

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

Shalev Itzkovitz<sup>1,2</sup>, Anna Lyubimova<sup>1,2,3</sup>, Irene C. Blat<sup>2,4</sup>, Mindy Maynard<sup>4</sup>, Johan van Es<sup>3</sup>, Jacqueline Lees<sup>2,4</sup>, Tyler Jacks<sup>2,4,5</sup>, Hans Clevers<sup>3</sup> and Alexander van Oudenaarden<sup>1,2,3,4,6</sup>

**Determining the molecular identities of adult stem cells requires technologies for sensitive transcript detection in tissues. In mouse intestinal crypts, lineage-tracing studies indicated 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-colour single-molecule fluorescent *in situ* hybridization to study a comprehensive panel of intestinal stem-cell markers during homeostasis, ageing and regeneration. We find that the expression of all markers overlaps at crypt-base cells. This co-expression includes *Lgr5*, *Bmi1* and *mTert*, genes previously suggested to mark distinct stem cells. Strikingly, *Dcamk1l* 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, indicating 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<sup>1,2</sup>. 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 postmitotic and differentiate

into either enterocytes—nutrient-absorbing cells that form the bulk of the tissue—or several types of secretory cell, including goblet cells, enteroendocrine cells and tuft cells<sup>3,4</sup>. The differentiated cells continue to migrate up, exiting the crypts towards larger invaginations into the lumen called villi. They are finally extruded from the tops of the villi about five 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 defence and stem-cell maintenance<sup>5</sup>.

Although 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, which seem mutually exclusive<sup>1,2</sup>. The ‘+4 hypothesis’, originally proposed in ref. 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 indicated specific expression of genes such as *Bmi1* (refs 7,8), *mTert* (ref. 9) and *Dcamk1l* (also known as *Dclk1*; ref. 10). Alternatively, the stem-cell zone hypothesis originally formulated in refs 11,12 posits that crypt-base columnar (CBC) cells residing at the very bottom of the crypts are the actual stem cells. Although independent lineage tracing studies using *Lgr5* (ref. 12), *Sox9* (ref. 13) and *Prominin-1* (ref. 14) have demonstrated stable labelling 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*<sup>15</sup>, lineage tracing with *Bmi1* (refs 7,8) and *mTert* (ref. 9) has implied the +4 cell as the stem cell of the small intestine. It was recently demonstrated that progenies of cells expressing *Bmi1* can give rise to new *Lgr5* stem cells following complete ablation of CBC cells<sup>8</sup>. These results pose the question of whether two or more distinct stem-cell populations uniquely marked by these genes coexist in mouse intestinal crypts<sup>1,2</sup>.

<sup>1</sup>Department of Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. <sup>2</sup>Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. <sup>3</sup>Hubrecht Institute–KNAW (Royal Netherlands Academy of Arts and Sciences) and University Medical Center Utrecht, Uppsalalaan 8, 3584 CT Utrecht, Netherlands. <sup>4</sup>Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. <sup>5</sup>Howard Hughes Medical Institute, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.

<sup>6</sup>Correspondence should be addressed to A.v.O. (e-mail: avano@mit.edu)

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 labelling 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 programme were based on methods such as quantitative PCR or microarray analysis of green fluorescent protein (GFP)-sorted cell populations<sup>15,16</sup> or laser-capture microdissected tissue<sup>17</sup>. Despite yielding important insights, these methods have several disadvantages, such as the use of knock-in mice, standardization issues related to the quantitative PCR process, insufficient sensitivity for the analysis of single cells and most importantly the loss of spatial information<sup>18</sup>. Immunohistochemistry and classic RNA *in situ* hybridizations<sup>19</sup> preserve tissue morphology, but sensitivity and specificity problems limit the generic use of these methods in yielding quantifiable co-expression data for several genes at the single-cell level. To overcome these limitations, several studies used multiply labelled fluorescent probes to detect single messenger RNA in fixed yeast and mammalian cells<sup>20–23</sup> as well as nuclear transcription sites in paraffin-embedded tissue<sup>24</sup>. 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 labelled fluorescent probes<sup>25</sup>, 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*<sup>25</sup> and *Caenorhabditis elegans* embryos<sup>26</sup>. Here we apply this method to mouse intestinal frozen sections, to obtain 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 fluorescence *in situ* hybridization (FISH) enables sensitive and specific *in situ* transcript detection in intestinal tissue

We designed a panel of 15 libraries of fluorescently labelled probes, each composed of 48 20-base pair (bp) oligos complementary to the coding sequences of previously indicated stem-cell markers (Fig. 1 and Supplementary Table S1). These included the R-spondin receptor *Lgr5* (refs 12,27,28), the Wnt targets *Ascl2* (ref. 16), *Cd44* (ref. 29), *Sox4* (ref. 30), *Sox9* (ref. 13), *Mmp7*, *EphB2* and *EphB3* (ref. 31), the RNA-binding protein *Musashi-1* (refs 32,33), *Olfactomedin-4* (*Olfm4*; ref. 16), *Prominin-1* (*CD133*; ref. 14), *Dcamk1l* (refs 10, 17), *Bmpr1a* (ref. 34) and mouse telomerase reverse transcriptase (*mTert*; refs 9,35) as well as the polycomb gene *Bmi1* (ref. 7). Hybridization of 6 µm cryosections 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 (Supplementary 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 on the basis of E-cadherin lateral membrane staining.

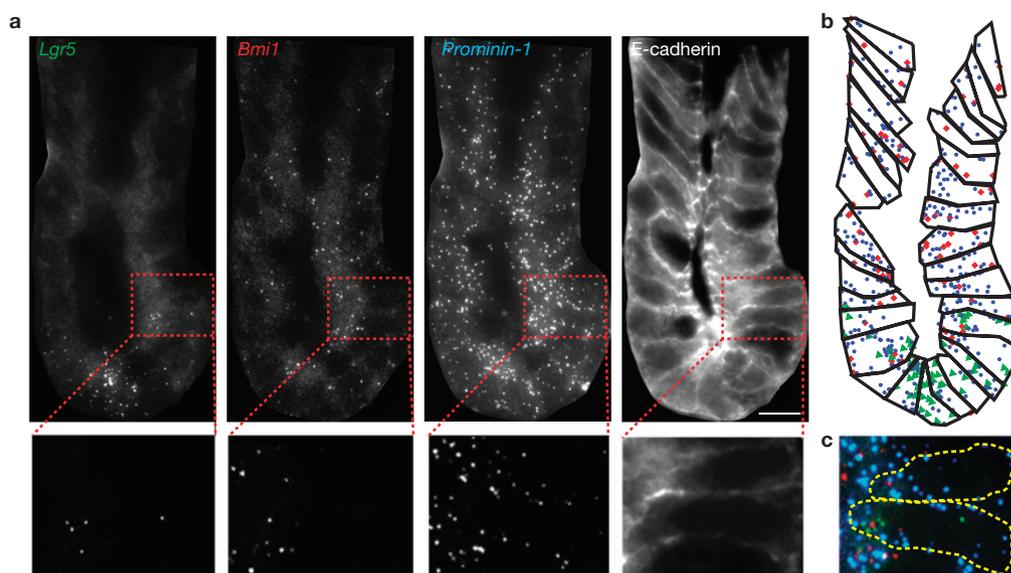
We first assessed whether our transcript-counting method correlates with the expression patterns in reporter mice. To this end we examined both fluorescence and transcript levels in the *Lgr5*-eGFP (enhanced GFP) reporter mouse model<sup>12</sup>. We detected cells with intense GFP signal, as well as eGFP transcripts at crypt bottoms in only one out of ten crypts on average, consistent with the pronounced variegated expression pattern previously reported (<http://jaxmice.jax.org/strain/008875.html>) (Fig. 2a). Unlike the patchy expression of the *Lgr5* reporter transgene, we uniformly detected the endogenous *Lgr5* transcripts in every crypt throughout the tissue. Importantly, the expression levels 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 analysed the expression of the intestinal differentiation markers *Gob5*, *Creb3l3* and *Lysozyme* and the proliferation marker *Ki67* (also known as *Mki67*). 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-month- and 11-month-old mice (Fig. 3a). The genes clustered into two groups (Fig. 3a and Supplementary Fig. S2c). The expression of *Lgr5*, *Musashi-1*, *Ascl2*, *Sox4*, *Sox9*, *Cd44*, *Olfm4* and *EphB3* was concentrated at crypt bottoms, levelling 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 and Supplementary Fig. S2c). Notably, all genes for which stable progeny labelling 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 and Supplementary Fig. S3). This co-expression explains the seemingly contradictory previously published results that demonstrated stable lineage tracing of progenies of cells expressing either of these genes<sup>7,9,12</sup>.

### Single-cell transcript correlations indicate regulatory connections

To infer the regulatory connections between the studied markers and to detect whether they are expressed in mutually exclusive



**Figure 1** Three-colour single-molecule FISH of intestinal stem-cell markers. **(a)** Small-intestinal fixed tissue sections were simultaneously hybridized with three differentially labelled probe libraries (here *Lgr5*–TMR (tetramethylrhodamine, green), *Bmi1*–cy5 (red) and *Prominin-1*–Alexa594 (blue)). Single transcripts appear as diffraction-limited spots under a fluorescent microscope. Fluorescein isothiocyanate (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  $0.3\mu\text{m}$  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 ten 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,  $5\mu\text{m}$ .

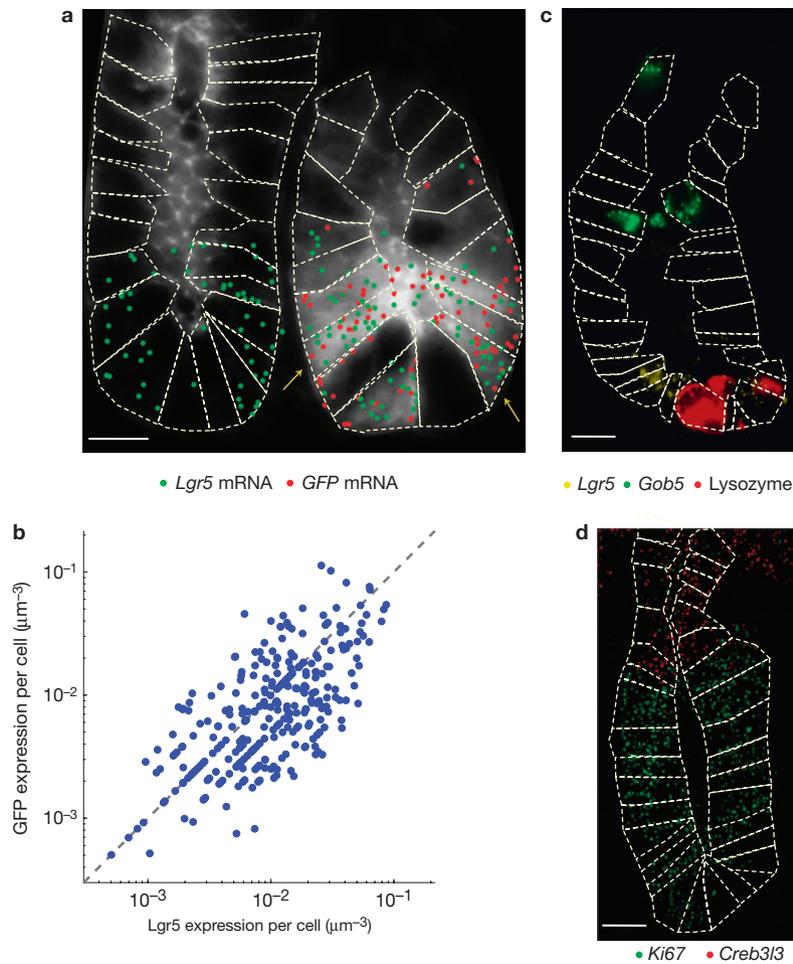
cells, as has been suggested for *Lgr5* and *Bmi1* (refs 1,2), and *Lgr5* and *mTert* (ref. 9), we calculated the single-cell correlation coefficients of pairs of genes (Fig. 4). Gene pairs that are highly correlated could be regulated by a common upstream gene or directly regulate each other, whereas 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 *Bmi1* knockout mouse (Fig. 4c), and a conditional knockout of the transcription factor *Ascl2* (ref. 16; Fig. 4d). The duodenum in *Bmi1* knockout mice was histologically similar to that in controls, as previously reported<sup>16</sup>. 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 (Figs 3c,e, 4h,  $R = 0.13$ ,  $P = 0.002$ ).

### *Dcamk1l* tuft cells co-express *Lgr5* and other stem-cell markers throughout the crypt axis

A unique expression pattern was exhibited by *Dcamk1l*. Unlike the broad expression patterns of the other stem-cell markers studied, we found that *Dcamk1l* transcripts were strongly concentrated in isolated cells appearing once every five crypt sections (Fig. 5). These cells were widely distributed from lower crypt positions to villi (Supplementary Fig. S4a) and specifically co-expressed the tuft-cell marker *Cox1* (ref. 4) (Fig. 5a,b). Strikingly, *Dcamk1l* cells at all crypt positions significantly co-expressed stem-cell markers that were otherwise confined to crypt bottoms. These included the Wnt targets *Lgr5* (median expression ratio with neighbouring cells of 4.99,  $P < 10^{-16}$ , Fig. 5a,c and Supplementary Fig. S4b) and *Sox9* (median ratio of 4.9,  $P < 10^{-16}$ , Fig. 5c). Other genes that were significantly expressed in *Dcamk1l* cells were *Musashi-1* (Supplementary Fig. S4c), *EphB2* and *EphB3* (Fig. 5c and Supplementary Fig. S4f). Although only a relatively small fraction of *Dcamk1l* 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 *Dcamk1l* cells (Supplementary Fig. S4d).

We next asked if the enrichment of stem-cell genes in *Dcamk1l* cells represents residual transcripts or rather a regulated expression signature (Fig. 5d). If *Dcamk1l* cells are quiescent and migrate very rapidly, transcript decay will be slower in these cells. Indeed, we found that *Dcamk1l* had low *Pcna* and *Ki67* expression, indicating quiescence<sup>36</sup> (Fig. 5c and Supplementary Fig. S6a). However these cells were enriched for *EphrinB1*, *EphB2* and *EphB3* transcripts (Fig. 5c and Supplementary Fig. S4e,f), the expression of which has been shown to correlate with lower rather than higher migration rates<sup>37</sup>. In addition, only a subset of stem-cell genes were enriched in *Dcamk1l*



**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). The grey scale reflects the GFP measurements. Green dots are automatically detected *Lgr5* endogenous transcripts, red dots are eGFP transcripts and dashed lines mark cell borders, on the basis of immunofluorescence with FITC-E-cadherin. Arrows point to cells with high GFP fluorescence. Unlike the transgene, which was expressed once every ten 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 on the basis of 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 image was filtered with a Laplacian of Gaussian filter (Methods). 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, 5  $\mu\text{m}$ .

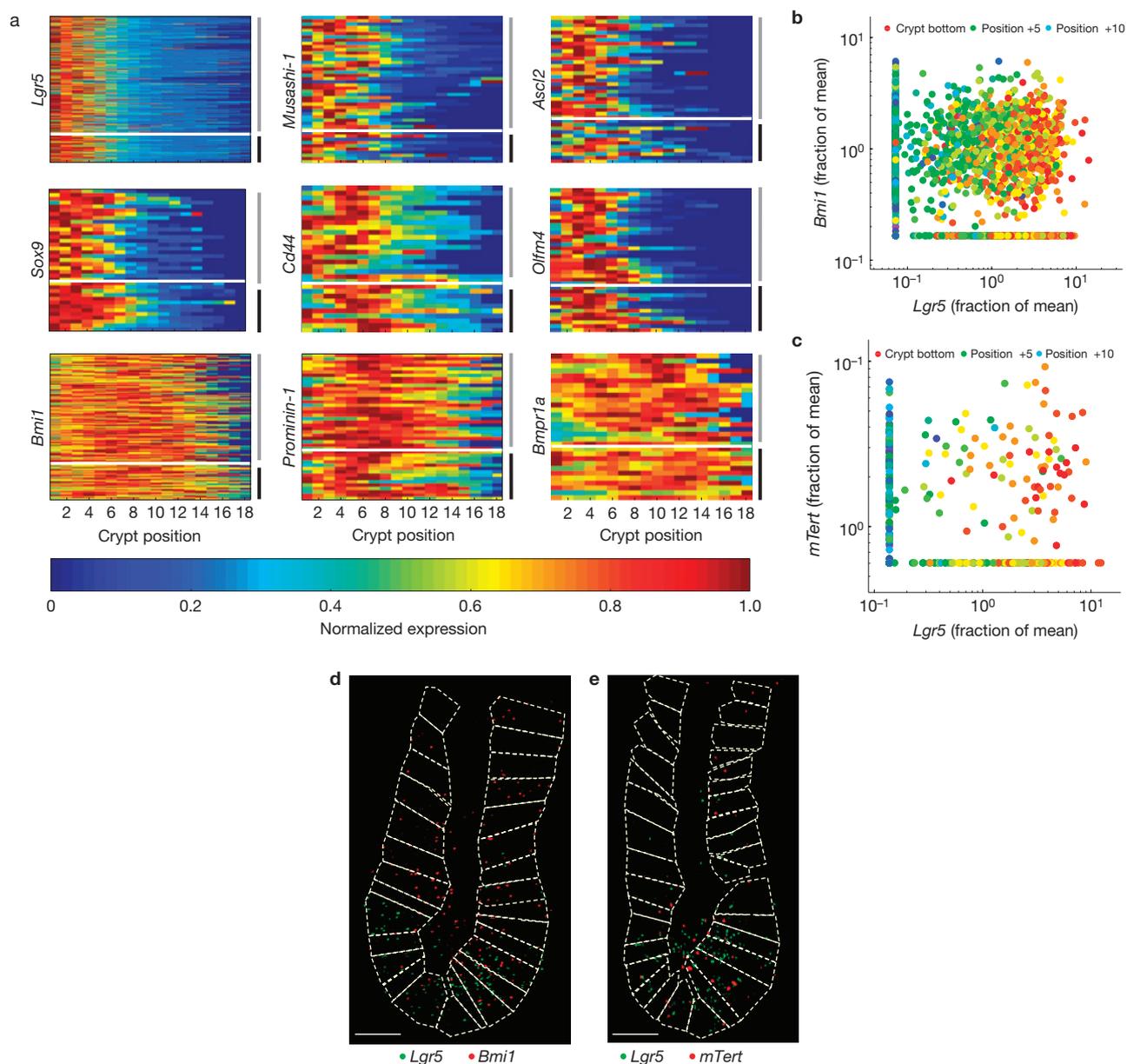
cells, whereas others, such as *Olfm4* and *Cd44*, were not (Fig. 5c and Supplementary Fig. S4g). Unlike *Lgr5*, *Olfm4* and *Cd44* transcripts did not spatially decay more slowly in *Dcamk1l* cells (Fig. 5e,f). Taken together, these findings indicate that some *Dcamk1l* cells exhibit a regulated expression signature that includes stem-cell markers otherwise confined to crypt bottoms.

***Dcamk1l* cells remain quiescent following irradiation**

Enrichment of *Lgr5* and other stem-cell markers in *Dcamk1l* cells could potentially implicate *Dcamk1l* cells as reserve stem cells. To address this possibility we repeated our measurements at different times following whole-body irradiation with 1 Gy, 6 Gy and 12 Gy (Supplementary Fig. S5). These perturbations have been shown to cause a massive cell death followed by regeneration in intestinal crypts<sup>38</sup>.

Indeed, irradiation with 12 Gy yielded a massive reduction in the number of crypts and their sizes, a phenomenon most prominent 48 h after irradiation (Supplementary Fig. S5b). Seven days after irradiation, we observed an increase in crypt sizes and extensive crypt fissions (Supplementary Fig. S5c).

We found that *Dcamk1l* cells did not enter cell cycle following irradiation, as apparent from their low *Ki67* expression (Supplementary Fig. S6a,b). The dynamics of *Dcamk1l* cell numbers closely followed that of the short-lived differentiated goblet cells, exhibiting a decrease up to 48 h after irradiation, followed by an increase at seven days (Supplementary Fig. S6c). Moreover, *Dcamk1l* cells did not exhibit increased death rates 6 h and 24 h after 1 Gy irradiation, as detected morphologically, regardless of whether they had *Lgr5* transcripts (Supplementary Fig. S6d). Taken together, these results do not



**Figure 3** Spatial expression profiles of stem-cell markers are broad and overlap at CBC cells. **(a)** Spatial expression profiles of stem-cell markers are invariant between crypts and to ageing. 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 grey 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

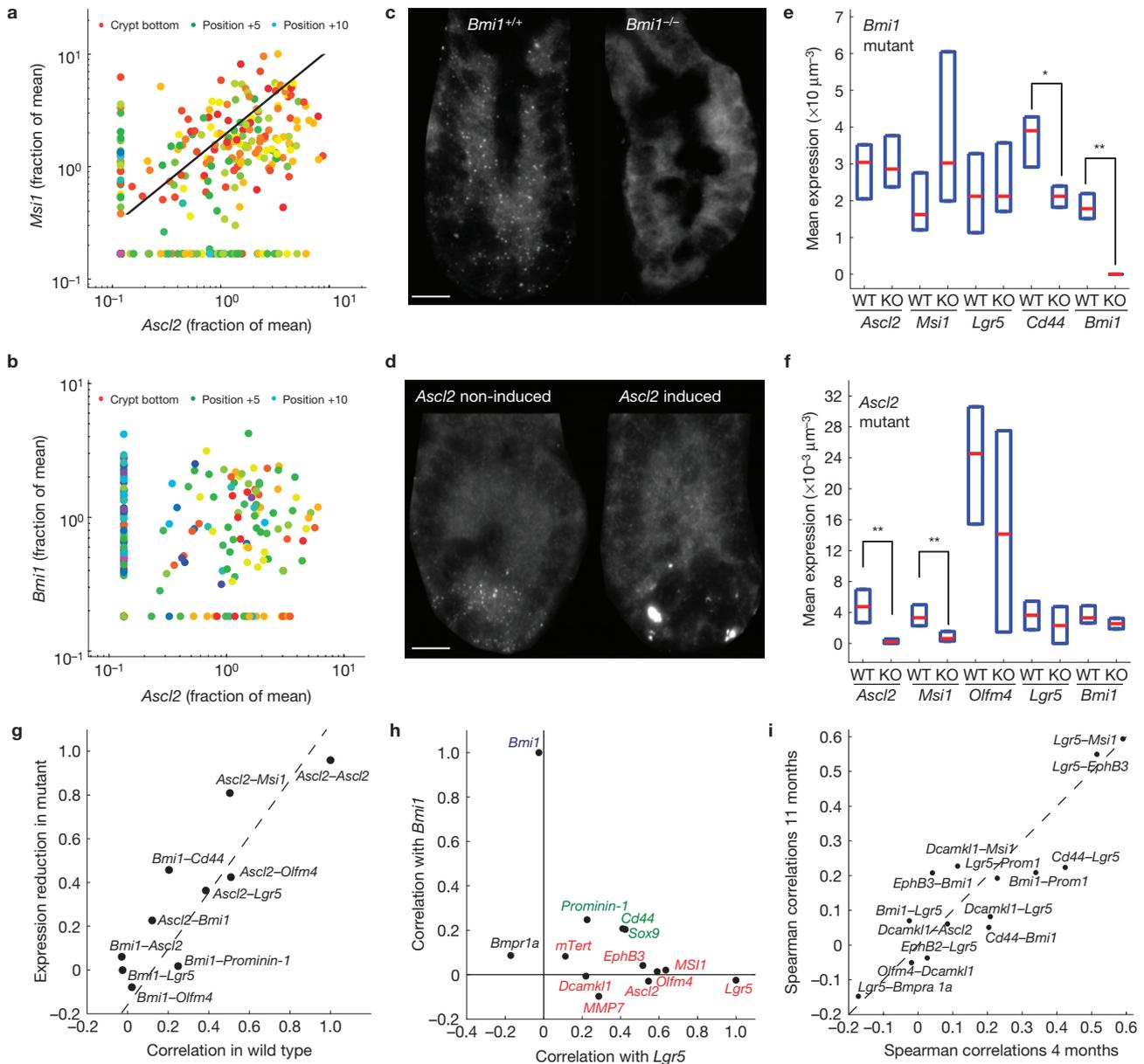
average concentration within the crypt from which the cell was sampled (cells with no transcripts were assigned the lowest non-zero concentration detected). Dot colours correspond to position along the crypt axis. 76% of *Lgr5*-positive cells contain *Bmi1* transcripts (1,073/1,417) whereas 48% of *Bmi1*-positive cells contain *Lgr5* transcripts (1,073/2,221). **(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  $0.3\mu\text{m}$  apart, filtered with a Laplacian of Gaussian filter (Methods). Scale bars,  $5\mu\text{m}$ .

support the possibility that *Dcamk1l* cells serve as reserve stem cells under this perturbation.

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

To obtain a comprehensive view of the expression changes following irradiation, and to detect genes that could be functionally important for

the tissue repair process, we measured the entire panel at different times after 12 Gy whole-body irradiation. We found striking differences in the spatial expression patterns of some stem-cell markers 48 h 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 and Supplementary Fig. S7). Most notable among these markers are *Cd44*

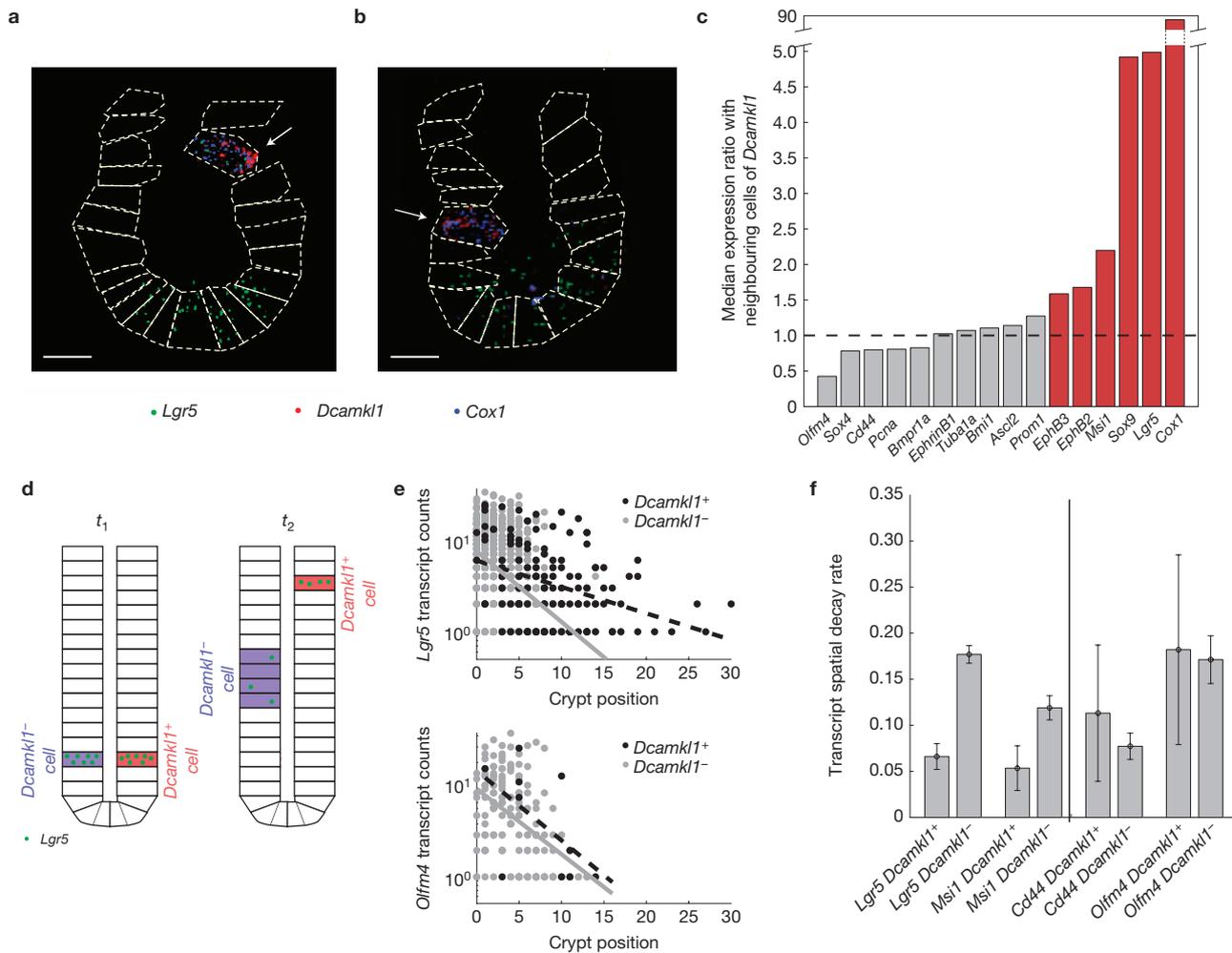


**Figure 4** Single-cell correlations of stem-cell markers are validated with mutants for key regulator genes. **(a)** *Musashi-1* (*Msi1*) and *Ascl2* are highly correlated at the single-cell level (Spearman correlation  $R = 0.7, P < 10^{-16}$ ). **(b)** *Bmi1* and *Ascl2* are not significantly correlated ( $R = -0.05, P = 0.74$ ). 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 colours correspond to position along the crypt axis. **(c)** *Bmi1* transcripts are detected in a wild-type mouse (left) but not in a *Bmi1* homozygous knockout mouse (right). **(d)** *Ascl2* transcripts are significantly reduced in an *Ah-Cre/Ascl2<sup>loxed/floxed</sup>* mouse 5 days after  $\beta$  NF induction (right) when compared with non-induced controls (left). Scale bars, 5  $\mu$ m. **(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 wild-type (WT, *Bmi1<sup>+/+</sup>*) and the

mutant (KO, *Bmi1<sup>-/-</sup>*) 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**, distributions of mean transcript concentrations per crypt cell for non-induced (WT, *Ascl2<sup>+/+</sup>*) and induced (KO, *Ascl2<sup>-/-</sup>*) 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 wild type and the mutant, divided by the wild-type 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 denotes significant correlation with *Lgr5*, blue denotes 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 ( $R = 0.88$ ).

(Fig. 6a,d,e), *Musashi-1* (Fig. 6c–e) and *Ascl2* (Fig. 6c–e). Interestingly, both the levels and spatial range of *Olfm4* first decreased after 48 h, and then significantly increased at 7 days (Fig. 6b,d,e). Although the

average spatial ranges of *Lgr5* and *Bmi1* slightly expanded following irradiation (Fig. 6d and Supplementary Fig. S7a,d), their transcript levels did not change significantly (Fig. 6e).



**Figure 5** *Dcamk1* 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 *Dcamk1* transcript levels (red dots). **(b)** Example of a *Dcamk1* cell with no *Lgr5* expression. Images in **a,b** are maximal projections of 15 optical sections spaced 0.3 μm apart, filtered with a Laplacian of Gaussian filter (Methods). Blue dots are *Cox1* transcripts, dashed lines mark cell borders and arrows mark the tuft cells. Scale bars, 5 μm. **(c)** Expression signature for *Dcamk1* cells. Shown are the median ratios of transcripts for different genes between *Dcamk1*<sup>high</sup> cells (cells with more than five transcripts) and the average levels in their immediate two neighbouring cells (above and below). Red bars are ratios that are significant relative to permuted crypts (see Methods), using a false discovery rate of 10%. **(d)** Passive-migration model for the elevated transcript levels

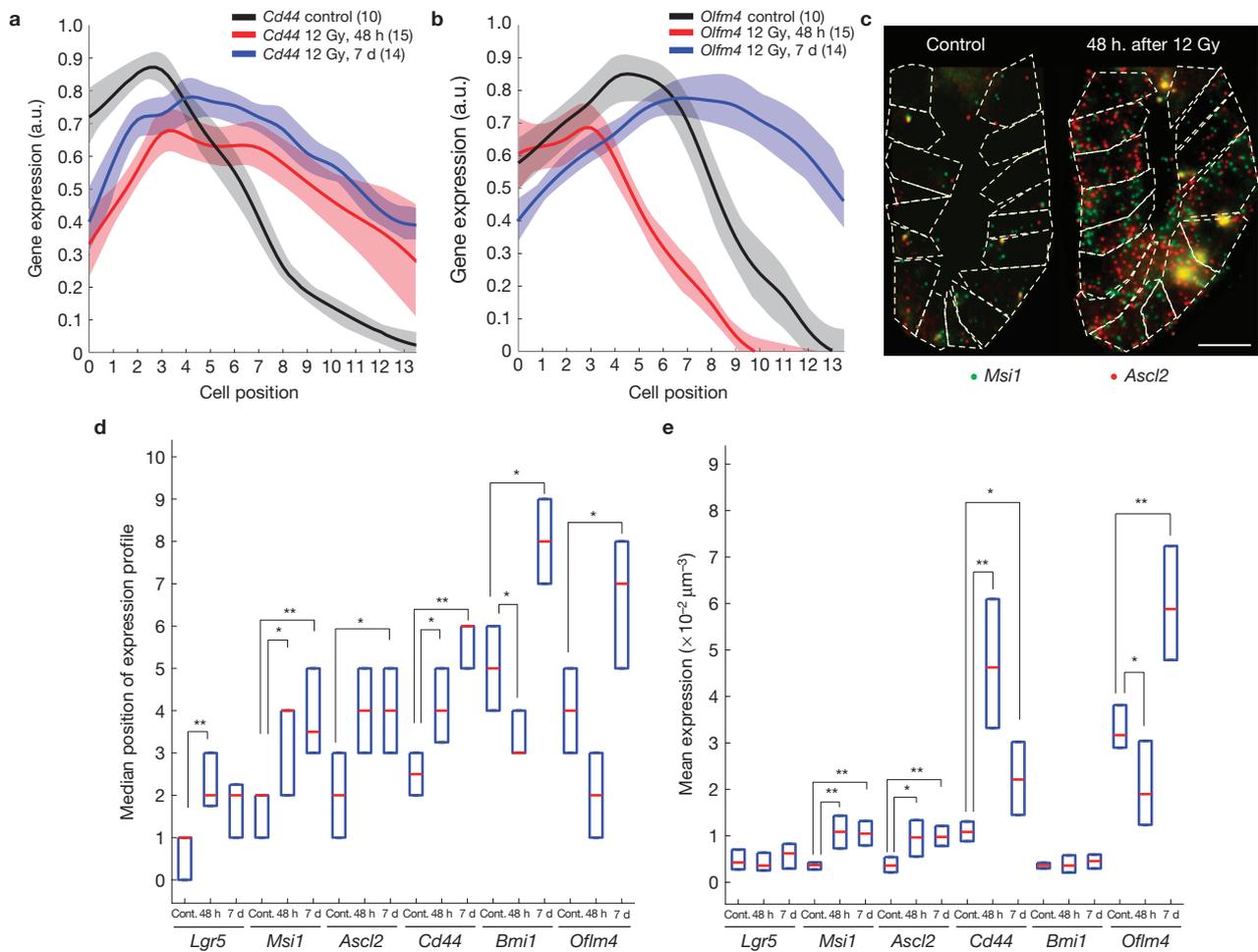
of *Lgr5* in *Dcamk1* cells. Green dots represent transcripts of *Lgr5* (or other stem-cell markers co-expressed in *Dcamk1* cells) in the progenies of two different cells migrating away from the stem-cell zone at the crypt bottoms—a *Dcamk1*-positive cell (red) and *Dcamk1*-negative cell (blue).  $t_1$  and  $t_2$  indicate successive times. If *Dcamk1*-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* will be lower. **(e)** *Lgr5* transcripts in *Dcamk1*<sup>high</sup> cells (black dots) decay more slowly with crypt position than *Lgr5* transcripts in *Dcamk1*-negative cells (grey dots). Transcripts of *Olfm4* in *Dcamk1*<sup>high</sup> cells (black) exhibit the same decay rate as in *Dcamk1*-negative cells (grey). Lines are exponential fits. **(f)** Transcript decay of *Lgr5* and *Musashi-1* is less than half as fast in *Dcamk1*<sup>high</sup> cells when compared with *Dcamk1*-negative cells, but comparable for *Cd44* and *Olfm4*.

**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, ageing 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—*Bmi1* (ref. 7), *Prominin-1* (ref. 14) and *mTert* (ref. 9). These were expressed throughout the crypt axis at almost constant

levels, and contrasted with *Lgr5* (ref. 12) and to a slightly lesser extent *Sox9* (ref. 13), the expression of which was concentrated at lower crypt positions. Importantly, all five genes were co-expressed in CBC cells<sup>11</sup>. 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. Although 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<sup>16,35</sup>, our measurements assess these co-expressions in a symmetric manner at the single-cell level in wild-type mice and



**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 killed after 48 h (red) and 7 days (blue). **(c)** Transcript levels of *Musashi-1* (green dots) and *Ascl2* (red dots) are significantly increased 48 h after whole-body 12 Gy irradiation. Images are maximal projections of six optical sections spaced 0.3  $\mu\text{m}$  apart. Scale bar, 10  $\mu\text{m}$ . **(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. \* $P < 0.01$ , \*\* $P < 0.001$ .

indicate the precise location of the cells co-expressing these stem-cell markers (Supplementary 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. Moreover, our results do not contradict the finding that some of the *Bmi1*-expressing cells can function as reserve stem cells that give rise to new CBC stem cells<sup>8</sup>.

We detected a unique expression signature for *Dcamk1l* cells, which includes significant co-expression with *Lgr5*. *Dcamk1l* has recently been shown to be a marker of tuft cells, a rare quiescent epithelial lineage of unknown function<sup>4,39</sup>. We found that, regardless of their *Lgr5* expression, *Dcamk1l* cells do not exhibit increased death rates following low dosage of gamma irradiation, as previously suggested for putative stem cells at higher crypt positions<sup>6,38</sup>. Following high dosage of gamma irradiation these cells did not enter cell cycle at any time and were depleted in proportion to goblet cells, a short-lived differentiated secretory cell type. Most importantly, all *Dcamk1l* cells,

both positive and negative for *Lgr5*, exhibited intense expression of the *Cox1* gene, a tuft-cell differentiation marker<sup>4</sup>. Although lineage tracing using a *Dcamk1l*-locus-driven Cre transgene would definitely resolve the possibility that some tuft cells could possess potential stem-cell function, our analysis indicates that such function is unlikely.

Our analysis indicates that during homeostasis the expression patterns of stem-cell markers are remarkably invariant between crypts within the same mouse and with ageing, with several markers such as *Lgr5*, *Olfm4*, *Cd44*, *Ascl2* and *Musashi-1* exhibiting spatially overlapping expression patterns and high single-cell correlations. The expression programme of these genes is however markedly different when the tissue is perturbed. This is evident from the marked 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 behaviour of the *Olfm4* expression pattern, which first retracts and then expands. These varying responses

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<sup>7,12</sup>, cell ablation<sup>8</sup> and *ex vivo* cultures<sup>15</sup>. 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 programme, controlled by niche-dependent signals, as well as intracellular transcriptional and signalling 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 ageing and tissue regeneration. Applying this technique to other tissues maintained by stem cells can provide important insights into the architecture of multicellular organisms, whereas similar studies in tumours can facilitate the detection of stem-cell-like signatures in cancer. □

## METHODS

Methods and any associated references are available in the online version of the paper at <http://www.nature.com/naturecellbiology>

*Note: Supplementary Information is available on the Nature Cell Biology website*

## ACKNOWLEDGEMENTS

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.

## AUTHOR CONTRIBUTIONS

S.I. and A.v.O. conceived the project. S.I., I.C.B. and A.L. carried out most of the experiments. S.I. analysed the data. M.M., J.L., J.v.E., T.J. and H.C. provided mice and assisted with experiments. S.I., H.C. and A.v.O. wrote the paper.

## COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at <http://www.nature.com/naturecellbiology>

Reprints and permissions information is available online at <http://www.nature.com/reprints>

- Potten, C. S., Gandara, R., Mahida, Y. R., Loeffler, M. & Wright, N. A. The stem cells of small intestinal crypts: where are they? *Cell Prolif.* **42**, 731–750 (2009).
- Li, L. & Clevers, H. Coexistence of quiescent and active adult stem cells in mammals. *Science* **327**, 542–545 (2010).
- van der Flier, L. G. & Clevers, H. Stem cells, self-renewal, and differentiation in the intestinal epithelium. *Annu. Rev. Physiol.* **71**, 241–260 (2009).
- 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.* **192**, 767–780 (2011).

- Sato, T. *et al.* Paneth cells constitute the niche for Lgr5 stem cells in intestinal crypts. *Nature* **469**, 415–418 (2010).
- Potten, C. S. Stem cells in gastrointestinal epithelium: numbers, characteristics and death. *Phil. Trans. R. Soc. Lond. B* **353**, 821–830 (1998).
- Sangiorgi, E. & Capecchi, M. R. Bmi1 is expressed *in vivo* in intestinal stem cells. *Nat. Genet.* **40**, 915–920 (2008).
- Tian, H. *et al.* A reserve stem cell population in small intestine renders Lgr5-positive cells dispensable. *Nature* **478**, 255–259 (2011).
- Montgomery, R. K. *et al.* Mouse telomerase reverse transcriptase (mTert) expression marks slowly cycling intestinal stem cells. *Proc. Natl Acad. Sci. USA* **108**, 179–184 (2011).
- 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* **26**, 630–637 (2008).
- Cheng, H. & Leblond, C. P. 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.* **141**, 537–561 (1974).
- Barker, N. *et al.* Identification of stem cells in small intestine and colon by marker gene Lgr5. *Nature* **449**, 1003–1007 (2007).
- Furuyama, K. *et al.* Continuous cell supply from a Sox9-expressing progenitor zone in adult liver, exocrine pancreas and intestine. *Nat. Genet.* **43**, 34–41 (2011).
- Zhu, L. *et al.* Prominin 1 marks intestinal stem cells that are susceptible to neoplastic transformation. *Nature* **457**, 603–607 (2009).
- Sato, T. *et al.* Single Lgr5 stem cells build crypt–villus structures *in vitro* without a mesenchymal niche. *Nature* **459**, 262–265 (2009).
- van der Flier, L. G. *et al.* Transcription factor achaete scute-like 2 controls intestinal stem cell fate. *Cell* **136**, 903–912 (2009).
- Giannakis, M. *et al.* Molecular properties of adult mouse gastric and intestinal epithelial progenitors in their niches. *J. Biol. Chem.* **281**, 11292–11300 (2006).
- Raj, A. & van Oudenaarden, A. Single-molecule approaches to stochastic gene expression. *Annu. Rev. Biophys.* **38**, 255–270 (2009).
- Gregorieff, A. & Clevers, H. Wnt signaling in the intestinal epithelium: from endoderm to cancer. *Genes Dev.* **19**, 877–890 (2005).
- Femino, A. M., Fay, F. S., Fogarty, K. & Singer, R. H. Visualization of single RNA transcripts *in situ*. *Science* **280**, 585–590 (1998).
- Levsky, J. M., Shenoy, S. M., Pezo, R. C. & Singer, R. H. Single-cell gene expression profiling. *Science* **297**, 836–840 (2002).
- Zenkhusen, D., Larson, D. R. & Singer, R. H. Single-RNA counting reveals alternative modes of gene expression in yeast. *Nat. Struct. Mol. Biol.* **15**, 1263–1271 (2008).
- Raj, A., Peskin, C. S., Tranchina, D., Vargas, D. Y. & Tyagi, S. Stochastic mRNA synthesis in mammalian cells. *PLoS Biol.* **4**, 1707–1719 (2006).
- Capodici, P. *et al.* Gene expression profiling in single cells within tissue. *Nat. Methods* **2**, 663–665 (2005).
- Raj, A., van den Bogaard, P., Rifkin, S. A., van Oudenaarden, A. & Tyagi, S. Imaging individual mRNA molecules using multiple singly labeled probes. *Nat. Methods* **5**, 877–879 (2008).
- Raj, A., Rifkin, S. A., Andersen, E. & van Oudenaarden, A. Variability in gene expression underlies incomplete penetrance. *Nature* **463**, 913–918 (2010).
- Carmon, K. S., Gong, X., Lin, Q., Thomas, A. & Liu, Q. R-spondins function as ligands of the orphan receptors LGR4 and LGR5 to regulate Wnt/β-catenin signaling. *Proc. Natl Acad. Sci. USA* **108**, 11452–11457 (2011).
- de Lau, W. *et al.* Lgr5 homologues associate with Wnt receptors and mediate R-spondin signalling. *Nature* **476**, 293–297 (2011).
- Zeilstra, J. *et al.* Deletion of the WNT target and cancer stem cell marker CD44 in Apc(Min+) mice attenuates intestinal tumorigenesis. *Cancer Res.* **68**, 3655–3661 (2008).
- Sinner, D. *et al.* Sox17 and Sox4 differentially regulate β-catenin/T-cell factor activity and proliferation of colon carcinoma cells. *Mol. Cell Biol.* **27**, 7802–7815 (2007).
- Gregorieff, A. *et al.* Expression pattern of Wnt signaling components in the adult intestine. *Gastroenterology* **129**, 626–638 (2005).
- 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.* **535**, 131–135 (2003).
- Potten, C. S. *et al.* Identification of a putative intestinal stem cell and early lineage marker; musashi-1. *Differentiation* **71**, 28–41 (2003).
- He, X. C. *et al.* BMP signaling inhibits intestinal stem cell self-renewal through suppression of Wnt-β-catenin signaling. *Nat. Genet.* **36**, 1117–1121 (2004).
- Schepers, A. G., 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.* **30**, 1104–1109 (2011).
- 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.* **123**, 315–327 (2005).
- Batlle, E. *et al.* β-catenin and TCF mediate cell positioning in the intestinal epithelium by controlling the expression of EphB/ephrinB. *Cell* **111**, 251–263 (2002).
- Potten, C. S., Booth, C. & Pritchard, D. M. The intestinal epithelial stem cell: the mucosal governor. *Int. J. Exp. Pathol.* **78**, 219–243 (1997).
- Gerbe, F., Brulin, B., Makrini, L., Legraverend, C. & Jay, P. DCAMK1 expression identifies Tuft cells rather than stem cells in the adult mouse intestinal epithelium. *Gastroenterology* **137**, 2179–2180 (2009) (author reply 2180–2171).

## 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* knockout experiments were carried out on 11-week-old female *Bmi1*<sup>-/-</sup> and a *Bmi1*<sup>+/+</sup> littermate control<sup>40</sup>. These mice were APC<sup>flxed/+</sup>, but as the mice had no Cre recombinase both were essentially wild type for APC. The *Ascl2* mutant was an 18-week-old male Ah-Cre/*Ascl2*<sup>flxed/flxed</sup>, 5 days after induction of the Cre enzyme by intraperitoneal injections of 200 ml β-naphthoflavone, as described in ref. 16. A non-induced Ah-Cre/*Ascl2*<sup>flxed/flxed</sup> littermate control was used for comparison. Whole-body gamma-irradiation dosages of 1 Gy, 6 Gy and 12 Gy were applied to 4-month-old C57bl6 mice as described in ref. 41. Mice were killed after 6 h, 24 h, 48 h and 7 days. Two mice per irradiation dosage and killing time were used. *Lgr5-IRES-eGFP-CreERT2* mice have been previously described<sup>12</sup>. All mice were fed ad libitum and killed in the morning. For all mice the duodenum was removed, flushed and fixed in 4% formaldehyde, incubated overnight with 30% sucrose in 4% formaldehyde and then embedded in OCT. Six-micrometre cryosections were used for hybridizations.

**Hybridizations and imaging.** Probe libraries were designed and constructed as described in ref. 25. Most libraries consisted of 48 probes of length 20 bp, complementary to the coding sequence of each gene (Supplementary Table S1). *Lgr5*, *Cd44* and *Ki67* libraries consisted of 96 probes. Hybridizations were done overnight with three differentially labelled probes using Cy5, Alexa594 and TMR fluorophores. A further FITC-conjugated antibody for E-cadherin (BD Biosciences) was added to the hybridization mix and used for protein immunofluorescence. 4,6-diamidino-2-phenylindole 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 (charge-coupled device) camera using MetaMorph software (Molecular Devices, Downington, PA). The image-plane pixel dimension was 0.13 μm. Quantification was done on stacks of six to 12 optical sections with Z spacing of 0.3 μm, in which not more than a single cell was observed.

Dots were automatically detected using a custom Matlab program, implementing algorithms described in ref. 25. Briefly, the dot stack images were first filtered with a three-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 (Supplementary 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

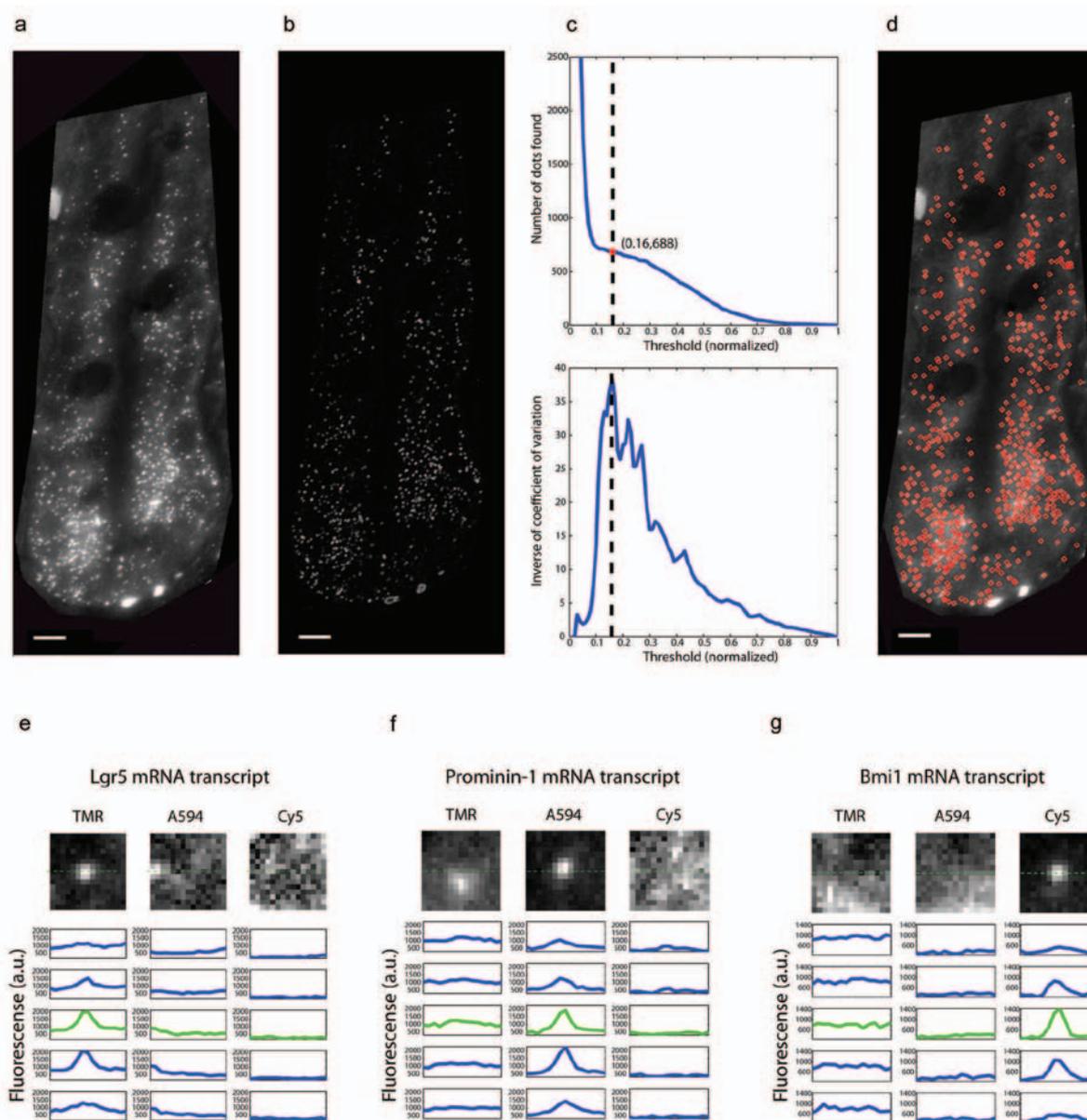
removed. Such dots occasionally appeared in the surrounding mesenchymal cells but were rare in the epithelial cells. Bleed-through of transcript signal between channels was minimal (Supplementary Fig. S1e–g). Cell segmentation was carried out manually on a maximal projection of the FITC channel. Transcript 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 multiplied by a voxel size of 0.13 μm × 0.13 μm × 0.3 μm. The crypt apex and outline were manually marked and used to determine cell position along the crypt axis. For four genes—*Olfm4*, *Dcamk1l*, *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 and Supplementary 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 three crypt positions and normalized to a maximum of unity. Co-expression analysis was carried out on pooled cells from all crypts. Transcript concentrations were first normalized by the average 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 with those obtained in randomized crypts in which the value of one of the genes was shuffled among cells, and *P* values reported were computed by transforming the *Z* score of the correlations compared with 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 *Dcamk1l* single-cell expression signature we computed for each *Dcamk1l*<sup>high</sup> cell the ratio between the transcript concentration in the cell and its immediate neighbours that were not *Dcamk1l*<sup>high</sup>. *Dcamk1l*<sup>high</sup> cells were defined as cells with more than five *Dcamk1l* 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 *Dcamk1l*<sup>high</sup> cells and one of the neighbouring 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 *Dcamk1l*<sup>high</sup> cells that were positive for *Lgr5* (*Lgr5*<sup>+</sup>), 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 versus crypt position.

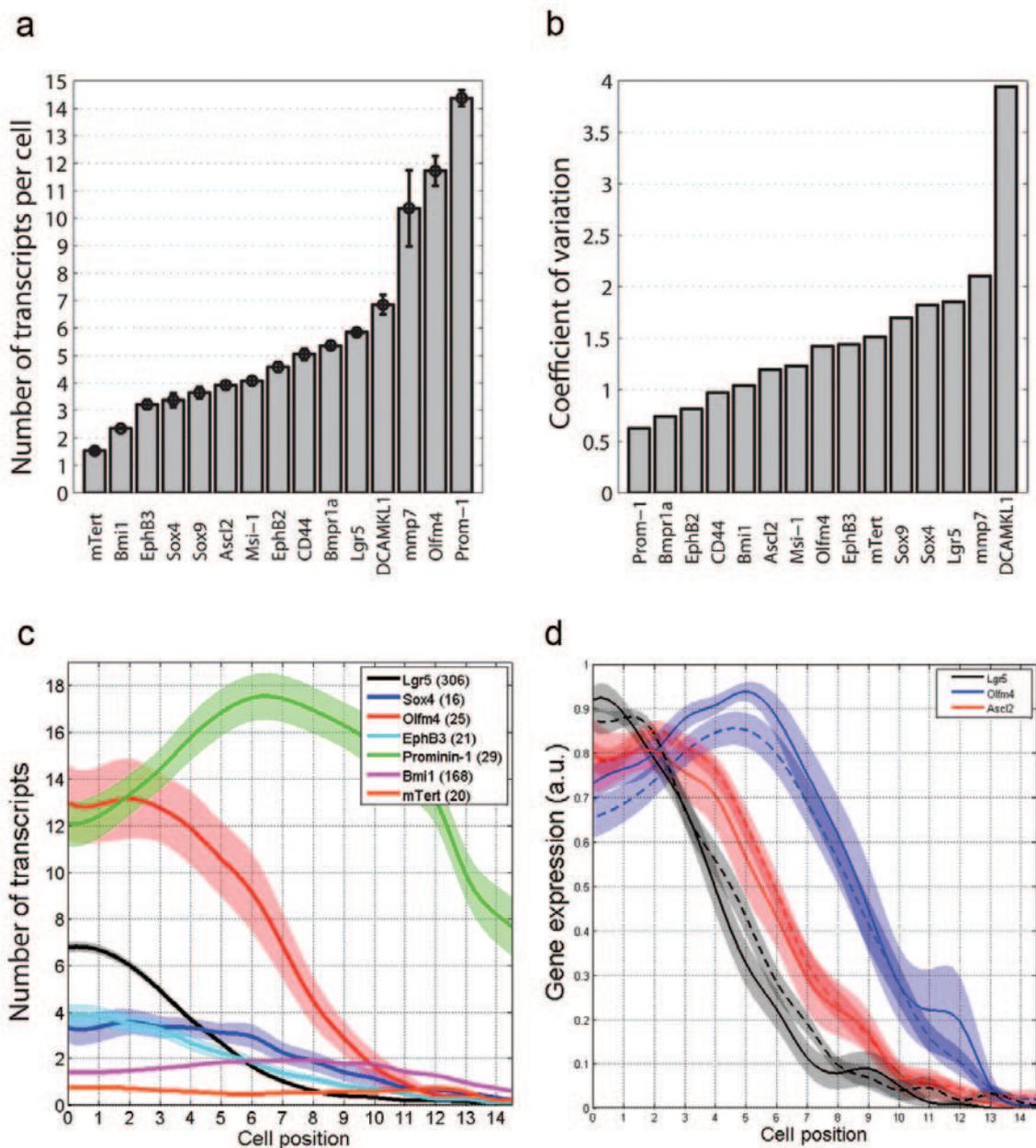
40. van der Lugt, N. M. *et al.* Posterior transformation, neurological abnormalities, and severe hematopoietic defects in mice with a targeted deletion of the *bmi-1* proto-oncogene. *Genes Dev.* **8**, 757–769 (1994).
41. Kirsch, D. G. *et al.* p53 controls radiation-induced gastrointestinal syndrome in mice independent of apoptosis. *Science* **327**, 593–596 (2010).

DOI: 10.1038/ncb2384



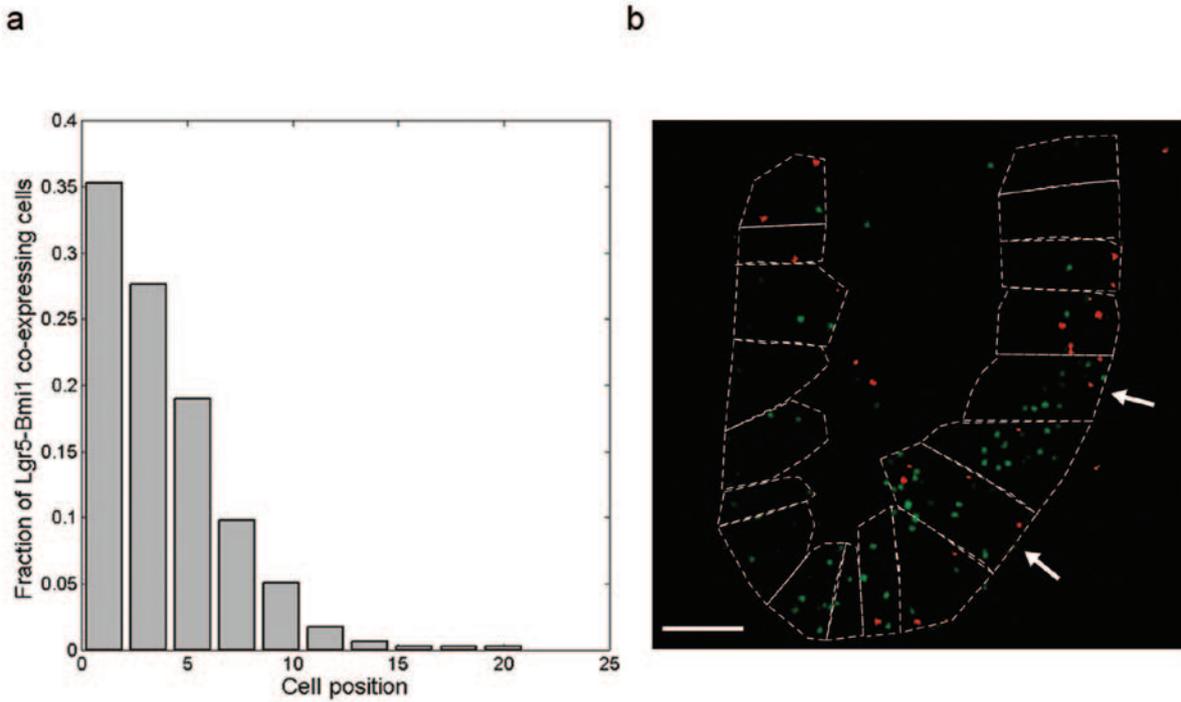
**Figure S1** Automatic detection of single dots in fluorescent images of intestinal crypts. **(a)** Original crypt image, dots are *Prominin-1* transcripts. **(b)** Image filtered with a 3-dimensional Laplacian of Gaussian filter of size 15 pixels and standard deviation of 1.5 pixels. **(c)** The number of dots (connected components in a thresholded image) is counted for each threshold  $T$  (top). The mean  $m$ , and standard deviation,  $s$ , of the number of dots in a sliding window of 5 thresholds centered at each threshold value  $T$  is determined and an inverse coefficient of variance is computed as  $f(T)=m/(s+\epsilon)$  (bottom, we use  $\epsilon=10$  to prevent divergence of  $f(T)$  at high threshold values, where both the mean and standard deviations are low). The threshold for which  $f(T)$  is maximal is automatically selected. This threshold, marked with a dashed vertical line, corresponding to 688 dots indicates a region in the threshold function over which the number of dots found is fairly insensitive to the threshold selected. **(d)** The original image and the overlaid detected dots (red diamonds), using the threshold of **(c)**.

Dots with an area of more than 400 pixels were excluded. All images are maximal projections of 20 stacks with a spacing of 0.3 microns. All scale bars are 5 microns. **(e-g)** Single molecule FISH on intestinal tissue exhibits minimal bleed-through between channels. Shown are images of duodenum tissue simultaneously co-hybridized with *Lgr5* probes labeled with TMR, *Prominin-1* probes labeled with A594 and *Bmi1* probes labeled with Cy5. **(e)** *Lgr5* mRNA dot labeled with TMR as seen through the TMR, Alexa 594 and Cy5 filter channels. Linescans of fluorescent intensity corresponding to the dashed green line through the image are given below, with the different linescans corresponding to measurements taken at increasing Z stacks (0.3 micron spacing). The green linescan corresponds to the z-slice shown in the image itself. Similar analysis is shown for a *Prominin-1* mRNA dot labeled with Alexa 594 **(f)** and a *Bmi1* mRNA dot labeled with Cy5 **(g)**. All images had their mean intensity level subtracted. Images are maximal projections of 12 optical sections.



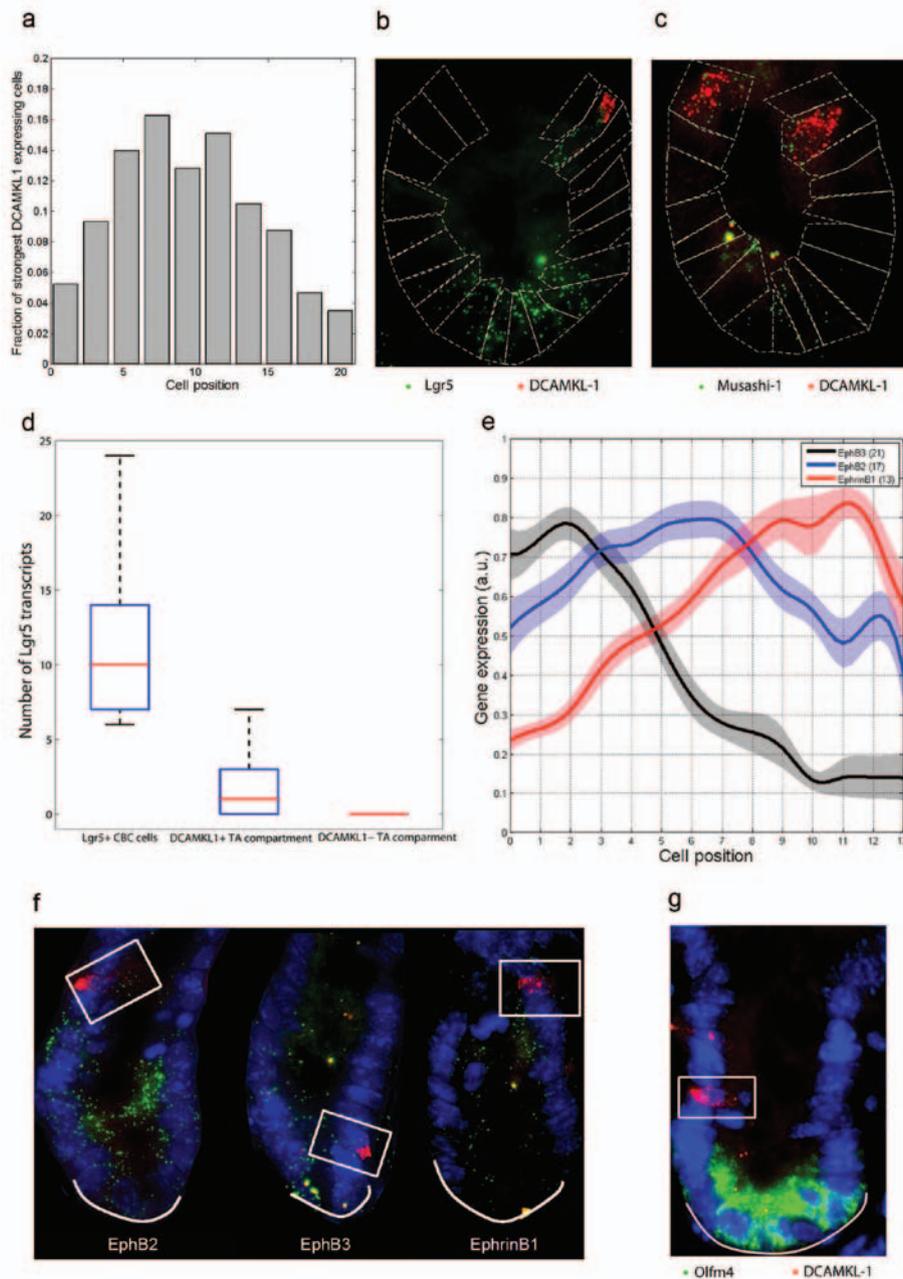
**Figure S2** Single cell expression of intestinal stem cell markers. **(a)** Absolute number of transcripts per cell for the different stem cell markers. Shown are means and standard errors of the number of dots in stacks of 12 optical sections. Levels of *Dcamkl-1* and *Olfm4* are underestimated as the abundance of these markers in some cells prohibited the detection of isolated dots. **(b)** Coefficient of variation (standard deviation divided by the mean) of number of transcripts per cell for the different stem cell markers. **(c)** Number of transcripts of different stem cell markers vs. position along

the crypt axis. Position 0 represents crypt bottoms. Profiles were generated by mapping the absolute transcript numbers in individual cells to the corresponding cell position and averaging over different crypts. Patches denote one standard error of the mean. Numbers of crypts used to generate the profiles are given in parentheses. Profiles were smoothed over 3 crypt positions. **(d)** Expression profiles are highly reproducible between adult mice. Shown are normalized expression profiles of selected markers for two 4-month old mice (dashed and solid lines).



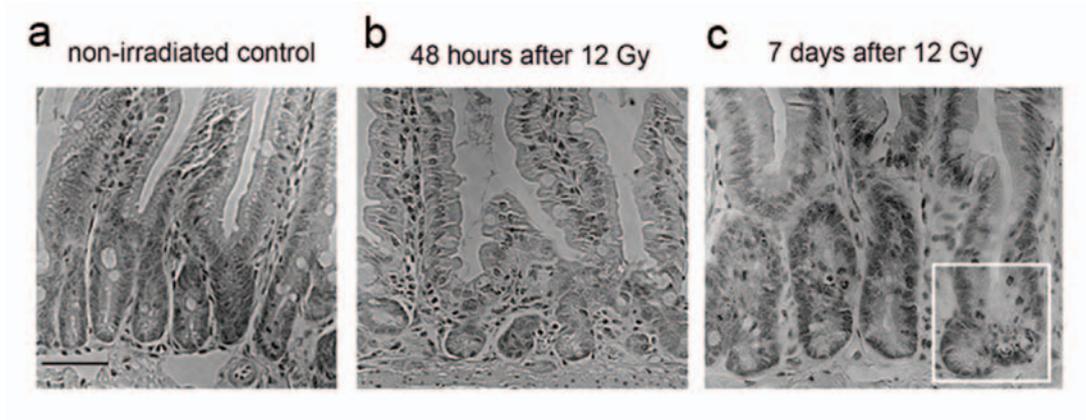
**Figure S3** *Bmi1* is broadly co-expressed with *Lgr5*. **(a)** Distribution of positions along the crypt axis in which cells co-expressing *Bmi1* and *Lgr5* are found. Position 0 denotes crypt bottom. 73% of co-expressing cells are at or below cell position +5. Analysis based on 117 crypts. **(b)** *Bmi1* and *Lgr5* are co-expressed in colon crypts. Green dots are single *Lgr5* transcripts and red

dots are single *Bmi1* transcripts. Arrows mark cells that are positive for both *Bmi1* and *Lgr5*. Images are maximal projections over stacks of 15 optical sections 0.3 microns apart, and were filtered with a 3-dimensional Laplacian of Gaussian filter of size 15 pixels and standard deviation of 1.5 pixels. Dashed lines mark cell borders. Scale bar is 5 microns.



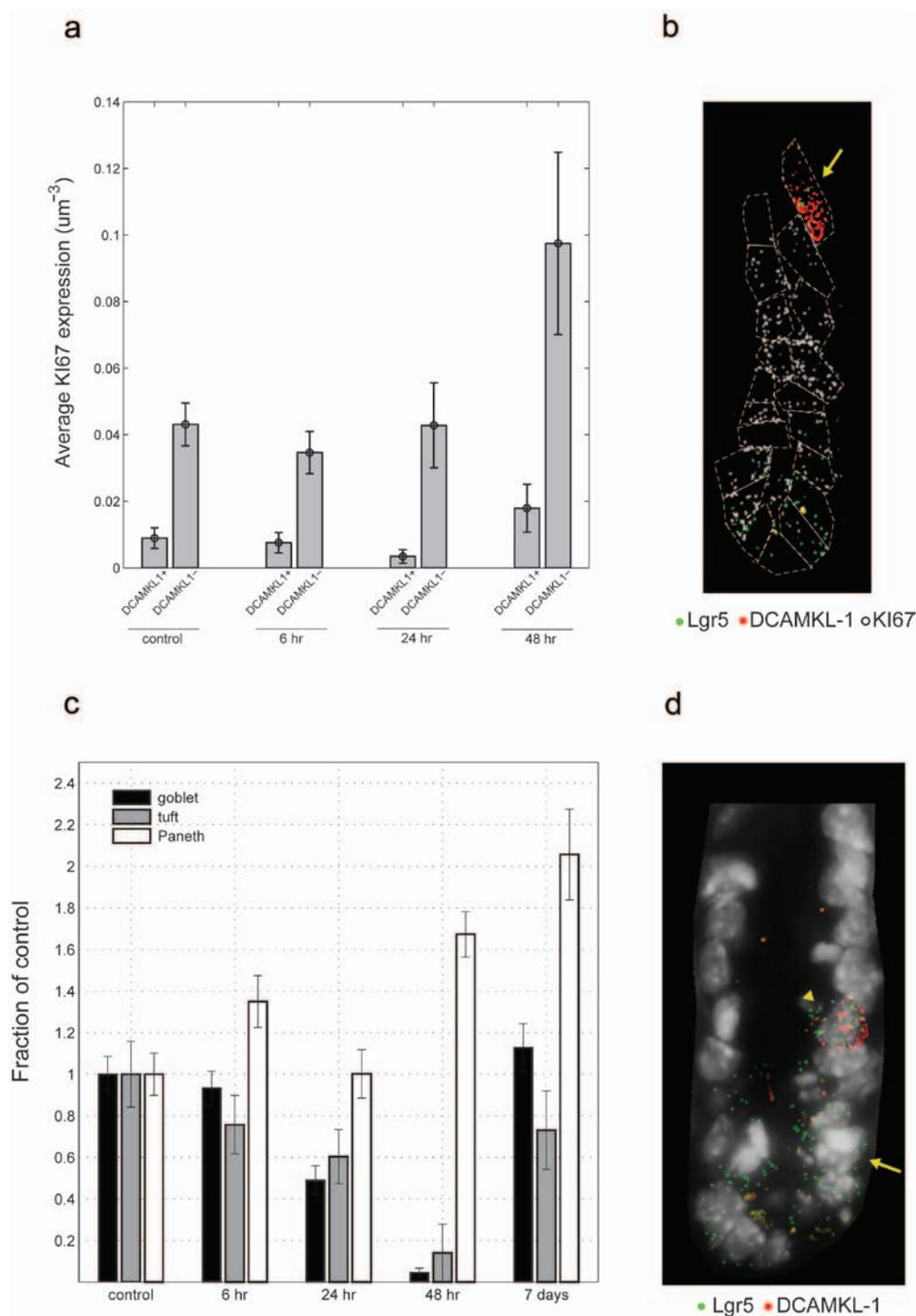
**Figure S4** Single-cell Gene expression signature of *Dcamkl-1* cells. **(a)** Distribution of cell position along the crypt axis where peak expression of *Dcamkl-1* is detected. Analysis based on 72 crypts. **(b-c)** *Dcamkl-1* cells at upper crypt positions significantly co-express *Lgr5* **(b)** and *Musashi-1* **(c)**. Red dots are *Dcamkl-1* transcripts; green dots are *Lgr5* **(b)** or *Musashi-1* **(c)** transcripts. Images are maximal projections over stacks of 15 optical sections 0.3 microns apart, and were filtered with a 3-dimensional Laplacian of Gaussian filter of size 15 pixels and standard deviation of 1.5 pixels. Dashed lines mark cell borders. **(d)** *Lgr5* is expressed in the transit amplifying compartment only in *DCAMKL1+* cells. Shown are the distributions of *Lgr5* transcript counts in three different cell types. *Lgr5+* crypt base columnar cells (*Lgr5+* CBC) were defined as cells at or below cell position 5 that had *Lgr5* expression that was larger than the median expression in *Lgr5*-positive cells at these positions. *DCAMKL1+* cells at the transit amplifying compartment (*DCAMKL1+* TA compartment) are cells above cell position 5 with more than 5 transcripts of *DCAMKL1*. *DCAMKL1-* cells at the transit amplifying compartment (*DCAMKL1-* TA compartment) are cells above cell position 5

with less than or equal to 5 transcripts of *DCAMKL1*. The distributions of *Lgr5* transcript counts in *DCAMKL1+* cells at the transit amplifying compartment (mean 2.5±0.3 transcripts) is significantly higher than in *DCAMKL1-* cells (mean 0.3±0.03, Kolmogorov-Smirnov  $p < 10^{-27}$ ) but significantly lower than the *Lgr5* transcript count distribution in CBC cells (mean 11.4±0.2, Kolmogorov-Smirnov  $p < 10^{-83}$ ). The red line denotes, the median, the blue box delimits 25–75 percentiles, and the outermost bars denote 99% of the data. **(e)** Expression profiles of *EphB3* (black), *EphB2* (blue) and *EphrinB1* (red) vs. crypt position. Position 0 represents crypt bottoms. **(f)** Cells with high levels of *Dcamkl-1* transcripts (red dots) exhibit equal or higher levels of the receptors *EphB2*, *EphB3* and the ligand *EphrinB1* (green dots) compared to their neighbors, indicative of slow, rather than fast migration. White curves denote crypt bottoms. White boxes highlight *Dcamkl-1* cells. **(g)** Cells that are high in *Dcamkl-1* transcripts (red dots) exhibit lower expression of *Olfm4* transcripts (green dots) compared to neighboring cells. Image is a maximal projection of stacks of 15 optical sections spaced 0.3 microns apart. White curve denotes crypt bottom. White box highlights a *Dcamkl-1* cell.



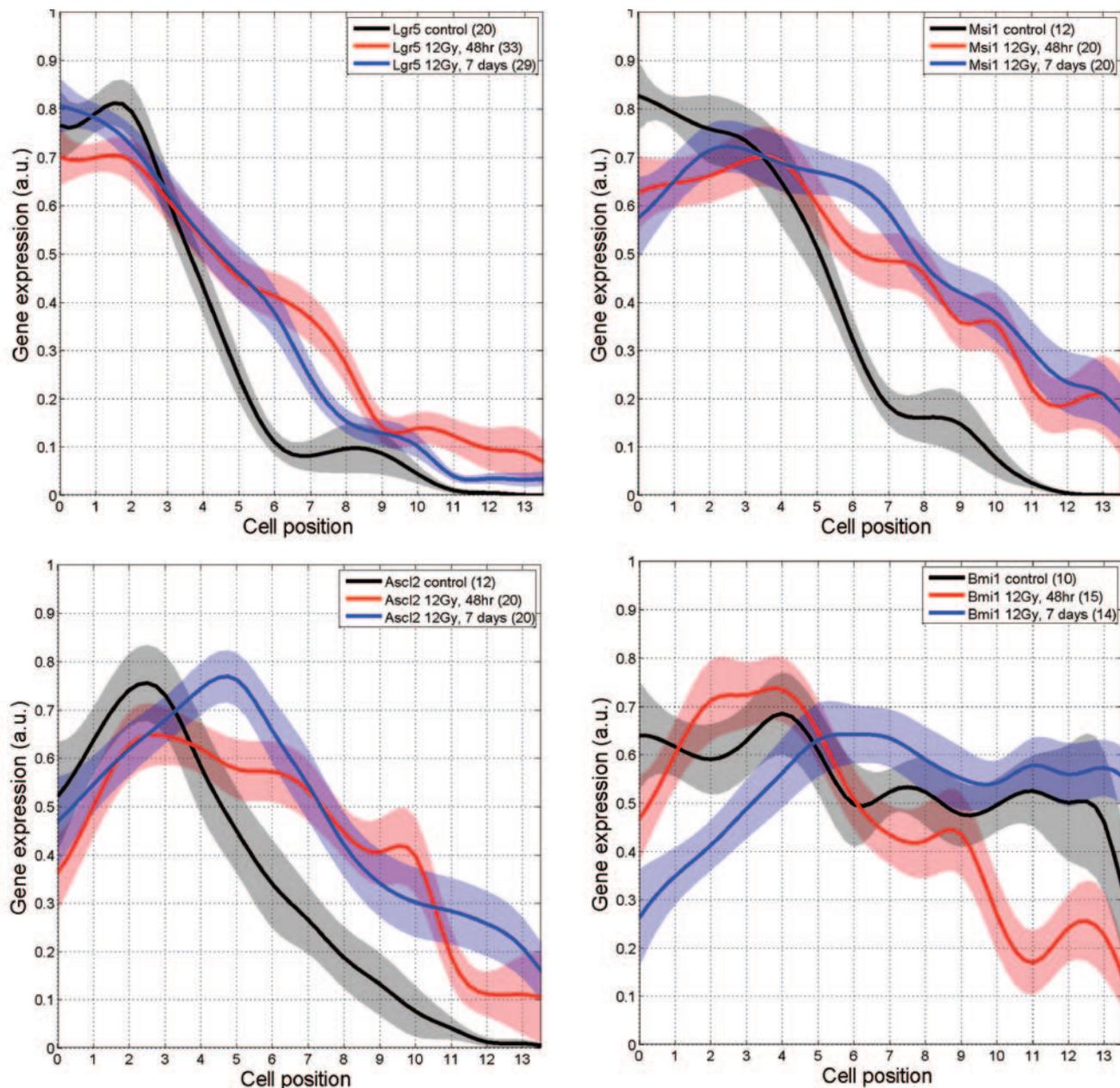
**Figure S5** Whole body irradiation effects on intestinal crypts. (a-c) Images of duodenum of a non-irradiated control mouse (a) and mice sacrificed 48 hour (b) and 7 days (c) after 12 Gy whole body gamma irradiation, stained

with hematoxylin and eosin. Crypts are considerably fewer and smaller after 48 hours. After 7 days crypt size significantly increases and extensive crypt fission is observed (white box). Scale bar is 50 microns.



**Figure S6** *Dcamkl-1* tuft cells do not enter cell cycle or exhibit increased sensitivity following irradiation. **(a)** Transcript levels of *Ki67* in *Dcamkl-1* positive tuft cells are significantly lower than in adjacent *Dcamkl-1* negative cells. Shown are *Ki67* transcript concentrations at different time points following 12Gy gamma irradiation. **(b)** An example of a crypt 6 hours after 12Gy gamma irradiation. Tuft cells, positive for *Dcamkl-1* (red dots) and *Lgr5* (green dots) remain quiescent, as evident from their low expression of *Ki67* (blue dots), compared to neighboring cells. Dashed lines mark cell borders. **(c)** *Dcamkl-1* cells decrease in numbers proportionally to goblet cells following 12 Gy gamma irradiation. Shown are the

fractions, compared to non-irradiated controls, of goblet cells (detected based on *Gob5* expression, black bars), tuft cells (detected based on *Dcamkl-1* expression, gray bars) and Paneth cells (detected based on *Lysozyme* expression, white bars). **(d)** *Dcamkl-1* tuft cells do not exhibit increased cell death 24 hours after 1Gy gamma irradiation. *Dcamkl-1* transcripts are in red, *Lgr5* transcripts in green, dapi nuclear staining is in white. Arrow denotes a dead cell. Arrowhead denotes a tuft cell positive for *Dcamkl-1* (red dots) and *Lgr5* (green dots). Images in **(b)** and **(d)** were filtered with a 3-dimensional Laplacian of Gaussian filter of size 15 pixels and standard deviation of 1.5 pixels.



**Figure S7** Whole body irradiation modifies the spatial expression profiles. **(a-d)** Average spatial expression profiles for *Lgr5* **(a)**, *Musashi-1* (*Msi1*) **(b)**, *Ascl2* **(c)** and *Bmi1* **(d)** in a non-irradiated control mouse (black) and in mice irradiated with 12 Gy and sacrificed after 48 hours

(red) and 7 days (blue). Spatial profiles were symmetrized, smoothed over 3 positions and normalized to 1. Patches are standard errors of the mean. Numbers of crypts used to generate the profiles are given in parentheses.