Published in final edited form as:
Nat Cell Biol. ; 14(1): 106–114. doi:10.1038/ncb2384.

Single molecule transcript counting of stem cell markers in the mouse intestine

Shalev Itzkovitz[1,2], **Anna Lyubimova**[1,2,3], **Irene Blat**[2,4], **Mindy Maynard**[4], **Johan van Es**[3], **Jacqueline Lees**[2,4], **Tyler Jacks**[2,4], **Hans Clevers**[3], and **Alexander van Oudenaarden**[1,2,3,4,†]

[1]Department of Physics, Massachusetts Institute of Technology, Cambridge MA 02139, USA
[2]Department of Biology, Massachusetts Institute of Technology, Cambridge MA 02139, USA
[3]Hubrecht Institute–KNAW (Royal Netherlands Academy of Arts and Sciences) and University Medical Center Utrecht, Uppsalalaan 8, 3584 CT Utrecht, Netherlands [4]Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

SUMMARY

Determining the molecular identities of adult stem cells requires novel technologies for sensitive transcript detection in tissues. In mouse intestinal crypts, lineage-tracing studies suggested that different genes uniquely mark spatially distinct stem-cell populations, residing either at crypt bases or at position +4, but a detailed analysis of their spatial co-expression has not been feasible. Here we apply three-color single molecule fluorescence *in-situ* hybridization to study a comprehensive panel of intestinal stem-cell markers during homeostasis, aging and regeneration. We find that the expression of all markers overlap at crypt-base-cells. This co-expression includes *Lgr5*, *Bmi1* and *mTert*, genes previously suggested to mark distinct stem cells. Strikingly, *Dcamkl-1* tuft cells, distributed throughout the crypt axis, co-express *Lgr5* and other stem cell markers that are otherwise confined to crypt bases. We also detect significant changes in the expression of some of the markers following irradiation, suggesting their potential role in the regeneration process. Our approach can enable the sensitive detection of putative stem cells in other tissues and in tumours, guiding complementary functional studies to evaluate their stem-cell properties.

Characterizing the physical locations and molecular identities of stem cells during tissue homeostasis and repair has been impeded by the lack of experimental tools for monitoring individual cells in intact tissue. The mouse small intestine is a prime example in which, despite decades of research, the molecular identities and precise locations of stem cells remain debatable[1, 2]. The epithelium in the mouse small intestine forms invaginations called crypts that protrude into the underlying connective tissue. Stem cells that reside in the lower parts of the crypts divide to give rise to transit amplifying cells, which rapidly migrate along the crypt axis while dividing a few more times. When the transit amplifying cells reach the upper crypt regions they become post-mitotic and differentiate into either enterocytes – nutrient absorbing cells that form the bulk of the tissue, or several types of secretory cells, including goblet cells, enteroendocine cells and tuft cells[3, 4]. The differentiated cells continue to migrate up, exiting the crypts towards larger invaginations into the lumen called

[†]To whom correspondence should be addressed: avano@mit.edu.

AUTHOR CONTRIBUTION S.I. and A.vO. conceived the project. S.I., I.B. and A.L. carried out most of the experiments. S.I. analyzed the data. M.M., J.L., J.vE., T.J. and H.C. provided mice and assisted with experiments. S.I. and A.vO. wrote the paper.

COMPETING FINANCIAL INTERESTS STATEMENT The authors declare that they have no competing financial interests.

villi. They are finally extruded from the tops of the villi about 5 days after their birth from stem cells. Paneth cells are longer–lived secretory progenies that migrate down towards the crypt bottoms where they are thought to play a role in crypt defense and stem cell maintanence[5].

While it is widely accepted that the intestinal stem cells that give rise to all epithelial lineages reside in the lower portions of crypts, different identities in terms of numbers, exact locations and genetic signatures have been proposed for these stem cells, that appear mutually exclusive[1, 2]. The "+4 hypothesis", originally proposed by Potten[6] posits that stem cells reside in cell position +4, just above the Paneth cells. This is based on unique characteristics of cells at these positions, including their high susceptibility to apoptosis, their non-random DNA strand segregation and suggested specific expression of genes such as *Bmi1*[7, 8], *mTert*[9] and *Dcamkl-1*[10]. Alternatively, the stem cell zone hypothesis originally formulated by Leblond[11, 12] posits that crypt base-columnar-cells (CBC) residing at the very bottom of the crypts are the actual stem cells. While independent lineage tracing studies using *Lgr5*[12], *Sox9*[13] and *Prominin-1*[14] have demonstrated stable labeling of the progenies of CBC cells, and a single *Lgr5*-high stem cell has been shown to reconstitute a long-lived and complete, self-renewing small intestinal organoid in vitro[15], lineage tracing with *Bmi1*[7, 8] and *mTert*[9] has implied the +4 cell as the stem cell of the small intestine. These results pose the question of whether two or more distinct stem cell populations uniquely marked by these genes co-exist in mouse intestinal crypts[1, 2].

Lineage tracing experiments provide functional proof that a gene of interest is expressed in stem cells, but are limited in detecting the precise location of the expressing cells and the expression pattern of other genes in these cells. Genes that are broadly expressed throughout the tissue in both stem cells and in their differentiated offspring would yield stable labeling of progenies, but would not be informative as to the location of stem cells and could not, on their own, be considered stem-cell markers. Thus detecting stem cell genes in mammalian tissues requires complementing lineage-tracing studies with sensitive methods to measure the precise location where candidate markers are expressed and to determine their co-expression patterns.

Previous attempts to characterize this co-expression program were based on methods such as qPCR or microarray analysis of GFP-sorted cell populations[15, 16] or laser-capture microdissected tissue[17]. While yielding important insights, these methods have several disadvantages, such as the use of knock-in mice, standardization issues related to the qPCR process, insufficient sensitivity for the analysis of single cells and most importantly the loss of spatial information[18]. Immunohistochemistry and classic RNA in-situ hybridizations[19] preserve tissue morphology, but sensitivity and specificity problems limit the generic use of these methods in yielding quantifiable co-expression data of several genes at the single cell level. To overcome these limitations, several studies used multiply labeled fluorescent probes to detect single mRNA in fixed yeast and mammalian cells[20–23] as well as nuclear transcription sites in paraffin-embedded tissue[24]. However detection of single mRNA in adult mammalian tissue, where single-cell resolution is crucial for identifying the distinct roles of individual cells, has not yet been demonstrated.

We have previously developed a sensitive method of transcript counting based on singly labeled fluorescent probes[25], enabling simultaneous detection of three different endogenous transcripts in individual cells. This technique was successfully applied to study expression in mammalian cells, as well as in *Drosophila*[25] and *C. elegans* embryos[26]. Here we apply this method to mouse intestinal frozen sections, to obtain, for the first time, a quantitative comprehensive *in-situ* description of the spatial patterns and combinatorial expression of stem cell markers at the single transcript level.

RESULTS

Single molecule FISH enables sensitive and specific *in-situ* transcript detection in intestinal tissue

We designed a panel of 15 libraries of fluorescently labeled probes, each composed of 48 20-bp oligos complementary to the coding sequences of previously suggested stem cell markers (Fig. 1). These included the R-spondin receptor *Lgr5*[12, 27, 28], the WNT targets *Ascl2*[16], *CD44*[29], *Sox4*[30], *Sox9*[13], *Mmp7*, *EphB2* and *EphB3*[31], the RNA binding protein *Musashi-1*[32, 33], *Olfactomedin-4* (*Olfm4*)[16], *Prominin-1* (*CD133*)[14], *Dcamkl-1*[10, 17], *Bmpr1a*[34], mouse telomarse reverse transcriptase (*mTert*)[9, 35] as well as the polycomb gene *Bmi1*[7]. Hybridization of 6-micron cryo-sections of small intestinal tissue with these libraries yielded bright diffraction limited dots, representing single transcripts (Fig. 1). These were automatically counted using custom image processing software (Fig. S1a–d). To study the co-expression of these genes at the single cell level we used three different fluorophores to simultaneously probe the expression of *Lgr5*, *Bmi1* and other genes from the panel and assigned their numbers to individual cells manually segmented based on E-cadherin lateral membrane staining.

We first assessed whether our transcript counting method correlates with the expression patterns in reporter mice. To this end we examined both fluorescence and transcript levels in the *Lgr5*-EGFP reporter mouse model[12]. We detected cells with intense GFP signal, as well as EGFP transcripts at crypt bottoms in only one out of ten crypts on average, consistent with the pronounced variegated expression pattern previously reported[36] (Fig. 2a). Unlike the patchy expression of the transgene we uniformly detected the endogenous *Lgr5* transcripts in every crypt throughout the tissue. Importantly the expression level of both *Lgr5* and EGFP transcripts, as well as GFP levels were highly correlated in the crypts that were positive for both (Fig. 2a,b, Spearman correlation $R = 0.68$, $p < 10^{-68}$). Thus our method is highly correlated to the transgene transcript and protein levels, but facilitates a much more comprehensive analysis of the tissue. To further test the specificity of our method we analyzed the expression of the intestinal differentiation markers *Gob5*, *Creb3l3* and *Lysozyme* and the proliferation marker *Ki67*. This yielded highly localized expression at the respective goblet, enterocyte, Paneth, and transit-amplifying cells, demonstrating the specificity of the technique (Fig. 2c,d).

Spatial expression patterns of intestinal stem-cell markers are broadly overlapping at crypt base cells

To facilitate our analysis of the expression patterns of the putative stem cell marker genes along the intestinal crypt, we created a spatial profile for each gene by projecting the single cell transcript counts on a vertical axis originating at the crypt apex. We found that the spatial expression profiles are remarkably invariant between crypts within the same mouse and almost identical between 4 months and 11 months old mice (Fig. 3a). The genes clustered into two groups (Fig. 3a, Fig. S2c) - the expression of *Lgr5*, *Musashi-1*, *Ascl2*, *Sox4*, *Sox9*, *CD44*, *Olfm4* and *EphB3* was concentrated at crypt bottoms, leveling off towards the upper crypt positions. In contrast, *Bmi1*, *Prominin-1*, *Bmpr1a* and *mTert* exhibited a broad expression pattern that was nearly constant throughout the crypt axis (Fig. 3a, Fig. S2c). Notably, all genes for which stable progeny labeling using lineage tracing has been demonstrated were broadly co-expressed in CBC cells at lower crypt positions. More than 75% of *Lgr5*-positive cells contained *Bmi1* transcripts and almost half contained transcripts of *mTert* (Fig. 3b–e). This co-expression indicates that *Lgr5*, *Bmi1* and *mTert* do not mark distinct stem cell populations coexisting within a crypt. Our measurements can therefore explain the seemingly contradictory previously published results that demonstrated stable lineage tracing of progenies of cells expressing either of these genes[7, 9, 12].

Single-cell transcript correlations suggest regulatory connections

To infer the regulatory connections between the studied markers and to detect whether they are expressed in mutually exclusive cells, as has been suggested for *Lgr5* and *Bmi1*[1, 2], and *Lgr5* and *mTert*[9] we calculated the single-cell correlation coefficients of pairs of genes (Fig. 4). Gene pairs that are highly correlated could be regulated by a common upstream gene or directly regulate each other, whereas pairs that are not correlated are predicted to belong to different regulatory modules. Significant negative correlation of genes would indicate that they tend to be expressed in mutually exclusive cells.

We found that some gene pairs such as *Ascl2* and *Musashi-1* were highly correlated ($R = 0.7$, $p < 10^{-16}$, Fig. 4a) whereas others, such as *Bmi1* and *Ascl2* were expressed in a non-coordinated fashion ($R = -0.05$, $p = 0.74$, Fig. 4b). We next measured our panel in mutants for the two main regulators among the studied genes – a knock-out mouse of the polycomb gene *Bmi1* (Fig. 4c), and a conditional knock-out of the transcription factor *Ascl2*[16] (Fig. 4d). The duodenum in *Bmi1* knock-out mice was histologically similar to that in controls, as previously reported[16]. We found that the higher the single-cell correlations between pairs of genes in the wild-type mice, the higher the expression reduction in the respective mutants (Fig. 4e–g, $R = 0.76$, $p = 0.0045$). Thus positive single-cell correlations between pairs of genes are indicative of regulatory connection between them. *Lgr5* and *Bmi1* did not exhibit significant correlation regardless of the cell position along the crypt axis from which cells were sampled (Fig. 3b $R = -0.025$, $p = 0.9$) and they exhibited significant positive correlations with a mutually exclusive subset of markers (Fig. 4h). Thus our analysis indicates that *Lgr5* and *Bmi1* are broadly co-expressed in CBC cells but that they do not affect each other's expression and belong to different regulatory modules. *mTert* was also broadly co-expressed with *Lgr5* and these markers exhibited a slight positive correlation (Fig. 3c,e Fig. 4h, $R=0.13$, $p=0.002$).

Dcamkl-1 tuft cells co-express *Lgr5* and other stem cell markers throughout the crypt axis

A unique expression pattern was exhibited by *Dcamkl-1*. Unlike the broad expression patterns of the other stem-cell markers studied, we found that *Dcamkl-1* transcripts were strongly concentrated in isolated cells appearing once every 5 crypt sections (Fig. 5). These cells were widely distributed from lower crypt positions to villi (Fig. S4a) and specifically co-expressed the tuft cell marker *Cox1*[4] (Fig. 5a,b). Strikingly, *Dcamkl-1* cells at all crypt positions significantly co-expressed stem cell markers that were otherwise confined to crypt bottoms. These included *Lgr5* (median expression ratio with neighboring cells of 4.99, $p < 10^{-16}$, Fig. 5a,c, Fig. S4b) and *Sox9* (median ratio of 4.9, $p < 10^{-16}$, Fig. 5c). Other genes that were significantly expressed in *Dcamkl-1* cells were *Musashi-1* (Fig. S4c) *EphB2* and *EphB3* (Fig. 5c, Fig. S4f). While only a relatively small fraction of *Dcamkl-1* cells at the transit amplifying compartment exhibited *Lgr5* expression comparable to the *Lgr5* expression in CBC cells (12%), the appearance of *Lgr5* above the crypt base was confined to *Dcamkl-1* cells (Fig. S4d).

We next asked if the enrichment of stem cell genes in *Dcamkl-1* cells represents residual transcripts or rather a regulated expression signature (Fig. 5d). If *Dcamkl-1* cells are quiescent and migrate very rapidly transcript decay would be slower in these cells. Indeed, we found that *Dcamkl-1* were depleted in *Pcna* and *Ki67*, suggesting quiescence[37] (Fig. 5c. Fig. S6a). However these cells were enriched for *EphrinB1*, *EphB2* and *EphB3* (Fig. 5c, S4e,f), the expression of which has been shown to correlate with slower rather than faster migration rates[38]. In addition, only a subset of stem cell genes were enriched in *Dcamkl-1* cells, whereas others, such as *Olfm4* and *CD44* were not (Fig. 5c, Fig. S4g). Unlike *Lgr5*, *Olfm4* and *CD44* transcripts did not spatially decay more slowly in *Dcamkl-1* cells (Fig. 5e–f). Taken together, these findings indicate that some *Dcamkl-1* cells, exhibit a regulated

expression signature that includes stem-cell markers, which are otherwise confined to crypt bottoms.

Dcamkl-1 cells remain quiescent following irradiation

Enrichment of *Lgr5* and other stem cell markers in *Dcamkl-1* cells could potentially implicate *Dcamkl-1* cells as reserve stem cells. To address this possibility we repeated our single molecule transcript counting measurements on intestines of mice at different time points following whole body irradiation with 1 Gy, 6 Gy and 12 Gy (Fig. S5). These perturbations have been shown to cause a massive cell death followed by regeneration in intestinal crypts[39]. Indeed, irradiation with 12 Gy yielded a massive reduction in the number of crypts and their sizes, a phenomenon most prominent 48 hours after irradiation (Fig. S5b). 7 days after irradiation, we observed an increase in crypt sizes and extensive crypt fissions (Fig. S5c).

We found that *Dcamkl-1* cells did not enter cell-cycle following irradiation, as apparent by their low *Ki67* expression (Fig. S6a,b). The dynamics of *Dcamkl-1* cell numbers closely followed that of the short-lived differentiated goblet cells, exhibiting a decrease up to 48 hours after irradiation, followed by an increase at 7 days (Fig. S6c). In addition *Dcamkl-1* cells did not exhibit increased death rates 6 hours and 24 hours after 1 Gy irradiation, as detected morphologically, regardless of whether they had Lgr5 transcripts (Fig. S6d). Taken together these results do not support the possibility that *Dcamkl-1* cells serve as reserve stem cells.

Expression of some stem-cell markers expand to include the entire crypt following irradiation

To obtain a comprehensive view of the expression changes that occur following irradiation, and to detect genes among our panel that could be functionally important for the tissue repair process we also measured the entire panel at different time points after 12 Gy whole body irradiation. We found striking differences in the spatial expression patterns of some stem-cell markers 48 hours and 7 days after 12 Gy whole body irradiation relative to non-irradiated controls. These differences included a considerable expansion in both the spatial range of expression and levels of some stem cell markers (Fig. 6, Fig. S7). Most notable among these markers are *CD44* (Fig. 6a,d,e), *Musashi-1* (Fig. 6c,d,e) and *Ascl2* (Fig. 6c,d,e). Interestingly, both the levels and spatial range of *Olfm4* first decreased after 48 hours, and then significantly increased at 7 days (Fig. 6b,d,e). While the average spatial ranges of *Lgr5* and *Bmi1* slightly expanded following irradiation (Fig. 6d, Fig. S7a,d), their transcript levels did not change significantly (Fig. 6e).

DISCUSSION

Revealing the molecular identity of stem cells in the mouse intestine has been impeded by lack of sensitive *in-situ* expression measurements. Here we applied single molecule transcript counting to establish a comprehensive database of expression patterns in the mouse intestine and demonstrated that these measurements can shed light on stem-cell identities during homeostasis, aging and repair.

Our study revealed broad spatial expression profiles for three of the five genes for which stable lineage tracing of progenies has been demonstrated in the mouse intestine – *Bmi1*[7], *Prominin-1*[14] and *mTert*[9]. These were expressed throughout the crypt axis at almost constant levels, and contrasted with *Lgr5*[12] and to a slightly lesser extent *Sox9*[13], the expression of which were concentrated at lower crypt positions. Importantly, all five genes were co-expressed in crypt base-columnar-cells[11]. Thus *Bmi1*, *Prominin-1* and *mTert*, while clearly

expressed in stem cells, do not on their own specifically mark intestinal stem cells. These results emphasize the importance of sensitive *in-situ* transcript detection in mammalian tissue as a complementary approach to lineage tracing in determining the precise location in which candidate stem cell markers are expressed. While previous studies showed co-expression of *Lgr5* and *Bmi1*, as well as mTert by comparing expression between fractions of dissociated low and high *Lgr5*-GFP cells[16, 35], our measurements assess these co-expressions in a symmetric manner at the single cell level in WT mice and indicate the precise location of the cells co-expressing these stem cell markers (Fig. S3a). It should be stressed however that our analysis does not imply that all crypt cells that express both *Bmi1* and *Lgr5* have equal stem cell potential.

We detected a unique expression signature for *Dcamkl-1* cells, which includes significant co-expression with *Lgr5*. *Dcamkl-1* has recently been shown to be a marker of tuft cells, a rare quiescent epithelial lineage of unknown function[4, 40]. We found that regardless of their *Lgr5* expression, *Dcamkl-1* cells do not exhibit increased death rates following low dosage of gamma irradiation, as previously suggested for putative stem cells at higher crypt positions[6, 39]. Following high dosage of gamma irradiation these cells did not enter cell cycle at any time point and were depleted in proportion to goblet cells, a short-lived differentiated secretory cell type. Most importantly, all *Dcamkl-1* cells, both positive and negative for *Lgr5*, exhibited intense expression of the *Cox1* gene, a tuft cell differentiation marker[4]. While Lineage-tracing utilizing a *Dcamkl-1*-locus driven Cre transgene would definitely resolve the possibility that some tuft cells could posses potential stem cell function, our analysis suggests that such function is unlikely.

The appearance of transcripts at higher crypt positions should not necessarily imply that active proteins are present and may simply represent residual transcripts decaying slower than the rates at which cells migrate. We found however that the expression profile of genes such as *Ki67* and *Creb3l3* exhibited a dramatic change in levels over one vertical cell position at the crypt-villus borders (Fig. 2d), suggesting that transcript decay rates in intestinal crypts are faster than cell migration rates. Transcript levels detected by our method were also highly correlated with protein levels detected using GFP (Fig. 2a).

Our analysis indicates that during homeostasis the expression patterns of stem cell markers are remarkably invariant between crypts within the same mouse and with aging, with several markers such as *Lgr5,Olfm4,CD44,Ascl2* and *Musashi-1* exhibiting spatially overlapping expression patterns and high single-cell correlations. The expression program of these genes is however markedly different when the tissue is perturbed. This is evident from the dramatic expansion in range and numbers of *Ascl2,Musashi-1* and *CD44* transcripts following irradiation, which contrasts with the almost constant levels of *Lgr5* and *Bmi1*, and the more intricate behavior of *Olfm4* expression pattern, which first retracts and then expands. These varying responses observed following perturbation are indicative of potential functional differences among the stem cell markers in damage repair.

Our transcript-counting method should be considered as a complementary approach to protein-expression assays as well as to functional techniques such as lineage tracing[7, 12], cell ablation[8] and ex-vivo cultures[15]. Our method can be combined with these functional methods in two ways. One would be to use lineage tracing or ex vivo cultures to first detect potential stem-cell markers. Our method can then be applied to characterize in detail the spatial co-expression patterns of these markers in wild-type tissue. Alternatively, unbiased gene-expression measurements using a panel of single-molecule FISH probes could detect potentially interesting gene-expression signatures in terms of spatial distribution in a tissue or an unusual co-expression pattern of a few genes in isolated cells. These genes could then

be followed up with other techniques to assess the functional importance of these gene-expression signatures.

The homeostasis of epithelial tissues is based on a complex expression program, controlled by niche-dependent signals, as well as intracellular transcriptional and signaling networks. Here we have shown that single molecule transcript counting combined with computational approaches can yield a detailed characterization of the spatial expression profiles and the single cell co-expression patterns of key genes, as well as the changes during aging and tissue regeneration. Applying this technique to other tissues maintained by stem cells can provide important insights into the architecture of multi-cellular organisms, while similar studies in tumors can facilitate the detection of stem cell like signatures in cancer.

METHODS

Mice *and* tissue

All animal studies were reviewed and approved by the Massachusetts Institute of Technology (MIT) Committee on Animal Care. Duodenum tissue was harvested from C57bl6 female mice at ages 4 months and 11 months. Each age group included between two and five mice. *Bmi1* knock-out experiments were performed on 11 week old female *Bmi1*−/− and a *Bmi*+/+ littermate control[41]. These mice were $APC^{floxed/+}$, but since the mice had no Cre recombinase both were essentially wild-type for APC. The *Ascl2* mutant was an 18 weeks old male Ah-Cre/$Ascl2^{floxed/floxed}$, 5 days after induction of the Cre enzyme by intraperitoneal injections of 200 ml b-napthoflavone, as described in[16]. A non-induced Ah-Cre/$Ascl2^{floxed/floxed}$ littermate control was used for comparison. Whole body gamma-irradiation dosages of 1Gy, 6Gy and 12Gy were applied to 4 months old C57bl6 mice as described in[42]. Mice were sacrificed after 6 hr, 24 hr, 48 hr and 7 days. Two mice per irradiation dosage and sacrification time were used. The *Lgr5*-EGFP mouse used (Fig. 2a,b) was described in[12]. All mice were fed ad libitum and sacrificed in the morning. For all mice duodenum was removed, fixed in 4% Formaldehyde, then incubated overnight with 30% sucrose in 4% formaldehyde and then embedded in OCT. 6-micron cryo-sections were used for hybridizations.

Hybridizations and imaging

Probe libraries were designed and constructed as described in Raj et al.[25]. Most libraries consisted of 48 probes of length 20bps, complementary to the CDS of each gene (Supplementary table). *Lgr5* and *Ki67* libraries consisted of 96 probes. Hybridizations were done overnight with three differentially labeled probes using Cy5, Alexa594 and TMR fluorophores. An additional FIT-C conjugated antibody for E-cadherin (BD Biosciences) was added to the hybridization mix and used for protein immunefluorescence. DAPI dye for nuclear staining was added during the washes. Images were taken with a Nikon Ti-E inverted fluorescence microscope equipped with a 100× oil-immersion objective and a Photometrics Pixis 1024 CCD camera using MetaMorph software (Molecular Devices, Downington, PA). The image-plane pixel dimension was 0.13 microns. Quantification was done on stacks of 6−12 optical sections with Z-spacing of 0.3 microns, in which not more than a single cell was observed.

Dots were automatically detected using a custom Matlab program, implementing algorithms described in Raj et al[25]. Briefly, the dot stack images were first filtered with a 3-dimensional Laplacian of Gaussian filter of size 15 pixels and standard deviation of 1.5 pixels. The number of connected components in binary thresholded images was then recorded for a uniform range of intensity thresholds and the threshold for which the number of components was least sensitive to threshold selection was used for dot detection (Fig. S1a–d). Automatic

threshold selection was manually verified and corrected for errors (<5% of crypts). Background dots were detected according to size and by automatically identifying dots that appear in more than one channel (typically <1% of dots) and were removed. Such dots occasionally appeared in the surrounding mesenchymal cells but were rare in the epithelial cells. Bleed through of transcript signal between channels was minimal (Fig. S1e–g). Cell segmentation was performed manually on a maximal projection of the FIT-C channel. Cells at the lower 15 positions of the crypt were typically segmented. Transcript concentrations were obtained by dividing the number of transcripts per cell by the cell volume estimated as the product of the segmented area and the number of vertical stacks times a voxel size of 0.13um *0.13um *0.3um. Crypt apex and outline were manually marked and used to determine cell position along the crypt axis. For four genes – *Olfm4,Dcamkl-1,Gob5* and *Lysozyme,* transcript abundance was too high in some of the cells to facilitate reliable dot counting. In these cells the cytoplasm was often uniformly fluorescent (Fig. 2c Fig. S4g). Thus for these genes our dot counting algorithm underestimated the number of transcripts per cell.

Statistical analysis

Spatial profiles were symmetrized by averaging identical cell positions on both sides of the crypt apex, smoothed over 3 crypt positions and normalized to a maximum of 1. Co-expression analysis was performed on pooled cells from all crypts. Transcript concentrations were first normalized by the mean for each crypt, to correct for possible variations in hybridization or imaging conditions. Cells with no transcripts were assigned the lowest transcript concentration measured in the mouse. Spearman correlation coefficients of pairs of genes were compared to those obtained in randomized crypts in which the values of one of the genes was shuffled among cells, and p-values reported were computed by transforming the Z-score of the correlations compared to those in randomized crypts using the normal distribution.

Kolmogorov-Smirnov tests were used to generate p-values for the comparison of expression distributions. To generate the *Dcamkl-1* single cell expression signature we computed for each *Dcamkl-1*^{high} cell the ratio between the transcript concentration in the cell and its immediate neighbors that were not *Dcamkl-1*^{high}. *Dcamkl-1*^{high} cells were defined as cells with more than 5 *Dcamkl-1* transcripts; similar results were obtained for other thresholds. P-values for the median ratios were computed by creating randomized datasets in which the transcript concentrations of the gene of interest were arbitrarily swapped between *Dcamkl-1*^{high} cells and one of the neighboring cells and ratios were recalculated. Z-scores for the median ratios were transformed to p-values based on a normal distribution. When estimating the fraction of *Dcamkl-1*^{high} cells that were positive for *Lgr5* (*Lgr5*+) we applied a threshold equal to the median of the *Lgr5* expression at cells at or below cell position 5 that had at least one *Lgr5* transcript. Transcript spatial decay rate was computed by linear regression of the logarithm of the transcript concentration vs. crypt position.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors would like to thank H. Youk, S. Semrau, S. Klemm and K. Hilgendorf for comments on the manuscript, and X. Wu, Z. Peng Fan and A. Yang for help with the cell segmentation software. This work was supported by the National Institutes of Health (NIH)/National Cancer Institute Physical Sciences Oncology Center at MIT (U54CA143874) and an NIH Pioneer award (1DP1OD003936) to A.v.O., and in part by Cancer Center Support (core) grant P30-CA14051 from the National Cancer Institute. S.I. acknowledges support from a European Molecular Biology Organization postdoctoral fellowship, the International Human Frontiers Science Program

Organization and the Machiah Foundation. I.C.B. acknowledges support from the Howard Hughes Medical Institute Gilliam fellowship. T.J. is the D. H. Koch Professor of Biology and a D. K. Ludwig Scholar.

REFERENCES

1. Potten CS, Gandara R, Mahida YR, Loeffler M, Wright NA. The stem cells of small intestinal crypts: where are they? Cell Prolif. 2009; 42:731–750. [PubMed: 19788585]
2. Li L, Clevers H. Coexistence of quiescent and active adult stem cells in mammals. Science. 2010; 327:542–545. [PubMed: 20110496]
3. van der Flier LG, Clevers H. Stem cells, self-renewal, and differentiation in the intestinal epithelium. Annu Rev Physiol. 2009; 71:241–260. [PubMed: 18808327]
4. Gerbe F, et al. Distinct ATOH1 and Neurog3 requirements define tuft cells as a new secretory cell type in the intestinal epithelium. J Cell Biol. 2011; 192:767–780. [PubMed: 21383077]
5. Sato T, et al. Paneth cells constitute the niche for Lgr5 stem cells in intestinal crypts. Nature. 2010; 469:415–418. [PubMed: 21113151]
6. Potten CS. Stem cells in gastrointestinal epithelium: numbers, characteristics and death. Philos Trans R Soc Lond B Biol Sci. 1998; 353:821–830. [PubMed: 9684279]
7. Sangiorgi E, Capecchi MR. Bmi1 is expressed in vivo in intestinal stem cells. Nat Genet. 2008; 40:915–920. [PubMed: 18536716]
8. Tian H, et al. A reserve stem cell population in small intestine renders Lgr5-positive cells dispensable. Nature. 2011
9. Montgomery RK, et al. Mouse telomerase reverse transcriptase (mTert) expression marks slowly cycling intestinal stem cells. Proc Natl Acad Sci U S A. 2011; 108:179–184. [PubMed: 21173232]
10. May R, et al. Identification of a novel putative gastrointestinal stem cell and adenoma stem cell marker, doublecortin and CaM kinase-like-1, following radiation injury and in adenomatous polyposis coli/multiple intestinal neoplasia mice. Stem Cells. 2008; 26:630–637. [PubMed: 18055444]
11. Cheng H, Leblond CP. Origin, differentiation and renewal of the four main epithelial cell types in the mouse small intestine. V. Unitarian Theory of the origin of the four epithelial cell types. Am J Anat. 1974; 141:537–561. [PubMed: 4440635]
12. Barker N, et al. Identification of stem cells in small intestine and colon by marker gene Lgr5. Nature. 2007; 449:1003–1007. [PubMed: 17934449]
13. Furuyama K, et al. Continuous cell supply from a Sox9-expressing progenitor zone in adult liver, exocrine pancreas and intestine. Nat Genet. 2011; 43:34–41. [PubMed: 21113154]
14. Zhu L, et al. Prominin 1 marks intestinal stem cells that are susceptible to neoplastic transformation. Nature. 2009; 457:603–607. [PubMed: 19092805]
15. Sato T, et al. Single Lgr5 stem cells build crypt-villus structures in vitro without a mesenchymal niche. Nature. 2009; 459:262–265. [PubMed: 19329995]
16. van der Flier LG, et al. Transcription factor achaete scute-like 2 controls intestinal stem cell fate. Cell. 2009; 136:903–912. [PubMed: 19269367]
17. Giannakis M, et al. Molecular properties of adult mouse gastric and intestinal epithelial progenitors in their niches. J Biol Chem. 2006; 281:11292–11300. [PubMed: 16464855]
18. Raj A, van Oudenaarden A. Single-molecule approaches to stochastic gene expression. Annu Rev Biophys. 2009; 38:255–270. [PubMed: 19416069]
19. Gregorieff A, Clevers H. Wnt signaling in the intestinal epithelium: from endoderm to cancer. Genes Dev. 2005; 19:877–890. [PubMed: 15833914]
20. Femino AM, Fay FS, Fogarty K, Singer RH. Visualization of single RNA transcripts in situ. Science. 1998; 280:585–590. [PubMed: 9554849]
21. Levsky JM, Shenoy SM, Pezo RC, Singer RH. Single-cell gene expression profiling. Science. 2002; 297:836–840. [PubMed: 12161654]
22. Zenklusen D, Larson DR, Singer RH. Single-RNA counting reveals alternative modes of gene expression in yeast. Nat Struct Mol Biol. 2008; 15:1263–1271. [PubMed: 19011635]
23. Raj A, Peskin CS, Tranchina D, Vargas DY, Tyagi S. Stochastic mRNA synthesis in mammalian cells. PLoS Biol. 2006; 4:e309. [PubMed: 17048983]

24. Capodieci P, et al. Gene expression profiling in single cells within tissue. Nat Methods. 2005; 2:663–665. [PubMed: 16118636]
25. Raj A, van den Bogaard P, Rifkin SA, van Oudenaarden A, Tyagi S. Imaging individual mRNA molecules using multiple singly labeled probes. Nat Methods. 2008; 5:877–879. [PubMed: 18806792]
26. Raj A, Rifkin SA, Andersen E, van Oudenaarden A. Variability in gene expression underlies incomplete penetrance. Nature. 2010; 463:913–918. [PubMed: 20164922]
27. Carmon KS, Gong X, Lin Q, Thomas A, Liu Q. R-spondins function as ligands of the orphan receptors LGR4 and LGR5 to regulate Wnt/{beta}-catenin signaling. Proc Natl Acad Sci U S A. 2011; 108:11452–11457. [PubMed: 21693646]
28. de Lau W, et al. Lgr5 homologues associate with Wnt receptors and mediate R-spondin signalling. Nature. 2011
29. Zeilstra J, et al. Deletion of the WNT target and cancer stem cell marker CD44 in Apc(Min/+) mice attenuates intestinal tumorigenesis. Cancer Res. 2008; 68:3655–3661. [PubMed: 18483247]
30. Sinner D, et al. Sox17 and Sox4 differentially regulate beta-catenin/T-cell factor activity and proliferation of colon carcinoma cells. Mol Cell Biol. 2007; 27:7802–7815. [PubMed: 17875931]
31. Gregorieff A, et al. Expression pattern of Wnt signaling components in the adult intestine. Gastroenterology. 2005; 129:626–638. [PubMed: 16083717]
32. Kayahara T, et al. Candidate markers for stem and early progenitor cells, Musashi-1 and Hes1, are expressed in crypt base columnar cells of mouse small intestine. FEBS Lett. 2003; 535:131–135. [PubMed: 12560091]
33. Potten CS, et al. Identification of a putative intestinal stem cell and early lineage marker; musashi-1. Differentiation. 2003; 71:28–41. [PubMed: 12558601]
34. He XC, et al. BMP signaling inhibits intestinal stem cell self-renewal through suppression of Wnt-beta-catenin signaling. Nat Genet. 2004; 36:1117–1121. [PubMed: 15378062]
35. Schepers AG, Vries R, van den Born M, van de Wetering M, Clevers H. Lgr5 intestinal stem cells have high telomerase activity and randomly segregate their chromosomes. EMBO J. 2011; 30:1104–1109. [PubMed: 21297579]
36. http://jaxmice.jax.org/strain/008875.html
37. Kohler T, Prols F, Brand-Saberi B. PCNA in situ hybridization: a novel and reliable tool for detection of dynamic changes in proliferative activity. Histochem Cell Biol. 2005; 123:315–327. [PubMed: 15616846]
38. Batlle E, et al. Beta-catenin and TCF mediate cell positioning in the intestinal epithelium by controlling the expression of EphB/ephrinB. Cell. 2002; 111:251–263. [PubMed: 12408869]
39. Potten CS, Booth C, Pritchard DM. The intestinal epithelial stem cell: the mucosal governor. Int J Exp Pathol. 1997; 78:219–243. [PubMed: 9505935]
40. Gerbe F, Brulin B, Makrini L, Legraverend C, Jay P. DCAMKL-1 expression identifies Tuft cells rather than stem cells in the adult mouse intestinal epithelium. Gastroenterology. 137:2179–2180. author reply 2180–2171 (2009). [PubMed: 19879217]
41. van der Lugt NM, et al. Posterior transformation, neurological abnormalities, and severe hematopoietic defects in mice with a targeted deletion of the bmi-1 proto oncogene. Genes Dev. 1994; 8:757–769. [PubMed: 7926765]
42. Kirsch DG, et al. p53 controls radiation-induced gastrointestinal syndrome in mice independent of apoptosis. Science. 2010; 327:593–596. [PubMed: 20019247]

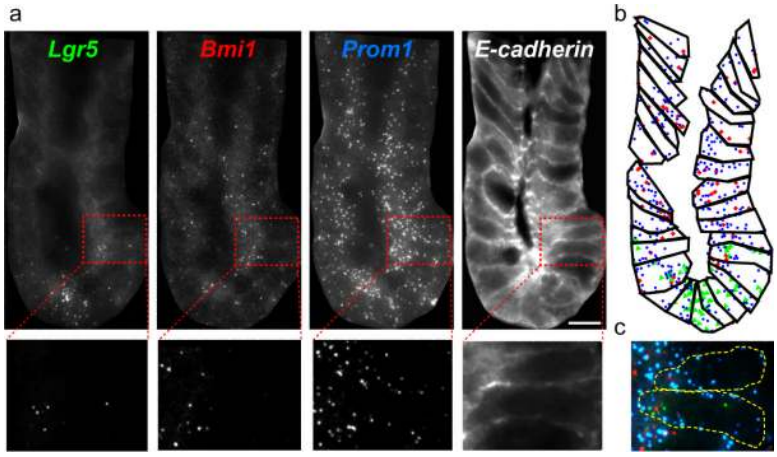

Figure 1. Three-color single molecule FISH of intestinal stem cell markers

(a) Small intestinal fixed tissue sections were simultaneously hybridized with three differentially labeled probe libraries (here *Lgr5* – TMR (green), *Bmi1* – cy5 (red), and *Prominin-1* – Alexa594 (blue)). Single transcripts appear as diffraction limited spots under a fluorescent microscope. Additionally, FITC-E-cadherin antibody labels cell membranes. Magnification of a representative area highlighted in red is shown below. Images are maximal projections of stacks of 20 optical sections spaced 0.3 microns apart. **(b)** Segmented crypt with transcripts for *Lgr5* (green triangles), *Bmi1* (red diamonds) and Prominin-1 (blue circles). Dots and cell borders are based on 10 optical sections from **(a)**. **(c)** Area highlighted in **(a)** showing the simultaneous detection of transcripts for *Lgr5* (green), *Bmi1* (red) and *Prominin-1* (blue). Dashed outlines denote cell borders. Scale bar is 5 microns.

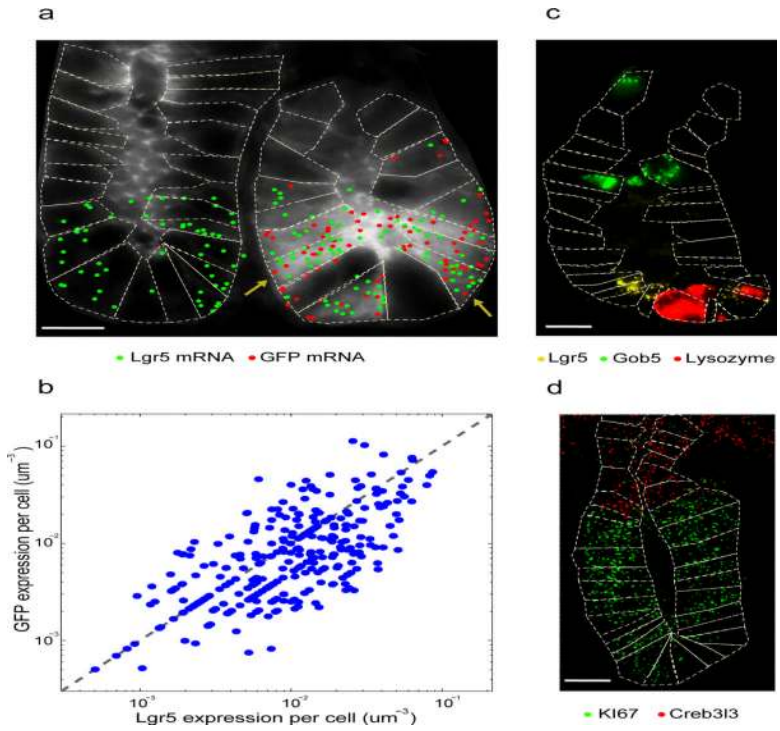

Figure 2. Single molecule FISH correlates with reporter expression in transgenic mice but provides a much broader sampling

(a) Expression analysis in the *Lgr5*-EGFP transgenic mice. Shown are two crypts, one positive for the transgene expression (right) and one negative (left). Gray scale reflects the GFP measurements. Green dots are automatically detected Lgr5 endogenous transcripts, red dots are EGFP transcripts, dashed lines mark cell borders, based on immunofluorescense with FITC-Ecadherin. Arrows point at cells with high GFP fluorescence. Unlike the transgene which was expressed once every 10 crypts, the endogenous transcripts were detected in each and every crypt. **(b)** Endogenous *Lgr5* transcript levels are highly correlated with EGFP transcript levels in the crypts in which the transgene is active (Spearman correlation R = 0.68, $p < 10^{-68}$). Analysis based on simultaneous single molecule FISH with probe libraries for *Lgr5* and EGFP. **(c–d)** Hybridization with single molecule FISH libraries yields highly localized and specific expression patterns. **(c)** Intestinal crypt hybridized with probes for Paneth cell marker *Lysozyme* (red), goblet cell marker *Gob5* (green) and stem cell marker *Lgr5* (yellow). **(d)** Intestinal crypt hybridized with the proliferation marker *Ki67* (green) and with the enterocyte marker *Creb3l3* (red). The sharp decline in expression at the crypt-villus border demonstrates that rates of transcript degradation are faster than cell migration rates in intestinal crypts. Dashed lines mark cell borders. Scale bars in (**a,c,d**) are 5 microns.

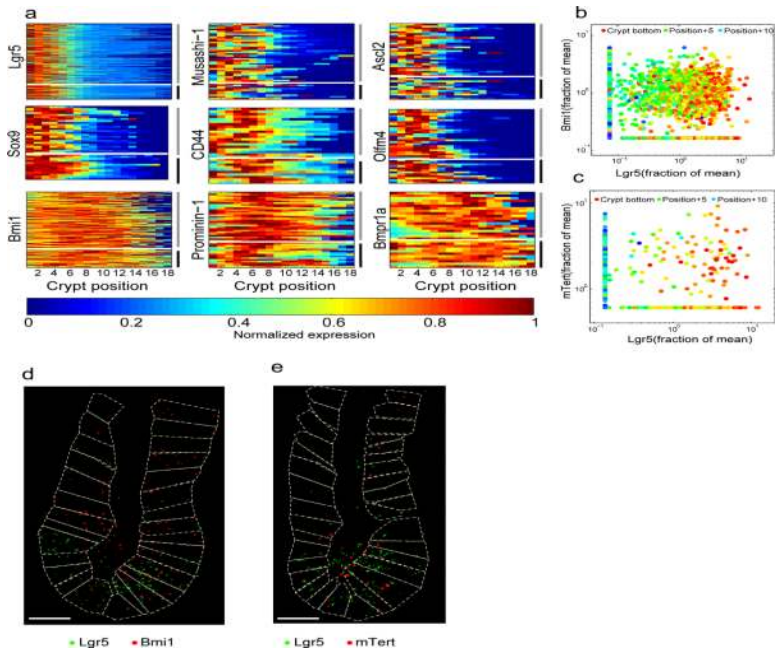

Figure 3. Spatial expression profiles of stem cell markers are broad and overlap at crypt-base-columnar cells

(a) Spatial expression profiles of stem-cell markers are invariant between crypts and to aging. Rows are different crypts; columns are crypt positions, position 0 is the crypt apex. All crypts above the white horizontal lines are from a 4-month-old mouse (marked with gray vertical bars); all crypts below the white lines are from an 11-month-old mouse (marked with black vertical bars). **(b)** *Bmi1* and *Lgr5* are extensively co-expressed in a non-correlated manner ($R = -0.025$, $p = 0.9$). Dots represent pooled single cells from crypts of a wild-type 4-month-old mouse. Coordinates are the transcript concentrations divided by the mean concentration within the crypt from which the cell was sampled (cells with no transcripts were assigned the lowest non-zero concentration detected). Dot colors correspond to position along the crypt axis. 76% of Lgr5-positive cells contain Bmi1 transcripts (1073/1417) whereas 48% of *Bmi1*-positive cells contain *Lgr5* transcripts (1073/2221). **(c)** *mTert* and *Lgr5* are co-expressed in CBC cells ($R=0.13$, $p=0.002$). **(d–e)** *Lgr5* (green dots) and *Bmi1* (red dots, **d**) as well as *mTert* (red dots, **e**) are co-expressed in crypt-base-cells. Dashed lines mark cell borders. Images are maximal projections of 15 optical sections spaced 0.3 microns apart, filtered with a Laplacian of Gaussian filter (Methods section). Scale bars in (**d,e**) are 5 microns.

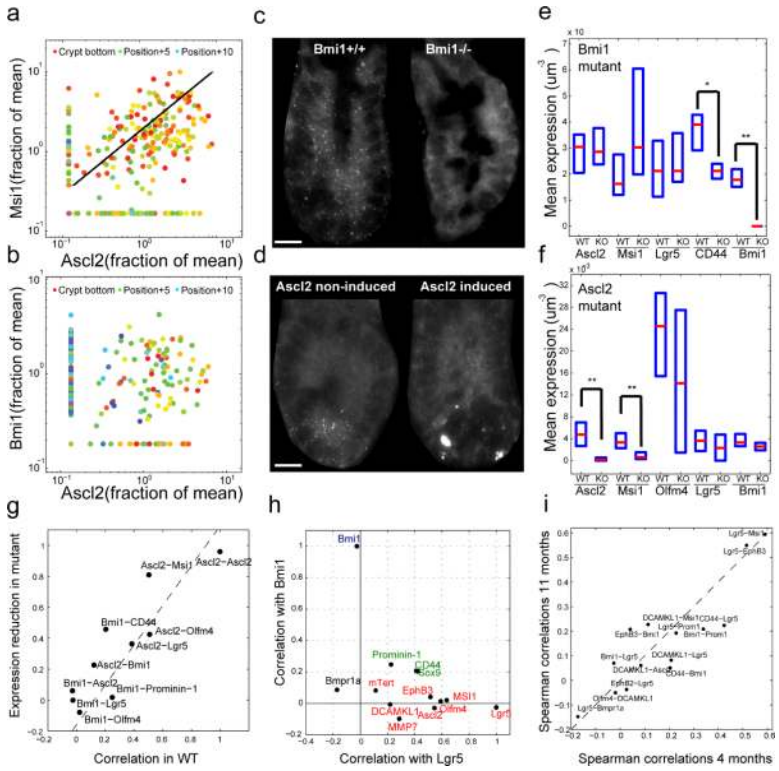

Figure 4. Single cell correlations of stem cell markers are validated with mutants for key regulator genes

(a) *Musashi-1* (*Msi1*) and *Ascl2* are highly correlated at the single cell level (Spearman correlation R = 0.7, $p < 10^{-16}$). **(b)** Bmi1 and *Ascl2* are not significantly correlated (R = −0.05, p = 0.74). Dots in **(a,b)** represent pooled single cells from crypts of a 4-month-old mouse. Coordinates are the transcript concentrations divided by the average concentration within the crypt from which the cell was sampled (cells with no transcripts were assigned the lowest non-zero concentration detected). Dot colors correspond to position along the crypt axis. **(c)** *Bmi1* transcripts are detected in a WT mouse (left) but not in a *Bmi1* homozygous knock-out mouse (right). **(d)** Ascl2 transcripts are significantly reduced in *Ah-Cre/Ascl2*$^{floxed/floxed}$ mouse 5 days after ßNF induction (right) compared to non-induced controls (left). Scale bars in (**c,d**) are 5 microns. **(e)** Deletion of *Bmi1* significantly reduces the expression of *CD44* and *Bmi1*. Shown are the distributions of the mean transcript concentrations per crypt cell for the WT (*Bmi1*+/+) and the mutant (*Bmi1*−/−) mice, where horizontal red lines are median concentrations and boxes delimit the 25–75 percentiles. * p < 0.02, ** p < 0.001. **(f)** As in **(e)**, shown are the distributions of mean transcript concentrations per crypt cell for non-induced (*Ascl2*+/+) and induced (*Ascl2*−/−) mice. **(g)** Reduction in expression of stem-cell markers in mice mutant for *Bmi1* and *Ascl2* is correlated with the Spearman correlation coefficients of these markers with either *Bmi1* or *Ascl2* respectively in the wild-type (R = 0.76, p = 0.0045). Expression reduction for each gene is the difference in median transcript concentration between the WT and the mutant, divided by the WT median levels. (**h**) Correlation map of stem-cell markers with *Bmi1* and *Lgr5*. Axes are the Spearman correlations between single-cell transcript concentration in either *Lgr5* (x-axis) or *Bmi1* (y-axis). Red denote significant correlation with *Lgr5*, Blue denote significant correlation with *Bmi1* and green with both. **(i)** Pairwise correlations are highly reproducible between a 4-month-old mouse and an 11-month-old mouse (R=0.88).

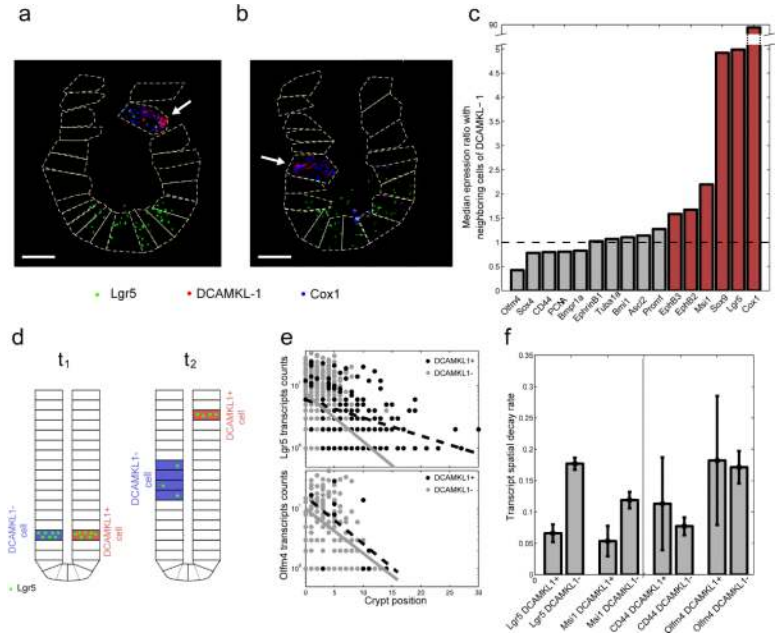

Figure 5. *Dcamkl-1* marks tuft cells occurring throughout the crypt axis that significantly co-express stem-cell markers otherwise confined to crypt bottoms

(a) Example of significant co-expression of *Lgr5* (green dots) in a cell with high *Dcamkl-1* transcript levels (red dots). **(b)** Example of a *Dcamkl-1* cell with no *Lgr5* expression. Images in **(a, b),** are maximal projections of 15 optical sections spaced 0.3 microns apart, filtered with a Laplacian of Gaussian filter (Methods section). Blue dots are *Cox1* transcripts, dashed lines mark cell borders and arrows mark the tuft cells. Scale bars are 5 microns. **(c)** Expression signature for *Dcamkl-1* cells. Shown are the median ratios of transcripts for different genes between *Dcamkl-1*^{high} cells (cells with more than 5 transcripts) and the average levels in their immediate two neighboring cells (above and below). Red bars are ratios that are significant relative to permuted crypts (see Methods), using false discovery rate of 10%. **(d)** Passive migration model for the elevated transcript levels of *Lgr5* in *Dcamkl-1* cells. Green dots represent transcripts of *Lgr5* (or other stem cell markers co-expressed in *Dcamkl-1* cells) in the progenies of two different cells migrating away from the stem-cell zone at the crypt bottoms – a *Dcamkl-1* positive cell (red) and *Dcamkl-1* negative cell (blue). t_1 and t_2 indicate successive times. If *Dcamkl-1* positive cells do not divide and migrate more rapidly than other cells, the spatial decay rate of the stem cell marker transcripts such as *Lgr5* would be lower. **(e)** *Lgr5* transcripts in *Dcamkl-1*^{high} cells (black dots) decay more slowly with crypt position than *Lgr5* transcripts in *Dcamkl-1* negative cells (gray dots). Transcripts of Olfm4 in *Dcamkl-1*^{high} cells (black) exhibit the same decay rate as in *Dcamkl-1* negative cells (gray). Lines are exponential fits. **(f)** Transcript decay rates of Lgr5 and Musashi-1 is more than twice slower in *Dcamkl-1*^{high} cells compared to*Dcamkl-1* negative cells, but comparable for *CD44* and *Olfm4*.

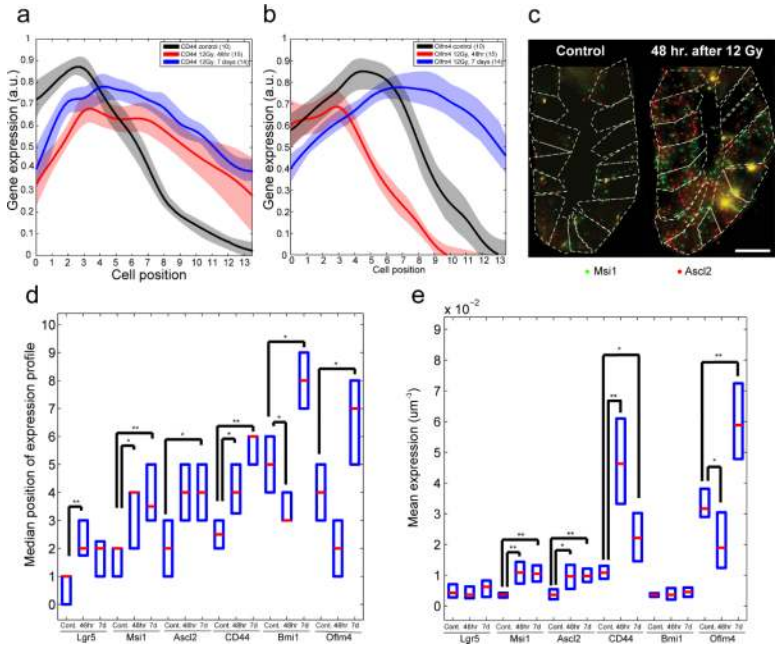

Figure 6. 12 Gy whole body irradiation results in significant changes in the levels and range of stem cell markers

(a–b), Average spatial expression profiles for *CD44* **(a)** and *Olfm4* **(b)** in a non-irradiated control mouse (black) and in mice irradiated with 12 Gy and sacrificed after 48 hours (red) and 7 days (blue). **(c)** Transcript levels of *Musashi-1* (green dots) and *Ascl2* (red dots) are significantly increased 48 hours after whole body 12Gy irradiation. Images are maximal projections of 6 optical sections spaced 0.3 microns apart. Scale bar is 10 microns. **(d)** The medians of the spatial expression profiles significantly increase for most stem cell markers following gamma irradiation. Shown are the distributions among different crypts of the median cell positions of the spatial expression profiles. Horizontal red lines are the medians of the distributions and boxes delimit the 25–75 percentiles. * - $p < 0.03$, ** $p < 0.001$. **(e)** Mean transcript concentration per crypt cell significantly increases for some stem cell markers (*Musashi-1*, *Ascl2*, *CD44*, *Olfm4*) but not for others (*Lgr5*, *Bmi1*). Horizontal red lines are the medians of the distributions and boxes delimit the 25–75 percentiles. * - $p < 0.01$, ** $p < 0.001$.