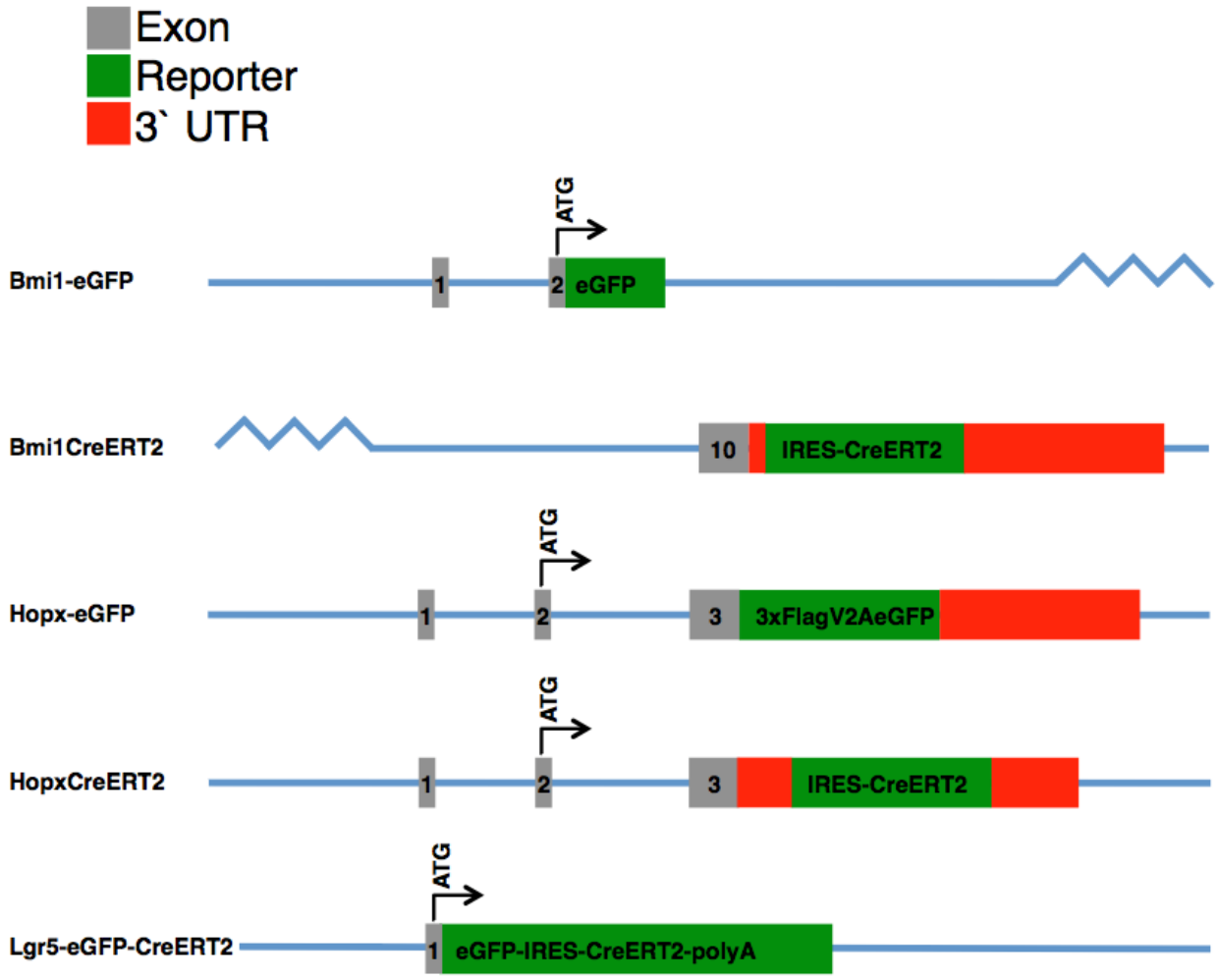
Stem Cell Reports, Volume *3*

Supplemental Information

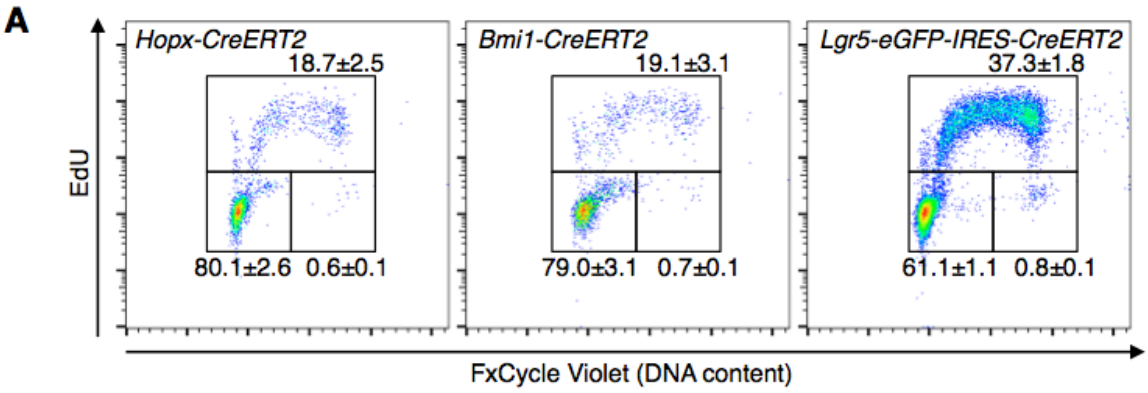
Single-Cell Analysis of Proxy Reporter Allele-Marked Epithelial Cells Establishes Intestinal Stem Cell Hierarchy

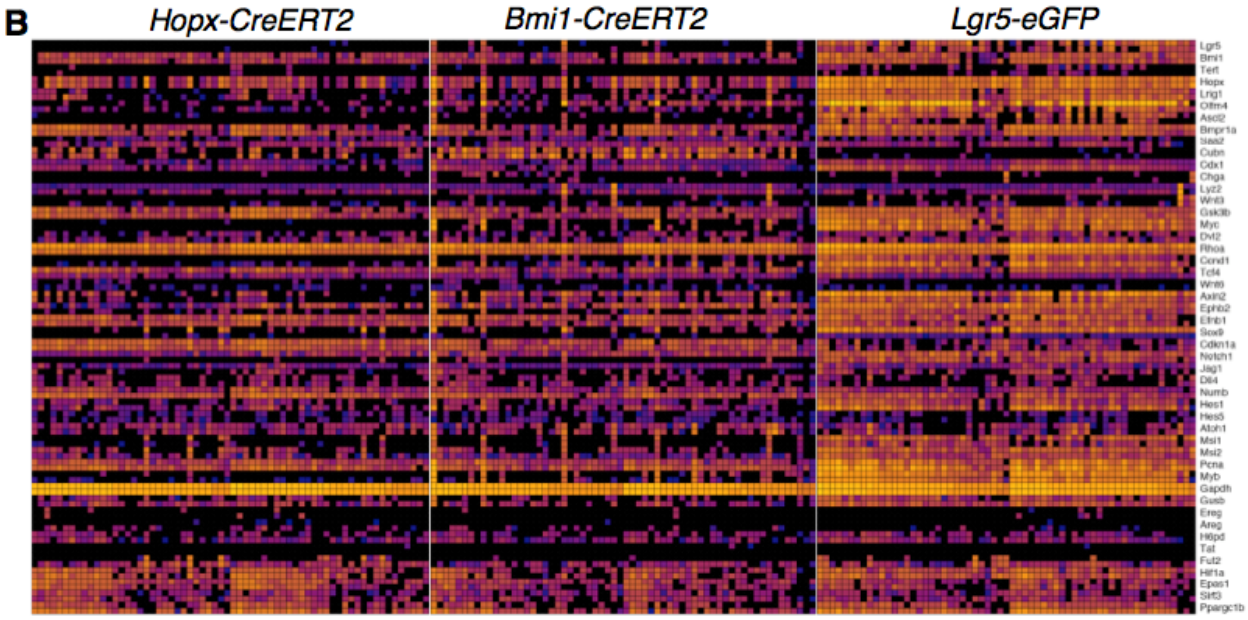
Ning Li, Maryam Yousefi, Angela Nakauka-Ddamba, Rajan Jain, John Tobias, Jonathan A. Epstein, Shane T. Jensen, and Christopher J. Lengner

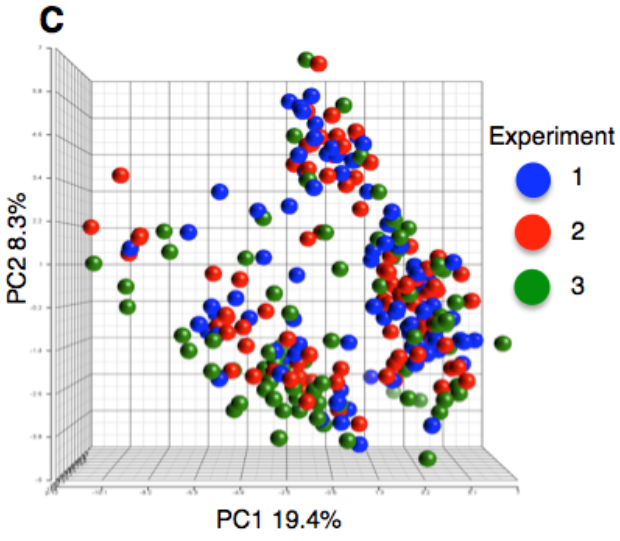
Supplemental Figure 1 *Li et al*

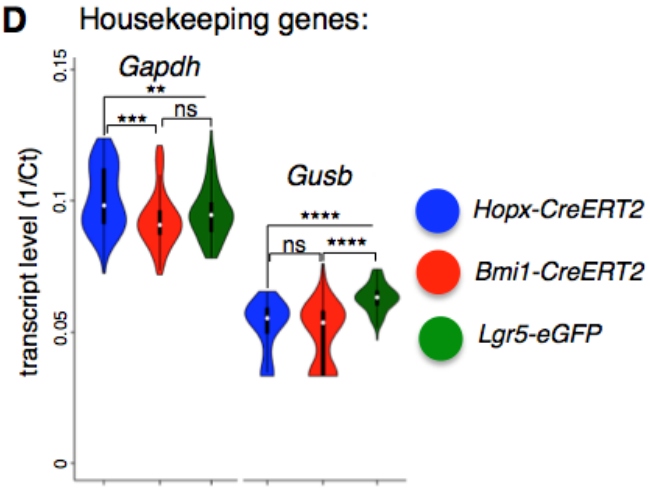


Supplemental Figure 2

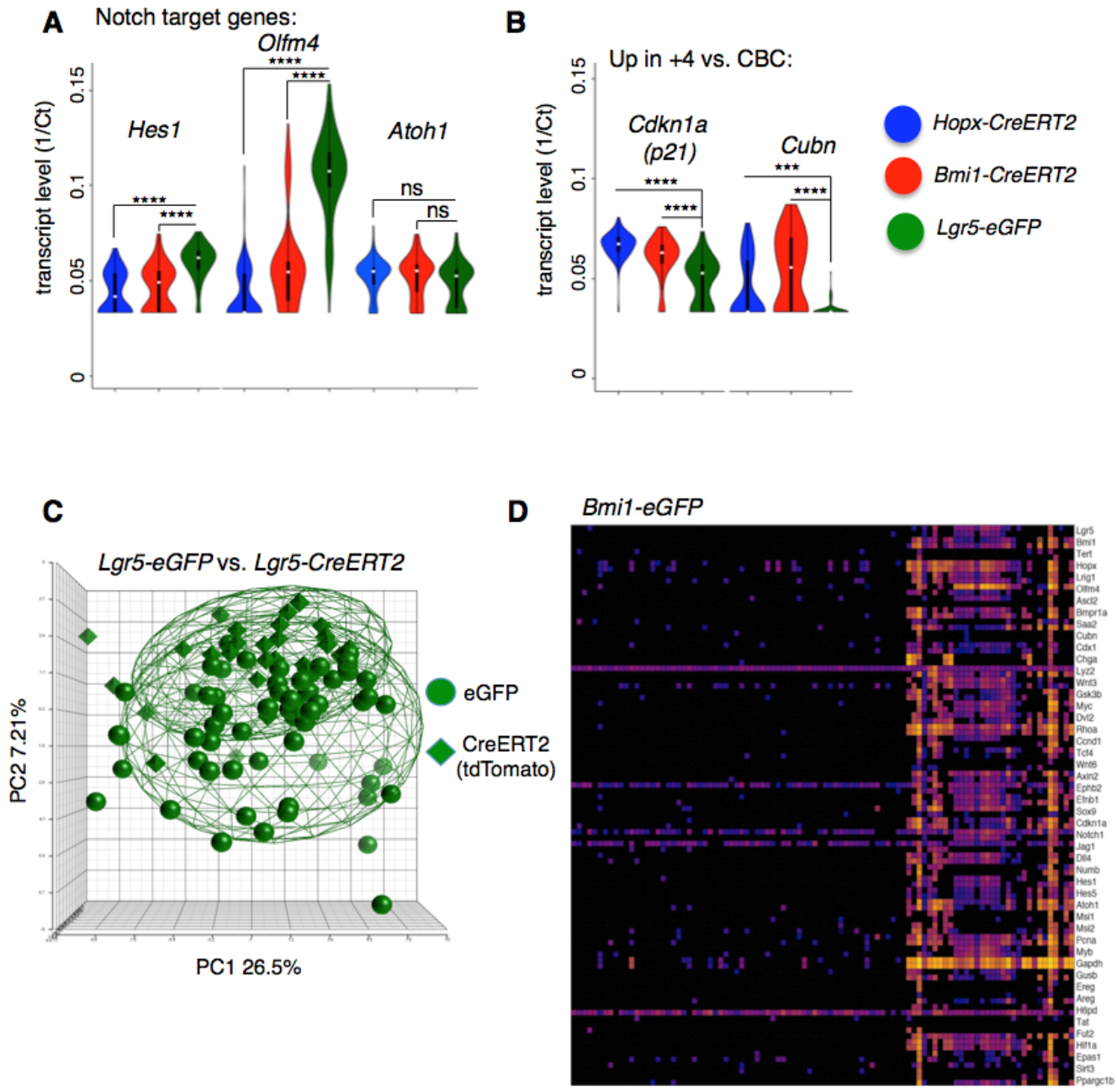




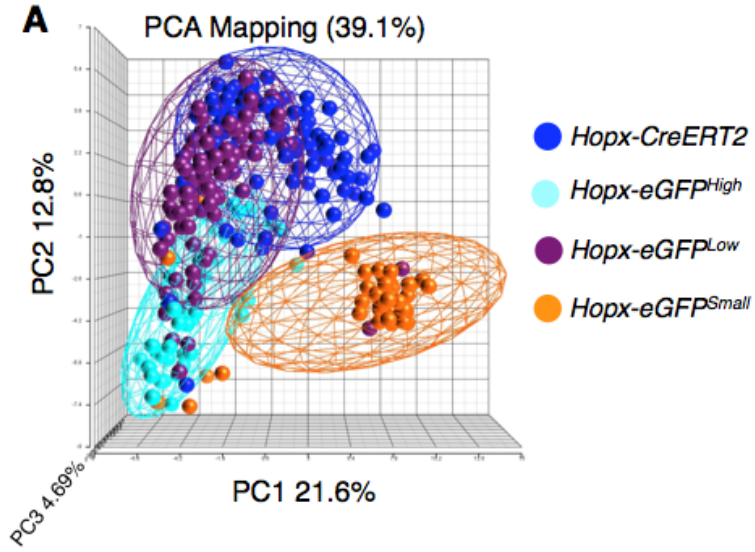


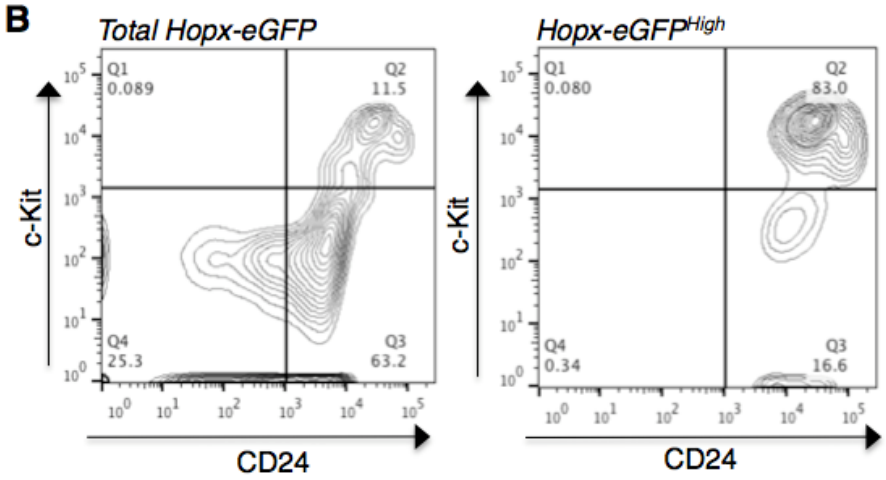


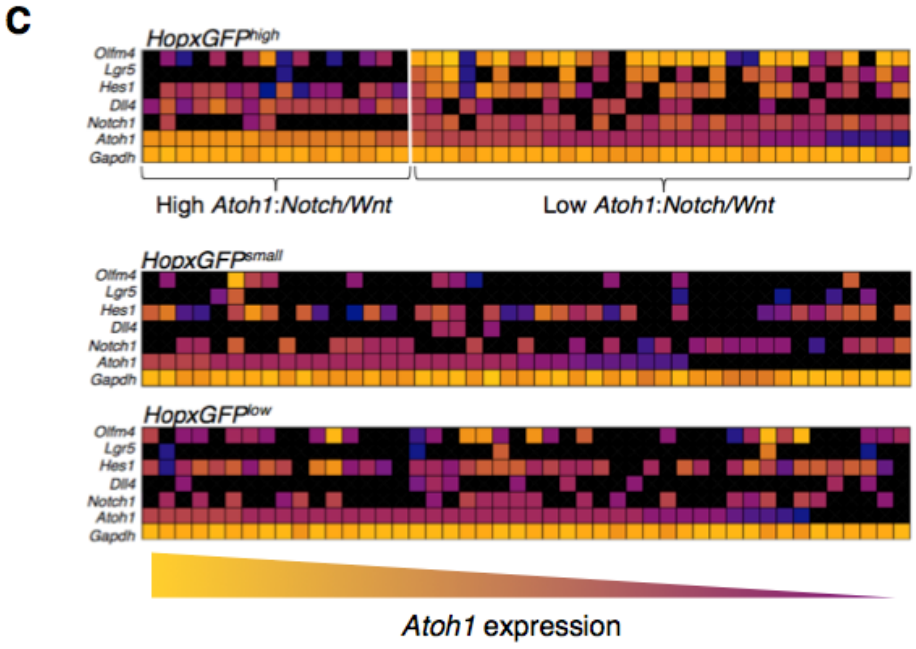
Supplemental Figure 3



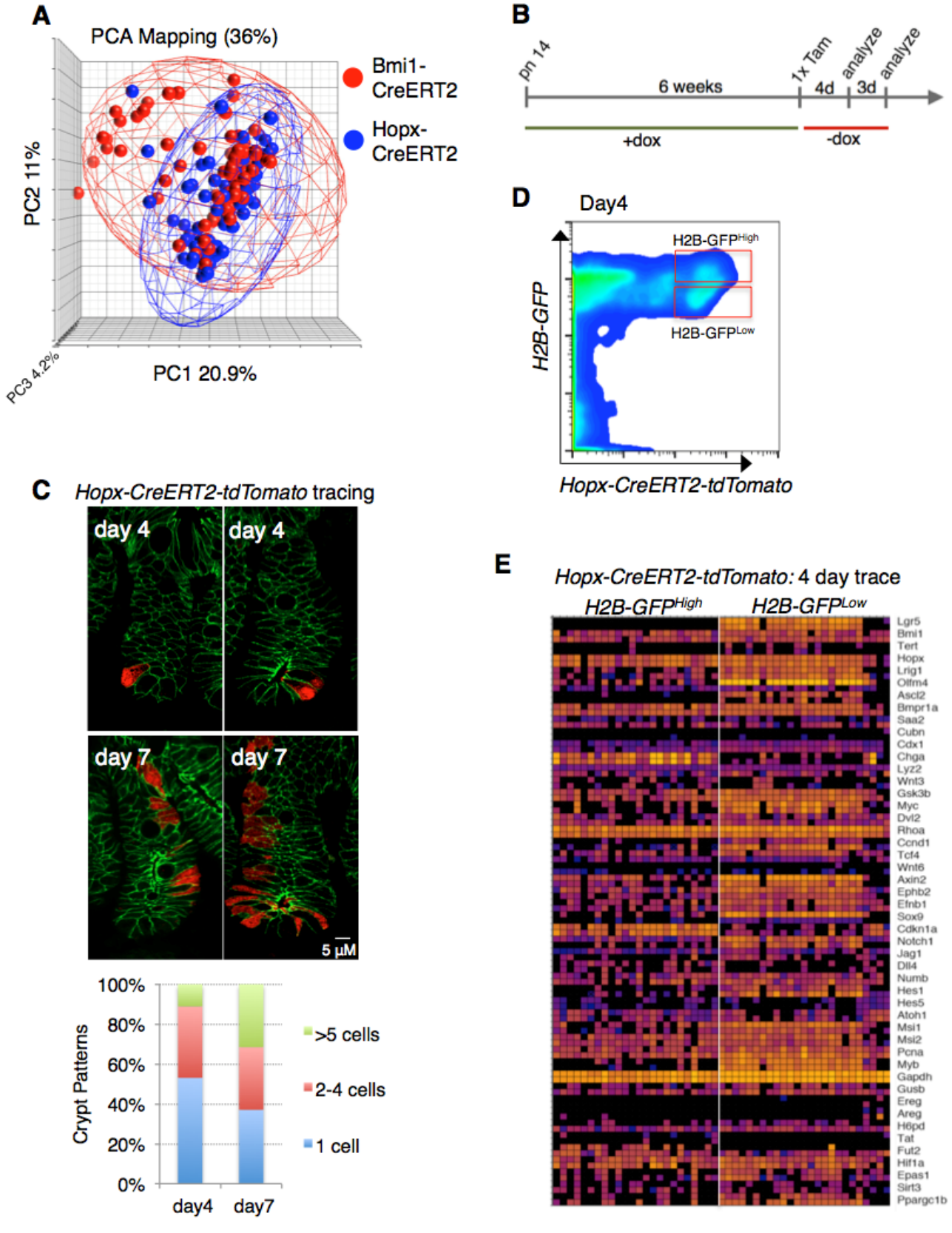
Supplemental Figure 4 *Li et al*



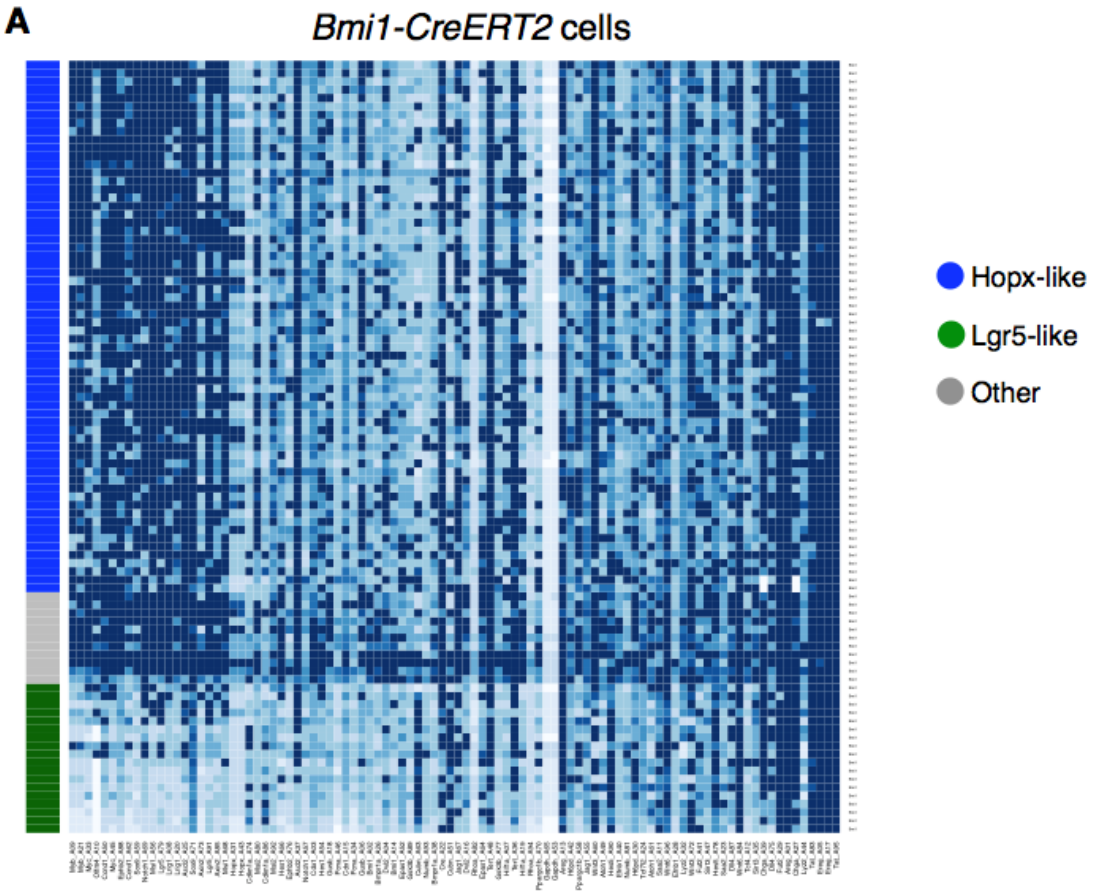




Supplemental Figure 5 *Li et al*



Supplemental Figure 6 *Li et al*



Supplemental Figure Legends

Supplemental Figure 1. Marker alleles analyzed. Site of insertion of *CreERT2* and *eGFP* proxy reporter alleles into endogenous *Bmi1*, *Hopx*, and *Lgr5* loci, relative to endogenous coding regions and untranslated regions (UTR). Refers to main Figure 1.

Supplemental Figure 2. Single cell analysis of bona fide ISC populations marked by *Bmi1-CreERT2*, *Hopx-CreERT2*, and *Lgr5-eGFP-IRES-CreERT2.* **A**. Flow cytometric analysis of cell cycle in the designated reporter-marked cell populations after a two-hour *in vivo* pulse labeling with EdU. **B.** Heatmap of gene expression across the three reporter-marked cell populations with a single cell in each column interrogated by 96 pairs of primers targeting 48 genes in duplicate. **C.** Principal component plot as in **Figure 2C**, color-coded by experiment number rather than reporter allele identity. **D.** Violin plots of the transcript levels of frequently used intestinal housekeeping genes *Gapdh* and *Gusb* in the designated reporter-marked cell populations. Asterisks indicate significance of differences in mean expression between indicated populations. ****: $p < 1x10^{-10}$, ***: $p < 1x10^{-5}$, **: $p < 0.005$. Refers to main Figure 2.

Supplemental Figure 3. Single cell analysis of ISC populations. A. Violin plots of the transcript levels Notch target genes in the designated populations. **B**. Violin plots transcript levels of the genes that best define reserve ISC identity versus active ISC identity, *Cdkn1a* and *Cubn* in the designated populations. Asterisks indicate significance of differences in mean expression between indicated populations. ****: $p < 1x10^{-10}$, ***: $p < 1x10^{-5}$. **C.** Principal component analysis of single *Lgr5-eGFP*$^{+}$ cells and tdTomato^{+} cells 18 hours after activation of the tdTomato reporter in *Lgr5-eGFP-IRES-CreERT2* mice with a single Tam dose. **D.** Heatmap of gene expression single FACS-purified *Bmi1-eGFP*$^{+}$ cells with a single cell in each column interrogated by 96 pairs of primers targeting 48 genes in duplicate (n=96). Refers to main Figure 2.

Supplemental Figure 4. Analysis of *Hopx-eGFP*-marked intestinal epithelial cells. A. Principal component analysis of the three *Hopx-eGFP*$^{+}$ populations versus *Hopx-CreERT2*$^{+}$ cells. **B.** Flow cytometric analysis of Paneth cells c-Kit^{High}, CD24^{High} in the total *Hopx-eGFP*$^{+}$ population (Top) and in the *Hopx-eGFP*^{High} population (Bottom). **C.** Heatmaps showing *Atoh1* expression relative to the designated Notch and Wnt-related genes in single cells isolated from the three *Hopx-eGFP*$^{+}$ populations, ranked by *Atoh1* transcript level. Refers to main Figure 3.

Supplemental Figure 5. Analysis of reserve ISCs and their progeny. A. Principal component analysis comparing *Hopx-CreERT2*$^{+}$ and *Bmi1-CreERT2*$^{+}$ cells. **B.** Strategy for pulse-chase labeling intestinal epithelial cells with doxycycline-inducible H2B-eGFP followed by chase and initiation of lineage tracing with a single Tam dose. **C.** Representative images of lineage tracing patterns in *Hopx-CreERT2-tdTomato* intestinal epithelium 4 or 7 days after a single Tam dose, with frequency of different observed patterns quantified below. Tomato is labeled in red, E-cadherin in green. **D.** FACS gating strategy for profiling H2B-eGFP^{High} versus H2B-eGFP^{Low} progeny of *Hopx-CreERT2*$^{+}$ cells. **E.** Heatmap of gene expression in single H2B-eGFP^{High} versus H2B-eGFP^{Low} progeny of *Hopx-CreERT2*$^{+}$ cells. Refers to main Figure 4.

Supplemental Figure 6. Cell-identity classification of Bmi1-CreER-marked cells. A. Single cells isolated from *Bmi1-CreERT2*$^{+}$*::LSL-Tomato* mice 18 hours after a single Tam injection were assigned identities with the algorithm trained on *Hopx-CreERT2*$^{+}$ and *Lgr5-eGFP*$^{+}$ cells. **B.** Representative micrographs of Lacz+ clones derived from *Lgr5-CreERT2::R26-LSL-Lacz* and *Hopx-CreERT2::R26-LSL-Lacz* mice 5 days and 6 months after initiation of lineage tracing with one dose of tamoxifen. Refers to main Figure 5.

Supplemental Methods

Flow Cytometry and Single-Cell Sorting

The intestine was cut open longitudinally and incubated with 5mM EDTA-HBSS solution at 4 °c for 30min to isolate epithelial cells. To generate a single cell suspension, cells were incubated with Accutase (BD Biosciences, San Jose, CA) at 37°c for 10min. Flow cytometry analysis was performed with BD LSRFortessa cell analyzer (BD Biosciences, San Jose, CA). DAPI negative cells were selected, then gated for single cell based on Forward-scatter height versus forward-scatter width (FSC-H vs FSC-W) and side-scatter height vs side-scatter width (SSC-H vs. SSC-W) profiles. Single-cell sorting experiments was performed with BD FACSAriaII cell sorter, each single cell was sorted into a different well of a 96-well PCR plate, using the FACSAriaII flow cytometer software package (FACSDiva) with single cell precision mode. Paneth cell isolation was done based on CD24 (eBioscience, 12-0242081)) and c-Kit (eBioscience, 25-1171-81) double staining. The size of the nozzle for all sorting is 100 µm (20 psi).

Intestinal Organoid Formation Assays.

Crypt organoid culture was performed as described previously (Sato et al., 2009). After intestinal crypt isolation and single cell digestion, a total of 1000 cells were sorted into one well of 96-well-plate coated with 50 µl of Matrigel (BD Bioscience). 100 µl of crypt culture medium (Advanced DMEM/F12 containing growth factors (50 ng/ml EGF (Invitrogen), 1 •g/ml R-spondin 1 (Wistar Institute protein production facility), 100 ng/ml Noggin (Peprotech) and 3 •M GSK-3 inhibitor (CHIR99021, Stemgent) was added. Pictures were taken after 10 days culture. The error bars represent the standard deviation from three biological replicates.

EdU Labeling and Radiation Injury

To assess the frequency and proliferation of *HopX-eGFP*$^{+}$ cells in response to injury, *HopX-eGFP* mice received 12 Gy whole body γ-irradiation. Irradiated mice and their littermate controls (non-irradiated *HopX-eGFP* mice) were injected with 0.3mg/kg body weight of 5-ethynyl-2´-deoxyuridine (EdU) (Life technologies) intraperitoneally 2 days after irradiation injury and 2 hours before euthanasia and isolation of the intestinal epithelium. After washing the longitudinally opened intestine in PBS, it was moved to 30 mM EDTA (EDTA, Sigma) and 1.5 mM DTT (Sigma) in HBSS at 4°C for 20 minutes. Then the intestine was incubated in 30 mM EDTA in HBSS at 37°C for 10 minutes. Vigorous pipetting was done to dissociate intestinal epithelium and single cell suspension was generated with 0.8 mg/ml Dispase (GIBCO). For

labeling proliferative cells Click-iT® Plus EdU Alexa Fluor® 647 Flow Cytometry Assay Kit (Life technologies) was used and proliferative cells were marked with Alexa fluor 647 azide dye according to the user manual. Analysis of frequency and proliferation of GFP^+ cells were done using flow cytometry on an LSR Fortessa and Flowjo software was used for data analysis. For histological analysis of EdU, intestinal tissues were formalin fixed and paraffin embedded, and EdU was labeled with the Click-iT®, using approximately 250 µL of reaction cocktail per slide. The slides were then washed and treated with mounting media containing DAPI.

Immunofluoresence

Intestines were fixed in 10% Formalin, paraffin-embedded and sectioned. Paraffin sections were pretreated in 0.01 M citrate buffer (pH 6) in a pressure cooker, incubated in primary antibodies, then incubated with Cy2- or Cy3- conjugated secondary antibodies (Jackson Laboratory) and counterstained with DAPI in mounting media (Vector labs). The following antibodies were used: DsRed (Clontech, 632496), GFP (Abcam, AB 6673). Images were acquired with a Zeiss Axioplan upright microscope and Leica TCS SP8 confocal microscope. Image processing was done using Fiji.

Reserve (Hopx-CreER+) vs. CBC (Lgr5+) Single Cell Classification

We trained a classification procedure for classifying any cell as either: 1. a *Hopx-CreER+* cell, 2. a *Lgr5+* cell or 3. a negative reference (other) cell type. This classification procedure was trained on a population of pure *Hopx-CreER+* cells, a population of pure *Lgr5+* cells and a negative reference population (all epithelial cells excluding *Hopx-eGFP+* cells which contain cells with both reserve and CBC stem cell identity (as seen in Figure 3).

Based on these populations, we calculated for each primer pair *g* (g = 1,...,96):

1. H(g,c), the proportion of pure *Hopx-CreER+* cells that had each possible cycle value c.
2. L(g,c), the proportion of pure *Lgr5+* cells that had each possible cycle value c.
3. N(g,c), the proportion of negative reference cells that had each possible cycle value c.

The possible cycle values were c = 0,1,..., 30.

As a running example, consider a pure *Hopx-CreER+* population with only four cells that had measured cycle values of 28, 29, 30, and 30 for a particular primer pair *g*. In this case,
H(g,30) = 0.50 and H(g,28) = H(g,29) = 0.25 and H(g,0) = H(g,1) = ... = H(g,27) = 0.0

These proportions H(g,c), L(g,c) and N(g,c) are used to compare the similarity of new cell to the *Hopx-CreER+*, Lgr5 and the negative reference populations. Specifically, let X be the cycle value for gene g in this new cell. We calculate the *Hopx-CreER+* similarity for gene *g* in this new cell as:

Hopx.Similarity(g) = H(g,X)/[H(g,X) + L(g,X) + N(g,X)]

In other words, if this new cell has a cycle value of X for primer pair *g* and if the pure *Hopx-CreER+* population also has many cells with that same cycle value (large H(g,X)), then we will give that new cell a high similarity to pure *Hopx-CreER+* for that primer pair *g*. We also calculate the similarity of the new cell to *Lgr5+* and the negative reference on primer pair *g*:

Lgr5.Similarity(g) = L(g,X)/[H(g,X) + L(g,X) + N(g,X)]

NegRef.Similarity(g) = N(g,X)/[H(g,X) + L(g,X) + N(g,X)]

Then, the total similarity of the new cell to *Hopx-CreER+* is the sum of the similarities for each primer pair *g* across all the primer pairs (g = 1,...,96):

Hopx.Similarity.Total = Sum_g Hopx.Similarity(g),

and the corresponding similarity of the new cell to *Lgr5+* or the negative reference is:

Lgr5.Similarity.Total = Sum_g Lgr5.Similarity(g)

NegRef.Similarity.Total = Sum_g Lgr5.Similarity(g)

Finally, we classify the new cell as *Hopx-CreER+*, *Lgr5+* or Negative Reference based on the maximum of these similarity scores, i.e.

New cell = *Hopx-like* if Hopx.Similarity.Total > Lgr5.Similarity.Total and Hopx.Similarity.Total > NegRef.Similarity.Total

or

New cell = *Lgr5*-like if Lgr5.Similarity.Total > Hopx.Similarity.Total and Lgr5.Similarity.Total > NegRef.Similarity.Total

or

New cell = NegRef if NegRef.Similarity.Total > Hopx.Similarity.Total and NegRef.Similarity.Total > Lgr5.Similarity.Total

Correlation Matrices

Within each population of cells (e.g. pure *Hopx-CreER*$^{+}$ cells, pure *Lgr5-eGFP*$^{+}$ cells, etc.), we calculated the Pearson correlation of the cycle values between each pair of genes. The Pearson correlation ranges between -1 and 1 and measures the degree of linear association in the cycle values between a pair of genes and is color coded, with the coding of color to numerical value presented in supplemental table 2. The R package 'corrplot' was used to calculate and visualize the correlations between each pair of genes.

Violin Plots

Violin plots were generated as follows. For each cell, we have a measure of the cycle time for 96 primer sets (48 genes with duplicate primer sets). A cycle value of 30 was imputed for any cycle values that did not amplify by 30 cycles (i.e., no signal). For each gene, violin plots were constructed using the statistical software R to compare the distribution of cycle times for that gene between the conditions. For PCA analysis, Fluidigm Ct values were averaged for each gene (across the two primer sets per gene) in each sample. Principal Components Analysis (PCA, using Partek Genomics Suite v6.6, Partek, Inc. St. Louis, MO) was used to visualize the global variation across the samples. Samples were colored to represent their condition. Statistical significance of differences between the mean expression values between populations was calculated using an independent sample t-test. Asterisks in figures denotes p-value for significance of differences in the mean expression value of the indicated gene across the indicated single cell populations.

Hierarchical Clustering

We calculated the Pearson correlation of the cycle values (across all 96 primer sets) between each pair of cells in all cell populations. These correlations were inputted into an agglomerative clustering algorithm to create a hierarchical clustering of all cells, with each population labeled with a different color. The R package 'hclust' was used (with the average linkage setting) to create the hierarchical clustering.

Supplemental References

Sangiorgi, E., and Capecchi, M.R. (2008). Bmi1 is expressed in vivo in intestinal stem cells. Nature genetics *40*, 915-920.

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.

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.

Steinhauser, M.L., Bailey, A.P., Senyo, S.E., Guillermier, C., Perlstein, T.S., Gould, A.P., Lee, R.T., and Lechene, C.P. (2012). Multi-isotope imaging mass spectrometry quantifies stem cell division and metabolism. Nature *481*, 516-519.

Takeda, N., Jain, R., Leboeuf, M.R., Padmanabhan, A., Wang, Q., Li, L., Lu, M.M., Millar, S.E., and Epstein, J.A. (2013). Hopx expression defines a subset of multipotent hair follicle stem cells and a progenitor population primed to give rise to K6+ niche cells. Development *140*, 1655-1664.

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.

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

van 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, F., Wang, J., Liu, D., and Su, Y. (2010). Normalizing genes for real-time polymerase chain reaction in epithelial and nonepithelial cells of mouse small intestine. Analytical biochemistry *399*, 211-217.

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*, 466-471.

Yui, S., Nakamura, T., Sato, T., Nemoto, Y., Mizutani, T., Zheng, X., Ichinose, S., Nagaishi, T., Okamoto, R., Tsuchiya, K.*, et al.* (2012). Functional engraftment of colon epithelium expanded in vitro from a single adult Lgr5(+) stem cell. Nature medicine *18*, 618-623.