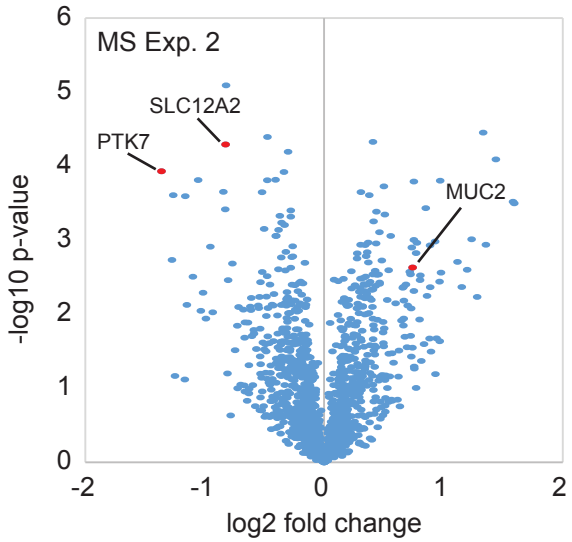
Stem Cell Reports, Volume *5*

Supplemental Information

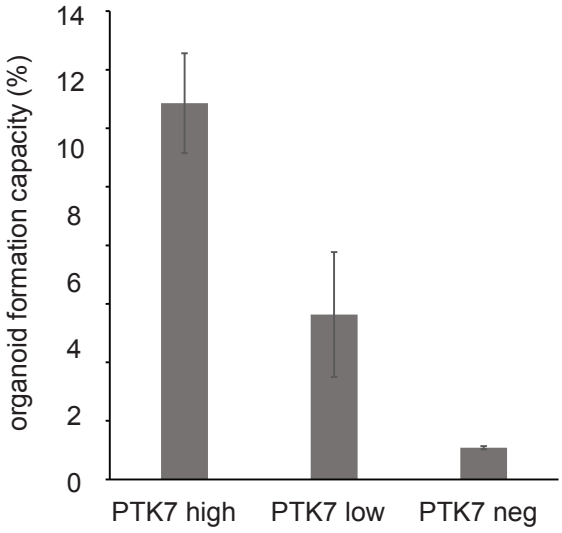
Isolation of Human Colon Stem Cells Using Surface Expression of PTK7

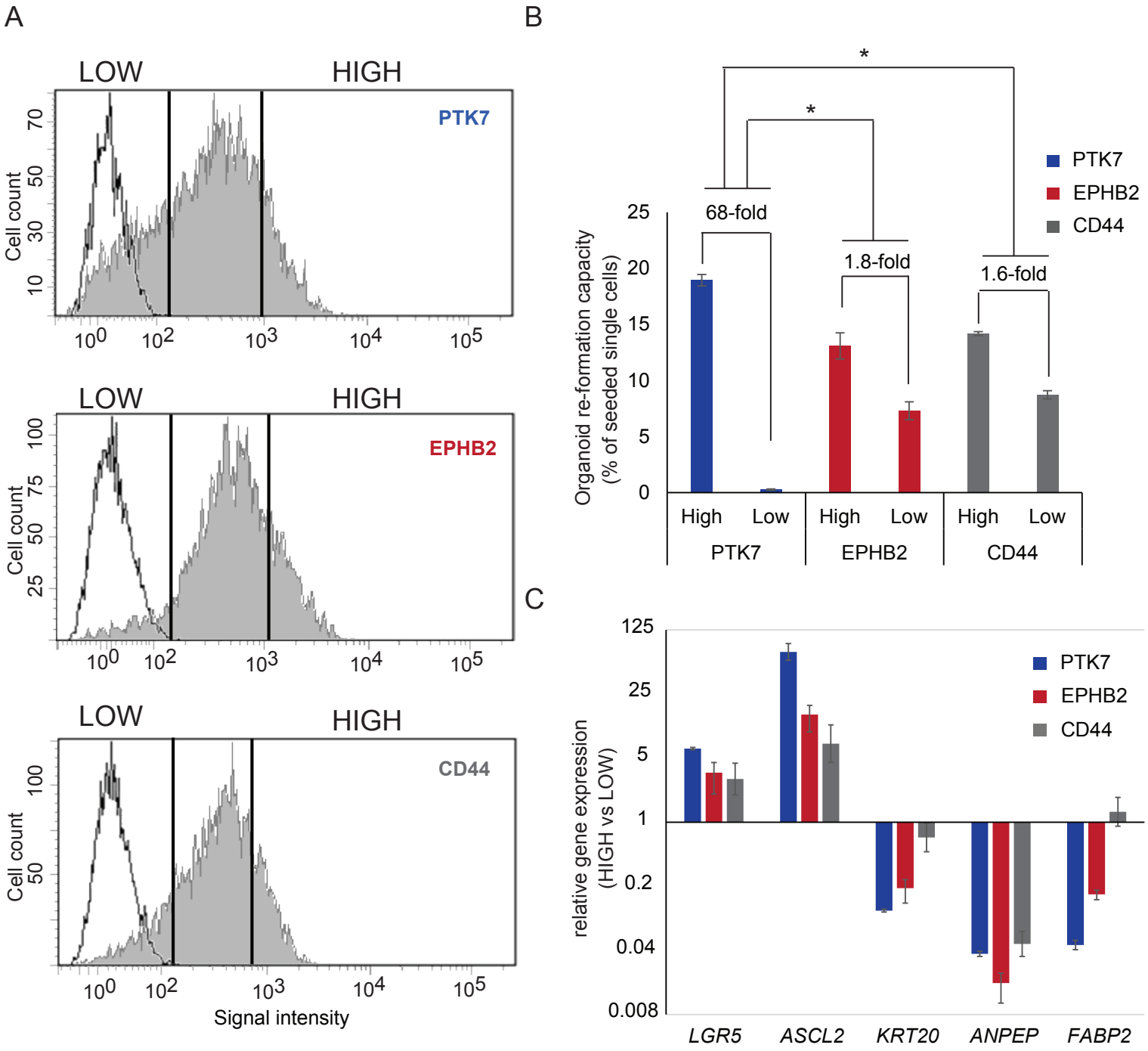
Peter Jung, Christian Sommer, Francisco M. Barriga, Simon J. Buczacki,, Xavier Hernando-Momblona, Marta Sevillano, Miquel Duran-Frigola, Patrick Aloy, Matthias Selbach, Douglas J. Winton, and Eduard Batlle

A

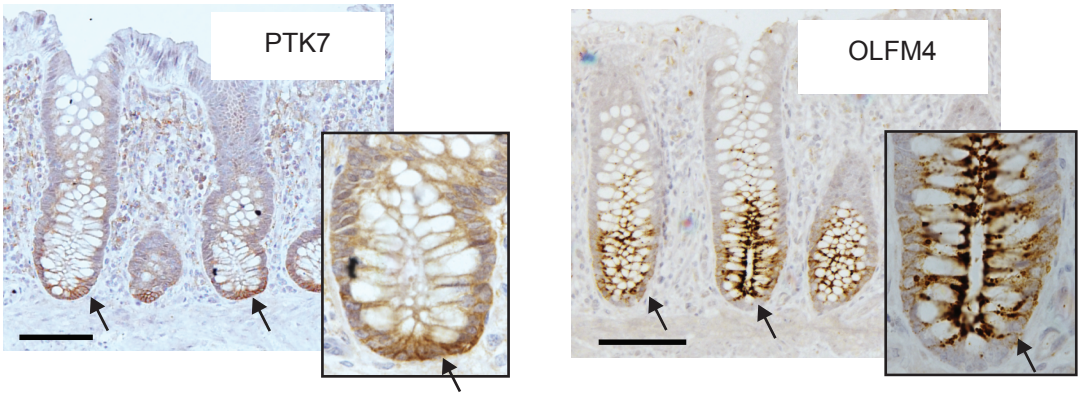

MS Exp. 2
SLC12A2
PTK7
MUC2
-log10 p-value
log2 fold change
6
5
4
3
2
1
0
-2
-1
0
1
2

B

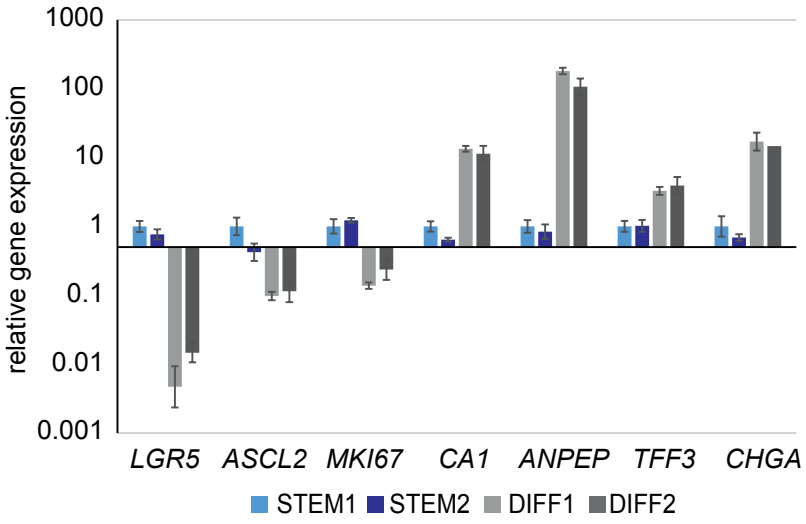

organoid formation capacity (%)
14
12
10
8
6
4
2
0
PTK7 high
PTK7 low
PTK7 neg


A
LOW
HIGH
PTK7
Cell count
EPHB2
CD44
Signal intensity
B
Organoid re-formation capacity
(% of seeded single cells)
68-fold
1.8-fold
1.6-fold
PTK7
EPHB2
CD44
High
Low
C
relative gene expression
(HIGH vs LOW)
LGR5
ASCL2
KRT20
ANPEP
FABP2

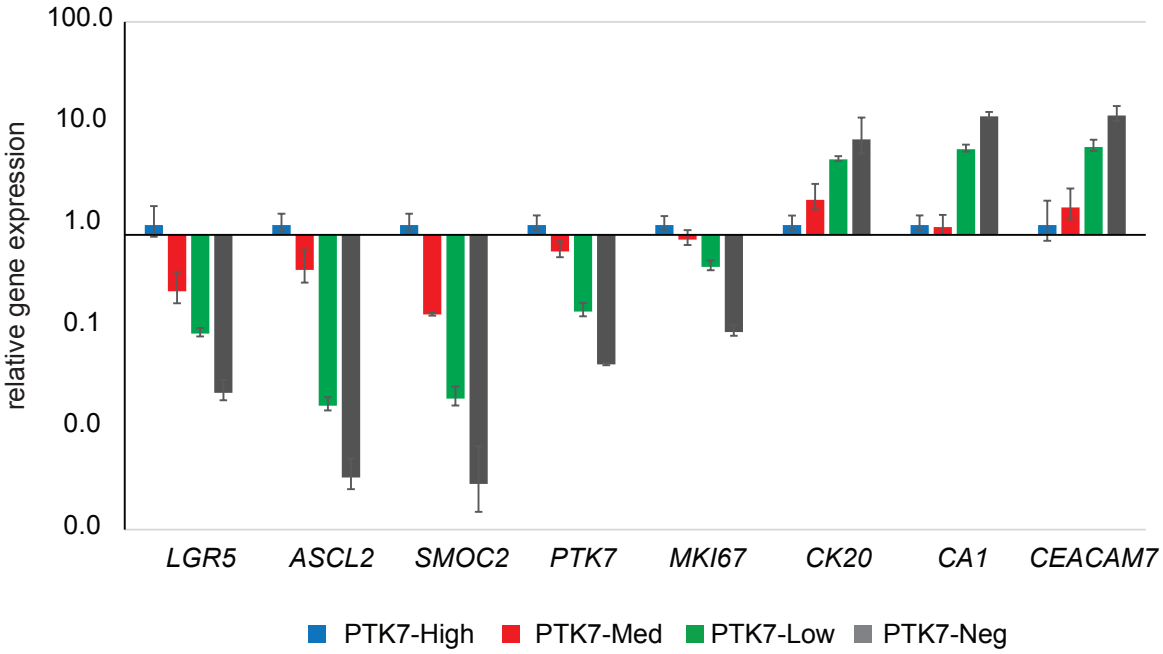
A


PTK7
OLFM4

B


1000
100
10
1
0.1
0.01
0.001
relative gene expression
LGR5
ASCL2
MKI67
CA1
ANPEP
TFF3
CHGA
STEM1
STEM2
DIFF1
DIFF2

C


100.0
10.0
1.0
0.1
0.0
0.0
relative gene expression
LGR5
ASCL2
SMOC2
PTK7
MKI67
CK20
CA1
CEACAM7
PTK7-High
PTK7-Med
PTK7-Low
PTK7-Neg

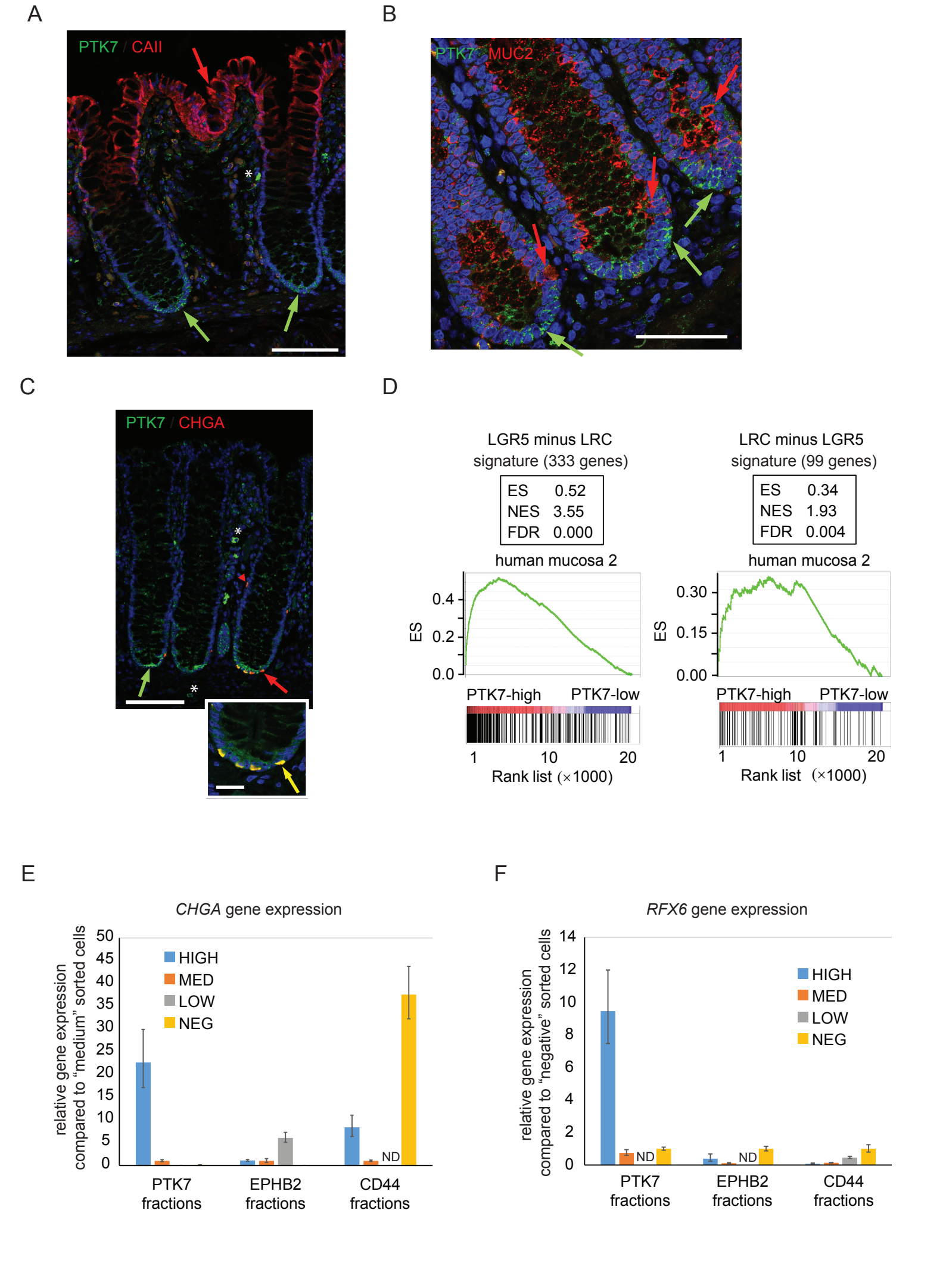

A
PTK7 / CAII
B
PTK7 / MUC2
C
PTK7 / CHGA
D
LGR5 minus LRC signature (333 genes)
ES 0.52
NES 3.55
FDR 0.000
human mucosa 2
ES
PTK7-high
PTK7-low
Rank list (×1000)
LRC minus LGR5 signature (99 genes)
ES 0.34
NES 1.93
FDR 0.004
human mucosa 2
E
CHGA gene expression
relative gene expression compared to "medium" sorted cells
HIGH
MED
LOW
NEG
ND
PTK7 fractions
EPHB2 fractions
CD44 fractions
F
RFX6 gene expression
relative gene expression compared to "negative" sorted cells
HIGH
MED
LOW
NEG
ND
PTK7 fractions
EPHB2 fractions
CD44 fractions

Table S2: Classification of proteins identified by mass spectrometry, related to Figure 1

Total rows	559		
HMPAS membrane protein annotations			
Concept	**Counts**	**P-value Rev Prot**	
Membrane protein	322	1.59E-13	
UniProt upper term subcellular location			
Concept	**Counts**	**P-value Rev Prot**	**Q-value Rev Prot**
Cell envelope	79	0.424	0.514
Cell junction	17	0.0413	0.0983
Cell projection	12	0.478	0.541
Chromosome	10	0.385	0.514
Cytoplasm	182	7.85E-10	1.33E-08
Cytoplasmic vesicle	12	0.0721	0.136
Endoplasmic reticulum	52	7.13E-08	5.98E-07
Endoplasmic reticulum	3	0.0463	0.0983
Endosome	16	0.00229	0.00742
Golgi apparatus	15	0.635	0.675
Lysosome	7	0.146	0.23
Melanosome	6	0.00218	0.00742
Membrane	226	0.00262	0.00742
Mitochondrion	60	1.06E-07	5.98E-07
Nucleus	147	0.149	0.23
Peroxisome	3	0.411	0.514
Secreted	38	0.965	0.965

SUPPLEMENTARY FIGURE LEGENDS

Figure S1. PTK7 is enriched in *ex vivo* cultured hCoSCs and specifies organoid cells with highest self-renewal capacity, related to Figure 1

A. Volcano plot of statistical significance (-Log10 p-value) against fold-change (log2 fold-change) between proteins derived from human colonic organoid cultures maintained in either stem (WREN) media or organoids switched to differentiation media (EN) for 72 hrs. Negative fold-changes represent stem cell-enriched proteins. Note that PTK7 shows strong enrichment in stem organoid cells. Shown are data from an initial mass spectrometry experiment (MS Exp 2) performed as biological triplicate but with lower cell numbers (see Table S1 for detailed results)

B. Quantification of organoid re-forming capacity of sorted cells in Figure 1C. (n = 2 biological replicates, error bars indicate standard error of the mean (SEM).

Figure S2. Comparison of PTK7 to other stem cell markers, related to Figure 2

A. Human colonic organoids (derived from complete primary human colonic crypts) maintained in WREN media were disaggregated and stained against the indicated surface markers by using APC fluorophore-conjugated antibodies. Via flow cytometry, 9-10% of cells showing the brightest staining intensity (HIGH, right sector) and cells showing weakest intensity set to the boundaries of antibody control (LOW, left sector) were isolated for re-plating and RNA isolation experiments. Note that PTK7 shows a broad distribution of staining intensity while EPHB2 and CD44 surface abundance of human colonic organoid cells is more homogenous. Control (transparent profile) and specific staining (grey profile) are shown in each histogram.

B. Organoid formation capacity of cellular fractions isolated via flow cytometry as shown in Figure S2A. Each cellular fraction was seeded 2 times (2 independent 3-D culture systems, biological replicate) and frequency of organoid-forming cells was assessed 8 days after seeding. Error bars indicate standard error of the mean (SEM). Note strongest enrichment of organoid-forming units using PTK7 surface marker (High vs. Low). * indicates significance: t-test p-value < 0.05.

C. Quantitative real-time PCR (qRT-PCR) analysis of stem cell and differentiation marker gene expression on organoid cells staining high compared to low for the indicated markers according to Figure S2A. Gene expression was assessed three times (technical replicate) and error bars indicate standard deviation. Note stronger enrichment of stem cell markers and de-enrichment of differentiation markers *KRT20*

and *FABP2* using PTK7 surface abundance as a read-out when compared to EPHB2 or CD44 surface marker abundance.

Figure S3. PTK7 specifies self-renewing, multi-potent stem cells of the human colon, related to Figure 3

A. Immunohistochemistry analysis of PTK7 (left panel) and Olfactomedin 4 (OLFM4), (right panel) expression on serial sections of normal human colonic mucosa. Black arrows indicate specific staining. Scale Bars, 50 µm.

B. Quantitative real-time PCR (qRT-PCR) analyses of *in vitro* propagated human colonic organoids derived from PTK7-high primary single cells. Data are represented as relative changes of gene expression between organoids cultured in stem cell (WREN) medium compared to organoids maintained in differentiation (EN) medium for 5 days. Experiment was performed as biological duplicate and data are shown separately (n=2, STEM1, 2; DIFF1, 2). Each sample was analyzed 3 times (n=3, technical replicate) and error bars indicate standard deviation.

C. qRT-PCR analysis of indicated gene expression on primary human colon epithelial cells (cells derived from a different individual than cells analyzed in Figure 3H) purified by FACS according to different PTK7 surface levels. Error bars indicate standard deviation (n = 3, technical replicate). As for the experiment performed in Figure 3E-H, only epithelial cells were included ($EpCAM^+$) and non-epithelial cells were excluded ($CD11^-/CD31^-/CD45^-$) from the analysis by selective FACS sorting.

Figure S4. PTK7+ human CoSCs show mixed features of canonical CoSCs and LRCs, related to Figure 4

A. Immunofluorescence staining on formalin-fixed paraffin sections (FFPS) of DAPI (blue), PTK7 (green), and Carbonic Anhydrase II (red). Scale bar: 100 µm. Arrows indicate specific staining. * indicates specific stromal PTK7 staining.

B. Immunofluorescence staining on formalin-fixed paraffin sections (FFPS) of DAPI (blue), PTK7 (green), and Muc2 II (red). Scale bar: 50 µm. Arrows indicate specific staining.

C. Immunofluorescence staining on formalin-fixed paraffin sections (FFPS) of DAPI (blue), PTK7 (green), and Chromogranin A (CHGA) (red). Scale bars: 100 µm (20 µm in zoomed-in image). Arrows indicate specific staining. Red arrow head: single positive

CHGA cell (PTK7-negative). Yellow: Overlaying staining of PTK7 (green) and CHGA (red). * indicates specific stromal PTK7 staining.

D. GSEA analysis (normal mucosa 2) was performed using the specific canonical (LGR5 minus LRC, left panel) and specific label-retaining (LRC minus LGR5, right panel) cell signatures (see Supplemental Experimental Procedures for details) for our PTK7 hCoSC microarray datasets (Table S3). ES: enrichment score, NES: normalized enrichment score, FDR: false discovery rate.

E. Quantitative real-time PCR (qRT-PCR) analysis of chromogranin (*CHGA*) gene expression on human organoid cells HIGH, MED, or LOW for the indicated marker surface abundance. Gene expression was assessed three times (technical replicate) and error bars indicate standard deviation. Similar results were obtained by analyzing colonic epithelial cells from an independent individual (data not shown). ND: not detectable

F. qRT-PCR analysis of *RFX6* gene expression on human organoid cells HIGH, MED, or LOW for the indicated marker surface abundance. Gene expression was assessed three times (technical replicate) and error bars indicate standard deviation. Similar results were obtained by analyzing colonic epithelial cells from an independent individual (data not shown). ND: not detectable.

Table S1. Proteomic data of differentially expressed proteins between WREN and EN media cultured human colonic organoids, related to Figure 1

Quantification of membrane-enriched protein fractions by mass spectrometry (qLC-MS-MS). Protein fractions were derived from *in vitro* grown human colonic organoids cultured under either stem cell-supporting (STEM) or differentiation-supporting (DIFF) media conditions (see Online Supplementary Methods). The experiment was performed as biological triplicates. “MS/MS count” equals number of identified peptides assigned to each protein. Only proteins showing a significant change between the two conditions according to the t-test (t-test significance “+”) were considered for statistical evaluation (see Supplementary Table 2). T-test difference is given in log2 values. A negative T-test difference implies protein enrichment in stem cells, a positive T-test difference labels proteins enriched in early (72 hrs) differentiated organoid cells. Membrane proteins were annotated using the HMPAS annotation system[1] and annotation derived from the UniProt database is shown for each identified protein.

Below this table, results are shown of an initial small-scale mass spectrometry experiment to identify proteins enriched in *in vitro* human colonic organoids cultured

under stem cell conditions. A negative T-test difference indicates stem organoid cell-enriched proteins.

Table S2. Classification of proteins identified by mass spectrometry, related to Figure 1

Proteins identified to be significantly enriched in stem or differentiated human colonic organoid cells are enriched in membrane proteins (the Uniprot subcellular location category “membrane” is significantly over-represented when compared to the human reviewed proteome, p-value: 0.00262, FDR q-value: 0.00742). Similarly, we observed an enrichment of membrane-interacting proteins according to the Human Membrane Protein Analysis System (HMPAS[1], fcode.kaist.ac.kr/hmpas), p-value: $1.59 \cdot 10^{-13}$.

Table S3. Microarray analysis of gene expression in different fractions of the human colonic epithelia according to their PTK7 surface abundance, related to Figure 3

Genome-wide expression profiling of normal human mucosa epithelial fractions sorted according to their PTK7-surface abundance (GEO database link: http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?token=wnqbwmsctxuvjuj&acc=GSE68340). Combined data from two mucosa samples (biological duplicate) are depicted. Shown are differentially expressed genes (PTK7 high versus low) showing an average fold change (fc) > 2 and p-value<0.05. Next to it, differentially expressed genes of PTK7 high versus low hCoSCs of the two independent samples (human mucosa 1 and 2) are shown in individual columns (Log2 fold changes are shown).

Table S4. Specific LGR5 and LRC stem cell signatures, related to Figure 4

Generation of the Label-retaining cell signature (LRC signature) from data published by Buczacki et al., Nature, 2013, and generation of the specific LGR5 stem cell signature (LRC genes removed) from data published by Muñoz, EMBO Journal, 2012. For detailed information, see Supplementary Experimental Procedures.

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

Quantitative Mass Spectrometry Analysis (q-LC-MS-MS)

Samples were processed by methanol-chloroform extraction, reduced, alkylated and digested with LysC and trypsin using standard protocols. After offline desalting, peptides were analyzed by LC-MS/MS on a Proxeon EASY-nLC 1000 system connected to a Q Excative Plus mass spectrometer (Thermo Scientific). Chromatography was performed using a 240 minute acetonitrile gradient on a 2000 mm long monolithic column kindly provided by Yasushi Ishihama (Kyoto University). The instrument was operated in the data dependent mode with the following settings for the full scans: resolution 70,000, AGC target value 3E6, maximum injection time 20 ms. The following settings were chosen for the MS2 scans: resolution 17,500, AGC target value 1E6, maximum injection time 60 ms. Raw files were processed with MaxQuant (version 1.5.1.2) for label free quantification (LFQ) using standard settings.

Enrichment and purification of membrane protein-fractions derived from 3-D human colonic organoid cultures

Isolation of membrane-enriched protein fractions from *in vitro* cultured organoids was performed as biological triplicate (3 stem cell organoid cultures versus 3 differentiated cultures). Around 10^7 human colonic organoid cells per sample, cultured in stem cell media (WREN) or differentiation media (EN) (differentiation was performed for 72 hrs only in order to yield comparable cell numbers for quantitative mass spectrometry analyses), were released from matrigel by using Cell Recovery Solution (1 ml / 50 µl drop)(Corning, NY, USA), and cells were incubated with Dispase (0.8 mg/ml, Gibco, Life Technologies) in PBS for 15 minutes at room temperature to liberate cells from extracellular matrix components. After this, membrane-enriched protein fractions were isolated from samples using the QProteome Plasma Membrane Isolation Kit (Qiagen, Hilden, Germany) according to the manufactures protocol except the following protocol adjustments: Completion of cell lysis was done by 25 syringe strokes using a G27 needle. Elution of enriched membrane proteins from Strep Tactin Magnetic beads was achieved by incubation with 8 M Urea/25 mM Hepes, pH7.0 for 10 min at RT. Elution was performed 4 times and the eluate fractions were combined.

For further concentration of eluted proteins, 4 volumes of ice-cold, pure acetone was added to samples, mixed carefully, and after incubation for 30 minutes on ice samples were centrifuged at 12,000×g for 15 min at 4°C. After washing with ice-cold acetone, precipitated pellets were air-dried and stored at -80°C.

Isolation of human colonic crypts and crypt-derived single cell preparation

Biological samples were obtained from individuals treated at the Hospital del Mar (Barcelona, Spain) or from Hospital Clinic (Barcelona, Spain) under informed consent. Experiments were approved by the ethics committee of IRB/Hospital Clinic (project ERC-208488/CRCprogramme).

Samples were received from individuals suffering from colorectal cancer disease with exception of experiments shown in Figure S2 and S4 with tumor-free individuals suffering from diverticulitis.

Muscle layer and sub-mucosa were removed from human fresh colectomy specimens and colonic mucosa was incubated with a mixture of antibiotics (Normocin (Invivogen, San Diego, California 92121, USA), Antibiotic-Antimycotic (Thermo Fisher Scientific, Waltham, Massachusetts, USA)) for 15 min at room temperature. Next, tissue was cut into small pieces and incubated twice in 10 mM DTT (Sigma St. Louis, MO 63103, USA) in PBS for 5 min at room temperature. Samples were then transferred to 8 mM EDTA in PBS and slowly rotated for 60 min at 4ºC. Supernatant was replaced by fresh PBS, and vigorous shaking of the sample yielded supernatants enriched in colonic crypts. Fetal bovine serum (FBS, Sigma) was added to a final concentration of 5% and fractions were centrifuged at 40×g in order to remove single cells. This washing procedure was repeated 3 times with Advanced DMEM/F12 (ADF, Invitrogen) medium supplemented with 2 mM GlutaMax (Invitrogen), 10 mM HEPES (Sigma), and 5% FBS (Sigma) (Washing buffer: WB). Purified crypts were either directly cultured (see below) or single cell suspensions were obtained as follows. Crypts were re-suspended in Disaggregation Medium (ADF, Glutamax, 10 mM HEPES, N-2, B-27 without retinoic acid (Invitrogen), 10 μM Y-27632 (Calbiochem, Gibbstown, NJ, USA) 2.5 μM PGE2 (Sigma), 0.4 mg/ml Dispase (354235, BD Biosciences, Two Oak Park, Bedford, MA, USA), and incubated for 15 min at RT with occasional re-suspension using a P1000 pipette. After this, crypt suspension was gently syringed using a 1.2 mm needle until single-cell enriched population was observed microscopically. Finally, cells were sequentially passed through 100, 70 and 40 μm mesh filters (BD Biosciences) and used for further analyses.

Immuno-labelling of human colonic crypt cells for Flow cytometry-assisted cell sorting (FACS)

Human colonic crypt cells were re-suspended in staining buffer (ADF, Glutamax, 10 mM HEPES, 5% FBS, 10 μM Y-27632). 10^6 cells/250 μl were incubated with either mouse anti-human PTK7 antibody coupled to APC fluorophore (dilution: 1:20, clone 188B, Miltenyi, Bergisch Gladbach, Germany, Cat. No. 130-091-366), anti-human/mouse EPHB2 antibody coupled to APC fluorophore (1.1 μg/ml, clone 2H9, BD Pharmingen, Cat. No.: 564699), or anti-human CD44 antibody coupled to APC fluorphore (dilution 1:100, BD Pharmingen, Cat. No. 559942). Control antibody (IgG from myeloma, M5284, Sigma, Cat.No. M5284) was coupled to Allophycocyanin (APC) using the Allophycocyanin (APC) conjugation kit (PJ25K, PROzyme, Hayward, CA, USA) and used to exclude non-specific staining signal and to define the negative cell fraction as has been also described previously[2].

At the same time, cells were stained using anti-Human EPCAM/TROP1 FITC-conjugated (0.2 μg/ml, R&D Systems, Minneapolis, USA, Cat.No. FAB9601F), anti-CD11-PE (Cat.No. 130-091-240), anti-CD31-PE (Cat.No.: 130-092-653), anti-CD45-PE (Cat.No.: 130-080-201) (dilution 1:100 each, Miltenyi Biotech) for 45 min on ice. After 2 washes with staining buffer, 4', 6-diamidino-2-phenylindole (DAPI, 1 μg/ml) was added and stained cells were sorted using a FACS Aria 2.0 (BD Biosciences, Two Oak Park, Bedford, MA, USA).

To obtain the four intestinal crypt cell populations, dead cells and debris were discarded by removing the DAPI+ subpopulation. Hematopoietic cells were excluded by removing CD11+/CD31+/CD45+ subpopulation. Epithelial cells were included by selecting for EpCAM+ staining and then further selected according to their differential PTK7 surface abundance. The brightest 4-5% PTK7+ cells were sorted as the PTK7-high fraction. The PTK7-medium population comprised the 12% PTK7+ cells adjacent to the PTK7-high population. The PTK7-low fraction was considered as the 18-19% adjacent to the PTK7-medium population. The PTK7-negative subpopulation did not stain for PTK7. Finally, these four fractions were further gated according to their forward scatter (FSC-A) to exclude purification of cell aggregates. EPHB2 and CD44 high sorted sub-populations were defined in the same way as the 4-5% brightest cells (data not shown).

***In vitro* culture of PTK7+ single cell-derived colonic organoids and recovery of 3-dimensional grown organoids for propagation or flow cytometry analysis.**

2000-3000 sorted cells were mixed with 50 μl matrigel and plated on pre-warmed 24-well culture dishes. After solidification (15 min at 37°C cell culture incubator), cells were overlaid with WREN medium (Wnt3a-conditioned medium and ADF 50:50, Glutamax, 10 mM HEPES, N-2 (1×), B-27 without retinoic acid (1×), 10 mM Nicotinamide, 1 mM N-Acetyl-L-cysteine, 1 μg/ml RSPO1 (in-house produced, 50 ng/ml human EGF (Invitrogen), 100 ng/ml human Noggin (Peprotech, Rocky Hill, NJ, USA), 1 μg/ml Gastrin (Sigma), and 0.05 μM PGE2 (Sigma), 0.1 μM A83-01 (Stemgent, Cambridge, MA, USA), 10 μM p38 inhibitor SB202190 (Sigma), 10 μM Y27632 (Sigma). Medium was replaced with fresh WREN medium ("Stem" medium) every other day.
For differentiation of colonic spheres, cells cultured in stem cell media were washed three times with ADF and cultured in crypt differentiation medium (EN) which did neither contain PGE2, Wnt3a, RSPO1, Nicotinamide, nor SB202190.
For further analysis, re-seeding, or FACS sorting, organoid cultures embedded in matrigel were washed with PBS, overlaid with Cell Recovery Solution (1 ml / 50 μl drop)(Corning, NY, USA), carefully mixed with a P1000 pipette, and transferred to a 15 ml reaction tube on ice for 45 minutes. Released intact organoids were precipitated and separated from cellular debris by low-speed centrifugation at 80×g at 4°C. After washing with ADF+HEPES+Glutamax, organoids were re-suspended in Disaggregation media (see above) and carefully syringed using a 1.2 mm (18G) needle. Single cell suspension was filtered using a 40 μm mesh filter and used for further analysis by flow cytometry assisted cell sorting.
Limiting dilution assays for assessment of organoid-forming units was done as follows. Different fractions of organoid cells were sorted according to their PTK7 surface abundance and various cell numbers (100, 500, and 2000 sorted events) were seeded in matrigel after dilution of sorted cell fractions. 24 hours after seeding, the amount of actually seeded/embedded cells in each drop was carefully assessed and this number was set to 100%. The percentage of these cells forming a bona-fide organoid was assessed microscopically, growth was documented by microscopy (data not shown) and organoid forming units were calculated by using the ELDA online tool[3].

To compare the growth efficiency of PTK7, CD44, EPHB2, and CD133 high versus low *ex vivo* cultured organoid cells under WREN culture conditions, organoids were dis-aggregated to obtain single cells and the cell suspension was stained with either APC-fluorophore coupled antibodies to PTK7 (Miltenyi, 1:20, Cat.No: 130-091-366), anti-

CD44 (BD Pharmingen, 1:100, Cat. No. 559942), anti-EPHB2 (BD Pharmingen, 1.1 ug/ml, No.: 564699), or anti CD133 (AC133, Miltenyi, 1:100, Cat.No. 130-098-829). Autofluorescence of viable (DAPI-stained) human organoid cells was measured in the APC-channel (670 nm/14 filter of the FACS Aria Cytometer) in order to define the "negative", not-staining cell population for FACS gating.

After cell sorting of different fractions, equal cell numbers for each sorted sub-population were seeded in 3-D matrigel matrix (BD Biosciences), number of viable cells was carefully assessed 24 hrs after cell seeding by microscopy, and organoid forming capacity was assessed 8 days after plating single cells.

Immunohistochemistry and Immunofluorescence labeling on Formalin fixed paraffin-embedded sections (FFPS)

Immunohistochemical staining was carried out using 3-µm tissue sections according to standard procedures. Briefly, after antigen retrieval (Citrate, pH6.0 in case of PTK7 staining), samples were blocked using Peroxidase-Blocking Solution (Dako, S202386) for 10 min at room temperature, and primary antibodies were then incubated with samples overnight at 4ºC. Slides were washed 3 times with EnVision FLEX Wash Buffer (Dako, Ref: K800721) for 5 minutes, and the corresponding secondary antibody was incubated with the sample for 30 min at room temperature. Samples were developed using 3,3′-diaminobenzidine (DAB) (DAKO), counterstained with hematoxylin and mounted. Antibodies used were against PTK7 (1:100, Sigma, Cat.No: HPA003222), EPHB2 (1:100, R&D Systems, Cat.No. AF467), KRT20 (1:200, DAKO, Cat.No: M7019), FABP1 (1:500, Sigma, Ref: HPA028275), and OLFM4 (1:500, Abcam, Cat.No: ab85046).

Immunofluorescence staining on FFPE was carried out using 3-µm tissue sections according to standard procedures. Antigen-retrieval was achieved using PTLink pH9 (DAKO) and samples were blocked using Peroxidase-Blocking Solution (Dako, S202386) for 10 min at room temperature and 10% normal donkey serum (Jackson ImmunoResearch Laboratories) for 20 min at room temperature. After incubation with primary antibodies in 1% bovine serum albumin (BSA) over night at 4ºC, slides were washed 3 times with EnVision FLEX Wash Buffer (Dako, K800721) for 5 minutes, and the corresponding secondary antibodies coupled to fluorescent dyes were incubated with the sample for 1 hour at room temperature in the dark. After washing, samples were incubated with Sudan Black B in order to reduce autofluorescence. Antibodies used were against PTK7 (1:15, Sigma, Cat.No.: HPA003222), CHGA (1:100, clone E-

5, Santa Cruz, Cat.No: sc-271738), MUC2 (1:100, Abcam, Cat No.: ab118964), and Carbonic Anhydrase 2 (CAII, 1:250, G-2, Santa Cruz, Cat.No: sc-48351). Secondary Antibodies Donkey anti-mouse Alexa 568 (Invitrogen, Cat.No. A10037) and Donkey-anti rabbit Alexa 488 (Invitrogen, Cat.No. A21206) were used at a dilution of 1:500.

Mouse models and induction of YFP-label in small intestinal cells

Cyp1a1-H2B-YFP mice[4] received 3 intraperitoneal injections (80 mg/kg) of β-Naphthoflavone within 24 hours. 10 days after injection, villi were removed and $CD24^+/CD11^-/CD31^-/CD45^-/Lectin^-$ small intestinal crypt cells positive (LRCs) or negative (Non-LRCs) for YFP label-retention were isolated for analysis as described previously [5]. DAPI was added (1 μg/ml) in order to discriminate against dead cells. Finally, fractions were further gated according to their forward scatter (FSC-A) to exclude purification of cell aggregates, and cells were sorted using a FACS Aria 2.0 (BD Biosciences, Two Oak Park, Bedford, MA, USA). Antibodies used were APC-coupled anti-mouse-Cd24 (Biolegend, Cat-No. 101814), PE-Cy7 coupled anti-mouse CD11 (BD Biosciences, Cat-No. 552850), PE-Cy7 coupled anti-mouse CD31 (Abcam, Cat.No. ab46733), and PE-Cy7 coupled anti-mouse CD45 (BD Biosciences, Cat.No:552848).

Pico-Profiling and Gene Expression Analysis

Small cell numbers obtained after Flow-cytometry-assisted cell sorting (1000-2000 cells) of *in vitro* organoid derived cells (Figure 2 and Figure S2) or primary tissue-derived cells used for comparative sorting of EPHB2, PTK7, CD44 cell fractions (Figure S4F,G) were subjected to pico-profiling as published previously[6]. From the derived amplified cDNA, 15-20 ng were used for each qRT-PCR reaction using commercial Taqman Assays (Applied Biosystems, ABI) and data were obtained from an Applied Biosystems StepOne Plus Real-time PCR instrument using standard device settings. The following Taqman assays were used:

Taqman gene expression assays for qRT-PCR analyses	
Gene name (Species)	**ABI assay number**
APCDD1 (human)	Hs_00537787_m1
B2M (human)	Hs_99999907_m1
LGR5 (human)	Hs_00173664_m1
ASCL2 (human)	Hs_00270888_S1
SMOC2 (human)	Hs_0159663_m1
PTK7 (human)	Hs_00897151_m1

CHGA (human)	Hs_00900370_m1
CEACAM7 (human)	Hs_00185152_m1
CK20 (human)	Hs_00300643_m1
ANPEP (human)	Hs_00174265_m1
TPH1 (human)	Hs_00188220_m1
RFX6 (human)	Hs_00543100_m1
CACNA1A (human)	Hs_01579431_m1
B2m (mouse)	Mm_00508106_m1
Lgr5 (mouse)	Mm_00438890_m1
Chga (mouse)	Mm_00514341_m1
Ptk7 (mouse)	Mm_00613362_m1
Krt20 (mouse)	Mm_00508106_m1
Rfx6 (mouse)	Mm_00624115_m1

Statistical Analysis

Basic statistical data were analyzed using MS Excel software. For quantitative RT-PCR data results are shown as means ± standard deviation (technical triplicate). Experiments using primary patient material were repeated three times (biological triplicates) and results were assessed and shown independently.
Statistical analysis of pooled data in Figure 3G was done using a two sample t-test (http://in-silico.net/tools/statistics/ttest) with permutation-based False discovery rate (FDR) p-value smaller than 0.05 used as a cut-off for significance (*).

In case of the Fold-change organoid formation significance in Figure 2B (pooled data from 4 independent culture experiments), a generalized linear model was fitted to the count data with Quasipoisson family and logarithmic link function using the "glm" function in the "stats" package of the R statistical language[10]. The Quasipoisson model was chosen above the Poisson model given the existence of over-dispersion in the data. P-value of the significance of the interaction term was computed through an "F-test" as implemented in the "drop1" function in R[11].

Annotation of Affymetrix microarray was performed using probe set information provided by Affymetrix in its product support web (http://www.affymetrix.com/support/ downloaded in 19/09/2014). Gene differential expression analysis was carried out

using linear models and empirical Bayes methods as implemented in R's 'Limma' package [7]

Mass spectrometry data analysis was performed with Perseus (1.5.0.31) following a previously published approach[8]. Briefly, after removing reverse database hits and potential contaminants, LFQ values provided by MaxQuant were log2 transformed. We required that proteins were at least quantified in three out of the six samples (that is, three replicates of the stem cells and differentiated cells each). Remaining missing values were imputed from a normal distribution (width 0.3, down shift 1.8). Proteins with significantly different abundance between stem cells and differentiated cells were identified using a two sample t-test with a permutation-based FDR of 0.05.

Statistical analyses for Gene Set Enrichment Analysis (GSEA) was performed as originally described in Subramanian A et al.[9]. Enrichment Score (ES) describes the maximum deviation from zero (x-axis of the graph). The normalized enrichment score (NES) accounts for differences in gene set size and therefore, the normalized enrichment scores (NES) can be used to compare analysis results across gene sets. It is defined as the actual ES divided by the mean (ES`s against all permutations of the dataset) The false discovery rate (FDR) is the estimated probability that a gene set with a given NES represents a false positive finding.

Gene signatures and Gene Set Enrichment Analyses (GSEA)

Intestinal stem cell (ISC) signature (*LGR5*-GFP high vs. low or LGR5 signature)

The LGR5 intestinal stem cell signature has been previously reported by Muñoz et al.[10] This signature was defined by comparing $Lgr5^{high}$ with $Lgr5^{low}$ cells in both Affymetrix and Agilent platforms, and selecting those genes enriched in $Lgr5^{high}$ cells in both platforms (Table S4).

Intestinal label-retaining cell (LRC) signature

The LRC signature was extracted from the microarray data published by Buczacki et al. By the use of the Cyp1a1:H2B-YFP allele, the authors identified intestinal label-retaining cells. In order to discriminate from other cell populations, they included the CD24 staining to select for crypt-base cells and stained for UEA Lectin to discriminate Paneth cells. Based on these markers, 3 populations were defined: intestinal LRCs ($CD24^{+}/UEA^{-}/ YFP^{+}$ cells), Paneth cells ($CD24^{+}/UEA^{+}/YFP^{-}$) and cycling lower crypt

cells (LCC) ($CD24^+$/UEA^-/ YFP^-).

To identify LRC-specific genes we compared H2B-YFP+; CD24+ (LRC) versus H2B-YFP-; CD24+ (LCC) cells as originally defined by Winton and colleagues [4] and selected those genes enriched over 3-fold with p-value < 0.05 (LRC vs LCC). This gene set was defined as the LRC signature. Of note, CD24 marks stem cells and TA cells at the bottom third of the crypt. Therefore, this comparison excluded most of the classical ISC markers from the LRC signature (SMOC2, ASCL2, etc.), which are expressed at equal levels in H2B-YFP+; CD24+ and H2B-YFP-; CD24+. We then compared this gene list with that of *LGR5*-GFP-high cells (Venn diagram in Figure 4E). As shown in Figure 4E and Table S4, there were 12 genes common between the LRC and LGR5 signatures, which were excluded in the GSEA analysis of each signature. It is worth mentioning that in this revised analyses, LGR5 passed the cut-off and is included in the H2B-YFP+ signature as well as in the LGR5 signature. Furthermore, this comparison rendered a set of 99 genes upregulated in LRCs but not in cycling LGR5-GFP-high cells. This list included CHGA, a few Paneth Cell markers plus other genes, which we suspect are related to the LRC phenotype. Note that both signatures were enriched in PTK7-high cells from the human colon (Figure 4F and Figure S4D).

Humanization of gene signatures

Both the LRC and LGR5 signatures were humanized by using the source database for homolog translation present in MGI (http://www.informatics.jax.org/). The genes present in both the LGR5 and LRC humanized signatures (12 genes) were removed and the unique genes in LRC and LGR5 signatures were used for all subsequent analysis (Table S4).

Gene set enrichment analysis

We performed GSEA analysis of the LGR5 and LRC signature on PTK7 populations. For this, we generated a ranked file for genes enriched in $PTK7^{high}$ compared to $PTK7^{low}$ cells for two independent patients. GSEA data were generated by using the GSEApreRanked tool of the Broad Institute's GenePattern.[9]

REFERENCES FOR SUPPLEMENTAL EXPERIMENTAL PROCEDURES

1. Kim MS, Yi GS. HMPAS: Human Membrane Protein Analysis System. Proteome Sci 2013;11:S7.

2. Jung P, Sato T, Merlos-Suarez A, et al. Isolation and in vitro expansion of human colonic stem cells. Nat Med 2011;17:1225-7.

3. Hu Y SG. ELDA: Extreme limiting dilution analysis for comparing depleted and enriched populations in stem cell and other assays. J Immunol Methods 2009;347:70-78.

4. Buczacki SJ, Zecchini HI, Nicholson AM, et al. Intestinal label-retaining cells are secretory precursors expressing Lgr5. Nature 2013;495:65-9.

5. Merlos-Suarez A, Barriga FM, Jung P, et al. The intestinal stem cell signature identifies colorectal cancer stem cells and predicts disease relapse. Cell Stem Cell 2011;8:511-24.

6. Gonzalez-Roca E, Garcia-Albeniz X, Rodriguez-Mulero S, et al. Accurate expression profiling of very small cell populations. PLoS One 2010;5:e14418.
7. Smyth GK. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Stat Appl Genet Mol Biol 2004;3:Article3.

8. Cox J, Hein MY, Luber CA, et al. Accurate Proteome-wide Label-free Quantification by Delayed Normalization and Maximal Peptide Ratio Extraction, Termed MaxLFQ. Molecular & Cellular Proteomics : MCP 2014;13:2513-2526.

9. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 2005;102:15545-50.

10. Munoz J, Stange DE, Schepers AG, et al. The Lgr5 intestinal stem cell signature: robust expression of proposed quiescent '+4' cell markers. EMBO J 2012;31:3079-91.

11. R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/