

A Versatile Strategy for Isolating a Highly Enriched Population of Intestinal Stem Cells

Christian M. Nefzger,^{1,2,3,7} Thierry Jardé,^{1,2,4,5,7} Fernando J. Rossello,^{1,2,3} Katja Horvay,^{1,2,4} Anja S. Knaupp,^{1,2,3} David R. Powell,⁶ Joseph Chen,^{1,2,3} Helen E. Abud,^{1,2,4,*} and Jose M. Polo^{1,2,3,*}

¹Department of Anatomy and Developmental Biology, Monash University, Wellington Road, Clayton, VIC 3800, Australia

²Development and Stem Cells Program, Monash Biomedicine Discovery Institute, Wellington Road, Clayton, VIC 3800, Australia

³Australian Regenerative Medicine Institute, Monash University, Wellington Road, Clayton, VIC 3800, Australia

⁴Cancer Program, Monash Biomedicine Discovery Institute, Wellington Road, Clayton, VIC 3800, Australia

⁵Centre for Cancer Research, Hudson Institute of Medical Research, 27-31 Wright Street, Clayton, VIC 3168, Australia

⁶Monash Bioinformatics Platform, Monash University, Wellington Road, Clayton, VIC 3800, Australia

⁷Co-first author

*Correspondence: helen.abud@monash.edu (H.E.A.), jose.polo@monash.edu (J.M.P.)

<http://dx.doi.org/10.1016/j.stemcr.2016.01.014>

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

SUMMARY

The isolation of pure populations of mouse intestinal stem cells (ISCs) is essential to facilitate functional studies of tissue homeostasis, tissue regeneration, and intestinal diseases. However, the purification of ISCs has relied predominantly on the use of transgenic reporter alleles in mice. Here, we introduce a combinational cell surface marker-mediated strategy that allows the isolation of an ISC population transcriptionally and functionally equivalent to the gold standard *Lgr5*-GFP ISCs. Used on reporter-free mice, this strategy allows the isolation of functional, transcriptionally distinct ISCs uncompromised by *Lgr5* haploinsufficiency.

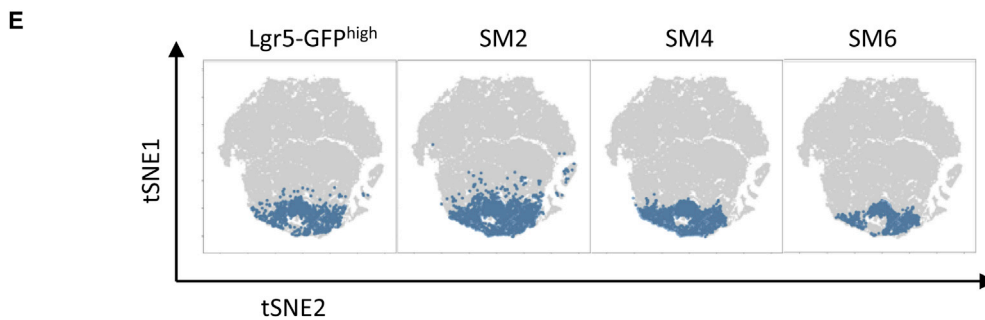
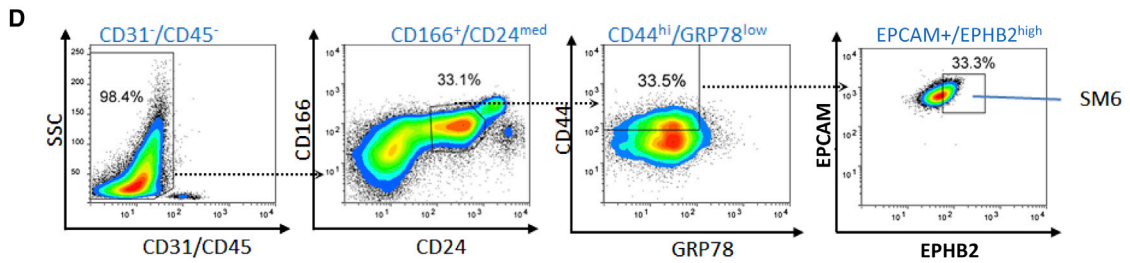
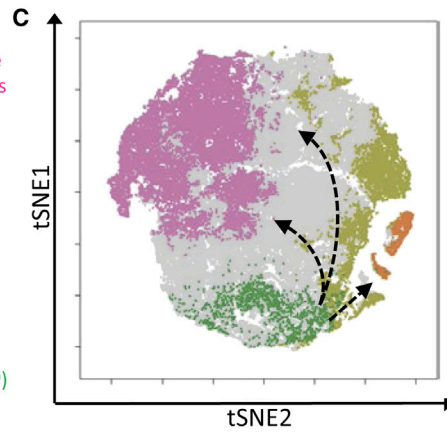
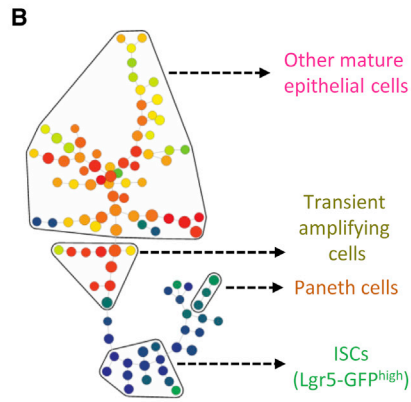
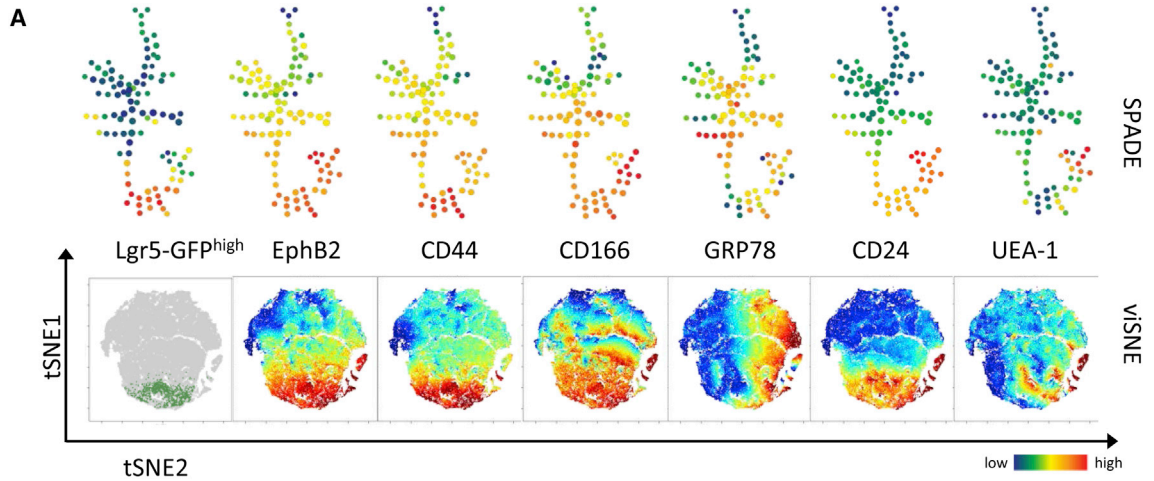
INTRODUCTION

The intestinal epithelium is a dynamic tissue that relies on integration of cell division, differentiation, migration, and apoptosis. Intestinal tissue homeostasis and regeneration are facilitated by multipotent tissue stem cells that have the ability to differentiate into multiple mature cell types. Two types of stem cells are currently proposed to reside in small intestinal crypts: cycling crypt base columnar (CBC) cells and +4 reserve cells (Barker, 2014; Clevers, 2013). CBC stem cells maintain daily homeostasis, while their reserve equivalents have been postulated to play a role in tissue regeneration upon injury (Barker, 2014; Clevers, 2013). The functional study of ISCs has been made possible by the recent characterization of ISC markers such as *Lgr5*, *Olfm4*, or *Sox9^{low}* for CBC cells, and *Bmi1*, *Hopx*, *Lrig1*, or *Sox9^{high}* for their presumed quiescent counterparts (Barker et al., 2007; Gracz and Magness, 2014; Gracz et al., 2010; Powell et al., 2012; Sangiorgi and Capecchi, 2008; Takeda et al., 2011).

Currently, the isolation of pure ISCs is primarily restricted to the use of targeted murine reporter alleles of ISC markers. However, the fidelity and specificity of these genes to mark ISCs is still controversial (Munoz et al., 2012; Tan and Barker, 2014). The most widely used reporter for CBC cell isolation is the *Lgr5*-Gfp knockin mouse model (Barker et al., 2007), which has facilitated the isolation and characterization of CBC stem cells in many studies (van der Flier et al., 2009).

However, this transgenic mouse model has several limitations: (1) the reporter cassette is prone to being silenced in over two-thirds of all crypts resulting in mosaic expression of the *Gfp* allele (Barker et al., 2007; Munoz et al., 2012); (2) LGR5 constitutes the receptor for R-SPONDINS (Carmon et al., 2011; de Lau et al., 2011; Glinka et al., 2011), potent WNT signal enhancers and stem cell growth factors, and the potential haploinsufficiency induced by the loss of one *Lgr5* allele (replaced by the *Gfp* reporter cassette) cannot be excluded; and (3) the extensive breeding required to cross genetically modified mouse models with the *Lgr5*-Gfp reporter strain.

Several strategies have been recently developed for CBC cell isolation via cell surface markers and fluorescence-activated cell sorting (FACS; Gracz et al., 2013; King et al., 2012; Merlos-Suarez et al., 2011; Wang et al., 2013). Although they represent considerable advances in the isolation of CBC cells independently of transgenic reporter alleles, these methodologies are suggested to be contaminated with other cell types and have not been fully characterized at the molecular level. The approach by Merlos-Suarez et al. (2011) mainly relies on extracting a subset of EPHB2 high cells from EPCAM+ epithelial cells (named SM2 in our study). However, the EPHB2 receptor is not only expressed at high levels in CBC cells but also in committed progenitor cells (Merlos-Suarez et al., 2011). In another study, Wang et al. (2013) used three crypt base markers (CD24/CD166/CD44) while depleting for GRP78+ progenitor cells (named SM4 in our study). Nonetheless, the resultant



(legend on next page)



population was found to be contaminated by endocrine cells (Wang et al., 2013).

RESULTS AND DISCUSSION

To investigate in a comprehensive way how these different cell surface markers are expressed in the different cell populations of the intestinal crypt, we employed two recently developed tools that allow mapping of high-dimensional cytometry data onto two dimensions, yet conserving its high-dimensional structure (Amir el et al., 2013; Qiu et al., 2011). Spanning-tree progression analysis of density-normalized events (SPADE) clusters phenotypically similar cells into nodes (Qiu et al., 2011), while viSNE displays individual cells on a map that preserves their multidimensional separation (Amir el et al., 2013). SPADE and viSNE have been used to interrogate, infer and visualize cellular hierarchies and transitions based on expression of cell surface markers in diverse systems including nuclear reprogramming (Lujan et al., 2015) and hematopoiesis (Qiu et al., 2011).

For the generation of high-dimensional flow cytometry data, intestinal epithelial cells from *Lgr5-Gfp* reporter mice were labeled with a broad range of intestinal crypt markers including markers of CBC cells (EPHB2, CD24^{med}, CD44, CD166), transit-amplifying cells (GRP78), Paneth cells (CD24^{high}, UEA-1), epithelial cells (EPCAM), and non-epithelial contaminating cells (CD45, CD31) (Figure S1A) (Merlos-Suarez et al., 2011; Wang et al., 2013; Wong et al., 2012). Analysis revealed that CBC cells, as identified by high levels of Lgr5-GFP expression (Lgr5-GFP^{high}) (Figures S1B and S1C), clustered together in SPADE trees and on viSNE maps (Figure 1A), and that the expression of EPHB2, CD44, CD166, and CD24 overlapped with this population to various degrees. Interestingly, when nodes/cells of the SPADE trees/viSNE maps were categorized into stem cells (Lgr5-GFP^{high}), Paneth cells (CD24^{high}, UEA1^{high}, SSC^{high}) (Sato et al., 2011), transient amplifying cells (GRP78^{high}), and other mature epithelial cell types (EPCAM-positive or EPCAM-negative, or low for CBC cell markers, negative for Paneth cell markers), the known intestinal cell hierarchy could be inferred (Figures 1B and 1C). The pool of Lgr5-GFP^{high} stem cells was closely associated with the niche cells (Paneth cells) and, via a stream of transient amplifying cells, was connected to the

other mature epithelial cell types. Therefore, this suggested that the combination of surface markers with multidimensional analysis could be used to identify sorting strategies for the purification of CBC cells.

As both SM2 (based on EPHB2 and EPCAM markers) and SM4 (a combination of CD24, CD44, CD166, and GRP-78 markers) strategies (Figures S1B, S1D, and S1E) failed to isolate a pure CBC cell population (Merlos-Suarez et al., 2011; Wang et al., 2013) and importantly as their key cell surface markers (EPHB2, CD44, CD166) have different expression patterns (Figure 1A), we utilized viSNE to explore whether a reporter-free sorting strategy combining the different intestinal crypt surface markers, termed SM6 (Figure 1D), could improve the purity of the CBC cell population to a level comparable with the Lgr5-GFP^{high} cells. Briefly, cells were depleted for contaminating CD31 and CD45 cells (endothelial and hematopoietic cells) and enriched for a specific population of CD166^{low} CD24^{med} cells. These cells were subsequently gated into CD44^{high} GRP78^{neg-low} cells and then only the EPCAM^{high}/EPHB2^{high} cells were sorted (see Supplemental Experimental Procedures for more details).

By using viSNE maps, the degree of overlap between SM2, SM4, and SM6 populations and the reference Lgr5-GFP^{high} cells was investigated (Figure 1E). Interestingly, the SM2 gating strategy was not able to exclude a considerable number of cells that clustered outside of the region occupied by the Lgr5-GFP^{high} population. However, both SM4 and SM6 strategies produced homogeneous appearing populations that overlapped well with Lgr5-GFP^{high} cells. As previously mentioned, the expression of the *Lgr5-Gfp* cassette is mosaic and, accordingly, many CBC cells are not labeled by GFP. To investigate whether the SM6 gating strategy was superior at purifying a homogeneous population of CBC cells, a Lgr5-GFP back gating analysis was conducted on SM2, SM4, and SM6 populations. The enrichment of both Lgr5-GFP^{high} cells and Lgr5-GFP^{low} cells within SM2, SM4, and SM6 cell populations was assessed. It is generally accepted that only Lgr5-GFP^{high} cells represent CBC cells, while Lgr5-GFP^{low} cells are committed progenitors of Lgr5-GFP^{high} cells. In agreement, single-cell PCR for Lgr5 demonstrated that nearly all Lgr5-GFP^{high} cells express the transcript in contrast to only a small fraction of Lgr5-GFP^{low} cells (Figures S1G and S1H). Our analysis showed that the SM6 strategy was better than SM2 and SM4 cell isolation strategies in enriching for Lgr5-GFP^{high} cells,

Figure 1. Multidimensional Analyses of Flow Cytometry Data and Isolation Strategy

- (A) Representative SPADE trees and viSNE maps colored for expression of Lgr5-GFP^{high}, EphB2, CD44, CD166, GRP78, CD24, and UEA-1. For ease of comparison and as a reference, the Lgr5-GFP^{high} population (green) was superimposed on a viSNE map (gray).
(B and C) SPADE tree (B) and viSNE map (C), both with superimposed intestinal hierarchy, denoted in (C) by arrows.
(D) Gating strategy used on live cells to isolate the SM6 population via cell surface markers.
(E) viSNE map with locations of Lgr5-GFP^{high}, SM2, SM4, and SM6 populations overlaid in blue.

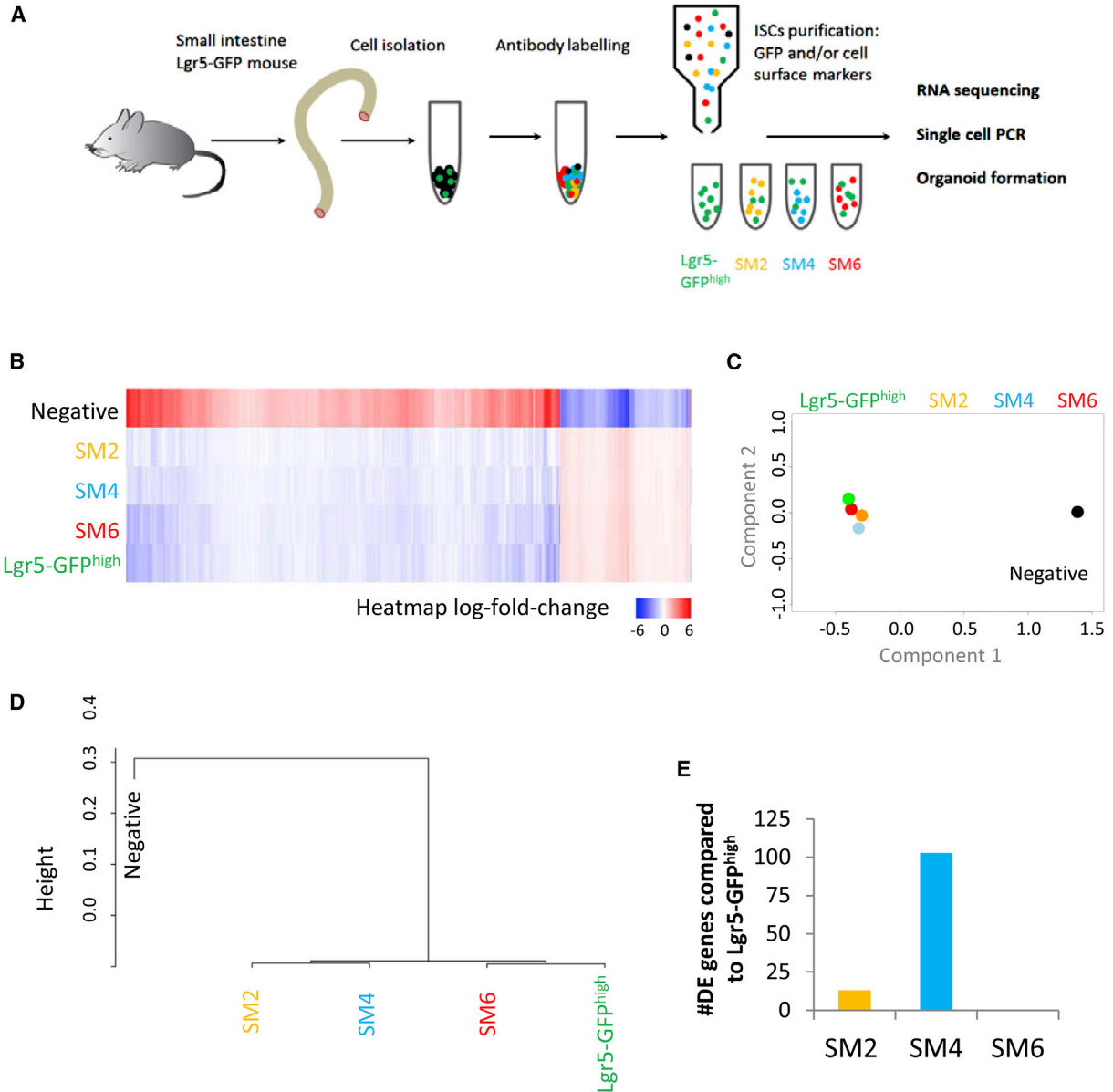


Figure 2. Bulk Profiling of Prospective CBC Cell Populations

(A) Schematic overview of the experimental procedure.

(B–D) Heatmap (B), principal component analysis (C), and unsupervised hierarchical clustering (D) for the RNA sequencing data derived from the five populations of interest: negative, SM2, SM4, SM6, and Lgr5-GFP^{high} (n = 2, experimental replicates). The displayed data are the average of two datasets for each group.

(E) Number of differentially expressed (DE) genes between Lgr5-GFP^{high} and SM2, SM4, or SM6 (n = 2, experimental replicates).

while depleting for Lgr5-GFP^{low} cells. However, these differences were only significant between SM6 and SM4 (Figures S1I and S1J). In order to adequately benchmark the quality of our method with the existing methods, we first performed RNA sequencing with the *Lgr5-Gfp* line on five FACS-purified groups: SM2, SM4, SM6, Lgr5-GFP^{high} reference population, and cells negative or low for all of the

cell surface markers used (negative) (Figure 2A, Figures S1B–S1F, Table S1). All the cell populations, with the exception of negative cells, had a similar transcriptional signature (Figure 2B). We used principal component analysis (PCA) to compare the sequencing data of the different isolation strategies. Importantly, the transcriptional signatures of SM6 and Lgr5-GFP^{high} cells overlapped, indicating



that these two populations were highly similar. SM2 and SM4 cell populations clustered further away and were therefore more different (Figure 2C) although still relatively close to the Lgr5-GFP^{high} population. Unsupervised hierarchical clustering on a population level also confirmed that CBC cell-enriched populations (SM2, SM4, Lgr5-GFP^{high}, and SM6) were clustered and distinct from the negative population (Figure 2D). Lgr5-GFP^{high} and SM6 cells formed a separate subgroup within this CBC cell-enriched branch (Figure 2D) confirming high similarity. Moreover, we could not find any genes that were significantly differentially expressed between the SM6 and Lgr5-GFP^{high} populations (Figure 2E) (2-fold, Benjamin-Hockberg correction). However, several genes were upregulated in SM2 and SM4 populations, mostly related to secretory cell lineage identity as already reported by Wang et al. (2013) for the SM4 approach (Figure 2E and Table S2). Together, these results indicate that cells isolated using our FACS sorting strategy are highly similar to the Lgr5-GFP^{high} cells from a transcriptional viewpoint.

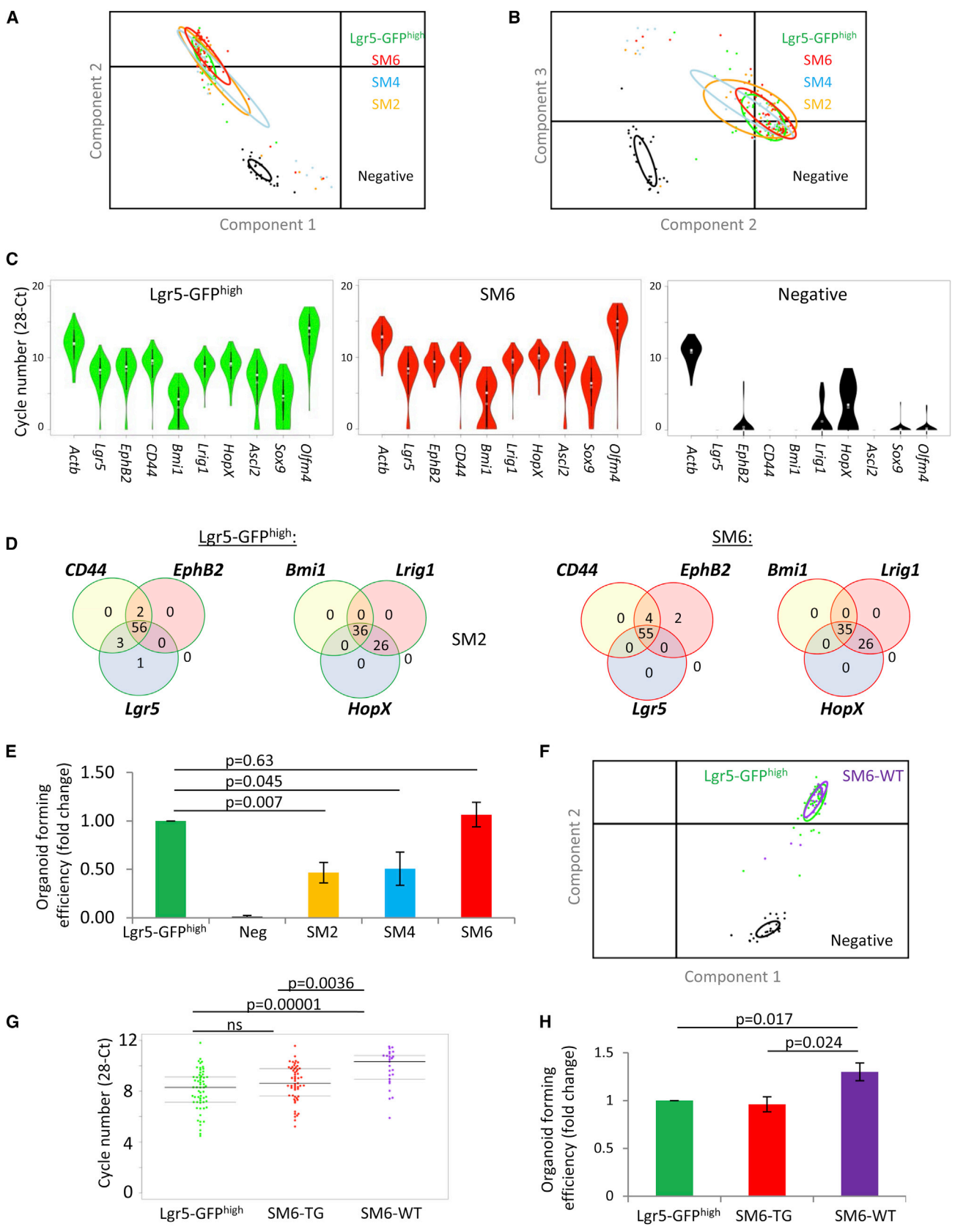
Expression of the *Lgr5-Gfp* reporter is mosaic in the intestine and only marks around a third of all CBC cells. The SM6 and Lgr5-GFP^{high} approaches allow the isolation of comparable cell numbers (SM6, 2.7% ± 0.4%; Lgr5-GFP^{high}, 2.6% ± 0.2% of all live cells) because the loss of a proportion of CBC cells via the SM6 method is a necessary trade-off between cell number and purity. In order to exclude the majority of the Lgr5-GFP^{low} cells, very stringent gating for EPHB2 is required (Figure 1D).

We performed single-cell transcriptional profiling for a broad panel of CBC and +4 reserve stem cell markers to determine the degree of homogeneity of the SM6 and Lgr5-GFP^{high} isolated cell populations (Figures S2A–S2F). PCA revealed that SM6 and Lgr5-GFP^{high} single-cell signatures overlapped and were highly homogeneous as indicated by the ellipses, representing 67% of the cells in each population (Figures 3A and 3B). The other strategies (SM2 and SM4) were more different. Violin plot analysis, which shows the distribution of gene expression per cell for any given population, demonstrated that all the different cell isolation methods were enriched for cells expressing ISC cell marker genes (*Lgr5*, *Olfm4*, *Bmi1*, *Lrig1*, *HopX*, *Sox9*, *CD44*, *EphB2*) (Figures 3C and S2G). Notably, this analysis also established that SM6 and Lgr5-GFP^{high} single cells had an analogous gene expression pattern at the individual cell level (Figures 3C and S2G). Co-expression of the key CBC markers *Lgr5*, *EphB2*, and *CD44* was detected in 90.1% and 90.3% of the individual cells from SM6 (n = 61) and Lgr5-GFP^{high} (n = 62) isolation methods, respectively, compared with only 61% for SM4 (n = 31) and 79% for SM2 (n = 29) (Figures 3D and S2H). Analysis of the co-expression of +4 ISC marker genes demonstrated a similar trend, where the majority of these genes were co-expressed

in each cell in the SM6 and Lgr5-GFP^{high} populations (Figures 3D and S2H), as previously described (Li et al., 2014). However, we noted slight differences between SM6 and Lgr5-GFP^{high} in the numbers of cells positive for the ISC marker *Sox9*. SM6 cells were more enriched for *Sox9*-positive cells (95.1%) compared with the Lgr5-GFP^{high} strategy (79%) (Figure S2H). In summary, our single-cell transcriptional analysis, based on these key genes, demonstrates that our isolation method gives rise to a homogeneous population of CBC cells which co-express key stem cell markers in a similar way to the well-established *Lgr5-Gfp* model.

Although SM6 and Lgr5-GFP^{high} CBC cell transcriptional signatures were highly similar, we wanted to confirm that these cells had similar functional capacities. Cells isolated using SM2, SM4, SM6, Lgr5-GFP^{high}, and negative strategies were assessed in an in vitro organoid assay (n = 3 for each cell population, five technical replicates per experiment). Although similar to the culture conditions described by Wang et al. (2013) for the growth of SM4 single cells, our culture conditions included the use of recombinant WNT3A, and the Rho kinase inhibitor Y-27632 was preferred to thiazovivin. All the cells, with the exception of negative cells, were capable of forming normal, round cystic organoids (Figure S2I), a classic architecture observed at day 4 after seeding when organoids are generated from single cells (Wang et al., 2013). However, there were significant differences in the number of organoids generated by the different sorting protocols. Cells isolated using SM6 and Lgr5-GFP^{high} sorting methods generated organoids at the same efficiency, which was almost 2-fold higher than the SM2 or SM4 strategies (Figure 3E). As the SM6 cell population is composed of both Lgr5-GFP-negative and Lgr5-GFP-positive cells (Figure S3A), we also investigated the organoid-forming potential of these two populations. At day 4, SM6-Lgr5-GFP^{negative} cells formed only marginally less organoids (0.91-fold) than SM6-Lgr5-GFP^{high} cells (Figure S3B). This demonstrates the capacity of the SM6 strategy to isolate Lgr5-positive stem cells that have silenced the GFP reporter. Together, these results confirm that SM6 and Lgr5-GFP^{high} cells are molecularly and functionally similar.

The establishment of a robust ISC isolation protocol that does not rely on the use of transgenic reporter alleles is critical to study ISCs in any transgene-free mouse strain. Therefore, we used our SM6 strategy to isolate cells from wild-type (WT) C57BL/6 animals (Figure S3C) and performed single-cell transcriptional analysis. The WT single cells (n = 30), isolated using the SM6 strategy (SM6-WT), had a transcriptional signature that was similar to the Lgr5-GFP^{high} cells (n = 62) (Figures S2A, S2F, and S3D). Moreover, PCA revealed that SM6-WT and Lgr5-GFP^{high} single-cell signatures overlapped and that the homogeneity of these cell populations was comparable (Figure 3F). However, a detailed analysis of the level of expression of the *Lgr5*



(legend on next page)



gene at the single-cell level in cells with detectable transcript levels revealed that at least 50% of cells isolated using the Lgr5-GFP^{high} strategy (38 of 60 cells) or SM6 strategy from *Lgr5-Gfp* mice (28 of 55 cells), which we will refer to now as SM6-TG to clearly differentiate it from SM6-WT, expressed half the amount of *Lgr5* compared with the cells isolated using the SM6 strategy from WT animals, SM6-WT (Figure 3G). These results suggest that the loss of one *Lgr5* allele due to the insertion of the reporter cassette is not fully compensated by the functional *Lgr5* allele at the individual cell level. In order to investigate in detail potential transcriptional differences between SM6-TG and SM6-WT cells, we performed RNA sequencing on freshly purified cells and found five genes to be differentially expressed (Figure 3E). Confirming our single-cell data, one of these genes was *Lgr5*, which was expressed at approximately 2-fold higher levels in SM6-WT cells compared with SM6-TG cells. The other four genes were the estrogen receptor *Esr1* (Cleveland et al., 2009), the immune-modulated *Erdr1*, a protective gene against cancer progression (Jung et al., 2011), the energy metabolism-associated gene insulin-degrading enzyme, *Ide*, and the fatty acid-binding protein 1, *Fabp1*. These four genes have not been reported to be WNT target genes, and we hypothesize that their differential expression is either a direct or indirect consequence of *Lgr5* haploinsufficiency. In order to address whether the observed transcriptional changes in these few genes, in particular *Lgr5* haploinsufficiency, induced functional defects, we isolated Lgr5-GFP^{high}, SM6-TG, and SM6-WT cells from littermate male animals (kept under the same housing conditions to minimize genetic and environmental differences) and subjected the cells to an organoid formation assay (n = 4, four experimental replicates with five technical replicates per experiment). At day 4, Lgr5-GFP^{high} cells gave rise to organoids at an efficiency of ~7% in contrast to SM6-WT cells, which gave rise to organoids at an efficiency of ~10% (Figures 3H and S3F; data are presented as fold change relative to Lgr5-GFP^{high}). In order to further characterize the organoids generated from

distinct cell populations, the expression of CBC stem cell markers (*Ascl2*, *Lgr5*, *Olfm4*), WNT signaling-related genes (*Axin-2*, *C-myc*, *Troy*), a niche marker (*Egf*), and differentiation markers (*Chromogranin A*, *Lysozyme*) were evaluated by quantitative RT-PCR (Figure S3G). This analysis revealed comparable expression levels of differentiation markers in organoids of all three groups. In SM6-WT organoids, a trend of higher expression of stem cell markers (*Lgr5*, *Olfm4*) and WNT target gene *Axin 2* was observed compared with Lgr5-GFP^{high} and SM6-TG (Figure S3G), but the differences were not statistically significant. *Esr1*, *Ide*, *Fabp1*, and *Erdr1* were expressed at comparable levels in organoid cultures of all three groups, and we speculate that the strong canonical WNT agonists in our culture media (CHIR, WNT3A, R-SPONDIN) might have compensated for direct or indirect effects of *Lgr5* haploinsufficiency. Taken together, these results suggest a potential functional deficiency within Lgr5-GFP cells with negative consequences on initial organoid establishment frequency.

In summary, we present a cell surface marker-mediated isolation protocol (a step-by-step protocol can be found in the Supplemental Experimental Procedures) for the purification of a highly enriched and homogeneous population of CBC stem cells molecularly and functionally comparable with ISCs extracted from *Lgr5-Gfp* mice. This strategy can also be utilized to isolate CBC cells from non-transgenic animals that express presumably normal physiological levels of *Lgr5*, *Esr1*, *Ide*, *Fabp1*, and *Erdr1*. The isolation strategy comprises a unique tool that should facilitate investigation of both intrinsic and extrinsic regulators of ISCs during normal homeostasis, age-related intestinal degeneration, and tumorigenesis.

EXPERIMENTAL PROCEDURES

Animals Used in This Study

Adult *Lgr5-eGFP-IRES-CreERT2* (courtesy of Professor Hans Clevers) and WT littermate male mice (6–12 weeks old, C57BL/6 background) were used in all experiments. Animals were housed in

Figure 3. Single-Cell Profiling and Functional Capacities of Prospective CBC Cell Populations

(A and B) Principal component analysis of the single-cell data for Lgr5-GFP^{high}, SM6, SM4, and SM2 cell populations*.

(C) Violin plots for key ISC marker genes for Lgr5-GFP^{high}, SM6, and negative cells*.

(D) Venn diagrams for some key ISC marker genes for Lgr5-GFP^{high} and SM6*.

(E) Organoid formation assay performed for Lgr5-GFP^{high}, negative, SM2, SM4, and SM6 single-cell populations (mean ± SEM, n = 3, experimental replicates, paired Student's t test, two-tailed).

(F) Principal component analysis.

(G) Beeswarm plot (one-way ANOVA with post hoc Bonferroni test; thick line, median; thin lines, quartiles)*. ns, not significant.

(H) Organoid formation assay performed for Lgr5-GFP^{high}, SM6-TG (SM6 strategy applied on transgenic *Lgr5-Gfp* animals), and SM6-WT (SM6 strategy applied on WT animals) single-cell populations (mean ± SEM, n = 4, experimental replicates, paired Student's t test, one-tailed).

*Replicates single-cell data: Lgr5GFP^{high} 62 cells pooled from two independent experiments, SM6/SM6-TG 61 cells pooled from two independent experiments, SM4 29 cells from one experiment, SM2 31 cells from one experiment, SM6-WT 30 cells from one experiment, Negative 31 cells from one experiment.



specific pathogen-free animal house conditions at the animal facility (Monash Animal Services) in strict accordance with good animal practice as defined by the National Health and Medical Research Council (Australia) Code of Practice for the Care and Use of Animals for Experimental Purposes. Experimental procedures were approved by the Monash Animal Research Platform Animal Ethics Committee. Animals were maintained under a 12/12 hr light/dark cycle at a temperature of 20°C with free access to food and water. For further information see, [Supplemental Experimental Procedures](#).

ACCESSION NUMBERS

The accession number for whole transcriptome sequencing experiments reported in this paper is SRA: SRP066815.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, three figures, and three tables and can be found with this article online at <http://dx.doi.org/10.1016/j.stemcr.2016.01.014>.

AUTHOR CONTRIBUTIONS

J.M.P., H.E.A., C.M.N., and T.J. designed the study and devised the FACS isolation protocol. C.M.N. performed FACS experiments, SPADE/viSNE, and the molecular experiments of the cells with support of A.S.K. and J.C.; T.J. performed cell preparation and organoid functional experiments with support from K.H. F.R. analyzed fluidigm and RNA sequencing data under the guidance of D.R.P. and J.M.P. C.M.N., T.J., H.E.A., and J.M.P. wrote the manuscript. All authors approved and contributed to the final version of the manuscript.

ACKNOWLEDGMENTS

We are grateful for the high-quality cell sorting service and technical input provided by the Monash Flowcore Facility. Furthermore, the authors thank the ACRF Centre for Cancer Genomic Medicine at the MHTP Medical Genomics Facility for assistance with next generation library preparation and Illumina sequencing. This work was supported by an NHMRC project grant APP1061883 to J.M.P. and H.E.A. J.M.P. was supported by a Silvia and Charles Senior Medical Viertel Fellowship.

Received: July 2, 2015

Revised: January 20, 2016

Accepted: January 20, 2016

Published: February 25, 2016

REFERENCES

Amir el, A.D., Davis, K.L., Tadmor, M.D., Simonds, E.F., Levine, J.H., Bendall, S.C., Shenfeld, D.K., Krishnaswamy, S., Nolan, G.P., and Pe'er, D. (2013). viSNE enables visualization of high dimensional single-cell data and reveals phenotypic heterogeneity of leukemia. *Nat. Biotechnol.* *31*, 545–552.

Barker, N. (2014). Adult intestinal stem cells: critical drivers of epithelial homeostasis and regeneration. *Nat. Rev. Mol. Cell Biol.* *15*, 19–33.

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

Carmon, K.S., Gong, X., Lin, Q., Thomas, A., and Liu, Q. (2011). R-spondins function as ligands of the orphan receptors LGR4 and LGR5 to regulate Wnt/beta-catenin signaling. *Proc. Natl. Acad. Sci. USA* *108*, 11452–11457.

Cleveland, A.G., Oikarinen, S.I., Bynote, K.K., Marttinen, M., Rafter, J.J., Gustafsson, J.A., Roy, S.K., Pitot, H.C., Korach, K.S., Lubahn, D.B., et al. (2009). Disruption of estrogen receptor signaling enhances intestinal neoplasia in *Apc(Min/+)* mice. *Carcinogenesis* *30*, 1581–1590.

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

de Lau, W., Barker, N., Low, T.Y., Koo, B.K., Li, V.S., Teunissen, H., Kujala, P., Haegebarth, A., Peters, P.J., van de Wetering, M., et al. (2011). *Lgr5* homologues associate with Wnt receptors and mediate R-spondin signalling. *Nature* *476*, 293–297.

Glinka, A., Dolde, C., Kirsch, N., Huang, Y.L., Kazanskaya, O., Ingelfinger, D., Boutros, M., Cruciati, C.M., and Niehrs, C. (2011). LGR4 and LGR5 are R-spondin receptors mediating Wnt/beta-catenin and Wnt/PCP signalling. *EMBO Rep.* *12*, 1055–1061.

Gracz, A.D., and Magness, S.T. (2014). Defining hierarchies of stemness in the intestine: evidence from biomarkers and regulatory pathways. *Am. J. Physiol. Gastrointest. Liver Physiol.* *307*, G260–G273.

Gracz, A.D., Ramalingam, S., and Magness, S.T. (2010). Sox9 expression marks a subset of CD24-expressing small intestine epithelial stem cells that form organoids in vitro. *Am. J. Physiol. Gastrointest. Liver Physiol.* *298*, G590–G600.

Gracz, A.D., Fuller, M.K., Wang, F., Li, L., Stelzner, M., Dunn, J.C., Martin, M.G., and Magness, S.T. (2013). Brief report: CD24 and CD44 mark human intestinal epithelial cell populations with characteristics of active and facultative stem cells. *Stem Cells* *31*, 2024–2030.

Jung, M.K., Park, Y., Song, S.B., Cheon, S.Y., Park, S., Houh, Y., Ha, S., Kim, H.J., Park, J.M., Kim, T.S., et al. (2011). Erythroid differentiation regulator 1, an interleukin 18-regulated gene, acts as a metastasis suppressor in melanoma. *J. Invest. Dermatol.* *131*, 2096–2104.

King, J.B., von Furstenberg, R.J., Smith, B.J., McNaughton, K.K., Galanko, J.A., and Henning, S.J. (2012). CD24 can be used to isolate *Lgr5+* putative colonic epithelial stem cells in mice. *Am. J. Physiol. Gastrointest. Liver Physiol.* *303*, G443–G452.

Li, N., Yousefi, M., Nakauka-Ddamba, A., Jain, R., Tobias, J., Epstein, J.A., Jensen, S.T., and Lengner, C.J. (2014). Single-cell analysis of proxy reporter allele-marked epithelial cells establishes intestinal stem cell hierarchy. *Stem Cell Rep.* *3*, 876–891.

Lujan, E., Zunder, E.R., Ng, Y.H., Goronzy, I.N., Nolan, G.P., and Wernig, M. (2015). Early reprogramming regulators identified



- by prospective isolation and mass cytometry. *Nature* 521, 352–356.
- Merlos-Suarez, A., Barriga, F.M., Jung, P., Iglesias, M., Cespedes, M.V., Rossell, D., Sevillano, M., Hernando-Momblona, X., da Silva-Diz, V., Munoz, P., et al. (2011). The intestinal stem cell signature identifies colorectal cancer stem cells and predicts disease relapse. *Cell Stem Cell* 8, 511–524.
- Munoz, J., Stange, D.E., Schepers, A.G., van de Wetering, M., Koo, B.K., Itzkovitz, S., Volckmann, R., Kung, K.S., Koster, J., Radulescu, S., et al. (2012). The Lgr5 intestinal stem cell signature: robust expression of proposed quiescent '+4' cell markers. *EMBO J.* 31, 3079–3091.
- Powell, A.E., Wang, Y., Li, Y., Poulin, E.J., Means, A.L., Washington, M.K., Higginbotham, J.N., Juchheim, A., Prasad, N., Levy, S.E., et al. (2012). The pan-ErbB negative regulator Lrig1 is an intestinal stem cell marker that functions as a tumor suppressor. *Cell* 149, 146–158.
- Qiu, P., Simonds, E.F., Bendall, S.C., Gibbs, K.D., Jr., Bruggner, R.V., Linderman, M.D., Sachs, K., Nolan, G.P., and Plevritis, S.K. (2011). Extracting a cellular hierarchy from high-dimensional cytometry data with SPADE. *Nat. Biotechnol.* 29, 886–891.
- Sangiorgi, E., and Capecchi, M.R. (2008). Bmi1 is expressed in vivo in intestinal stem cells. *Nat. Genet.* 40, 915–920.
- Sato, T., van Es, J.H., Snippert, H.J., Stange, D.E., Vries, R.G., van den Born, M., Barker, N., Shroyer, N.F., van de Wetering, M., and Clevers, H. (2011). Paneth cells constitute the niche for Lgr5 stem cells in intestinal crypts. *Nature* 469, 415–418.
- Takeda, N., Jain, R., LeBoeuf, M.R., Wang, Q., Lu, M.M., and Epstein, J.A. (2011). Interconversion between intestinal stem cell populations in distinct niches. *Science* 334, 1420–1424.
- Tan, D.W., and Barker, N. (2014). Intestinal stem cells and their defining niche. *Curr. Top. Dev. Biol.* 107, 77–107.
- van der Flier, L.G., van Gijn, M.E., Hatzis, P., Kujala, P., Haegebarth, A., Stange, D.E., Begthel, H., van den Born, M., Guryev, V., Oving, I., et al. (2009). Transcription factor achaete scute-like 2 controls intestinal stem cell fate. *Cell* 136, 903–912.
- Wang, F., Scoville, D., He, X.C., Mahe, M.M., Box, A., Perry, J.M., Smith, N.R., Lei, N.Y., Davies, P.S., Fuller, M.K., et al. (2013). Isolation and characterization of intestinal stem cells based on surface marker combinations and colony-formation assay. *Gastroenterology* 145, e1–e21.
- Wong, V.W., Stange, D.E., Page, M.E., Buczacki, S., Wabik, A., Itami, S., van de Wetering, M., Poulosom, R., Wright, N.A., Trotter, M.W., et al. (2012). Lrig1 controls intestinal stem-cell homeostasis by negative regulation of ErbB signalling. *Nat. Cell Biol.* 14, 401–408.

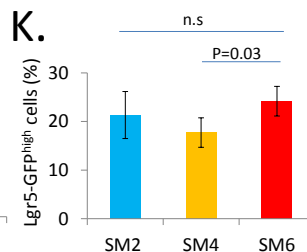
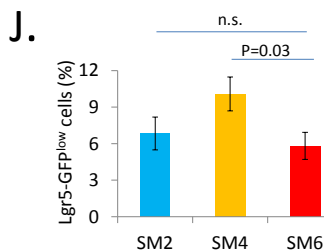
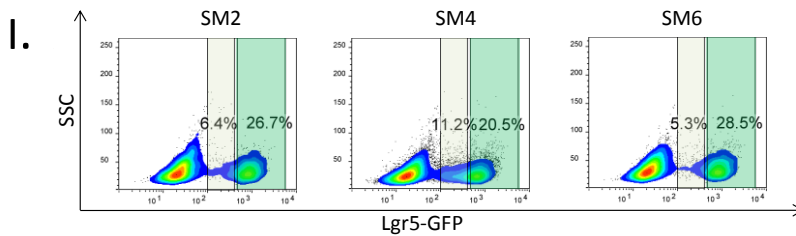
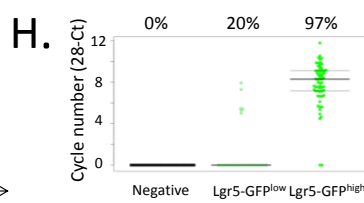
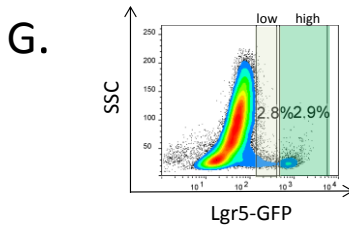
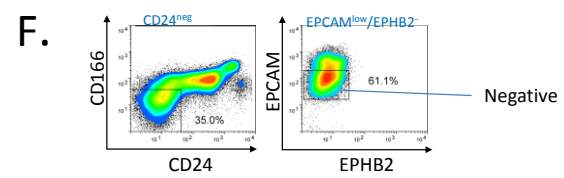
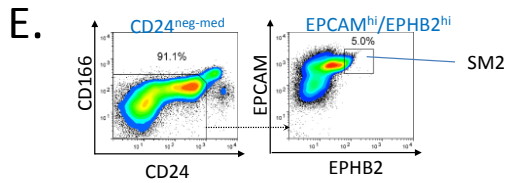
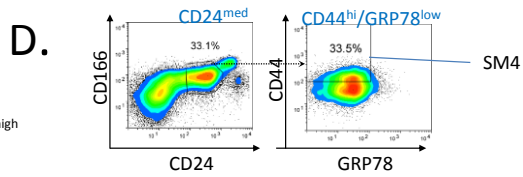
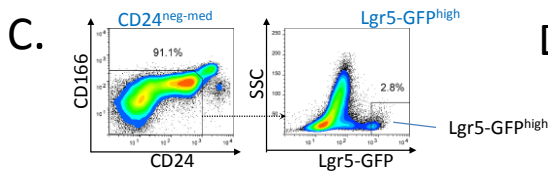
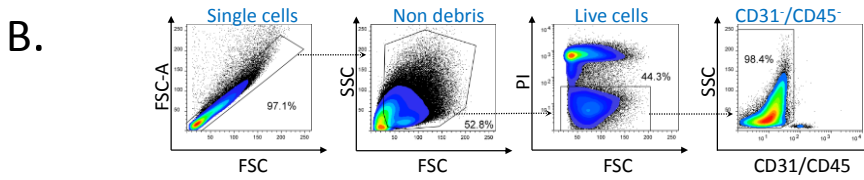
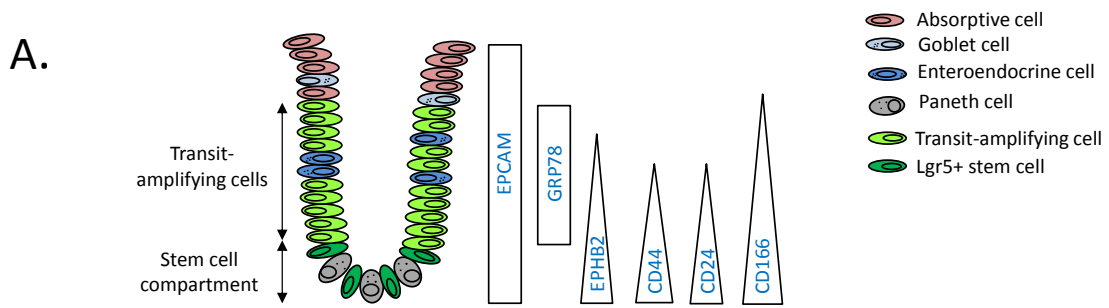
Stem Cell Reports, Volume 6

Supplemental Information

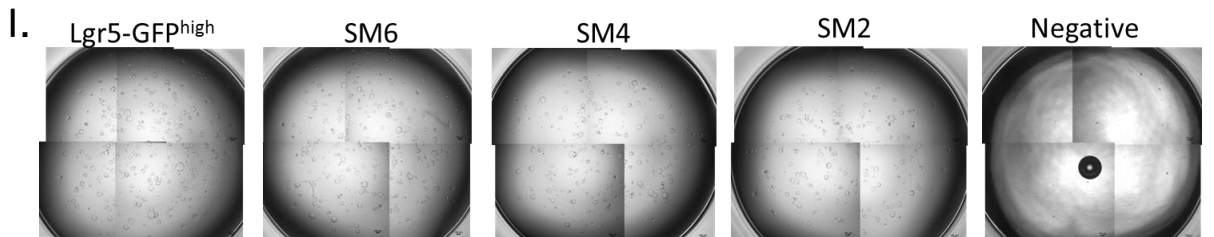
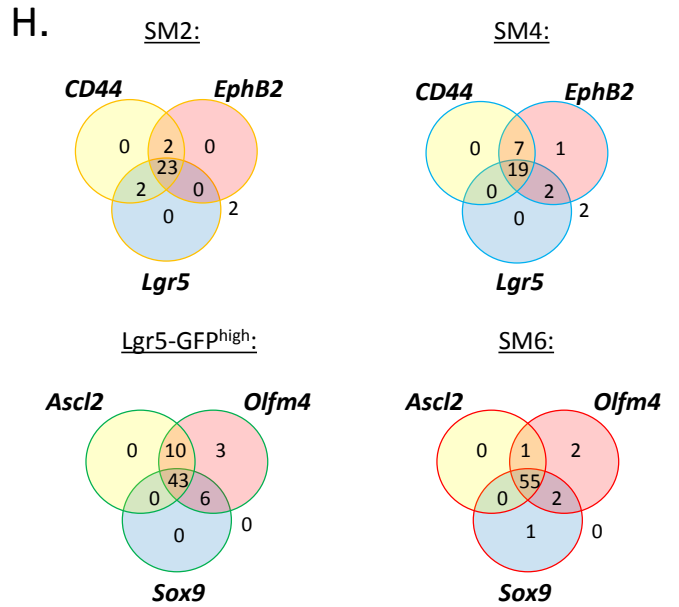
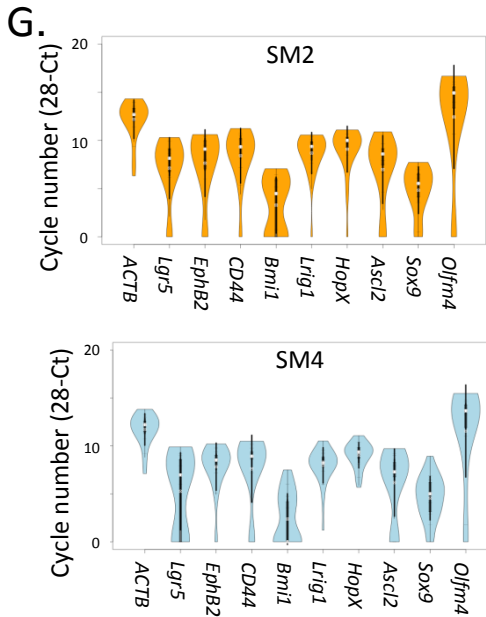
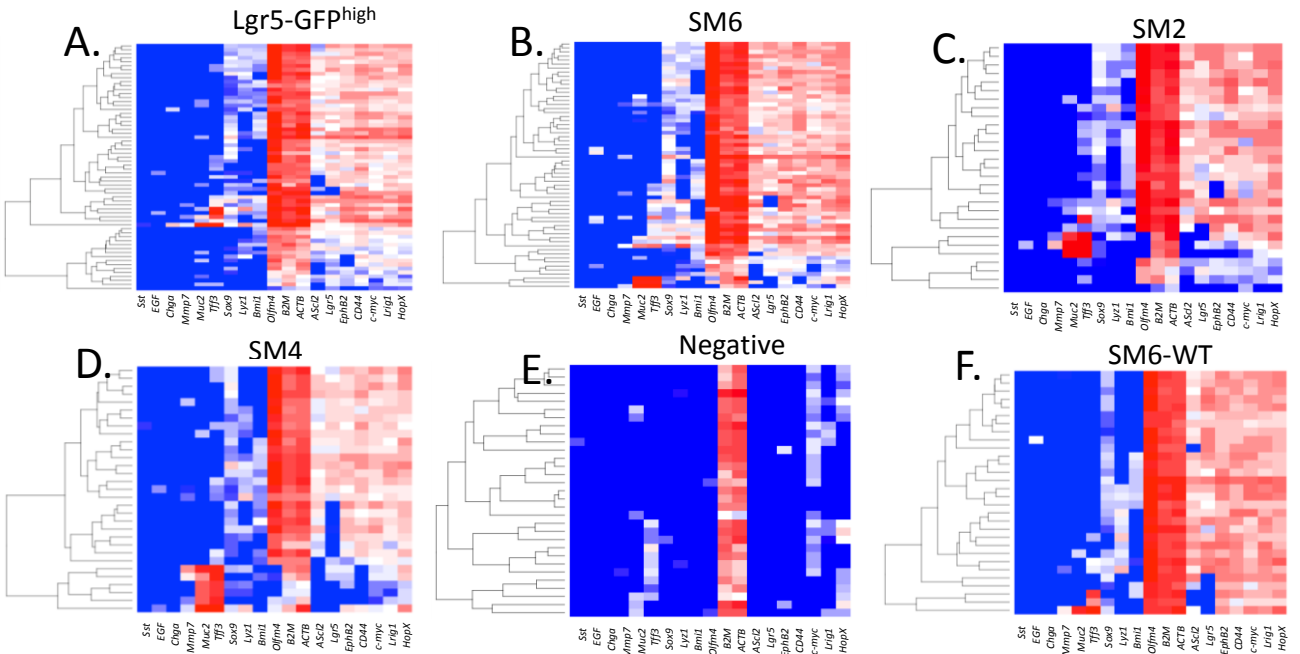
A Versatile Strategy for Isolating a Highly Enriched Population of Intestinal Stem Cells

Christian M. Nefzger, Thierry Jardé, Fernando J. Rossello, Katja Horvay, Anja S. Knaupp, David R. Powell, Joseph Chen, Helen E. Abud, and Jose M. Polo

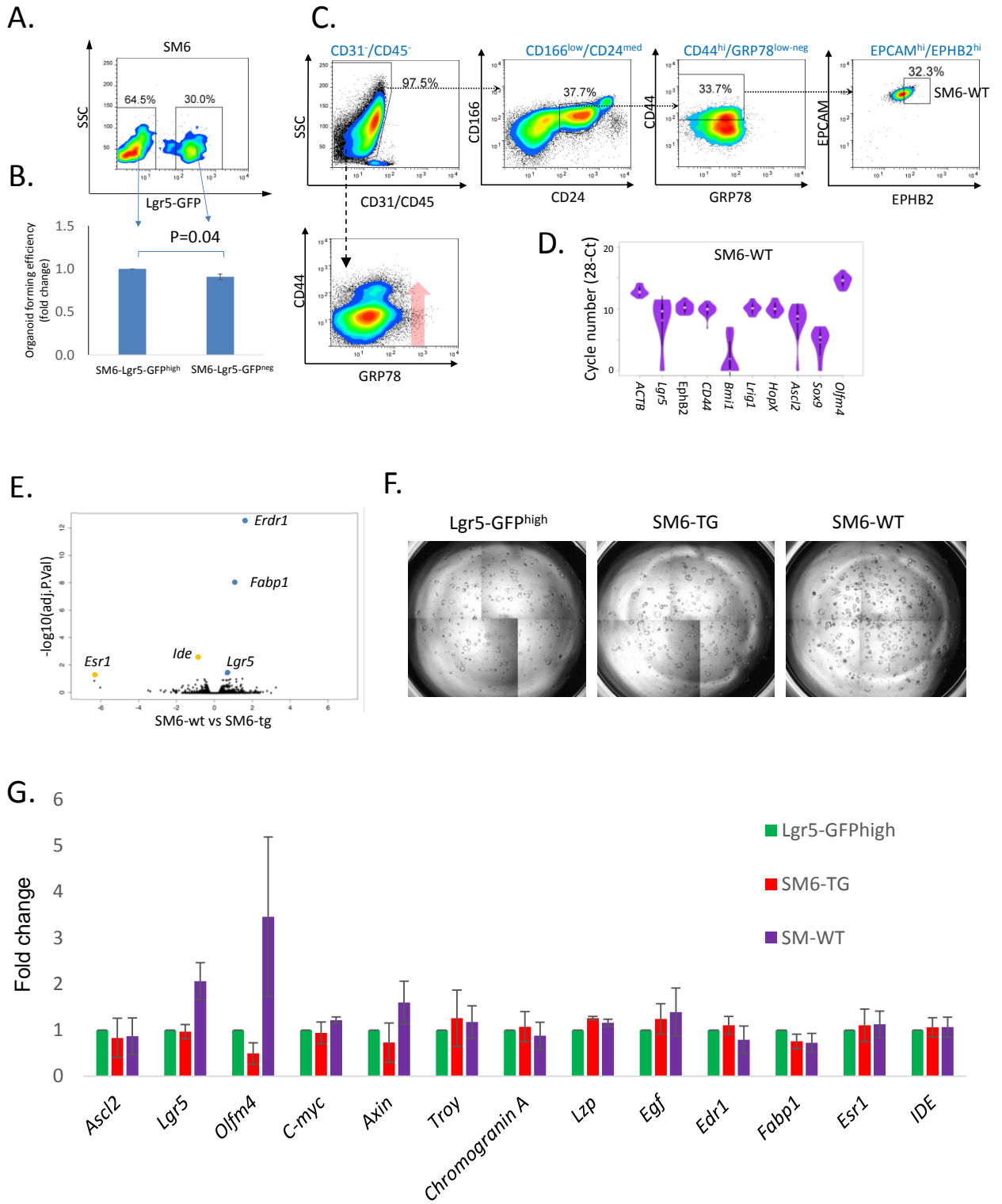
Supplementary Figure 1:



Supplementary Figure 2:



Supplementary Figure 3:



Supplementary figure 1: Gating strategies, related to Figure 1. (A) Localisation of cell surface markers in the intestinal crypt. (B) Depletion for aggregates, debris, PI positive events and CD31/CD45 positive cells before gating in on the (C) Lgr5-GFP^{high} (depleted for CD24 positive Paneth cells), (D) SM4, (E) SM2 and (F) Negative populations. (G) Representative FACS blots depicting Lgr5-GFP^{low} and Lgr5-GFP^{high} cells from an *Lgr5-Gfp* reporter animal. (H) Beeswarm plot for single cell Lgr5 expression for Negative, Lgr5-GFP^{low} and Lgr5-GFP^{high} cells; percentage values of cells with detectable Lgr5 transcripts are indicated above the blot (Negative 31 cells from one experiment, Lgr5GFP^{low} 30 cells from one experiment, Lgr5GFP^{high} 62 cells pooled from 2 independent experiments). (I) Representative FACS blots depicting Lgr5-GFP^{low} and Lgr5-GFP^{high} cells for SM2, SM4 and SM6. Quantification of (J) Lgr5-GFP^{low} cells and (K) Lgr5-GFP^{high} cells for SM2, SM4 and SM6 strategies; a paired Wilcoxon test was performed (mean \pm s.e.m, n=6, experimental replicates).

Supplementary figure 2: Additional profiling data, related to Figure 3. (A-F) Heat maps of the single cell data for Lgr5-GFP^{high}, SM6, SM2, SM4, Negative and SM6-WT*. (G) Violin plots for SM2 and SM4*. (H) Venn diagrams for SM2, SM4, Lgr5-GFP^{high} and SM6*. (I) Composite images of whole 96-wells at day 4 of culture for Lgr5-GFP^{high}, SM6, SM4, SM2 and the Negative population (scale bars, 100 μ m).

*replicates Single cell data: Lgr5GFP^{high} 62 cells pooled from 2 independent experiments, SM6/SM6-TG 61 cells pooled from 2 independent experiments, SM4 29 cells from one experiment, SM2 31 cells from one experiment, SM6-WT 30 cells from one experiment, Negative 31 cells from one experiment.

Supplementary figure 3: Additional profiling data, related to Figure 3. (A) Representative FACS blot depicting SM6-Lgr5^{negative} and SM6-Lgr5^{high} cells to subfractioning of SM6 according to Lgr5-GFP expression. (B) Organoid formation frequency (fold change compared to SM6-Lgr5^{high}) for SM6-Lgr5^{negative} and SM6-Lgr5^{high} cells; a 2-tailed unpaired Student's T test was performed (mean \pm s.e.m., n=3, experimental replicates). (C) Isolation of wild-type CBC cells using our combination of 6 cell surface markers (SM6, FACS plots 1st lane); robust shift in CD44 expression characterizes successful cell isolation (FACS plot, 2nd lane). (D) Violin plots for key ISC marker genes for SM6-WT cells (30 cells from one experiment). (E) Volcano plot depicting differentially expressed genes between SM6-TG and SM-WT (n=2, experimental replicates). (F) Organoid culture for prospective ISC populations. Composite images of whole 96-wells at day 4 of culture for Lgr5-GFP^{high}, SM6-TG and SM6-WT populations (scale bars, 100 μ m). (G) qPCR performed on day 4 organoid cultures (mean \pm s.e.m., n=3, experimental replicates).

Supplementary Table 1: Cells surface marker profile of cell populations of interest, related to Figure 1

	Negative cells	SM2	SM4	SM6	Lgr5-GFP
CD31/CD45	neg	neg	neg	neg	neg
CD24	neg	neg-med	med	med	neg-med
CD166	neg	n/a	low	low	n/a
CD44	neg	n/a	high	high	n/a
GRP78	neg	n/a	neg-low	neg-low	n/a
EPCAM	low	high	n/a	high	n/a
EPHB2	neg	high	n/a	high	n/a

Supplementary Table 2: List of differentially expressed genes, related to Figure 2.

SM2 versus Lgr5-GFP	SM4 versus Lgr5-GFP		
<i>Gsdmc2</i>	<i>Bcas1</i>	<i>Sst</i>	<i>Rep15</i>
<i>Gsdmc4</i>	<i>Chgb</i>	<i>Tph1</i>	<i>Plcb1</i>
<i>Npc1l1</i>	<i>Muc2</i>	<i>Rab3c</i>	<i>Srgap1</i>
<i>Spink4</i>	<i>Spink4</i>	<i>Ern2</i>	<i>Lrrc26</i>
<i>Muc2</i>	<i>Tff3</i>	<i>Mptx1</i>	<i>Trpa1</i>
<i>Gsdmc3</i>	<i>Ano7</i>	<i>Spdef</i>	<i>Gal3st2</i>
<i>Slc43a2</i>	<i>Creb3l1</i>	<i>Hpd</i>	<i>Tpsg1</i>
<i>Anxa13</i>	<i>Gsdmc4</i>	<i>Pcsk1</i>	<i>Gfi1</i>
<i>Dclk1</i>	<i>Gsdmc2</i>	<i>Lmx1a</i>	<i>Sphkap</i>
<i>Apol10a</i>	<i>Agr2</i>	<i>Vcan</i>	<i>Ptprr</i>
<i>Sis</i>	<i>Rgs2</i>	<i>Dbpht2</i>	<i>Myof</i>
<i>Ano7</i>	<i>Sytl2</i>	<i>Cacna1h</i>	<i>Rgs4</i>
<i>Ms4a18</i>	<i>Hepacam2</i>	<i>Cck</i>	<i>Insm1</i>
	<i>Atoh1</i>	<i>Cpe</i>	<i>Dnahc9</i>
	<i>Atp2a3</i>	<i>Kcnh6</i>	<i>Enpp2</i>
	<i>Map1b</i>	<i>Ptprn2</i>	<i>Fhl1</i>
	<i>Pde2a</i>	<i>Scn3a</i>	<i>Scnn1a</i>
	<i>Serpina1b</i>	<i>Cadps</i>	<i>Nfasc</i>
	<i>Clca3</i>	<i>Gch1</i>	<i>St8sia3</i>
	<i>Rfx6</i>	<i>Itga4</i>	<i>Capn9</i>
	<i>Guca2a</i>	<i>Fcgbp</i>	<i>Aplp1</i>
	<i>P2ry4</i>	<i>Kcnk6</i>	<i>Gnao1</i>
	<i>Selm</i>	<i>Pclo</i>	<i>Cacna2d1</i>
	<i>Chga</i>	<i>Afp</i>	<i>Map2</i>
	<i>Gcnt3</i>	<i>Bace2</i>	<i>Nkx2-2</i>
	<i>Anxa13</i>	<i>Gcg</i>	<i>Rcan3</i>
	<i>Scin</i>	<i>Adcy1</i>	<i>Sv2c</i>
	<i>Tac1</i>	<i>Gpr116</i>	<i>Gip</i>
	<i>Gsdmc3</i>	<i>Sec24d</i>	<i>Disp2</i>
	<i>Zg16</i>	<i>Klk1</i>	<i>Gfra3</i>
	<i>Sct</i>	<i>Pdia5</i>	<i>Kcnj5</i>
	<i>S100a6</i>	<i>Anxa2</i>	<i>Scg2</i>
	<i>Ccl9</i>	<i>Cbfa2t3</i>	<i>St18</i>
	<i>Vim</i>	<i>Rap1gap</i>	
	<i>Pyy</i>	<i>Amigo2</i>	

Supplementary Table 3: Antibodies used in this study, related to detailed multi-step protocol.

Antibody	Dilution Factor			Company	Clone	Catalog#	Excitation laser	Detection filter
	1°	2°	3°					
Mouse anti-EphB2	1:100			Genentech	2H9	courtesy of Genentech*		
rabbit anti-GRP78	1:100			Sigma	polyclonal	G9043		
donkey-anti-mouse-IgG AF555 antibody		1:200		Thermo-Fischer	polyclonal	A31570	561nm	555-633nm
goat-anti-rabbit-APC-Cy7		1:100		Santa Cruz Biotechnology	polyclonal	sc-3847	635nm	750-810nm
rat-anti-Epcam-eFluor450			1:100	eBioscience	G8.8	48-5791-82	405nm	425-475nm
rat-anti-CD45-BV510			1:200	BDBiosciences	30-F11	563891	405nm	500-550nm
rat-anti-CD31-BV510			1:200	BDBiosciences	MEC 13.3	563089	405nm	500-550nm
rat-anti-CD44-BV650			1:100	Biolegend	IM7	103049	405nm	640-680nm
rat-anti-CD24-PeCy7			1:100	eBioscience	M1/69	25-0242-82	561nm	750nm long pass
rat-anti-CD166-APC			1:100	eBioscience	eBioALC48	17-1661-82	635nm	655-685nm

*This antibody is now also available from BDBiosciences and is provided at the same concentration as used in this study.

Supplemental Experimental Procedures:

Crypt isolation and cell dissociation

Mice were culled by cervical dislocation. As previously described (Horvay et al., 2015; Jarde et al., 2013; Wang et al., 2013), the small intestinal tube was dissected out and flushed with PBS to remove faeces. Small intestinal tracts were opened longitudinally, scraped with a glass coverslip to remove villi, cut into 5-mm pieces and washed with PBS five times to remove unattached epithelial fragments, mucus and faeces. Following incubation for 30 min at 4°C in 3mM EDTA-PBS, intestinal crypts were released from small intestine tissue fragments by mechanically pipetting with a 10ml pipette in PBS and repeating this step three times. Isolated intestinal crypts were strained (70-µm cell strainer, BD Biosciences) and pelleted by centrifugation three times at 1500 rpm for 2 minutes at 4°C. The collected crypts were incubated for 30 minutes at 4°C in DMEM/F12 – 10% serum (Gibco) and then dissociated in TrypLE Express (Invitrogen) supplemented with 10 µM Rock inhibitor (Y-27632, Abcam) and 2.5µg/ml DNase 1 (Sigma-Aldrich) for 5 minutes at 37°C. Cell clumps and mucus were removed using a 70-µm cell strainer (BD Biosciences) and the remaining dissociated cells were washed twice with PBS and collected by centrifugation at 4°C at 1500 rpm for 3 minutes.

Flow cytometry

All antibody labelling steps as well as the final resuspension of the samples were performed with PBS supplemented with 2mM EDTA, 2% FBS and 10 µM Rock inhibitor (Y-27632). Cellularised crypts were submitted to a three step sequential antibody labelling procedure: (I) mouse-anti-EPHB2 antibody (1:100 dilution, clone 2H9, courtesy of Genentech); (II) donkey-anti-mouse-IGG AF555 antibody (1:200, polyclonal, LifeTechnologies, cat# A31570) rabbit-anti-GRP78 antibody (1:100, polyclonal, Sigma, cat# G9043); anti-UEA-1-Biotin (1:1000,

Vectorslabs, cat# B-1065) (III) Streptavidin-BUV395 (1:100, BD Biosciences cat# 564176), rat-anti-EPCAM-eFluor450 (1:100, clone: G8.8, eBioscience, cat# 48-5791-82), rat-anti-CD31-BV510 (1:200, clone: MEC 13.3, BD Biosciences, cat# 563089) rat-anti-CD45-BV510 (1:200, clone: 30-F11, BD Biosciences, cat# 563891), rat-anti-CD44-BV650 (1:100, clone IM7, Biolegend, cat# 103049), rat-anti-CD24-PeCy7 (1:100, clone: M1/69, eBioscience, cat# 25-0242-82), rat-anti-CD166-APC (1:100, clone: eBioALC48, eBioscience, cat# 17-1661-82), and secondary antibody goat-anti-rabbit-APC-Cy7 (1:100, polyclonal, Santa Cruz Biotechnologies, cat# sc-3847). All antibody labelling steps were carried out (for the cells of one animal) in a 500µl volume for 15 minutes on ice; after each antibody labelling step, cells were washed with 10ml cold PBS and pelleted at 400xg for 3 minutes. The cells for each animal were then resuspended in a final volume of 1ml, passed through a 70µm strainer and transferred into appropriate FACS tubes where propidium iodide (PI) was added to a concentration of 2µg/ml. Cell sorting was carried out with a 100 µm nozzle on an Influx instrument (BD Biosciences). The gating strategies to isolate SM2 and SM4 were adapted from Merlos-Suarez *et al.* and Wang *et al.* (Merlos-Suarez *et al.*, 2011; Wang *et al.*, 2013). For all populations of interest (SM2, SM4, SM6, Lgr5-GFP^{high} and Negative), aggregates, debris, dead cells (PI+) and CD45+/CD31+ hematopoietic/endothelial contaminants were depleted. Before isolating SM2 and Lgr5-GFP^{high} cells, Paneth cells were excluded by depleting for CD24^{hi} cells. For Lgr5-GFP^{high}, 2.5-3% of the Lgr5-GFP brightest cells were selected; for SM2 the top 5% EPCAM^{high}/EPHB2^{high} were selected. For SM4 and SM6, the CD24^{med}/CD166^{low} population was subgated into CD44^{high}/GRP78^{neg-low} cells (the gate was set to encompass ~ 25% of the population). For SM6, an additional step was included where ~33% of the top EPCAM^{high}/EPHB2^{high} cells were collected (please note the % value of the final SM6 gate was set to approximate/emulate the position of the final SM2 gate). Purity of collected fractions

was confirmed by reanalysis of a small fraction of the sorted cells. For single cell applications cells were double sorted.

Multidimensional analyses of flow cytometry data

We used the Cytobank platform (Fluidigm, South San Francisco, California) to generate viSNE maps and SPADE trees from Flow Cytometry Standard files. Analyses were performed on live cells depleted for CD31 and CD45 positive cells and EPCAM negative cells. To generate viSNE maps, 10^5 events in total were used for sampling. SPADE trees were generated with a target number of 100 nodes; the down sampled events target was set to 100%. For both viSNE and SPADE six fluorescent channels were used for dimensional reduction (EPHB2, CD44, CD166, GRP78, CD24 and UEA-1).

RNA sequencing

RNA was extracted with Qiagen's RNeasy micro kit from $2-3 \times 10^4$ FACS isolated cells as per instructions. For generation of sequencing libraries, 25ng of RNA (RIN value >9) were submitted to SPIA amplification (NuGen). Two biological replicates per condition were sequenced using the HiSeq 2000 sequencing platform (Illumina, San Diego, CA, USA). Each library was pair-end with a 100nt read length (350nt average insert size). The targeted number of sequencing reads per sample was 15 million. Raw sequencing reads were assessed for overall quality using FASTQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Sequencing specific adaptors and low quality reads (Phred score of 6 consecutive bases below 15, minimum read length of 36nt) were filtered and hard trimmed using Trimmomatic [v 0.30] (Bolger et al., 2014). Sample reads were aligned to the mouse genome [complete mm10 (UCSC version, December 2011)] using Tophat2 [v 2.0.13, default parameters] (Kim et al., 2013). Transcript quantification was performed using HTSeq [v 0.6.1, default parameters] (Anders et

al., 2015) and transcripts with more than ten sequencing reads in at least one sample were used for further analysis. Sample library size was normalized using the TMM method (Robinson and Oshlack, 2010). The sequences reported in this paper are available at the NIH Short Reads Archive, (www.ncbi.nlm.nih.gov/sra), accession number SRP066815.

Single Cell PCR

Single cell PCR was performed as previously described (Polo et al., 2012) with LifeTechnologies Single Cell to Ct kit. In brief, 96-well plates for qPCR were filled with 10ul lysis solution and single cells were deposited with a cell sorter into each well. As per kits instructions, cDNA was produced from the lysate and submitted to 18 cycles of pre-amplification with TaqMan probes (Life Technologies) of the 19 genes of interest (*Actb*, *Ascl2*, *B2M*, *Bmi1*, *c-myc*, *Cd44*, *Chga*, *Egf*, *EphB2*, *HopX*, *Lgr5*, *Lrig1*, *Lyz1*, *Mmp7*, *Muc2*, *Olfm4*, *Sox29*, *Sst*, *Tff3*). Pre-amplified templates that were positive for housekeeper *Actb* (manually tested with qPCR) were then used for Single-cell PCR data collection with a Biomark instrument (Fluidigm). Results are expressed as $\text{Log}_2\text{Ex} = \text{LOD (Limit of Detection) Cq} - \text{Cq [Gene]}$. The limit of detection was set to 28. If Log_2Ex value is negative, $\text{Log}_2\text{Ex} = 0$. For SM2, SM4, SM6-WT, Lgr5GFP^{low} and Negative approximately 30 cells per group from one experiment were used for analysis. For key populations SM6 and Lgr5-GFP^{high} in total around 60 cells (from two separate experiments) were used for analysis.

Quantitative RT-PCR

After 4 days in culture, organoids generated from single SM6-TG, SM6-WT or Lgr5-GFP^{high} cells were homogenised and total RNA extracted using a RNeasy micro kit (Qiagen), as previously described (Jarde et al., 2015). RNA was reverse transcribed using the QuantiTect Reverse Transcription kit (Qiagen). Quantitative reverse transcriptase polymerase chain

reaction (qRT-PCR) was performed using Brilliant II SYBR Green QPCR Master Mix (Agilent technologies). Triplicate samples were analysed on a LightCycler 480 machine (Roche Diagnostics). Gene expression levels were calculated using the 2^{-DDCt} method using *Gapdh* as a normaliser. The following primer sequences (depicted 5'-3') were used: *Ascl2* (F: CAGGAGCTGCTTGACTTTTCCA, R: GGGCTAGAAGCAGGTAGGTCCA), *Axin2* (F: GCAGCTCAGCAAAAAGGGAAAT, R: TACATGGGGAGCACTGTCTCGT), *Chromogranin A* (F: TCCCCACTGCAGCATCCAGTTC, R: CCTTCAGACGGCAGAGCTTCGG), *C-myc* (F: CTAGTGCTGCATGAGGAGACAC, R: GTAGTTGTGCTGGTGAGTGGAG), *Egf* (F: GTTCAGTGCTTGGGAGAGATG, R: CCTGGGAATTTGCAAACAGTA), *Esr1* (F: CCCGCCTTCTACAGGTCTAAT, R: CTTTCTCGTTACTGCTGGACAG), *Erd1* (F: GGTCAAGATGTATGTGCCACC, R: GCTTCTACGTGTGTGCTTTTCG), *Fabp1* (F: GGAATTGGGAGTAGGAAGAGCC, R: TGGACTTGAACCAAGGAGTCAT), *Ide* (F: AATCCGGCCATCCAGAGAATA, R: GGGTCTGACAGTGAACCTATGT), *Lgr5* (F: CCTTGGCCCTGAACAAAATA, R: ATTTCTTTCCCAGGGAGTGG'), *Lzp* (F: GAGACCGAAGCACCGACTATG, R: CGGTTTTGACATTGTGTTCGC), *Olfm4* (F: AACATCACCCCAGGCTACAG, R: TGTCCACAGACCCAGTGAAA), *Troy* (F: GACTGCCTGCCAGGATTTTAC, R: CAGTGTGGTTCGTAGGGAGG), *Gapdh* (F: CTCGTCTCATAGACAAGATGGTGAAG, R: AGACTCCACGACATACTCAGCACC).

Cell culture

Following FACS isolation, single epithelial cells were collected in DMEM/F12 supplemented with 10% serum and 10 μ M Y-27632 (Abcam). Intestinal cells were centrifuged at 4°C for 5 minutes at 1500 rpm. The cell pellet was resuspended in growth-factor reduced Matrigel (1000 cells per μ l, BD Biosciences) containing 10 μ M JAGGED-1 (Anaspec). 5000 cells were seeded per well in a 96 well plate. Following Matrigel polymerisation, 100 μ l of crypt culture medium

per well was overlaid (DMEM/F12 supplemented with N2, B27, penicillin/streptomycin, glutamax, 10mM HEPES, fungizone, 50 ng/ml EGF (Peprotech), 100 ng/ml NOGGIN (Peprotech), 1 µg/ml R-SPONDIN 1 (R&D Systems), 10 µM Y-27632 (Abcam), 100 ng/ml WNT-3a (R&D) and 2.5 µM CHIR (Stemgent)). Intestinal cells were maintained in a 37°C humidified atmosphere under 5% CO₂. After 3 days, the culture medium was entirely replaced by freshly made culture medium without Y-27632 and WNT-3a. After 4 days in culture, images of wells (5 wells per condition, 3-5 biological replicates) were taken and organoids were manually counted using FIJI image analysis cell counter software.

Statistical analysis and visualization

Descriptive statistics and plots were analyzed and produced using made4 (Culhane et al., 2005), caroline (Schruth, 2013), limma (Ritchie et al., 2015), gplots (Warnes et al., 2015) and beeswarm. Principal component and unsupervised hierarchical clustering (Pearson's correlation) analyses were performed using limma (Ritchie et al., 2015), bioDist (Ding et al.) and hclust (<http://CRAN.R-project.org/package=gplots>) respectively. Other statistical tests were performed as indicated in the figure legends.

Detailed multi-step protocol for SM6 isolation from C57/Bl6 wild type animals

Part A: Isolation of intestinal epithelial cells

- 1) Cull mice by cervical dislocation
- 2) Generously spray the animals abdomen with alcohol before removing the small intestine and collecting it in 30ml of ice cold PBS
- 3) Flush the intestinal tube with ice cold PBS with a 20ml Syringe to remove faeces.
- 4) Cut open the small intestinal tube longitudinally
- 5) With the inside of the intestinal tube facing up, very gently scrape the surface with a glass coverslip to remove villi

- 6) Cut the intestinal tract into 5mm long pieces and wash 5 times with 30ml PBS to remove unattached epithelial fragments, mucus and faeces (Note: the washing steps are crucial for the final quality of the preparation)
- 7) Incubate for 30 min at 4°C in 30mls of 3mM EDTA-PBS with gentle agitation
- 8) In 30ml fresh ice cold PBS, release intestinal crypts from small intestinal tissue fragments by mechanically pipetting them vigorously with a 10ml pipette and repeating this step three times
- 9) Strain isolated intestinal crypts through a 70-µm cell strainer (BD Biosciences) and pellet by centrifugation three times at 1500 rpm for 2 minutes at 4°C to enrich for crypt fragments
- 10) Incubate the collected crypts for 30 minutes at 4°C in 10ml DMEM/F12 plus 10% fetal bovine serum (FBS)
- 11) Fill tube to 30ml with ice cold PBS and pellet at 1500 rpm for 3 minutes at 4°C
- 12) Cellularize crypts in 1ml TrypLE Express supplemented with 10uM Rock inhibitor and 2.5µg/ml DNase I for 5 minutes at 37°C
- 13) Immediately add 200 ul FBS and gently pipette up and down (~20 times) with a 1000ul pipette to break up clumps
- 14) Fill tube with 30ml ice cold PBS, pass through a 70um strainer and pellet for 3 minutes by centrifugation at 1500rpm

Part B: FACS purification of intestinal stem cells:

- 15) Resuspend cells in 10ml ice cold PBS and put aside 500 ul as unlabelled control (Note: the control cells are to be strained, supplemented with Propidium iodide @ 2ug/ml and 10 µM Rock inhibitor before use); pellet the remaining cells by centrifugation for 3 min at 1500rpm
- 16) Label epithelial cells via a 3-step labelling protocol (note: preparation of the Antibody labelling solutions is outlined in supplementary Table 3 and the methods section).
- 17) Resuspend the cell pellet in 500µl primary antibody labelling solution per mouse and incubate on ice for 15minutes.
- 18) Add 10ml ice cold PBS and pellet for 3 minutes at 1500rpm.
- 19) Repeat steps 17 and 18 for the secondary and the tertiary antibody labelling solutions
- 20) Resuspend the fully labelled pellet in 1ml of solution supplemented with Propidium iodide (2ug/ml), pass through a 70um strainer and transfer into appropriate FACS sample tubes
- 21) Note, compensation controls are essential for this multicolour protocol. Cells (from step 15) labelled with the individual, conjugated antibodies (or via a secondary approach for EPHB2/GRP78) are ideal, but antibody capture beads from BD Bioscience can also be used (except for the PeCy7 channel where the use of a labelled cell control is required).
- 22) Use the unlabelled cell sample (step 15) and the compensation tubes (step 21) to calibrate the cell sorter (100µm nozzle)

- 23) Gate out debris, aggregates and dead cells and set gates to capture the SM6 population as described in Figure 1D and Supplementary Fig 1B. (Crucial: Successful cell preparations with a high number of intestinal stem cells are defined by a robust shift of CD44 expression in a subset of all live cells as depicted in Supplementary Figure 3C)
- 24) Sort cells into collection tubes with DMEM/F12, 10% FBS and 10 μ M Rock inhibitor
- 25) Note that once the sorting process has commenced it is crucial that the gates for the CD44^{high}/GRP78^{low} population and the EPCAM⁺/EPHB2^{high} population are checked on regular basis to ensure that only $\leq 33\%$ of these populations are gated for. (Note: while sorting, if possible, display ≥ 100000 live events, this will make it easier to establish relatively stable gates)
- 26) If sorting larger samples it is advisable to resuspend the sort sample every 15-20 minutes by gentle pipetting.

After sorting has been completed, it is important to routinely perform re-analysis on a small fraction of the sorted cells (20-40ul) to verify purity and viability of the target population.

Supplementary References

- Anders, S., Pyl, P. T., and Huber, W. (2015). HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* *31*, 166-169.
- Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* *30*, 2114-2120.
- Culhane, A. C., Thioulouse, J., Perriere, G., and Higgins, D. G. (2005). MADE4: an R package for multivariate analysis of gene expression data. *Bioinformatics* *21*, 2789-2790.
- Ding, B., Gentleman, R., and Carey, V. bioDist: Different distance measures. R package version 1380.
- Horvay, K., Jarde, T., Casagrande, F., Perreau, V. M., Haigh, K., Nefzger, C. M., Akhtar, R., Gridley, T., Berx, G., Haigh, J. J., *et al.* (2015). Snai1 regulates cell lineage allocation and stem cell maintenance in the mouse intestinal epithelium. *The EMBO journal* *34*, 1319-1335.
- Jarde, T., Evans, R. J., McQuillan, K. L., Parry, L., Feng, G. J., Alvares, B., Clarke, A. R., and Dale, T. C. (2013). In vivo and in vitro models for the therapeutic targeting of Wnt signaling using a Tet-ODeltaN89beta-catenin system. *Oncogene* *32*, 883-893.
- Jarde, T., Kass, L., Staples, M., Lescesen, H., Carne, P., Oliva, K., McMurrick, P. J., and Abud, H. E. (2015). ERBB3 Positively Correlates with Intestinal Stem Cell Markers but Marks a Distinct Non Proliferative Cell Population in Colorectal Cancer. *PloS one* *10*, e0138336.
- Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S. L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome biology* *14*, R36.
- Merlos-Suarez, A., Barriga, F. M., Jung, P., Iglesias, M., Cespedes, M. V., Rossell, D., Sevillano, M., Hernando-Momblona, X., da Silva-Diz, V., Munoz, P., *et al.* (2011). The intestinal stem cell signature identifies colorectal cancer stem cells and predicts disease relapse. *Cell stem cell* *8*, 511-524.

Polo, J. M., Anderssen, E., Walsh, R. M., Schwarz, B. A., Nefzger, C. M., Lim, S. M., Borkent, M., Apostolou, E., Alaei, S., Cloutier, J., *et al.* (2012). A molecular roadmap of reprogramming somatic cells into iPS cells. *Cell* *151*, 1617-1632.

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., and Smyth, G. K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic acids research* *43*, e47.

Robinson, M. D., and Oshlack, A. (2010). A scaling normalization method for differential expression analysis of RNA-seq data. *Genome biology* *11*, R25.

Schruth, D. (2013). caroline: A Collection of Database, Data Structure, Visualization, and Utility Functions for R. R package version 076 <http://CRANR-projectorg/package=caroline>.

Wang, F., Scoville, D., He, X. C., Mahe, M. M., Box, A., Perry, J. M., Smith, N. R., Lei, N. Y., Davies, P. S., Fuller, M. K., *et al.* (2013). Isolation and characterization of intestinal stem cells based on surface marker combinations and colony-formation assay. *Gastroenterology* *145*, 383-395 e381-321.

Warnes, G. R., Bolker, B., Bonebakker, L., Gentleman, R., Liaw, W. H. A., Lumley, T., Maechler, M., Magnusson, A., Moeller, S., Schwartz, M., and Venables, B. (2015). gplots: Various R Programming Tools for Plotting Data. R package version 2160 <http://CRANR-projectorg/package=gplots>.