

SUPPLEMENTARY MATERIAL FOR:

The *Lgr5* Intestinal Stem Cell Signature: Robust Expression of Proposed Markers of Quiescent '+4' Cells.

Supplementary Table S1	Affymetrix stem cell signature
Supplementary Table S2	Agilent stem cell signature
Supplementary Table S3	The mRNA stem cell signature
Supplementary Table S4	Genes unique to the “Agilent stem cell signature”
Supplementary Table S5	Genes unique to the “Affymetrix stem cell signature”
Supplementary Table S6	Details about MS experiments
Supplementary Table S7	Identified proteins
Supplementary Table S8	Quantified proteins
Supplementary Table S9	The protein stem cell signature
Supplementary Table S10	Genes in the “mRNA stem cell signature” detected but not enriched in MS
Supplementary Table S11	Proteins unique to the “protein stem cell signature”
Supplementary Table S12	The intestinal stem cell signature
Supplementary Table S13	Probe libraries for single molecule mRNA in-situ
Supplementary Table S14	qPCR primer sequences

Supplementary Figure S1	Workflow
Supplementary Figure S2	Evaluation of protein identification
Supplementary Figure S3	Evaluation of protein quantification
Supplementary Figure S4	Technical and biological reproducibility of proteomic analysis
Supplementary Figure S5	mRNA and protein correlation
Supplementary Figure S6	In situ hybridizations of genes showing a gradient along the crypt axis
Supplementary Figure S7	Bmi1 short term lineage tracing

SUPPLEMENTARY TABLE LEGENDS

Table S1. The Affymetrix stem cell signature. Expression for 20,819 genes was detected from which 379 were found to be significantly enriched in the stem cells (see Materials and Methods)

Table S2. The Agilent stem cell signature. Expression for 13,967 genes was detected from which 291 were found to be significantly enriched in the stem cells (see Materials and Methods)

Table S3. The mRNA stem cell signature. 384 genes were found to be significantly enriched in stem cells in both transcriptomics platforms or significant in one and enriched >1.5-fold in the other.

Table S4. Genes unique to the “Agilent stem cell signature”. Details for the 71 genes found to be unique to the Agilent stem cell signature.

Table S5. Genes unique to the “Affymetrix stem cell signature”. Details for the 54 genes found to be unique to the Affymetrix stem cell signature.

Table S6. Details about MS experiments. MS information regarding the two biological experiments.

Table S7. Identified proteins. After combining the two biological replicates, 7,967 unique proteins groups were identified at a FDR of 1.4%. Uniprot accession numbers were submitted to PANTHER Protein classification system and classified according to their Molecular Function, Biological Process, Cellular Localization and Protein Class (464 proteins could not be annotated). Further details regarding the protein identification can be found in Supplementary Figure S1.

Table S8. Quantified proteins. Overall, 6,077 unique proteins groups were quantified. The average of protein ratios (log₂ scale) between both biological is presented. Sequence Coverage, number of quantified peptides per protein and the Relative Standard Deviations (RSD) are also provided for each separate biological replicate.

Table S9. The protein stem cell signature. Using a cut-off of 1.5-fold consistently in both proteomic replicas, 278 proteins were found to be enriched in the intestinal stem cells.

Table S10. Genes in “mRNA stem cell signature” detected but not enriched in MS. Details for the 27 genes that were subtracted from the final intestinal stem cell signature due to their equal expression on protein level between stem cells and daughter cells as measured by MS.

Table S11. Proteins unique to the “protein stem cell signature”. Details for the 59 proteins found to be unique to the protein stem cell signature.

Table S12. The intestinal stem cell signature. Combining the two transcriptomics data sets and the proteomic data set resulted in the definition of 510 genes showing a significant enrichment in Lgr5^{+ve} stem cells.

SUPPLEMENTARY FIGURES

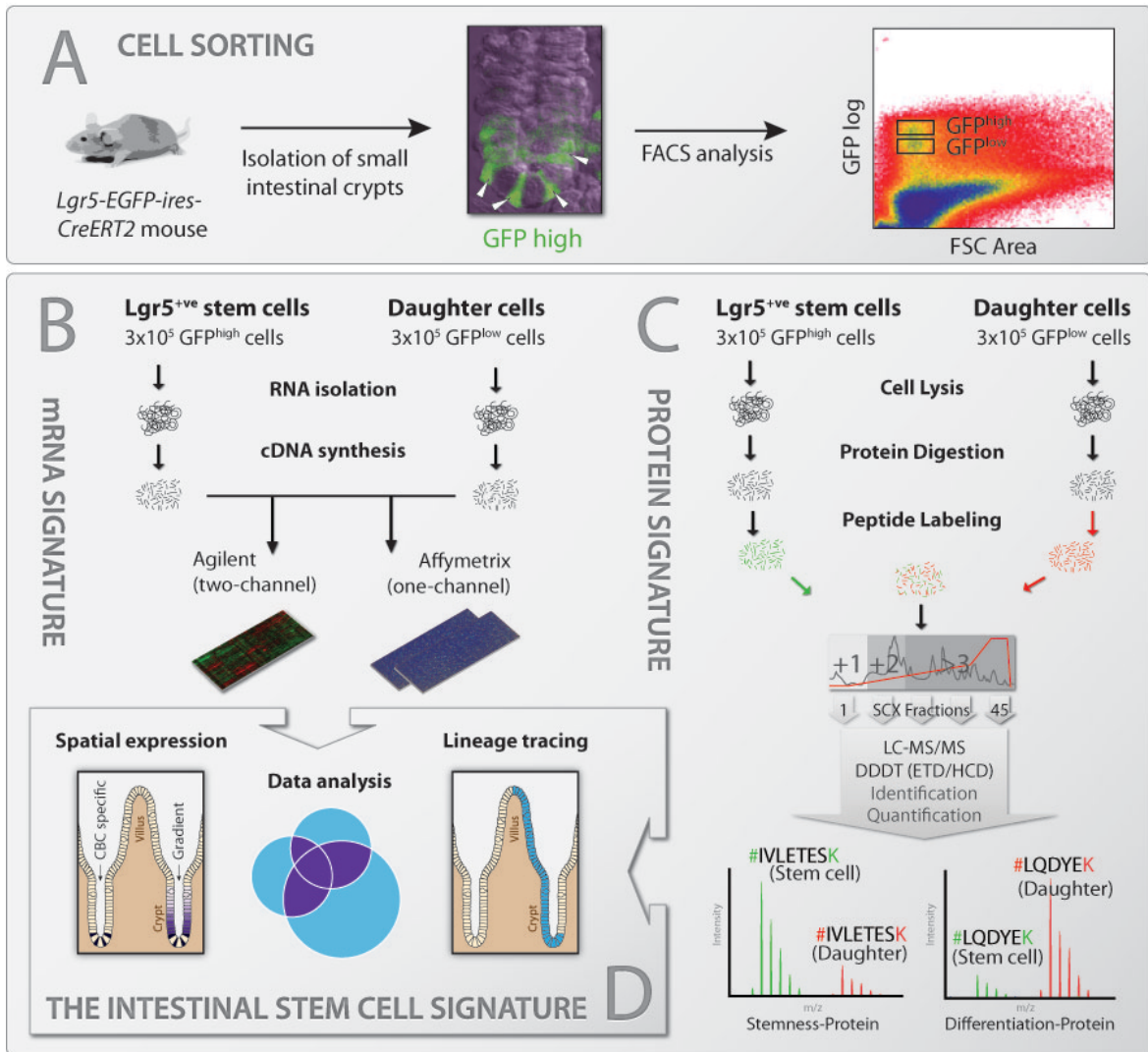


Figure S1. Experimental design. A) *Lgr5*-GFP^{high} (stem cells) and *Lgr5*-GFP^{low} cells (daughter cells) were FACS sorted from small intestine of *Lgr5-EGFP-ires-CreERT2* mice. B) Two independent platforms, i.e. Affymetrix and Agilent, were used for gene expression profiling of the two cell populations C) Using the same cell populations, peptides were differentially labeled, pre-fractionated by using strong cation exchange (SCX) and analyzed by LC-MS/MS. A second biological replica was performed where labels were swapped. D) By combining both approaches, a comprehensive list of genes specifically enriched in the stem cells was obtained. Spatial expression of candidates found by all three methods was examined by RNA in situ hybridization. For one gene found to be specifically expressed in stem cells, *Smoc2*, a knock-in mouse for lineage tracing was generated.

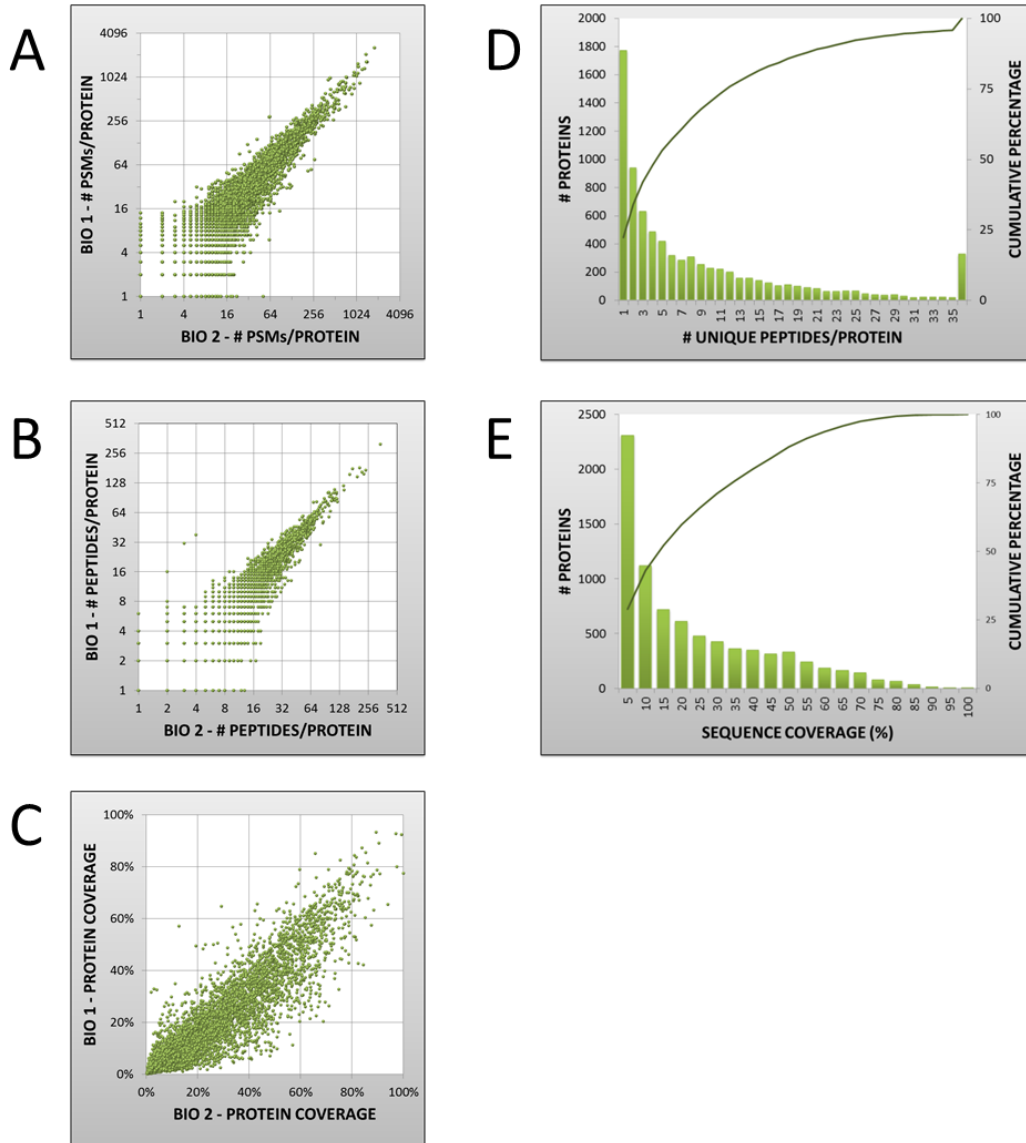


Figure S2. Evaluation of protein identification. The number of PSMs (A), number of peptides (B) and sequence coverage (C) per protein was compared in the two biological experiments and a good correlation was found. In total, 7,967 unique protein groups could be identified (FDR = 1.4 %, Mascot Ion Score>20). More than 75% of the protein groups were identified on the basis of at least two unique peptides (D) and more than 5% sequence coverage (E). Detailed information regarding the identification can be found in [Supplementary Table S5](#).

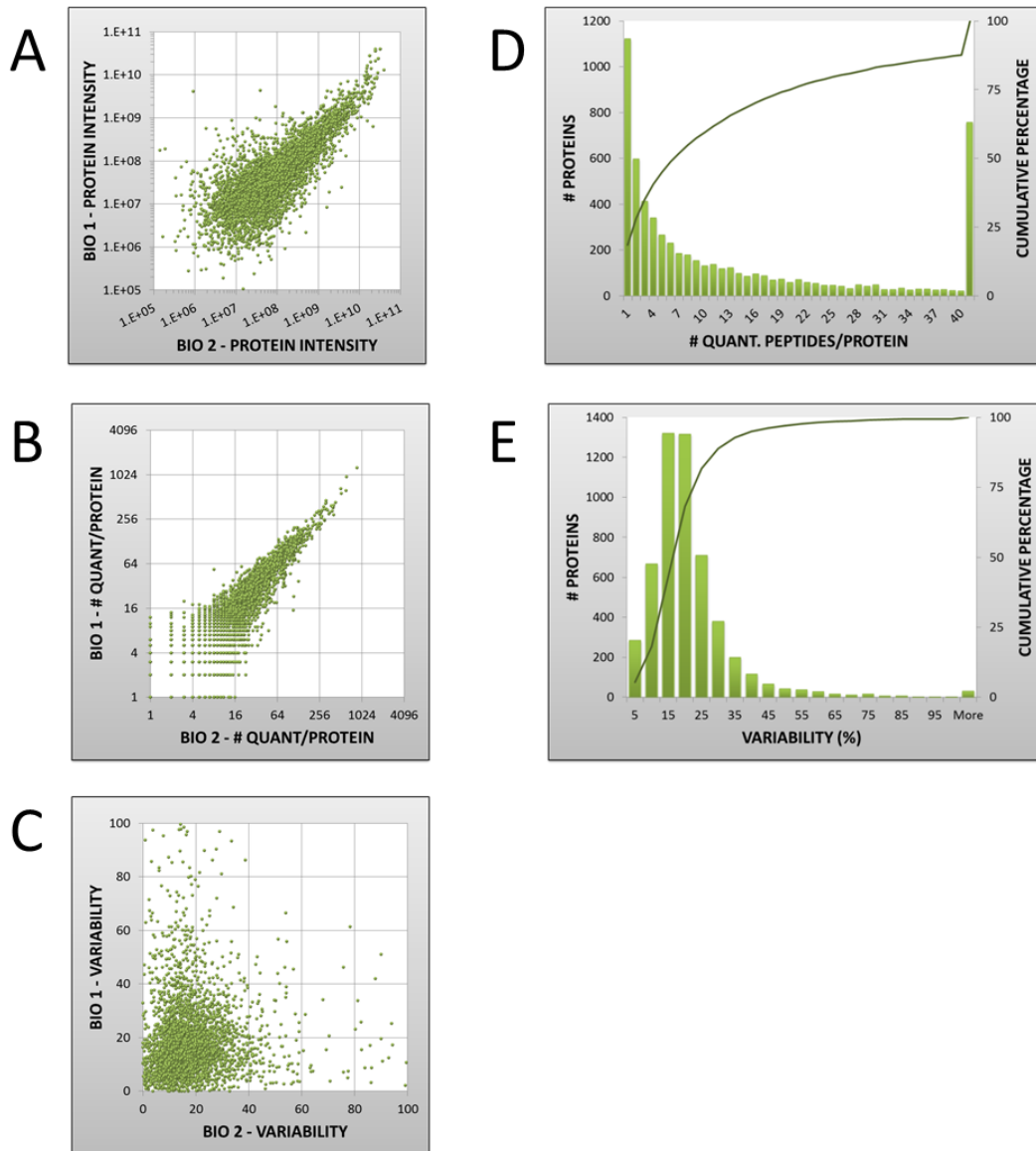


Figure S3. Evaluation of protein quantification. The protein intensity (A) (calculated based on the peptide XICs), number of quantified peptides per protein (B) and the variability (C) (calculated as the relative standard deviation) were compared between the two independent biological experiments and a good correlation was found. This resulted in the robust and precise quantification of 6,077 unique protein groups (4,817 in both replicates). More than 70% of the proteins were quantified on the basis of at least 2 unique peptides (D) and more than 85% showed errors below 35% (E). Detailed information regarding the quantification can be found in [Supplementary Table S5](#).

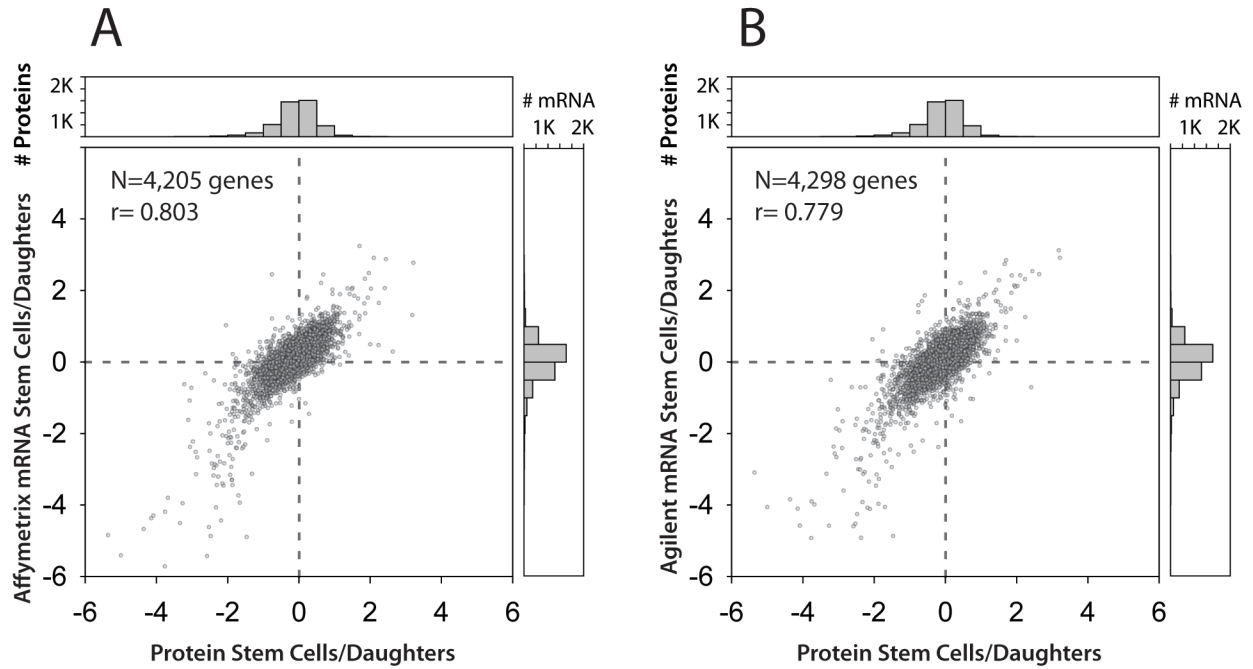


Figure S5. mRNA and protein correlation. Protein ratios were matched to their corresponding mRNA levels (log₂ scale) using the official gene symbols for Affymetrix (A) and Agilent (B) platforms. The number of genes that could be correlated and the Pearson product-moment correlation coefficient are indicated for each case.

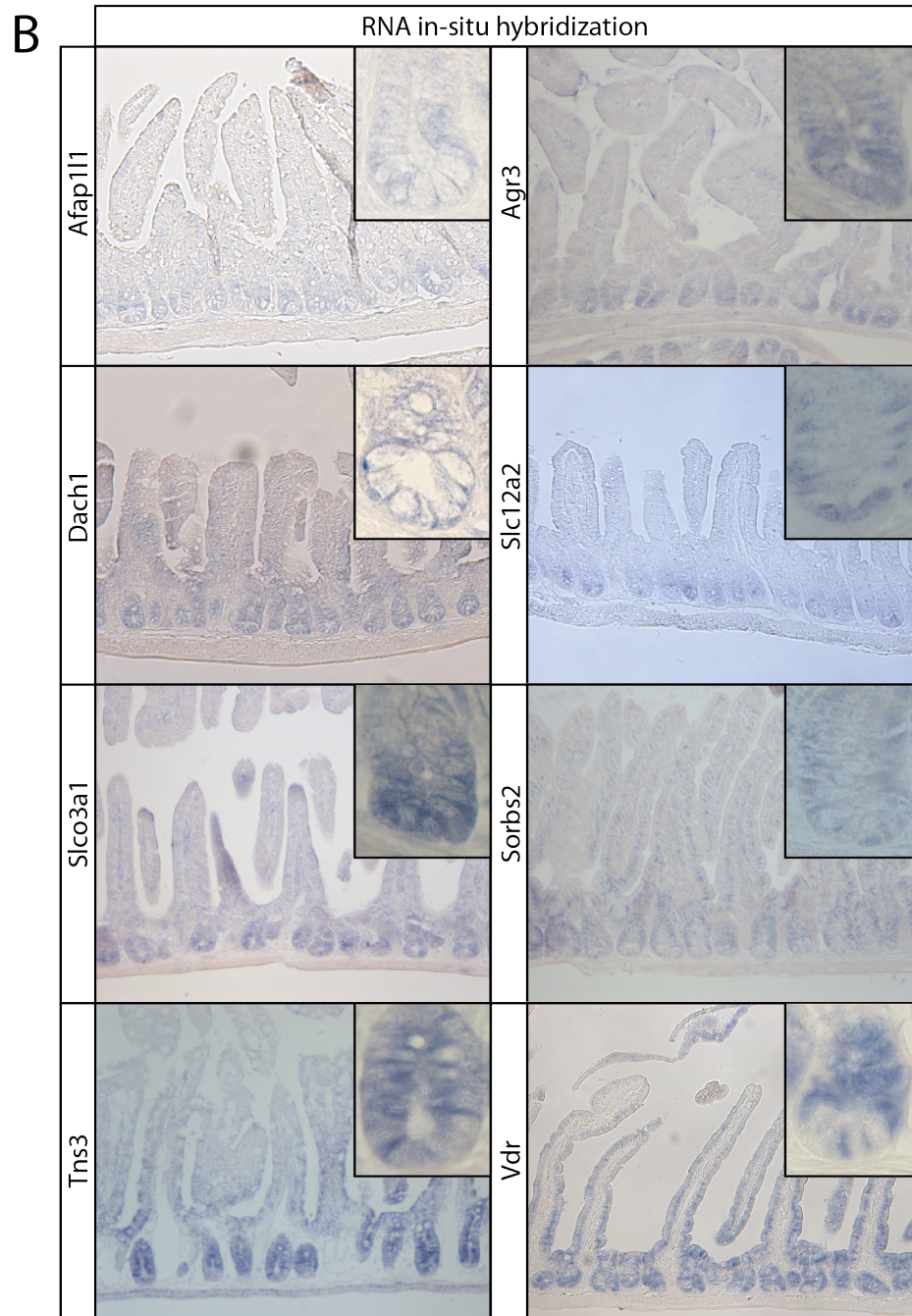
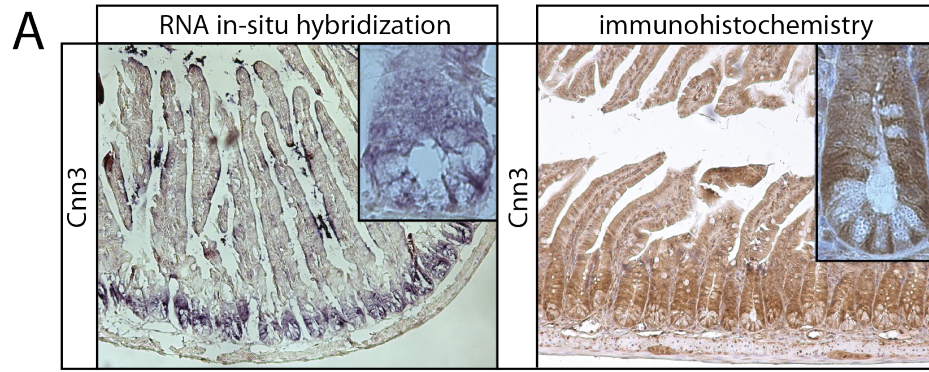


Figure S6.
mRNA in situs of genes showing a gradient along the crypt axis

A mRNA in situ hybridization screen was performed for the 31 signature genes in the central overlap (Figure 2B) to explore their spatial expression pattern. A subset of genes showed a gradual expression from the bottom to the top of the intestinal crypts. For Cnn3 the gradient could also be confirmed on protein level by immunohistochemistry.

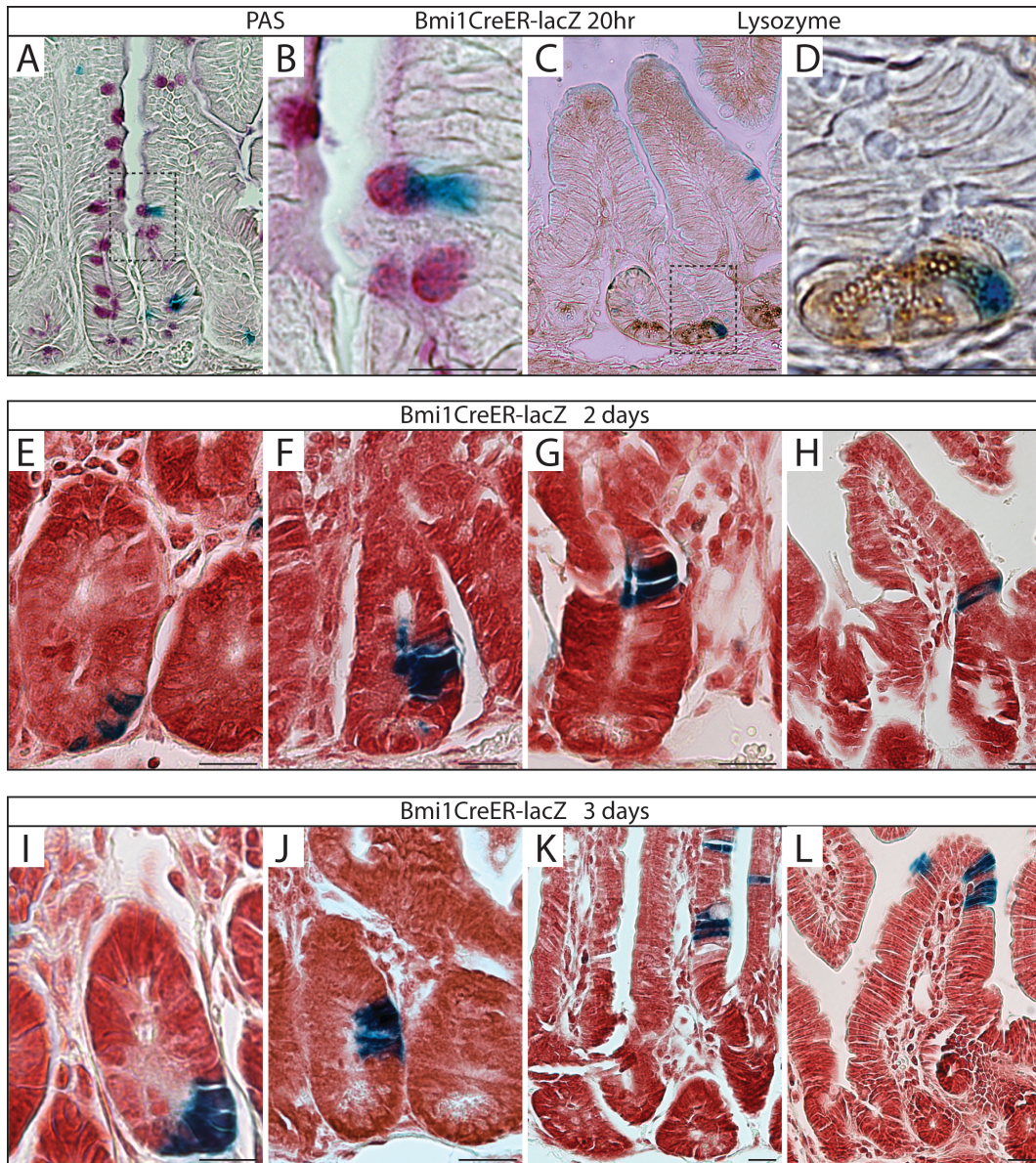


Figure S7. *Bmi1* short term lineage tracing. A-D) LacZ staining of *Bmi1*-CreER/R26R-LacZ mice, 20hrs after induction combined with: A) Periodic acid-Schiff staining for goblet cells. B) Magnification of A showing a LacZ-positive goblet cell. C) Lysozyme staining for Paneth cells. D) Magnification of C showing a LacZ-positive Paneth cell. E-H) LacZ staining of *Bmi1*-CreER/R26R-LacZ mice 2 days after induction. Small trains of cells are visible in the CBC compartment, the TA compartment and the villus compartment. I-L) LacZ staining of *Bmi1*-CreER/R26R-LacZ mice 3 days after induction