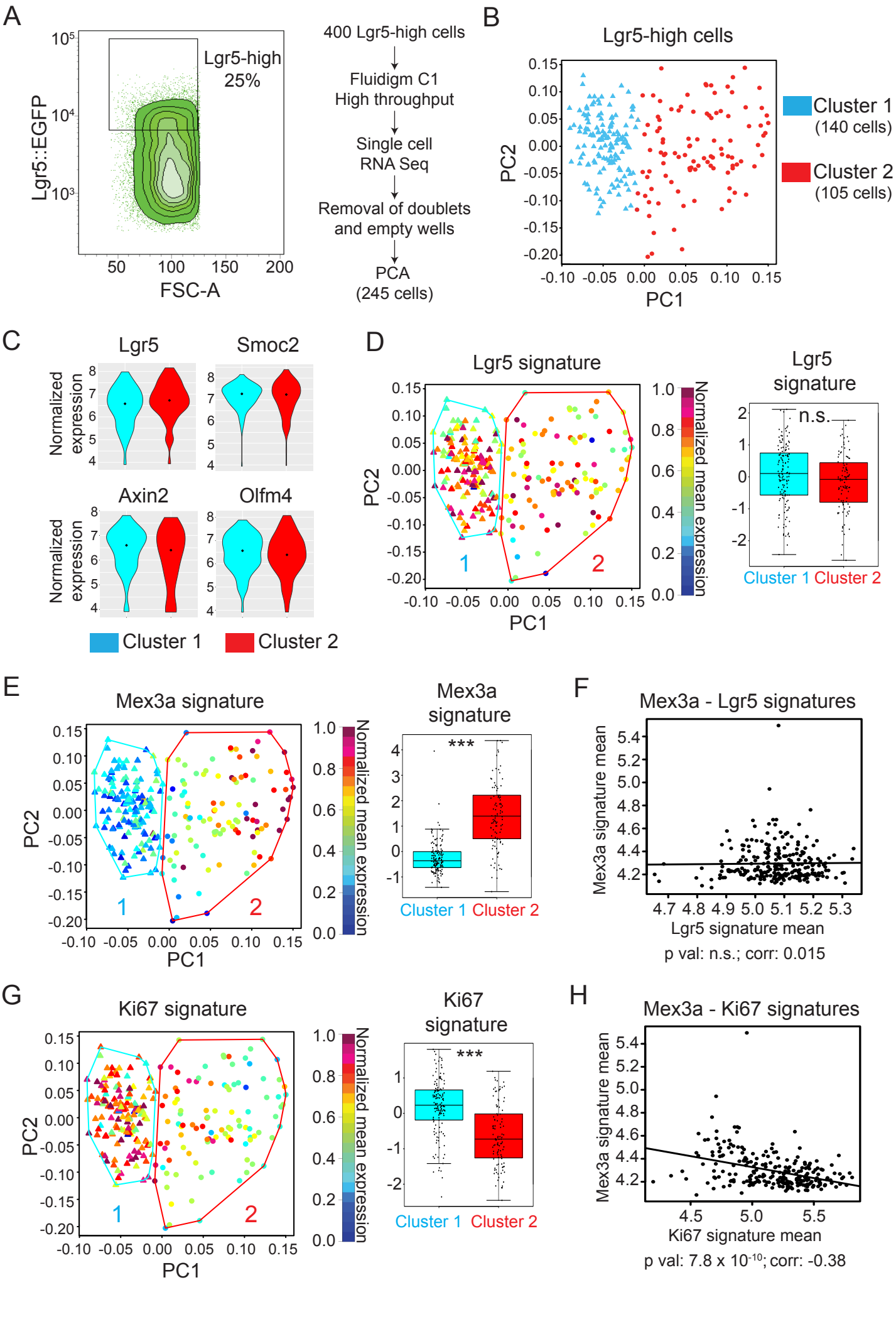
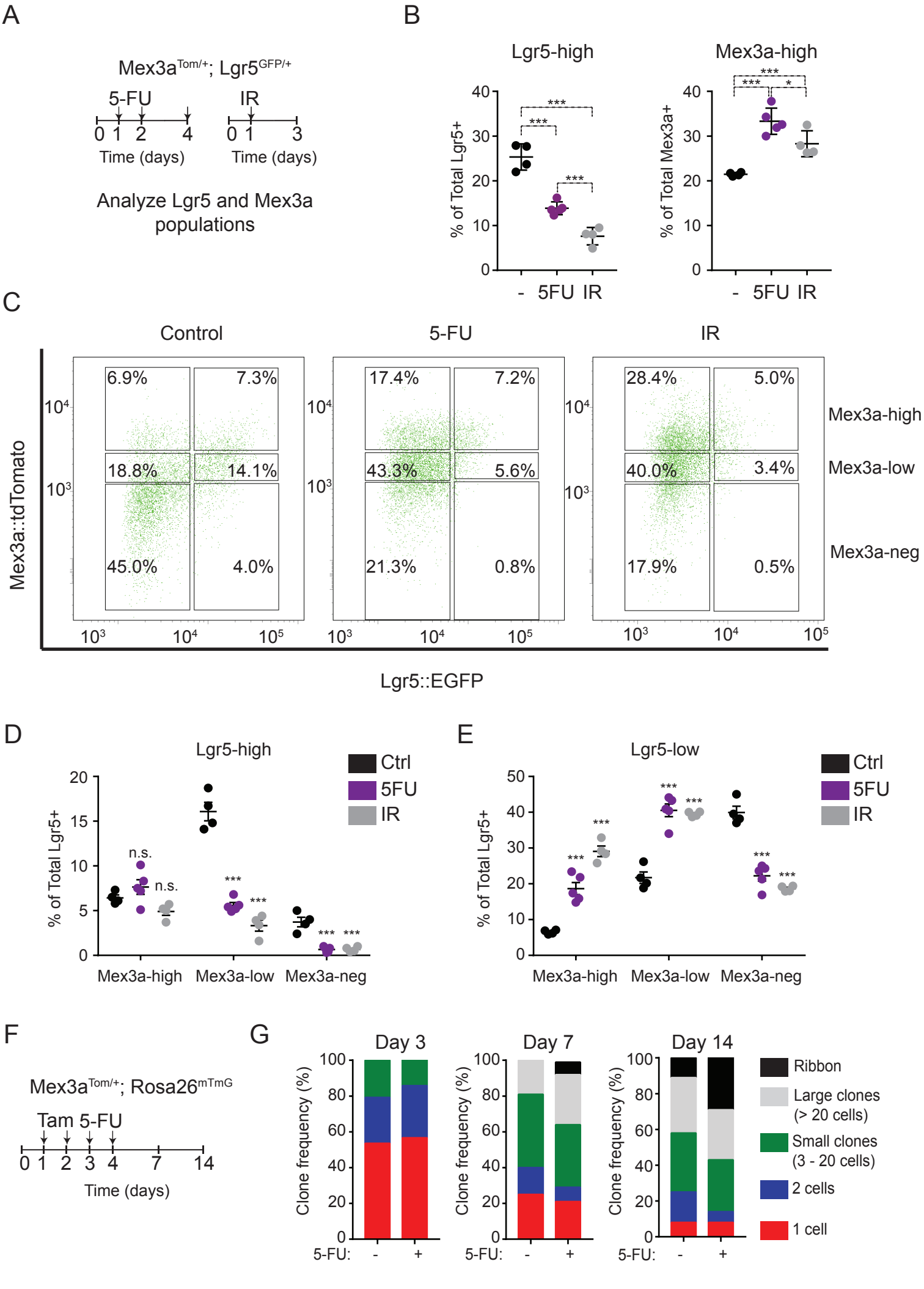

A
Lgr5-high
25%
Lgr5::EGFP
FSC-A
400 Lgr5-high cells
Fluidigm C1
High throughput
Single cell
RNA Seq
Removal of doublets
and empty wells
PCA
(245 cells)
B
Lgr5-high cells
PC1
PC2
Cluster 1
(140 cells)
Cluster 2
(105 cells)
C
Lgr5
Smoc2
Axin2
Olfm4
Normalized expression
Cluster 1
Cluster 2
D
Lgr5 signature
Normalized mean expression
Lgr5
signature
n.s.
Cluster 1 Cluster 2
E
Mex3a signature
Mex3a
signature
***
Cluster 1 Cluster 2
F
Mex3a - Lgr5 signatures
Mex3a signature mean
Lgr5 signature mean
p val: n.s.; corr: 0.015
G
Ki67 signature
Ki67
signature
***
Cluster 1 Cluster 2
H
Mex3a - Ki67 signatures
Mex3a signature mean
Ki67 signature mean
p val: 7.8 x 10-10; corr: -0.38

Figure 7


A
Mex3a^Tom/+; Lgr5^GFP/+
5-FU
IR
0 1 2 4
0 1 3
Time (days)
Time (days)
Analyze Lgr5 and Mex3a populations
B
Lgr5-high
Mex3a-high
% of Total Lgr5+
% of Total Mex3a+
- 5FU IR
C
Control
5-FU
IR
Mex3a::tdTomato
Lgr5::EGFP
6.9%
7.3%
18.8%
14.1%
45.0%
4.0%
17.4%
7.2%
43.3%
5.6%
21.3%
0.8%
28.4%
5.0%
40.0%
3.4%
17.9%
0.5%
Mex3a-high
Mex3a-low
Mex3a-neg
D
Lgr5-high
Ctrl
5FU
IR
% of Total Lgr5+
n.s.
Mex3a-high
Mex3a-low
Mex3a-neg
E
Lgr5-low
F
Mex3a^Tom/+; Rosa26^mTmG
Tam 5-FU
0 1 2 3 4 7 14
Time (days)
G
Day 3
Day 7
Day 14
Clone frequency (%)
5-FU:
Ribbon
Large clones (> 20 cells)
Small clones (3 - 20 cells)
2 cells
1 cell

# Supplemental Information

## Mex3a marks a slowly dividing subpopulation of Lgr5+ intestinal stem cells

Francisco M. Barriga, Elisa Montagni, Miyeko Mana, Maria Mendez-Lago, Xavier Hernando-Momblona , Marta Sevillano, Amy Guillaumet-Adkins,Gustavo Rodriguez-Esteban, Simon J. A. Buczacki, Marta Gut, Holger Heyn, Douglas J. Winton, Omer H. Yilmaz, Camille Stephan-Otto Attolini, Ivo Gut and Eduard Batlle

**Supplementary Table List**

**Table S1** – Gene expression signatures used to define robust ISC genes (Related to Figure 1)

**Table S2** – Genes enriched in intestinal Mex3a-high cells (Related to Figure 1)

**Table S3** – GO categories enriched in intestinal Mex3a-high and Lgr5-high cells (Related to Figure 1)

**Table S4** – Gene enriched in intestinal label-retaining cells (Related to Figure 2)

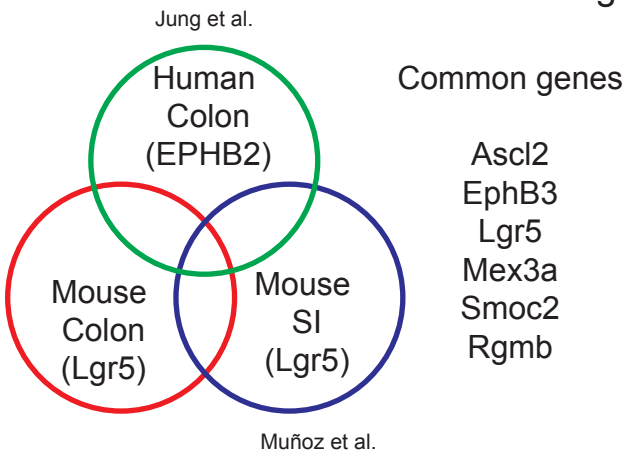
**Table S5** – Full RT qPCR relative expression data from Mex3a/Lgr5 subpopulations (Related to Figure 5)

**Table S6** – Gene expression signatures used for transcriptional analysis in single cell RNA Seq (Related to Figures 6 and S6)

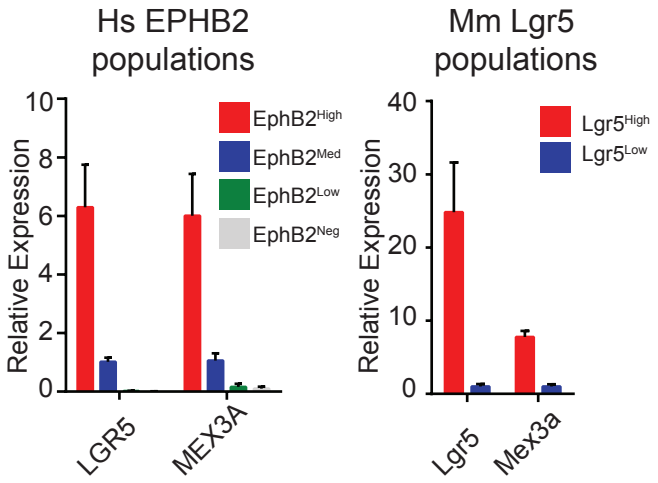
**Table S7** – List of primers, Taqman probes and antibodies (Related to STAR methods)

Figure S1 - Related to Figure 1

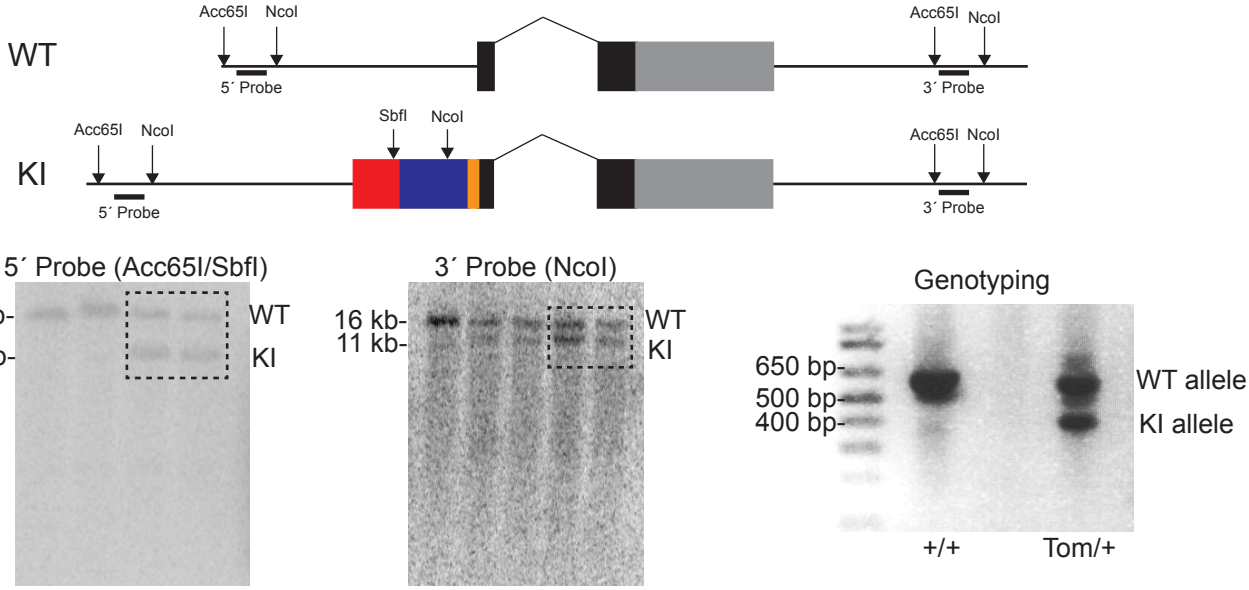
A




B




C




D

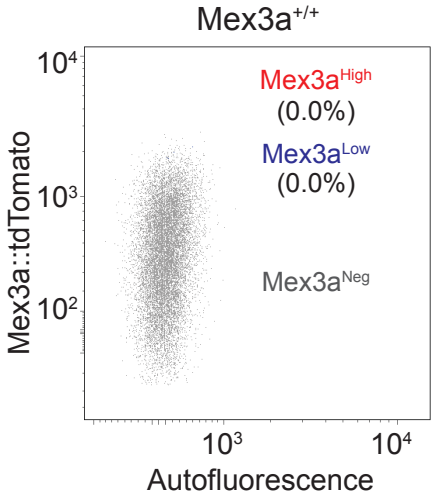


Figure S1 – Identification of Mex3a as a robust ISC gene and generation of the Mex3a Tom/+ allele (Related to Figure 1).

(A) Overlap of intestinal signatures identifies Mex3a as a novel and robust ISC-specific gene. Comparison of Lgr5- derived signatures of mouse small intestine, mouse colon and EphB2-derived human colonic stem cell signatures (Refer to Table S1).

(B) RT qPCR analysis of Lgr5 and Mex3a expression in intestinal populations defined by EPHB2 (left panel) and Lgr5 expression (right panel). Bars depict the mean and upper/lower limits of relative expression obtained from a representative sorting experiment.

(C) Generation of the Mex3a knock-in allele reporter mice. Depiction of the WT and targeted locus of mouse Mex3a, as well as Southern blots for the 5´and 3´arms of the insertion. A representative genotyping PCR is shown.

(D) Representative FACS plot of small intestine crypt cells obtained from a Mex3a+/+ mouse.

A

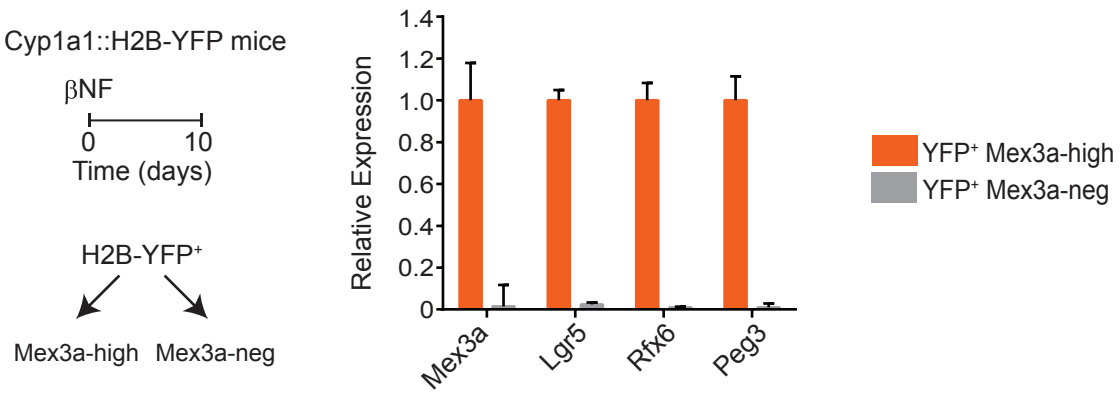



Figure S2 – Mex3a-high cells enrich for LRC genes within the H2B-YFP+ retaining population (Related to Figure 2).

(A) Mex3a identifies LRCs within the YFP+ intestinal population. Cyp1a1::H2B-YFP mice were induced and chased for 10 days. YFP+ cells were divided in Mex3a-high and Mex3a-neg populations and analyzed by RT qPCR for ISC and LRC markers. Bars depict the mean and upper/lower limits of relative expression obtained from a representative sorting experiment.

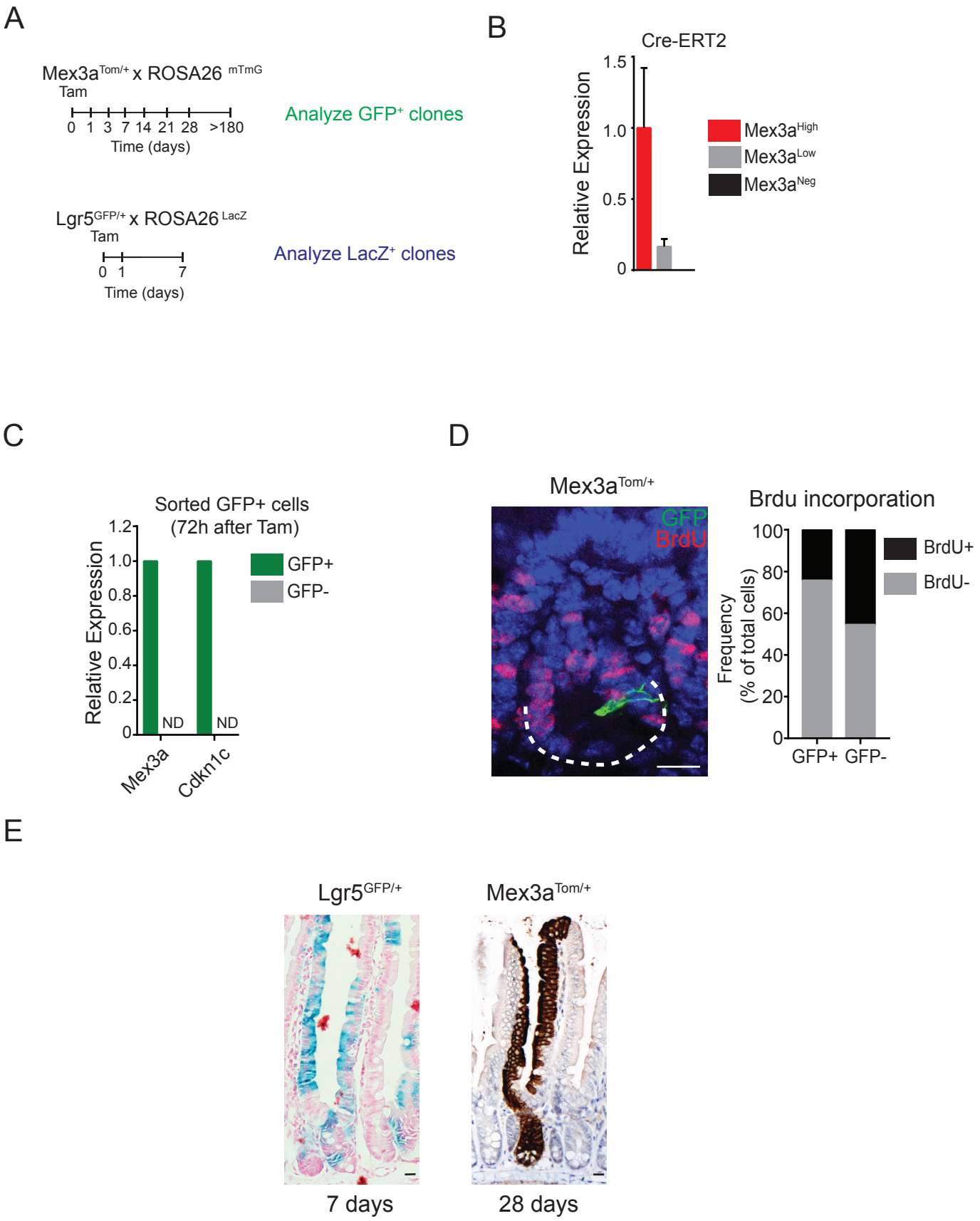

A
Mex3a Tom/+ x ROSA26 mTmG
Tam
0 1 3 7 14 21 28 >180
Time (days)
Analyze GFP+ clones
Lgr5 GFP/+ x ROSA26 LacZ
Tam
0 1 7
Time (days)
Analyze LacZ+ clones
B
Cre-ERT2
Relative Expression
1.5
1.0
0.5
0
Mex3a High
Mex3a Low
Mex3a Neg
C
Sorted GFP+ cells
(72h after Tam)
Relative Expression
1.2
1.0
0.8
0.6
0.4
0.2
0
ND
ND
Mex3a
Cdkn1c
GFP+
GFP-
D
Mex3a Tom/+
GFP
BrdU
Brdu incorporation
Frequency
(% of total cells)
100
80
60
40
20
0
GFP+ GFP-
BrdU+
BrdU-
E
Lgr5 GFP/+
Mex3a Tom/+
7 days
28 days

Figure S3 – Lineage tracing from Mex3a Tom/+ allele is derived from Mex3a-high cells (Related to Figure 3).

(A) Experimental setup to study lineage tracing in Mex3aTom/+; ROSA26mTmG mice and Lgr5GFP/+; ROSA26LacZ mice.

(B) Relative expression of CreERT2 levels in Mex3a populations. Bars depict the mean and upper/lower limits from a representative RT qPCR.

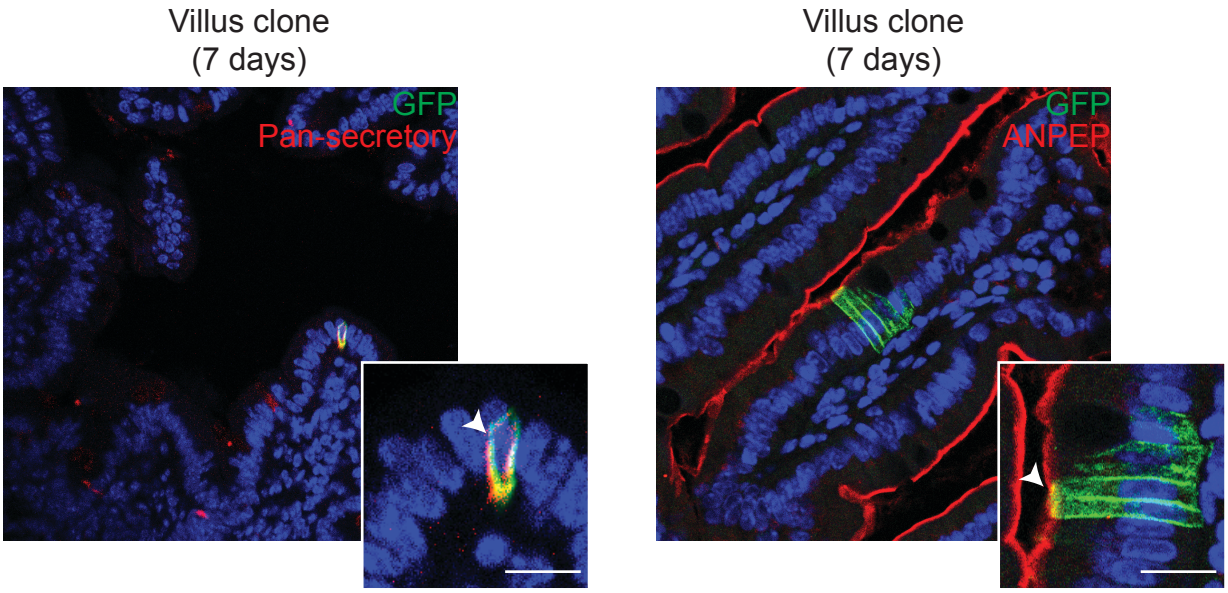
(C) Relative expression of Mex3a and Cdkn1c in GFP+ and GFP- cells 72 hours after tracing. Mex3aTom/+; ROSA26mTmG mice were treated with tamoxifen and 72 hours after induction GFP+ and GFP- cells were FACS sorted and RNA was extracted f or low cell number expression assays. Bars depict mean ± upper/lower limit.

(D) Analysis of BrdU incorporation in Mex3a-derived clones 3 days after tracing. (Left) Representative picture of a GFP+ clone that is negative for BrdU. (Right) Quantification of Brdu+ cells within GFP+ (n > 30 cells) and GFP- (n >150 cells) populations of crypts after 3 days of tracing. Scale bar represents 20 μm.

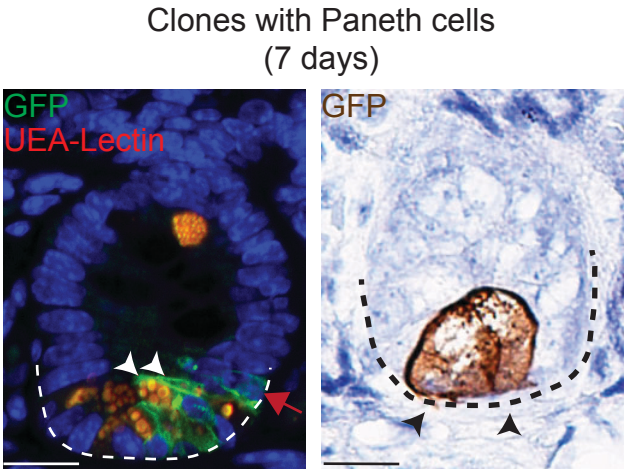
(E) Examples of ribbon clones from a 7 day Lgr5-driven tracing experiment (left) and a 28 day Mex3a-driven tracing experiment (right). Scale bar represents 20 μm.

Figure S4 - Related to Figure 4

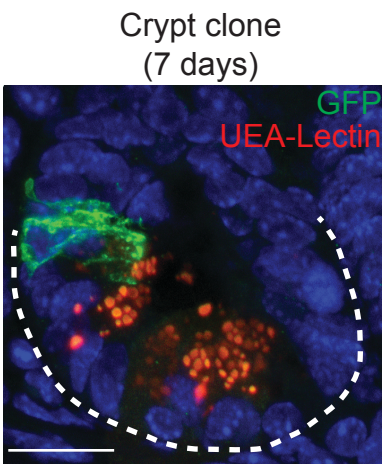
A




B




C




D

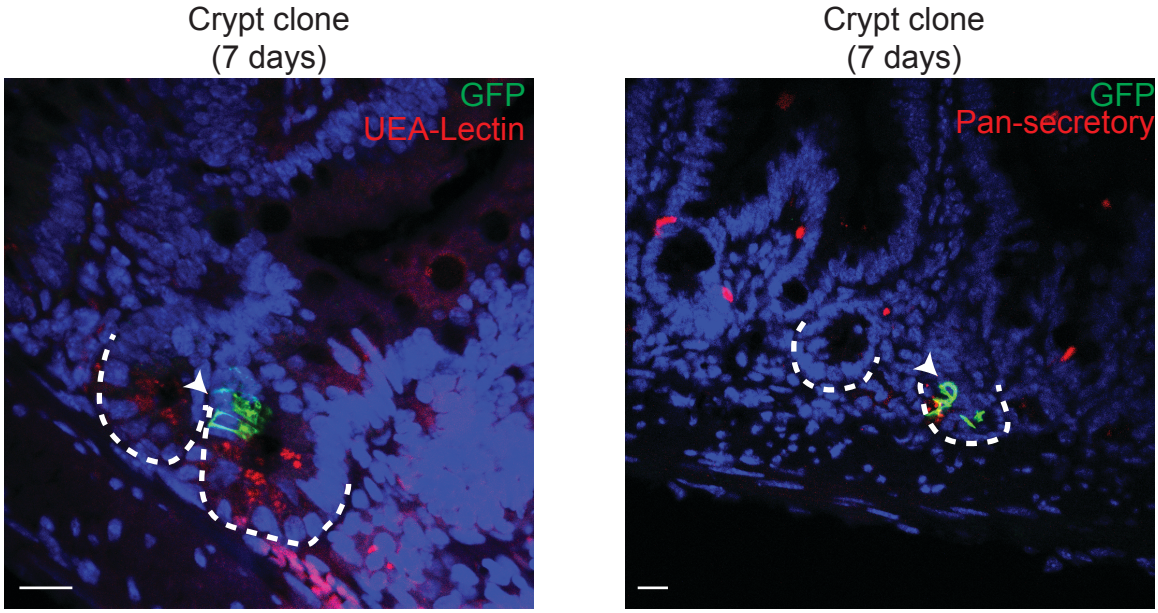


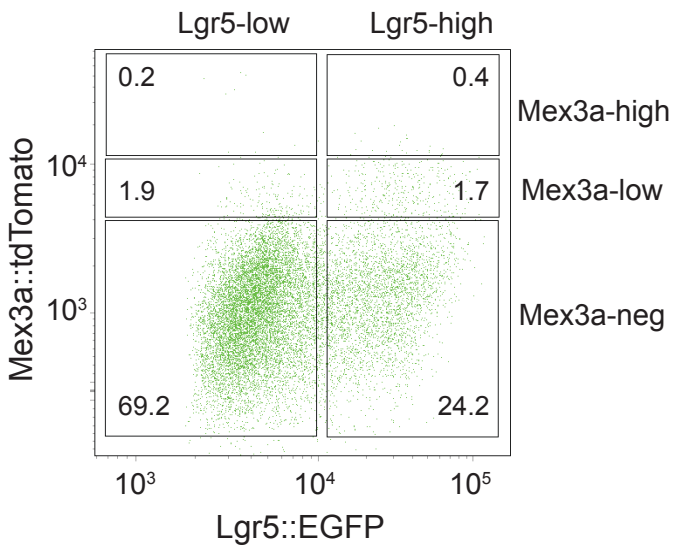
Figure S4 – Mex3a-derived clones are comprised of all intestinal lineages
(Related to Figure 4)

(A) Low magnification images of villus clones positive for pan-secretory (left) and absorptive (right) markers. Inset shows cells in higher magnification and arrowhead indicates the positive staining for the respective markers. Scale bar represents 10 μm.

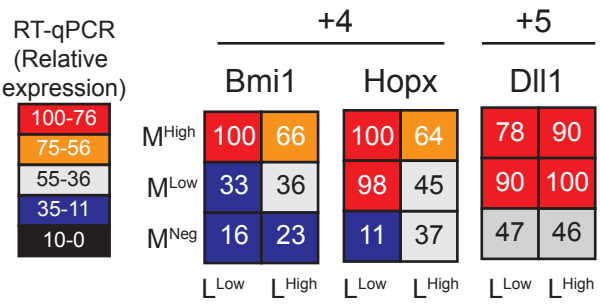
(B) Mex3a-high cells give rise to Paneth cells. (Left) Multicellular 7 day clone containing Paneth cells (white arrowheads) and undifferentiated cells (red arrow). (Right) 7 day clone composed by two Paneth cells (black arrowheads). Scale bar represents 20 μm.

(C-D) A fraction of Mex3a-derived clones remains undifferentiated after 7 days.
(C) Representative image of an undifferentiated small clone 7 days after tracing.
(D) Low magnification images of undifferentiated crypt clones 7 days after tracing. Arrowheads indicate the GFP+ clone. Note that all are located in position +3/+4. Scale bar represents 20 μm.

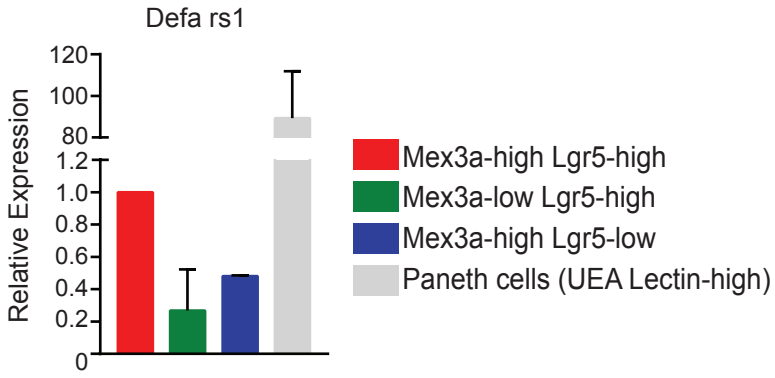
A


Mex3a+/+ ; Lgr5GFP/+
Lgr5-low
Lgr5-high
0.2
0.4
1.9
1.7
69.2
24.2
Mex3a-high
Mex3a-low
Mex3a-neg
Mex3a::tdTomato
10^4
10^3
10^3
10^4
10^5
Lgr5::EGFP

B


RT-qPCR
(Relative
expression)
100-76
75-56
55-36
35-11
10-0
+4
+5
Bmi1
Hopx
Dll1
MHigh
MLow
MNeg
100
66
100
64
78
90
33
36
98
45
90
100
16
23
11
37
47
46
LLow
LHigh
LLow
LHigh
LLow
LHigh

C


Defa rs1
Relative Expression
120
100
80
1.2
1.0
0.8
0.6
0.4
0.2
0
Mex3a-high Lgr5-high
Mex3a-low Lgr5-high
Mex3a-high Lgr5-low
Paneth cells (UEA Lectin-high)

D

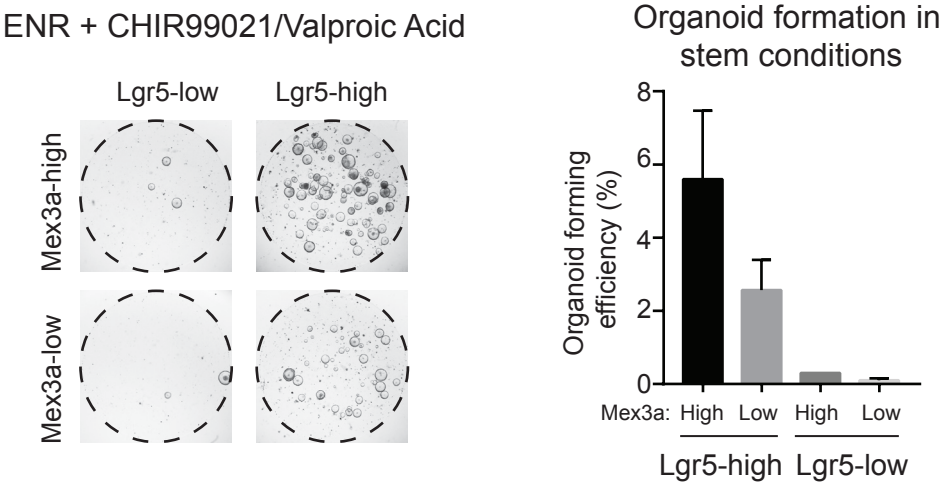

ENR + CHIR99021/Valproic Acid
Lgr5-low
Lgr5-high
Mex3a-high
Mex3a-low
Organoid formation in
stem conditions
Organoid forming
efficiency (%)
8
6
4
2
0
Mex3a: High Low High Low
Lgr5-high Lgr5-low

Figure S5 – Mex3a-high cells express lower levels of Defa rs1than mature Paneth cells and retain their differences in clonogenic potential even under high Wnt/Notch culture conditions (Related to Figure 5).

(A) Representative plot of small intestine cells from Lgr5GFP/+; Mex3a+/+ mouse. The gates are the ones used to identify Mex3a-Tomato subpopulations in compound reporters. Frequencies shown are relative to total Lgr5-GFP cells.

(B) Relative expression of Bmi1, Hopx and Dll1 in Mex3a/Lgr5 populations. Values are normalized to the subpopulation with the highest expression of each gene. Data represent the mean of two independent sorting experiments.

(C) Defa rs1 expression is highly enriched in differentiated Paneth cells. Small intestine cells were stained with UEA-Lectin to sort Paneth cells as well as the different Mex3a/Lgr5 populations. Bars depict the mean and upper/lower limits of expression.

(D) Organoid forming capacity of Mex3a/Lgr5 subpopulations in stem cell conditions. (Left) Representative images of 7 day organoids derived from Mex3a/Lgr5 populations. Note that in this conditions organoids are spheres and do not present differentiated cells. (Right) Quantification of organoid forming efficiency of Mex3a/Lgr5 populations in stem conditions. Bars depict mean ± SEM (n = 3).

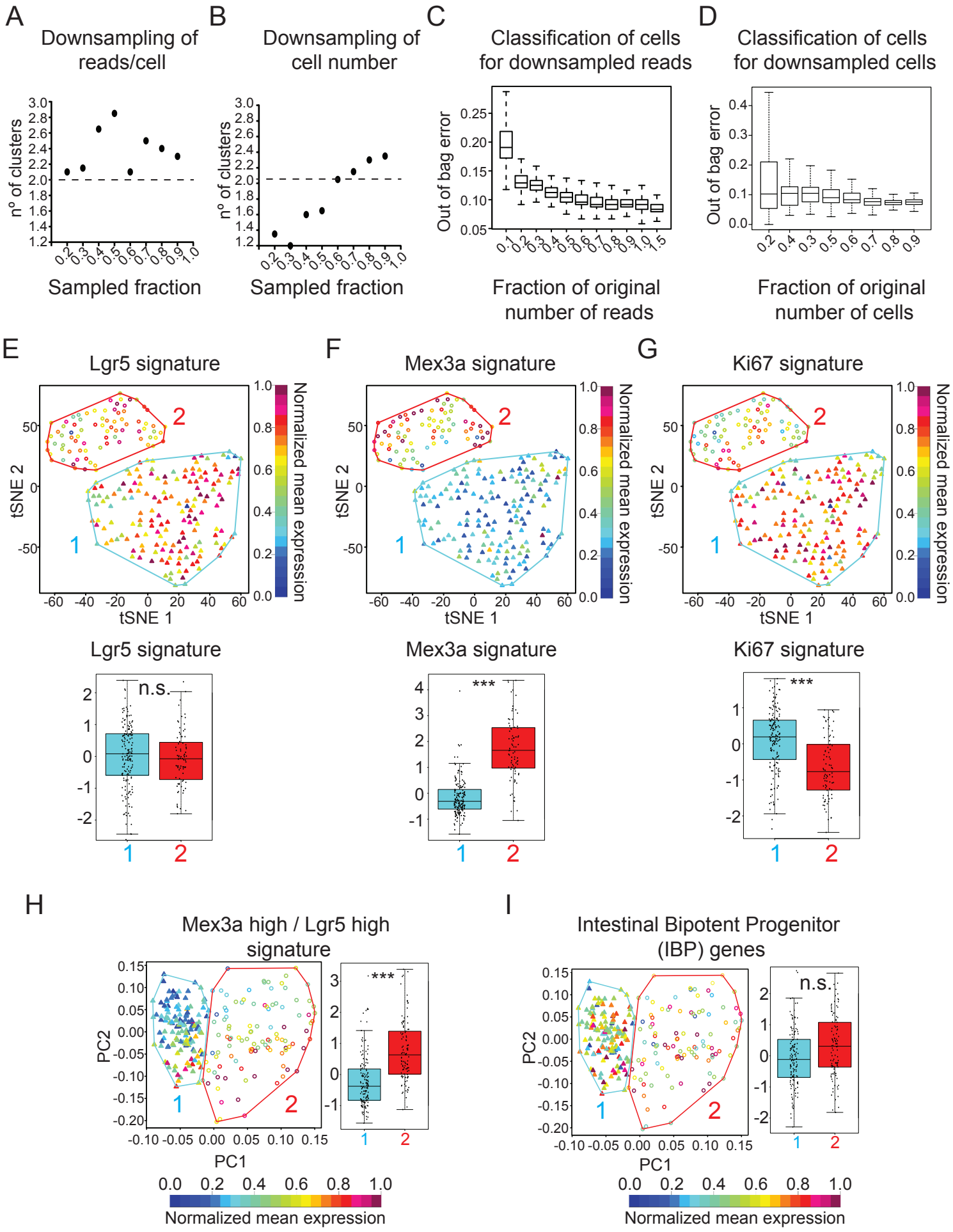

A
Downsampling of reads/cell
n° of clusters
Sampled fraction
B
Downsampling of cell number
n° of clusters
Sampled fraction
C
Classification of cells for downsampled reads
Out of bag error
Fraction of original number of reads
D
Classification of cells for downsampled cells
Out of bag error
Fraction of original number of cells
E
Lgr5 signature
tSNE 2
tSNE 1
Normalized mean expression
Lgr5 signature
n.s.
F
Mex3a signature
tSNE 2
tSNE 1
Normalized mean expression
Mex3a signature
***
G
Ki67 signature
tSNE 2
tSNE 1
Normalized mean expression
Ki67 signature
***
H
Mex3a high / Lgr5 high signature
PC2
PC1
***
Normalized mean expression
I
Intestinal Bipotent Progenitor (IBP) genes
PC2
PC1
n.s.
Normalized mean expression

Figure S6 – Robustness tests, tSNE visualization and IBP signature expression in Lgr5 single cell profiling dataset (Related to Figure 6).

(A) Analysis of the effect of reads/cell on cluster prediction. Predicted cluster number is stable upon downsampling of the number of reads/cell.

(B) Analysis of the effect of cell number on cluster prediction. Predicted cluster number is affected by a downsampling of over 50% of total cells analyzed.

(C) Analysis of the effect of reads/cell on cell assignation to Cluster 1 or 2. The ability to correctly classify a cell into Cluster 1 or 2 is stable even when 20% of total reads/cell is used.

(D) Analysis of the effect of total cells on cell assignation to Cluster 1 or 2.

(E) The Lgr5-signature is equally expressed in Clusters 1 and 2 defined by tSNE. (Upper) Mean expression of the Lgr5-signature plotted in each Lgr5-high cell across clusters in tSNE dimensions. Lines are drawn on the edges of clusters to facilitate their visualization. (Bottom) Box-plots of relative expression of the Lgr5-signature in Cluster 1 and Cluster 2. n.s., $p > 0.05$ in Kruskal-Wallis test.

(F) The Mex3a-signature is enriched in Cluster 2 defined by tSNE. (Upper) Mean expression of the Mex3a-signature plotted in each Lgr5-high cell across clusters in tSNE dimensions. Lines are drawn on the edges of clusters to facilitate their visualization. (Bottom) Box-plots of relative expression of the Mex3a-signature in Cluster 1 and Cluster 2. ***, $p < 0.001$ in Kruskal-Wallis test.

(G) The Ki67-signature is enriched in Cluster 1 defined by tSNE. (Upper) Mean expression of the Ki67-signature plotted in each Lgr5-high cell across clusters in tSNE dimensions. Lines are drawn on the edges of clusters to facilitate their visualization. (Bottom) Box-plots of relative expression of the Ki67-signature in Cluster 1 and Cluster 2. ***, $p < 0.001$ in Kruskal-Wallis test.

(H) Mex3a-high/Lgr5-high signature is enriched in Cluster 2. (Left) Mean expression of the Mex3a-high/Lgr5-high-signature plotted in each Lgr5-high cell. (Right) Box-plots of relative expression of the Mex3a-high/Lgr5-high-signature in Cluster 1 and Cluster 2. ***, $p < 0.001$ in Kruskal-Wallis test with a minimal fold change of 10%.

(I) The intestinal bipotent progenitor (IBP) is not differentially expressed in Clusters 1 and 2. (Left) Mean expression of the IBP-signature plotted in each Lgr5-high cell. (Right) Box-plots of relative expression of the IBP-high-signature in Cluster 1 and Cluster 2.

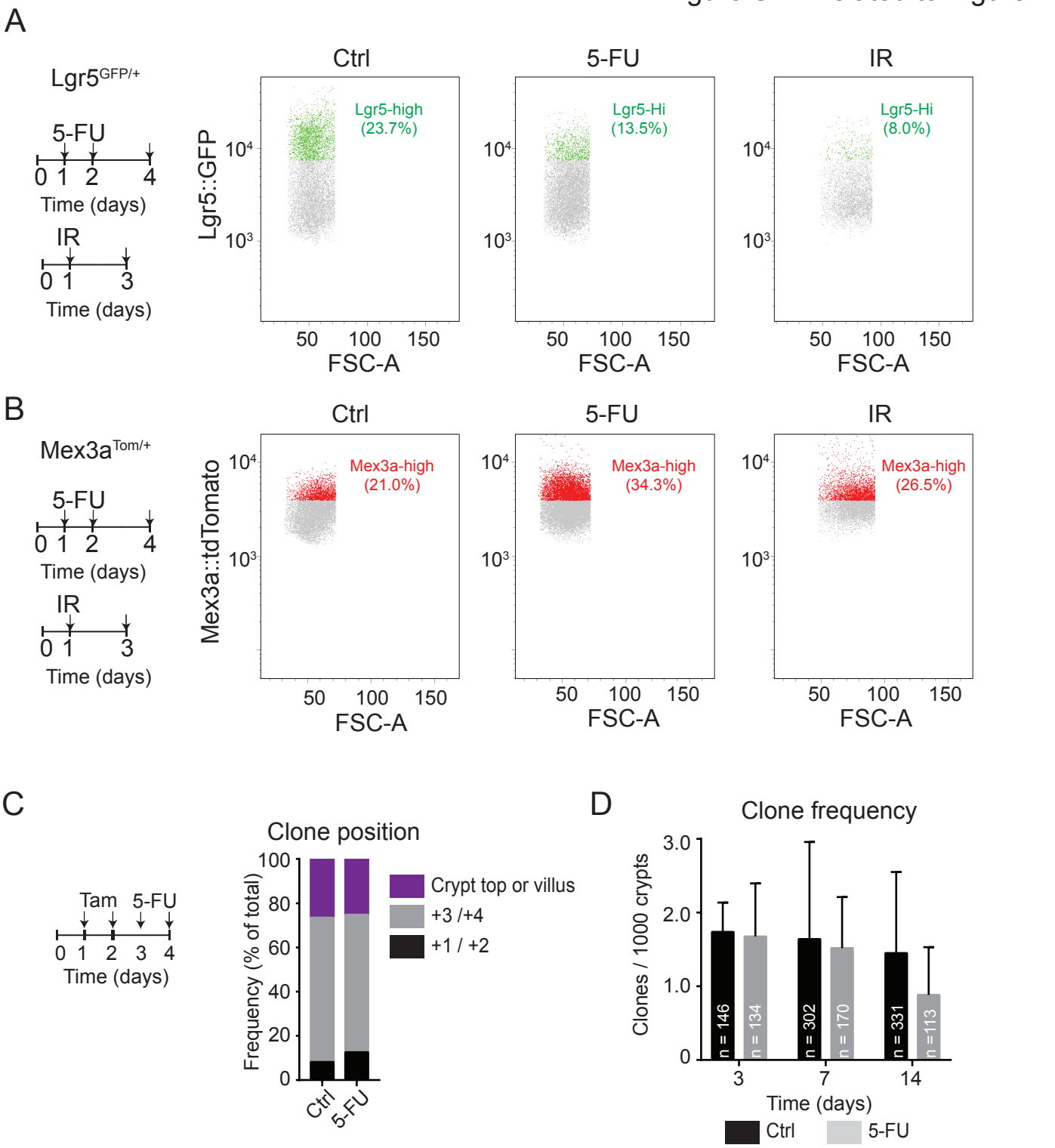

A
Lgr5GFP/+
5-FU
0 1 2 4
Time (days)
IR
0 1 3
Time (days)
Ctrl
5-FU
IR
Lgr5::GFP
Lgr5-high (23.7%)
Lgr5-Hi (13.5%)
Lgr5-Hi (8.0%)
FSC-A
B
Mex3aTom/+
Mex3a::tdTomato
Mex3a-high (21.0%)
Mex3a-high (34.3%)
Mex3a-high (26.5%)
C
Tam 5-FU
0 1 2 3 4
Time (days)
Clone position
Frequency (% of total)
Crypt top or villus
+3 /+4
+1 / +2
Ctrl
5-FU
D
Clone frequency
Clones / 1000 crypts
n = 146
n = 134
n = 302
n = 170
n = 331
n = 113
Time (days)
Ctrl
5-FU
E
Control
5-FU
GFP

Figure S7 – Behavior and lineage tracing quantification of Mex3a-high cells upon chemotherapy and radiotherapy treatments (Related to Figure 7).

(A) Representative FACS plots of small intestine Lgr5-GFP+ cells from control (left panel), 5-FU treated (middle panel) and IR treated (right panel) mice. The frequency of Lgr5-high cells is shown relative to total Lgr5-GFP cells.

(B) Representative FACS plots of small intestine tdTomato+ cells from control (left panel), 5-FU treated (middle panel) and IR treated (right panel) mice. The frequency of Mex3a-high cells is shown relative to total Mex3a-tdTomato cells.

(C) Initial clone position in Mex3a-derived tracing is unaffected after 5FU treatment. Clone position was assessed at 3 days in control (n = 146 clones) and 5-FU (n = 134 clones) treated mice.

(D) Clone frequency is not affected by 5-FU in Mex3a-derived tracing. Clones / 1000 crypts were assessed at 3 (Control, n = 146; 5-FU, n = 134), 7 (Control, n = 302; 5-FU, n = 170) and 14 (Control, n = 331; 5-FU, n = 113) days. Bars depict ± SD.

(E) Mex3a-derived clones, excluding ribbons, are largely located at position +4 after 1 week regardless of 5-FU treatment. Mex3aTom/+ mice were induced with Tamoxifen and then treated with two consecutive doses of 5-FU. Representative images in control (left panels) and 5-FU treated (right panels) of Mex3a-derived clones 1 week after tamoxifen induction. Scale bar represents 20 μm.