A rapid method for generating transplantable and biologically responsive colonic tissue from human induced pluripotent stem cells.

Authors

William Dalleywater(1,2)*, Alexander V. Predeus(3)**, Batuhan Cakir(3)**, Pavel Mazin(3)**, Jayakumar Vadakekolathu(4)**, Sergio Rutella(4)**, Marian L. Meakin (1), Alison A. Ritchie(1), Shamir Montazid(5), Sara Cuevas Ocaña(1), Nadine Holmes(6), Victoria Wright(6), Fei Sang(6), Adam Bills(1), Declan Sculthorpe(1), Rasa Elmentaite(3), Sarah A. Teichmann(3,7), Shazia Irshad(5), Ian Tomlinson(5), Andrew Silver(8), Ricky D. Wildman(9), Nicholas R.F Hannan(1)*, Felicity R.A.J. Rose(10)*, Mohammad Ilyas(1,2)*

*Corresponding authors

**These authors made equal contributions

Affiliations

1: Unit of Translational Medical Sciences, School of Medicine, Biodiscovery Institute, University of Nottingham, Nottingham, UK

2: Department of Cellular Pathology, Queen's Medical Centre, Nottingham University Hospitals NHS Trust, Nottingham, UK

3: Wellcome Sanger Institute, Hinxton, Cambridge, UK

4: John van Geest Cancer Research Centre, Nottingham Trent University, Nottingham, UK

5: Nuffield Department of Medicine, Oxford, UK

6: School of Life Sciences, University of Nottingham, Nottingham, UK

7: Theory of Condensed Matter Group, Cavendish Laboratory, Department of Physics, University of Cambridge, Cambridge, UK

8: Centre for Genomics and Child Health, Blizard Institute, Barts and The London School of Medicine and Denistry, Queen Mary University of London, London, UK

9: Faculty of Engineering, University of Nottingham, Nottingham, UK

10: School of Pharmacy, Biodiscovery Institute, University of Nottingham, Nottingham, UK

Abstract

Colonic disease causes significant morbidity and an accurate model of the human colon is urgently needed. Here we describe a 15-day protocol which simultaneously generates intestinal epithelial and mesenchymal cell populations from human induced pluripotent stem cells. Cells were seeded on collagen to create colonic patches (CoPs) and cultured *in vitro*. Single-cell sequencing of CoPs identified similar cell populations to those seen in normal colon. Engraftment of CoPs into mouse subcutis showed development of mucosa containing epithelial crypts (with enterocytes, goblet cells and neuroendocrine cells), multiple stromal populations, smooth muscle and human blood vessels anastomosed to murine vasculature.

We also demonstrate the versatility of our *in-vitro* model in studies of fibrosis and epithelial-mesenchymal interaction. Stimulation of CoPs with different cytokines resulted in cytokine-specific fibrogenic activity. When iPSC-derived mesenchyme was isolated and co-cultured with different epithelial cancer cell lines, there was cell line-specific alteration of mesenchymal gene expression. As well as utility in disease modelling, the transplantability of CoPs raises their possible use as therapeutic autologous grafts for damaged colon.

Main

Diseases of the colon are common in the human population. Inflammatory diseases (such as ulcerative colitis and Crohn's disease) and neoplastic diseases cause significant morbidity and mortality (1-3). These conditions may arise due to abnormality of the epithelium or mesenchyme (including the supporting stroma) (4). Furthermore, the stroma – once thought to be homogeneous – has recently been shown to contain several cell populations, each of which has a different relationship with the epithelium (4, 5). To investigate the biology of colonic diseases and to develop new therapeutics, accurate biological models are required. The most popular are primary organoid *in vitro* models although these require rapid access to primary tissue (usually surgical resection specimens) and they are not able to replicate the complex cellular population profile and 3D architecture of the colon (6, 7). Mouse models have certain advantages but they are expensive, can be difficult to manipulate and observations may not translate to the human condition.

We set out to develop a human induced pluripotent stem cell (hiPSC)-based differentiation platform that would overcome these limitations and allow interrogation of the epithelium, the mesenchyme and the interaction between these compartments. (8, 9). In this study we present a rapid and robust protocol for generating a responsive and transplantable model of colonic mucosa from hiPSCs (Fig. 1A, Ext. Fig. 2).

Firstly, to replicate the complexity of the mucosa, we aimed to drive hiPSCs simultaneously to dual endoderm and mesoderm specification (Fig. 1A-D). Many published protocols rely on poorly defined components such as conditioned media, serum and antibiotics (10-15). To increase reproducibility, we devised a protocol which limited the use of recombinant proteins, replaced serum with defined products and was conducted in antibiotic-free conditions. Our protocol generates intestinal progenitors in eight days which resulted in the up-regulation of both endoderm markers, Sox17 and FoxA2 and the mesoderm marker, Brachyury which was not seen in the

comparative protocols (Fig. 1B). (16). Flow cytometry confirmed the presence of a dual population and intestinal differentiation was confirmed by nuclear Cdx-2 expression (Fig. 1C).

From Day 9 onwards, the co-differentiated cells were placed on collagen 1 hydrogels to create "colonic patches" (CoPs) (Fig. 1D) (17). By Day 22 in *in vitro* culture the epithelial cells in the CoPs had formed a monolayer on the surface of the gel and assembled into crypt-like structures. These cells expressed the intestinal epithelial markers, E-cadherin, Cdx-2 together with villin (which notably demonstrated cell polarisation with expression limited to the apical cell membrane). Crypt-like organization was demonstrated by enrichment of cells expressing Ki67 (denoting the transit amplifying compartment) in the basal regions of the crypts. Vimentin-positive mesenchymal cells in the CoPs were stratified beneath the surface epithelium and around the crypt structures.

We compared bulk transcriptomic profiles in cell populations present at Day 8 of co-differentiation protocol with the expression profiles of cells cultured with a comparative protocol (also lasting 8 days) (11, 12) Differentiated cells derived using both protocols showed similar levels of expression of intestinal epithelial markers but cells grown under our protocol were also strongly enriched for stroma-related genes including HAND1, LUM, BMP4, ActA2, SOX6 and KDR (Fig. 1E) (4, 18).

We further validated the identity of the cells by separating Day 8 cultures into enriched "epithelial" and "mesenchymal" groups using cell sorting based on EpCAM expression (a marker of epithelium) (Fig. 1F). Expression profiling by RNA-Seq confirmed intestinal epithelial identity amongst the EpCAM+ group including enrichment of general epithelial markers (such as SOX17, LGR5, SHH/IHH) and specific intestinal markers (CDH17, ALPI, VIL1). The EpCAM- group showed enrichment for mesenchymal markers including ActA2, BMP4, LUM and SOX6 (4, 18) (Fig. 1F).

To characterise the cellular populations in the epithelial and mesenchymal compartments, single-cell mRNA sequencing was performed on sorted day 8 and day 15 cells (Fig. 1G/H). The single-cell expression profiles of the EpCAM- cells were compared to publicly available single-cell reference atlases generated as part of the human cell atlas project (Fig. 1I) (5, 18, 19). This demonstrated a diversity of emerging intestinal stromal populations indicative of mesoderm differentiation towards the cell types seen in the mature intestinal stroma (19, 20). The identity of these populations was further confirmed by demonstrating enrichment of marker genes for intestinal epithelium, endothelium, mesoderm and more mature mesenchyme (Fig. 1J). Furthermore, we demonstrated a posterior homeobox programme (important for specifying the fate of the mesoderm towards an appropriate intestinal mesenchymal fate), dominated by HoxA10/11, HoxB9 and HoxC6-9 (Fig. 1K). Interestingly, we replicate the recent finding that Cdx-2 is required for both intestinal endoderm and mesoderm specification, by showing expression of Cdx-2 in both populations at day 8 (20). We used RNA velocity to model trajectories of early mesoderm (EpCAM- cells at day 8) into endothelial progenitors (high FLI1, ESAM) and fibroblast/smooth muscle progenitors (high LUM, SOX6, ActA2) (Ext. Fig. 5). We showed a gradual decrease in early markers (LIX1, PRRX1) across pseudotime in both progenitor populations corresponding with an increase in lineage specific markers. To demonstrate further diversity and maturation of mesenchymal populations, we studied specific lineage marker genes. We showed that within later mesenchymal clusters (3+4, Fig. 1L) subclusters of cells show enriched smooth muscle markers (Myh11, ActA2, CNN1, ActG2) as well as markers of specific mature fibroblast subsets (CXCL14, F3, Rspo2, NPY)(18). Within the epithelial population, genes characteristic for intestinal progenitor populations dominated (Fig. 1M)(18). Finally, we looked at the response of mesenchymal cells to treatment with the Hedgehog agonist, purmorphamine (Fig. 1N/2B). We showed enrichment with Gli1 and PTCH1, confirming a Hedgehog response, as well as upregulation of Wnt4 and sFRP1 and downregulation of BMP ligands, which recapitulates the

118 phenotype of Wnt-secreting crypt-niche mesenchyme described in mice by Degirmenci et al. (21).
119 (22).

120 Given the wide array of cellular populations at Day 15, we were prompted to wonder whether the
121 CoPs – even at this early stage – were transplantable and capable of maturation to adult tissue.
122 CoPs were modified by chemically cross-linking the lower part of the collagen gel (to increase
123 strength for handling) while the upper part was not cross-linked (Fig. 1O). Day 15 CoPs were
124 engrafted into the sub-cutis of immunocompromised (R2G2) mice and harvested after 2, 3 or 4
125 weeks (Fig. 1P). Across all experiments we found viable tissues consistently in around 55% mice
126 indicating high transplant efficiency. There was progressive increase in tissue organisation and cell
127 maturation and, by four weeks, luminal spaces lined by epithelium with concentric layers of
128 specialised mesenchyme emerged (Fig. 1Q). The epithelium showed intestinal differentiation
129 including multiple cell types (neuroendocrine, goblet cells, and proliferative cells) (Fig. 1R). The
130 underlying mesenchyme resembled lamina propria and smooth muscle fibres (identified by
131 Smooth Muscle Actin (SMA) and desmin expression) formed a boundary beneath this,
132 recapitulating the muscularis mucosae (Fig. 1S). The tissues became highly vascularised and, using
133 antibodies highly specific for human mitochondria, the vascular channels were shown to be lined
134 with human endothelial cells (Fig. 1T). The vascular channels contained blood indicating that the
135 human blood vessels had undergone anastomosis with murine dermal blood vessels and had
136 become functional. Digital Spatial RNA Profiling was performed on histologically normal human
137 colon mucosa and compared with the subcutaneous transplants (Fig. 1U). Endothelial transcripts
138 were demonstrated in transplanted tissues (Fig. 1V). The mesenchyme showed a progressive
139 transcriptional shift from an "immature" phenotype (Fig. 1W: decreasing TWIST1, mesoderm
140 marker) to a more mature phenotype. The mesenchyme retained Gli1 expression (Fig. 1X)
141 important for maintaining the stem cell niche, while epithelium was enriched for the Hedgehog
142 ligand, IHH (Fig. 1Y). Of particular note, was a gradient of increasing mesenchymal PDGFRA
143 expression from the base of the lamina propria to the mucosal surface(Fig. 1Z) which coincided

144 with epithelial maturation along the crypt axis. In normal tissues we showed an opposing apex-
145 base gradient of both Gli1 (Fig.1X) and PDGFRA (Fig.1Z) corroborating other studies which
146 have demonstrated PDGFRA+ cells at the crypt apex (23). Our data suggest that the emergence
147 of this population may be important for intestinal epithelial maturation. That transplantable tissue
148 can be generated from hiPSCs in just 15 days, raises the exciting prospect of developing autologous
149 tissue grafts to treat damaged intestinal mucosa following intestinal inflammation.

150 Our data, repeated with four different hiPSC lines (Ext. Fig 2), showed that the CoPs can replicate
151 the architecture of the colorectal mucosa. They will facilitate investigation of disorders of the
152 epithelium and of the mesenchyme. For any investigative tool to be useful for hypothesis testing,
153 it should be responsive to experimental manipulation. We firstly tested the utility of our system
154 to model the behaviour of the stromal populations *in vitro* by stimulating CoPs (at Day 15) with
155 specific cytokines (Fig. 2A-E). When treated with TGF-β (Fig. 2A), stromal cells took on a well-
156 described myofibroblast molecular profile characterised by matrix-synthesis and a contractile
157 protein expression pattern (FN^{high}/$Col1A1^{high}$/$Myh11^{high}$/$SOX6^{low}$/ $Wnt2b^{low}$) (4, 5, 18). This was
158 accompanied by deposition of extracellular matrix and contraction of the CoP – features
159 resembling the fibrous scarring seen in inflammatory bowel disease (Fig. 2D). Treatment of CoPs
160 with A83-01, a small molecule inhibitor of TGF-β, abrogated the response to TGF-β. A similar
161 response was seen when CoPs were treated with combined TNFα/interferon-γ (Fig. 2E). The
162 similarity may be explained by the significant increase in TGF-β1 gene expression induced by
163 TNFα/interferon-γ stimulation. In contrast, treatment of CoPs with the Hedgehog agonist
164 purmorphamine (Fig. 2B) resulted in strong upregulation of Gli1 and Wnt2b, reflecting a
165 phenotype suggested by Degirmenci et al. as characteristic of stromal cells surrounding the colonic
166 crypt niche (21). In response to sustained CHIR-99021 stimulation (Fig. 2C), both epithelial and
167 stromal compartments showed proliferation whilst treatment with either BMP4 or the Wnt
168 inhibitor, ICG-001 showed a reduction in cell number. ICG-001-treated cultures showed extensive
169 apoptosis demonstrating the requirement of Wnt activity for cell viability. Interestingly, when CoPs

were treated with alternative cytokine networks (IL-6, IL-13 or IL-17), a Wnt-like phenotype was instead observed (Fig. 2C/E). Finally, we showed that the cells can be cultured successfully as intestinal organoids when encapsulated within Matrigel® with similar results to those observed when cells were differentiated initially with a comparative protocol widely used in the literature (protocol 2) (Ext. Fig 2/3).

We tested whether our system could be used to investigate epithelial-mesenchymal interaction by separating the hiPSC-induced epithelium and mesenchyme and reconstituting the mesenchyme with epithelial cancer cell lines to create "Cancer-CoPs". Day 8 cells were separated by cell sorting based on EpCAM expression. The mesenchymal cells were grown as monocultures on Matrigel-coated plates for up to 7 days and were then established as co-cultures either with the cell line HCT116 or LS1034. Following four days of co-culture, the mesenchymal cells were separated and compared with the mesenchymal cells which had remained as monocultures (Fig. 2F). HCT116 and LS1034 are epithelial cancer cell lines derived from colorectal cancers. HCT116 contains a β-catenin mutation and shows microsatellite instability whilst LS1034 contains an *APC* mutation and is microsatellite stable (24). Comparative profiling showed that HCT116 and LS1034 each caused altered expression of genes in the co-cultured mesenchymal cells which were specific to each cell line (Fig. 2F) as well as a set of genes commonly altered by both cell lines (Ext. Figure 5/Suppl. Table 1). Reciprocal stromal signalling to epithelium (Fig.2G) induced cell line-specific changes in BMP and Wnt signalling pathways and an increase in the proliferation markers in HCT116 but not in LS1034. (25)

In conclusion, we present a rapid protocol to generate a biologically responsive model of colonic tissue using human induced pluripotent stem cells which can be used to investigate both neoplastic and non-neoplastic diseases of epithelium and mesenchyme. Further development of the model will include introduction of neurones and immune cells (as these are not generated *in-situ* and migrate in during colonic development). Critical to the development of hiPSC-based therapies is

the elimination of the risk of teratoma induction through residual pluripotency; we have extensively tested our model and we have not found any residual pluripotency (Ext. Fig 1). The transplantability of our model raises the possibility of generating novel autologous tissue-engineered cell therapies. Next steps will involve scaling up the differentiation protocol to generate sufficient tissue for therapeutic applications in addition to further demonstration of the utility of the model to investigate bowel diseases.

Acknowledgements

We are grateful for the support of all members of the research groups led by Profs Ilyas, Rose, Wildman and Dr Hannan. We acknowledge the support of University of Nottingham core facilities, including the Nottingham Biodiscovery Institute, Biosupport Unit, DeepSeq facility and Flow cytometry facility, as well as the cellular pathology department at Nottingham University Hospitals NHS Trust. We thank Prof Chris Denning for the kind donation of ReblPat human iPSCs for use in these studies. We are also grateful for the support of facilities at the John van Geest Centre, Nottingham Trent University and the CellGenIT team at the Wellcome Sanger Institute. WD was supported by a Clinical Research Training Fellowship (Medical Research Council MR/R017336/1) and NIHR Academic Clinical Fellowship and Academic Clinical Lectureships as well as grants from the Pathological Society of Great Britain and Ireland TSGS 1019 (WD) and CLSG 1021 (AB, MI, WD). NRFH was supported by UKRI (MRC MR/S009930/1, NC3Rs NC/X002101/1 & NC/Y000838/1 and Bowel Research UK. SCO was supported by MRC (MR/S009930/1). DS was supported by a Bowel Research UK PhD studentship.

Contributions

WD, NRFH, FRAJR, RDW and MI conceived the study, contributed to experimental design and data analysis, and co-wrote the manuscript. WD performed and analysed the experiments. SM, AB

and SI aided in or performed the experiments. MM and AR performed *in vivo* experiments. NH and VW performed single-cell RNA sequencing experiments. DS aided with optimising and performing immunohistochemistry. JV and SR performed Nanostring Digital Spatial Profiling experiments. AP, BC, PM, JV, SR, FS, RE and ST aided with bioinformatic analyses and interpretation of single-cell and spatial profiling studies. NRFH, SI and IT provided iPSC lines. SI, AW and IT aided in interpretation of stromal cell modulation experiments. All authors contributed to the editing of the manuscript.

References

1. Bassi A, Dodd S, Williamson P, Bodger K. Cost of illness of inflammatory bowel disease in the UK: a single centre retrospective study. Gut. 2004;53(10):1471.
2. Yeshi K, Ruscher R, Hunter L, Daly NL, Loukas A, Wangchuk P. Revisiting Inflammatory Bowel Disease: Pathology, Treatments, Challenges and Emerging Therapeutics Including Drug Leads from Natural Products. Journal of Clinical Medicine. 2020;9(5):1273.
3. Nivedita G, Purushothaman P. A UK cost of care model for inflammatory bowel disease. Frontline Gastroenterology. 2015;6(3):169.
4. Kinchen J, Chen HH, Parikh K, Antanaviciute A, Jagielowicz M, Fawkner-Corbett D, et al. Structural Remodeling of the Human Colonic Mesenchyme in Inflammatory Bowel Disease. Cell. 2018;175(2):372-86.e17.
5. Elmentaite R, Ross ADB, Roberts K, James KR, Ortmann D, Gomes T, et al. Single-Cell Sequencing of Developing Human Gut Reveals Transcriptional Links to Childhood Crohn's Disease. Developmental Cell. 2020;55(6):771-83.e5.
6. Hayden PJ, Harbell JW. Special review series on 3D organotypic culture models: Introduction and historical perspective. In Vitro Cellular & Developmental Biology - Animal. 2021;57(2):95-103.
7. Sprangers J, Zaalberg IC, Maurice MM. Organoid-based modeling of intestinal development, regeneration, and repair. Cell Death & Differentiation. 2021;28(1):95-107.
8. Doss MX, Sachinidis A. Current Challenges of iPSC-Based Disease Modeling and Therapeutic Implications. Cells. 2019;8(5).
9. Takahashi K, Tanabe K, Ohnuki M, Narita M, Ichisaka T, Tomoda K, et al. Induction of Pluripotent Stem Cells from Adult Human Fibroblasts by Defined Factors. Cell. 2007;131(5):861-72.
10. Múnera JO, Sundaram N, Rankin SA, Hill D, Watson C, Mahe M, et al. Differentiation of Human Pluripotent Stem Cells into Colonic Organoids via Transient Activation of BMP Signaling. Cell stem cell. 2017;21(1):51-64.e6.
11. Spence JR, Mayhew CN, Rankin SA, Kuhar MF, Vallance JE, Tolle K, et al. Directed differentiation of human pluripotent stem cells into intestinal tissue in vitro. Nature. 2011;470(7332):105-9.
12. Wells JM, Spence JR. How to make an intestine. Development. 2014;141(4):752-60.

13. Macedo MH, Araújo F, Martínez E, Barrias C, Sarmento B. iPSC-Derived Enterocyte-like Cells for Drug Absorption and Metabolism Studies. Trends in Molecular Medicine. 2018;24(8):696-708.
14. Mithal A, Capilla A, Heinze D, Berical A, Villacorta-Martin C, Vedaie M, et al. Generation of mesenchyme free intestinal organoids from human induced pluripotent stem cells. Nature Communications. 2020;11(1):215.
15. Múnera JO, Kechele DO, Bouffi C, Qu N, Jing R, Maity P, et al. Development of functional resident macrophages in human pluripotent stem cell-derived colonic organoids and human fetal colon. Cell Stem Cell. 2023;30(11):1434-51.e9.
16. Tamminen K, Balboa D, Toivonen S, Pakarinen MP, Wiener Z, Alitalo K, et al. Intestinal Commitment and Maturation of Human Pluripotent Stem Cells Is Independent of Exogenous FGF4 and R-spondin1. PLOS ONE. 2015;10(7):e0134551.
17. Dalleywater W, Wheat F, Sculthorpe D, Hyland G, Ilyas M. In Vitro Culture and Histological Evaluation of 3D Organotypic Cultures. In: Ordóñez-Morán P, editor. Intestinal Differentiated Cells: Methods and Protocols. New York, NY: Springer US; 2023. p. 155-70.
18. Elmentaite R, Kumasaka N, Roberts K, Fleming A, Dann E, King HW, et al. Cells of the human intestinal tract mapped across space and time. Nature. 2021;597(7875):250-5.
19. Chuan X, Martin P, Simone W, Laura J, Benjamin JS, Regina H, et al. Automatic cell type harmonization and integration across Human Cell Atlas datasets. bioRxiv. 2023:2023.05.01.538994.
20. Yu Q, Kilik U, Holloway EM, Tsai Y-H, Harmel C, Wu A, et al. Charting human development using a multi-endodermal organ atlas and organoid models. Cell. 2021;184(12):3281-98.e22.
21. Degirmenci B, Valenta T, Dimitrieva S, Hausmann G, Basler K. GLI1-expressing mesenchymal cells form the essential Wnt-secreting niche for colon stem cells. Nature. 2018;558(7710):449-53.
22. La Manno G, Soldatov R, Zeisel A, Braun E, Hochgerner H, Petukhov V, et al. RNA velocity of single cells. Nature. 2018;560(7719):494-8.
23. McCarthy N, Manieri E, Storm EE, Saadatpour A, Luoma AM, Kapoor VN, et al. Distinct Mesenchymal Cell Populations Generate the Essential Intestinal BMP Signaling Gradient. Cell stem cell. 2020;26(3):391-402.e5.
24. Berg KCG, Eide PW, Eilertsen IA, Johannessen B, Bruun J, Danielsen SA, et al. Multi-omics of 34 colorectal cancer cell lines - a resource for biomedical studies. Molecular Cancer. 2017;16(1):116.
25. Orzechowska-Licari EJ, Bialkowska AB, Yang VW. Sonic Hedgehog and WNT Signaling Regulate a Positive Feedback Loop Between Intestinal Epithelial and Stromal Cells to Promote Epithelial Regeneration. Cellular and Molecular Gastroenterology and Hepatology. 2023.

Methods

Standard cell culture conditions

Unless otherwise specified, all cell culture was performed using aseptic technique under normoxic conditions at 37 °C with 5% CO_2 in a humidified incubator. All cells used were

tested routinely for Mycoplasma at least monthly through the Nottingham Biodiscovery Institute testing programme. All media constituents and supplements are detailed in Supplementary Table 2.

Cell culture conditions for human induced pluripotent stem cells

Human iPSCs were generated within the Nottingham Biodiscovery Institute (formerly the Centre for Biomolecular Sciences), University of Nottingham from fibroblasts harvested by punch biopsy of axillary dermis, using methods previously described (9) and were gifted to this project by Professor Chris Denning, University of Nottingham. The REBLPAT line was used for most studies from passage 27 and used for experiments between passage numbers 30 and 36. In addition, BE31 and BE32 hiPSC lines were also used for validation. All hiPSC lines were approved and derived under University of Nottingham ethics committee number 09/H0408/74.

hiPSCs were cultured in Essential 8™ medium (Life Technologies, Bleiswijk, Netherlands) to maintain pluripotency, at 37 °C, 5% CO_2 in a humidified incubator. Medium was changed daily by aspirating spent medium, washing once with phosphate-buffered saline (PBS) and replacing with fresh Essential 8™ medium. Cells were cultured for three to four days to reach at least 80% confluence before passaging. In order to passage, cells were washed once in PBS, after which TrypLE Select (Life Technologies) was added for up to 5 minutes, until cleavage between cells could be seen upon microscopy, but before complete detachment from the culture surface. TrypLE Select was then quickly, but carefully aspirated, fresh Essential 8™ medium containing 10μM Y-27632 dihydrochloride (ROCK-inhibitor [ROCKi]) was added and the detached cells counted. Upon passage, approximately 30,000 iPSCs cm^{-2} were added to each culture vessel, which had been previously coated with Matrigel® at 11.1 μg cm^{-2} for at least 90 minutes incubated under standard cell culture conditions. For the first 24 hours of culture, cells were grown in medium containing 10 μM ROCKi.

Co-differentiation to posterior endoderm and mesoderm

Three protocols were tested for co-differentiating iPSCs to endoderm and mesoderm. Our final optimised protocol (protocol 1) is a novel method and is used for all experiments unless otherwise specified. Two other protocols (protocol 2 + 3) were based on widely used formulations described elsewhere in the literature and were used for comparison. Please contact the corresponding authors for further details of the protocol.

Neutralisation of collagen

Type 1 rat-tail collagen (Life Technologies) was used at a stock concentration of 3 or 10 mg mL^{-1}. To achieve a concentration of 2 mg mL^{-1} for each 1 mL of collagen gel required, 667 µL of stock was added to a tube on ice. To this, 100 µL of 10X PBS were added. To neutralise the solution, 0.025 µL of sterile 1M sodium hydroxide were added for each microlitre of collagen. The solution was topped up to 1 mL with medium (DMEM/F12), which if neutralised correctly displayed a slight pink colour of the phenol red pH indicator.

Culture on collagen hydrogels

A method for demonstrating the growth of posterior endoderm subsequent to initial differentiation was performed on collagen hydrogels using previously published methods (17). The procedure was performed in either 12-well Transwell® plates (Corning) with 0.4 µm pore size membranes. The plates were prepared by adding 500 µL of ice-cold neutralised collagen (see above) into each Transwell® insert. The plates were then incubated for 1 – 4 hours in standard culture conditions, to allow the collagen to form a semi-solid gel.

Basal maturation medium consisted of DMEM / F12 (1:1) with L-glutamine (Life Technologies), 1x B27, 1x N2 (Life Technologies). This was initially supplemented with

CHIR-99021 5μM, FGF-4 (Peprotech) 100 ng mL^{-1} and recombinant human Noggin 100 ng mL^{-1} (Peprotech), as well as IGF-1 (100 ng mL^{-1}, Peprotech) and FGF-2 (50 ng mL^{-1}, Peprotech). For the initial period of culture, all media were supplemented with ROCKi 10μM.

In the bottom of each well with a Transwell® insert 1.5mL of supplemented basal maturation medium was added. Cells were detached and re-suspended in supplemented basal maturation medium; 500,000 cells (12-well plate) were added in 500 μL of medium onto the surface of each Transwell® insert containing gelled collagen.

Cell medium was changed in both chambers every 3-4 days for the first two weeks, but with only IGF-1 and FGF-2 supplemented. Following this, the medium was aspirated from the top chamber and only replaced in the basal chamber, to create an air-liquid interface. Cultures were continued for up to 4 weeks, with media replacement in the bottom chamber every 3 – 4 days.

Following culture, medium was aspirated from the wells and replaced with 10% (*w*/*v*) neutral buffered formalin (NBF; Sigma Aldrich) to fix the tissues. Gels were carefully removed from the Transwell® insert and allowed to float in formalin and fixation took place overnight at room temperature, for at least 18 hours. After fixation, gels were lifted from the culture plate, blotted on tissue paper to remove excess formalin, bisected and then placed in a plastic mould. Low melting point agarose (Sigma Aldrich) was warmed to allow melting and was then laid over the gel within the plastic mould. The mould was then placed on ice to solidify the agarose. After solidifying, the agarose was trimmed and the embedded gel placed in a tissue cassette. The agarose/gel composites were fixed in NBF for a further 24 hours, before automated tissue processing (Leica) to dehydrate and take the tissues to paraffin wax. In brief, tissue cassettes were loaded onto the automated tissue processor and submerged into 50% (*v*/*v*) methanol in distilled water. Tissue cassettes were moved to sequentially higher concentrations of alcohol (50%, 75%, 95% and 100% [*v*/*v*]) for two hours each and then through three 100% xylene baths for two hours each. Finally, tissue cassettes were submerged in molten paraffin wax for two hours before being removed from the processor for embedding.

Following processing, the agarose/gel composites were embedded cut edge facing down in paraffin wax and mounted onto a cassette. Sections of cooled tissue blocks were cut using a microtome (Leica) at 3 – 4 μm, floated and collected onto poly-L-lysine coated slides. Slides were then either stored to be used later or warmed to 50 – 60 °C to melt the tissues onto the slide, before being used in histological assays described in subsequent sections.

Organoid culture

Organoid culture was performed in either 24-well plate Transwell® inserts using a mixture of Matrigel® and neutralised collagen (1:1) or with drops of Matrigel® formed on a plate. For both methods, 150,000 posterior endoderm differentiated iPSCs were added in a small volume (up to 10% of total volume) of DMEM/F12 medium to cooled hydrogel solution. After rapid, but careful mixing, to avoid both solidification and air bubble formation, the hydrogel/cell mixture was transferred to the Transwell® insert (500 μL) or onto the surface of a well on a 6-well plate (100 μL – 2 drops per well). The 6-well plates containing Matrigel® drops were inverted and incubated for 2 hours at standard cell culture conditions, while plates with Transwell® inserts were incubated at normal orientation for 2 hours before adding culture medium.

Culture medium consisted of basal maturation medium (as described above) with the growth factors EGF 50 ng mL^{-1}, R-spondin 50 ng mL^{-1} and noggin 100 ng mL^{-1}, and ROCKi at 10 μM. The same media constituents were used throughout culture, except for ROCKi, which was not used in subsequent media changes. Media were changed every 3-4 days, and organoids were passaged into new Matrigel® drops every 10 days. For passage, Matrigel® drops were overlaid with 800 μL Gentle Cell Dissociation reagent per well (StemCell Technologies, Vancouver, Canada) and incubated for 5 minutes. A micropipette was used to disrupt each Matrigel® drop by repeatedly pipetting the drop up and down until disaggregated. The contents of each well were transferred to a 1.5 mL

406 microtube (one per well). The well was washed with 400 µL of dissociation reagent which
407 was then added to the microtube. The organoids were incubated for a further 10 minutes
408 at room temperature, before centrifugation at 5 °C (300g, 5 minutes). Supernatant was discarded
409 and organoids were washed with 1 mL cold culture medium and centrifuged under the same
410 conditions again. The supernatant was discarded and organoids resuspended in thawed 400 µL
411 Matrigel® at 4 °C. New organoid drops were then formed in 6 well plates as described above. At
412 the time of passage, ROCKi was supplemented in medium.
413 Matrigel® drops were gently disaggregated by incubation in cold PBS followed by tituration.
414 Organoids were then embedded in agarose, using similar principles to those described above and
415 analysed using histological techniques.
416
417 **Stromal modulation experiments**
418 Stromal cells were cultured on Matrigel® coated 6-well plates using standard coating parameters.
419 The baseline media consisted of DMEM/F12 (1:1) with IGF-1 (100ng mL^{-1}), FGF-2 (50ng mL^{-1})
420 and PDGFbb (2ng mL^{-1}). For TGF-β experiments, media was supplemented with TGF-β1 (1ng
421 mL^{-1}, Peprotech) or A83-01 (Sigma Aldrich, 500 nM). Cultures were maintained for up to 6 days.
422 The same conditions were used for 3D collagen cell cultures described elsewhere but for longer
423 periods. For purmorphamine experiments, baseline media was supplemented with
424 purmorphamine (10 µM, Sigma Aldrich) and cultures were maintained for up to 6 days.
425
426 **Cancer cell line co-cultures with stroma.**
427 HCT116 and LS1034 colorectal cancer cell lines were maintained in DMEM with 10% FBS and
428 L-glutamine prior to co-culture. Cells were passaged when near confluent and grown on tissue
429 culture plastic in flasks. Prior to co-cultures, separated stromal cells (EpCAM-) from iPSCs were
430 cultured in DMEM/F12 (1:1) with B27 (1X) and N2 (1X) supplements and supplemented with

rhIGF-1 (100ng mL^{-1}), rhFGF-2 (50ng mL^{-1}), A83-01 (500nM) and rhPDGFbb (2ng mL^{-1}) on plates/flasks coated with Matrigel® using standard coating parameters.

To establish co-cultures, plates were coated with Matrigel® using standard coating parameters. Stromal cells were seeded at 300,000 per well on a 6 well plate and further cultured until 90% confluent. Co-cultures were established by seeding 150,000 or 300,000 HCT116 or LS1034 cells in stromal cell media as above with IGF-1/FGF-2 but without PDGFbb or A83-01. Monocultures of stromal cells were maintained by not seeding cell lines but switching to the same media without PDGFbb or A83-01. Likewise, cell lines were established as monocultures on Matrigel® coated plates with the same media as used for stroma monoculture/co-cultures. Monocultures/co-cultures were maintained for 72 hours and then harvested by separating the cells using EpCAM cell separation as described above. The cells could then be used in downstream assays.

RNA extraction and quantification

All RNA was extracted and purified from 6-well or 12-well plates, using a total mammalian RNA mini-prep kit (Sigma Aldrich) according to manufacturer's instructions. Following media aspiration, wells were washed with PBS once. RNA lysis buffer containing 1% 2-mercaptoethanol was added to wells and incubated for 2 minutes at room temperature and the lysate was transferred to an Eppendorf tube, on ice. After addition of an equal volume of 70% ethanol, the mixture was transferred to an RNA binding column, and all steps for purification of RNA were performed according to the kit protocol, including an ancillary on-column DNA digestion for 15 minutes (Sigma Aldrich). Following purification, RNA was quantified using a Nanodrop™ 2000 spectrophotometer (Life Technologies).

Up to 2 µg RNA was reverse transcribed to cDNA using the Omniscript RT Kit (Qiagen, Manchester, UK) or High-Capacity Reverse Transcription Kit (ThermoFisher) according to manufacturer's protocol. In brief, per 20 µL reaction, up to 12.5 µL of RNA template was added to 7.5 µL mastermix. The mastermix consisted of 1 µL reverse-transcriptase

enzyme, 2 µL 10X RT buffer, 2 µL RT random primers, 0.8 µL 25X dNTP mix (100mM), 1 µL RNase inhibitor. All reactions were topped up to 20 µL with nuclease-free water. All procedures were performed in a decontaminated UV-light PCR hood using nuclease free reagents and plastic consumables.

Quantitative real-time PCR was carried out on the resulting cDNA. Reactions were prepared using PowerUp™ SYBR™ Green Master Mix (Applied Biosystems, Waltham, MA, USA) using custom primer pairs (Table 3; Eurofins, Ebersburg, Germany). Primers (see supplementary table 3) were custom designed such as to span an exon-exon junctions and with a PCR product size of 70 – 150bp. All PCR sample preparation was performed in a UV-sterilised hood. Equal quantities of cDNA were added in each reaction series and at least 2 replicates per cDNA sample were performed, as well as appropriate negative controls. PCR was performed on ViiA™ 7 Real-Time PCR machine using a fast cycling protocol consisting of an initial hold stage of two minutes at 50 °C (for activation of UNG), followed by five minutes at 95 °C (to activate hot-start Taq polymerase) followed by 40 cycles of denaturation for 1 second at 95 °C followed by annealing at an optimised temperature between 56 °C and 65 °C for 30 seconds, a separate extension step was not required. Fluorescence was read during the annealing step of each cycle. After cycling stages, a melt curve stage was included to verify the specificity of PCR amplification. To compare relative gene expression between samples, an average CT value of replicates was taken if the values were within 0.5 of each other; otherwise the average of the 2 nearest values was taken, if the difference between the values was less than 0.5. Data points with larger discrepancies were excluded. CT values for individual primers were compared with reference to a housekeeping gene using Livak's $2^{-\Delta CT}$ method or $2^{-\Delta\Delta CT}$ method if an appropriate reference condition was available for comparison.

Flow cytometry (EpCAM)

Cells were washed once with PBS and detached with brief TrypLE treatment. Following aspiration of the TrypLE, cells were resuspended in warm RPMI-1640 and transferred to

Eppendorf tubes. The tubes were centrifuged at 300 xg for 5 minutes and then washed once with PBS, followed by further centrifugation. Reactions were performed in Eppendorf tubes and each step was followed by centrifugation at 300 xg for 5 minutes unless otherwise specified. Primary Anti-EpCAM APC conjugated antibody with isotype control was used for all experiments (Miltenyi Biotec, Surrey, United Kingdom). Cells were blocked in 3% (*w*/*v*) BSA in PBS for 15 minutes, followed by washing and addition of primary antibody diluted in 3% (*w*/*v*) BSA in PBS, with incubation for 30 minutes at 4 °C.

Cells were analysed using either an FC500 or MoFlo (Beckman Coulter, Indianapolis, IN, USA) flow cytometer, with all procedures kindly optimised by the staff of the School of Life Sciences Flow Cytometry Facility, University of Nottingham (Dr David Onion and Mrs Nicola Croxall).

Data were analysed using the Kaluza Analysis software package (Beckman Coulter). Gating parameters on forward/side scatter were used consistently across all samples, and non-singlet cells were excluded from fluorescent intensity analyses to avoid overestimation. Negative and isotype controls were performed, which determined the threshold fluorescent intensity values for positive staining. For single marker expression, data were plotted on a histogram, while for dual labelling studies, data were plotted on scatter plots. Where possible, dual labelling studies used fluorophores with minimal spectral overlap. If appropriate, quantitative summary data were calculated.

Bulk mRNA sequencing

Cells lysates were prepared directly from plates and RNA was extracted as described in section 2.14. After quantification by Nanodrop, samples were frozen and stored at -80 °C. Samples were diluted in nuclease-free water to obtain 2 µg total RNA in 20 µL sample volume. All samples were sent on dry ice via overnight courier to Novogene Limited for sample QC and subsequent library preparation and sequencing. Results were returned in

raw FastQC format, BAM alignment files and raw and normalised count matrices for bioinformatic analysis. Subsequent analysis was performed using open-source SeqMonk and R software. Briefly, normalised gene matrices were generated using SeqMonk and initial hierarchical clustering was performed using dual approach based on intensity difference (based on log-transform) and DESeq2 packages (raw counts) to compare gene expression differences across multiple datasets. For comparison between two conditions, R was used to run DESeq2 on raw gene count matrices followed by Volcano plots based on log10 adjusted p-value and log2 fold change for individual genes. All DESeq2 analyses were performed with default parameters including correction for multiple hypothesis testing using the Benjamini-Hochberg method.

Single-cell mRNA sequencing

Cells were dissociated as described in section 2.4, but with a prolonged dissociation time of 8 minutes to ensure a fully dissociated cell population. Cells were resuspended in dPBS with 1% (*w/v*) BSA to prevent intercellular adhesion. Cells were counted within the cell culture facility, using an automated counter, and resuspended at 1 million cells mL^{-1}. A second count was performed within the DeepSeq facility as described below. All subsequent steps were performed with the kind collaboration of Dr Nadine Holmes, Miss Victoria Wright and other members of the DeepSeq facility, and a summary is given below.

Single cell 3' whole transcriptome sequencing libraries were prepared from dissociated cell suspensions using the Chromium Next GEM Single Cell 3' Library and Gel Bead Kit v3.1, the Chromium Next GEM Chip G Single Cell Kit and the Dual Index Kit TT Set A (10X Genomics; PN-1000147, PN-1000127 and PN-1000215). Cell counts and viability estimates were obtained using the LUNA-II Automated Cell Counter (Logos Biosystems), Trypan Blue Stain, 0.4 % *(w/v)* and Luna Cell Counting Slides (Logos Biosystems; T13001 and L12001). Live cell counts were used to calculate cell input, rather than total cell count,

as visual inspection of cell field on the LUNA II and the gating histogram, showed that > 90% of cells were viable and that the cell counter appeared to be counting some extracellular debris as non-viable cells. The number of input cells targeted was 3,300 cells per sample, with the aim of generating sequencing libraries from ~ 2,000 single cells. All steps, including GEM Generation and Barcoding, Post GEM-RT Cleanup and cDNA Amplification and Library Construction were performed according to the Chromium Next GEM Single Cell 3' Library and Gel Bead Kit v3.1 User Guide, Rev B (CG000315). Variable steps of this protocol included using 12 cycles of cDNA amplification and 8-12 cycles of library amplification. Amplified cDNA was quantified using the Qubit Fluorometer and the Qubit dsDNA HS Assay Kit (ThermoFisher Scientific; Q32854) and fragment length profiles were assessed using the Agilent 4200 TapeStation and Agilent High Sensitivity D5000 ScreenTape Assay (Agilent; 5067-5592 and 5067-5593). Completed sequencing libraries were quantified using the Qubit Fluorometer and the Qubit dsDNA HS Assay Kit and fragment length distributions assessed using the Agilent 4200 TapeStation and the High Sensitivity D1000 ScreenTape Assay (Agilent; 5067-5584, 5067-5585).

Libraries were pooled in equimolar amounts and the final library pool was quantified using the KAPA Library Quantification Kit for Illumina Platforms (Roche; KK4824). Libraries were sequenced on the Illumina NextSeq 500 over two NextSeq 500 High Output v2.5 150 cycle kits (Illumina; 20024907) to generate > 25,000 raw reads per cell for each sample, using custom sequencing run parameters described in the 10X protocol.

Following sequencing, raw outputs from the sequencer were converted to FastQC files, aligned to the reference genome (hg38) and outputted to raw and filtered count matrices using the CellRanger (v6.1.1) pipeline from 10X genomics.

Individual analysis was performed using jupyter-lab tools on Ubuntu Linux using the following packages: Scanpy (1.8.2), anndata (0.7.8), UMAP (0.5.1), NumPy (1.20.3), SciPy (1.7.3), Pandas

(1.5.2), scikit (1.1.3), statsmodels (0.12.2), igraph (0.10.8), PyNNDescent (0.5.5), scvelo, cellrank, MatPlotLib, Scrublet and CellTypist (1.6.2). Briefly, samples were imported and merged. Predicted doublets were removed using Scrublet using a threshold of 0.25. Basic filtering for gene detection included presence in at least 5 cells and all cells with a minimum of 200 and maximum of 7000 genes. Genes with greater than 20% mitochondrial reads were filtered out. Cells were normalised and log-transformed and highly variable genes were then identified. Normalised expression levels next underwent principal component analysis (PCA), nearest-neighbour analysis (neighbours=15 and PCs = 10, optimised using the PCA elbow plot) and uniform mapping and approximation projection (UMAP), followed by clustering using the leiden algorithm (resolution=0.3).

Next clusters were annotated using unbiased CellTypist automated labelling. The "Developing_Human_Organs.pkl" model was selected as it was based on cells at the most similar differentiation stage. Annotations were made over leiden clusters using the majority-voting method. Feature plots, dot plots and matrix plots (heatmaps) were generated using scanpy functions. For scoring of the epithelial cluster for specific cell types, marker gene lists for each epithelial cell type were derived from Elmentaite et al. (x). The epithelial leiden cluster was extracted from the main anndata object and then scored in turn using the scanpy score_genes function. These scores were then plotted together using the stacked violin plotting tool.

In vivo teratoma assay

Undifferentiated iPSCs and cells differentiated according to protocols 1 and 2 were dissociated and pelleted by centrifugation at 300 xg for 5 minutes. For initial experiments, either one million or two million cells were implanted to optimise conditions. Either CD1 nude or R2G2 immunosuppressed mouse models were used for optimisation. In final conditions, one million cells were implanted into R2G2 mice. Cell pellets were transported on ice to the Biosupport Unit (BSU), University of Nottingham as well as thawed Matrigel ® on ice. Immediately prior to implantation, 50

μL of Matrigel was added to the cell pellet and cells were resuspended within the matrix. This cell suspension was then injected into the testicular subcapsular space. Mice were monitored by BSU staff throughout the period for weight and signs of ill health. Seven weeks after implantation, the mice were sacrificed and a post-mortem examination of the testicular space and abdominal cavity was undertaken. Any tumour tissue or evidence of cell growth was excised. All tissues were photographed and then placed in formalin in preparation for further histological examination. All surgical procedures described here and in the following section were carried out by Alison Ritchie and Marian Meakin in the University of Nottingham Biosupport Unit following approval by the AWERB under the home office project licence P435A9CF8.

In vivo subcutaneous engraftment of populated collagen scaffolds

Tissue scaffolds were prepared from collagen following an optimised protocol in which 250 μL collagen (2mg mL^{-1}) from rat's tail was neutralised to form a hydrogel within a Transwell ® mould. After gelation, collagen gels were chemically crosslinked using a protocol adapted from Kim et al. (168). After extensive washing, a further 250 μL of neutralised collagen (2mg mL^{-1}) was added to the surface of the crosslinked collagen and allowed to gel. Next, 500,000 iPSC-derived intestinal cells at day 8 were passaged to the surface of the gel in the presence of CHIR (5μM), Noggin (100ng mL^{-1}), ROCKi (10μM), IGF-1 (100ng mL^{-1}) and FGF2 (50ng mL^{-1}) in DMEM/F12 medium supplemented with B27 and N2 supplements. The following day, medium was replaced with basal medium with additional IGF-1, FGF-2 and Rspo-1 (50ng mL^{-1}). Medium was replaced again on day 4 and then on day 6 (one day before implantation). This culture medium contained IGF1 (200ng mL^{-1}), FGF-2, Rspo-1 and Wnt-3a (50ng mL^{-1}) to enhance to proliferation of intestinal epithelial stem cells. On the day of implantation, tissue grafts were carefully lifted out of the Transwell and placed in medium with the same constituents as well as ROCKi (10μM) to improve cell

613 viability during engraftment.

614 Tissues were engrafted by making a small incision into the flank of R2G2 immunosuppressed mice
615 within a sterile hood. The tissue graft was inserted into the pouch formed by incision, and the
616 incision was closed with a surgical clip. After two, three or four weeks, mice were sacrificed and
617 the surgical site opened. Tissue from the site was removed and photographed, and fixed in
618 formalin prior to histological examination.

619

620 **Nanostring GeoMX DSP RNA profiling**

621 Slides for Nanostring analysis were prepared from formalin-fixed paraffin-embedded tissues.
622 Multiple tissues were combined on a single slide to allow simultaneous analysis. Nanostring data
623 was collected from two separate slides consisting of representative tissues collected from different
624 *in vivo* transplantation timepoints (slide1) and a reference normal adult human colon. Each
625 Nanostring slide consisted of two to six different pieces of tissue.

626 Slides were stained with pan-cytokeratin (Nanostring) and DAPI stains and then hybridised with
627 Cancer Transcriptome Atlas probes. The slides were kept hydrated at all times. The slides were
628 then loaded onto the Nanostring GeoMX data spatial profiler (DSP). Areas of illumination (AOI)
629 were selected based on areas of relevant morphology, such as apical and basal crypt regions.
630 Automated segmentation was also performed based on cytokeratin expression. Each AOI was
631 selected to contain at least 100 predicted cells (nuclei) based on DAPI staining. Having selected
632 the AOIs for each slide, automated harvesting was performed on the DSP and probes were
633 collected in 96 well plates each with well-specific barcodes. Library preparation was performed
634 according to manufacturer's instructions and the pre-made libraries were sequenced by NovoGene
635 Limited using Illumina NovoSeq 6000 using a PE-150 sequencing strategy on a single flow-cell
636 lane. After sequencing, fastq files were loaded into the DSP for pairing with AOIs.

To process the Nanostring data, we utilised the SpatialDecon package. Prior to deconvolution, we generated cell type expression profiles from the integrated foetal gut cell atlas object, considering cells with more than 10 genes and cell types with more than 5 cells. Background counts were determined using negative probes, and the SpatialDecon algorithm was applied to estimate cell type abundances for each area of illumination (AOI).

Subsequently, we employed the SpatialOmicsOverlay package to visualise the data on the image data. Initially, we made necessary modifications to package functions to resolve compatibility issues. We then imported the image and annotation data along with counts into the SpatialOverlay object. AOIs were grouped based on gene expression profiles using hierarchical clustering to identify similarities, resulting in four main clusters. Then, the cell types were assigned to each AOI based on the highest cell type abundance. Since each main slide consisted of 2 to 6 different sub slides, AOIs from each main slide were annotated according to their location, such as top-right or bottom-left. To focus on individual sub slides, the SpatialOverlay object was cropped accordingly, and marker expression levels were visualised by plotting them onto the AOIs.

Basic histology

Slides were dewaxed and rehydrated by sequential immersion in xylene, methanol and distilled water. After rehydration, slides were laid flat and tissue sections were covered with Shandon instant haematoxylin (Life Technologies) for 3 minutes. Slides were then washed in distilled water, differentiated as necessary by rapid immersion and emersion in acid-alcohol (3% HCl in 95% ethanol) and washed again in distilled water. Submersion in Scott's tap water (1% w/v magnesium sulphate and 0.067% w/v sodium bicarbonate in distilled water) was performed for 5 minutes to blue the haematoxylin, followed by further washing in distilled water. The slides were stained with Shandon instant eosin (Life Technologies) for 30 seconds, briefly washed in distilled

water and then dehydrated by sequential immersion in ethanol and xylene. Slides were mounted with DPX and a coverslip was placed over the tissue.

For Alcian blue staining, tissues were similarly rehydrated. Slides were incubated in 3% (*v*/*v*) acetic acid solution for 1 minute to acidify the tissue and then immersed in Alcian blue solution for 15 minutes. After washing in distilled water, slides were differentiated as necessary in acid-alcohol, washed in distilled water and then tissues were covered with nuclear fast red solution (Sigma Aldrich) for 5 minutes for counterstaining. Following washing, slides were dehydrated and mounted as described above.

Immunohistochemistry

Slides were dewaxed and rehydrated as described above. Owing to previous fixation, antigen retrieval was performed by incubating slides in simmering sodium citrate buffer (pH6) for 20 minutes; a microwave set on low power was used to maintain temperature (approximately 95 °C). Once cooled, slides were mounted under water onto Shandon Sequenza® coverplates (Life Technologies) and placed into a Sequenza® rack. Slides were washed three times with 200 µL TBST (Tris-buffered saline with 0.01% Tween-20). All immunohistochemistry was performed using the Novolink™ polymer detection system according to manufacturer's instructions; all volumes used were 100 µL and washing was performed with TBST. Briefly, following protein and peroxidase blocking steps and washing, 100 µL of diluted antibody solution (in TBST) was added per slide; see Table 4 for details of antibodies used and concentrations. Following overnight incubation in the primary antibody solution at 4 °C, slides were washed and exposed to the post-primary solution for 1 hour at RT, polymer-peroxidase for 30 minutes, 3',3'-diaminobenzidine (DAB) for 5 minutes and modified haematoxylin for 5 minutes. After this, slides were dehydrated and mounted as described above. Some IHC was performed by the Nottingham University Hospitals

687 Cellular Pathology Department, according to standard operating procedures, which are
688 available upon request;any such assays are indicated within the text.
689
690 **Statistical analysis and software**
691 Statistical analyses were performed if appropriate to determine statistically significant
692 differences. The student's t-test (unpaired, two-tailed) or one-way ANOVA tests were used for
693 continuous data. Unless otherwise specified, the following symbols are used to indicate statistical
694 significance: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$. Numbers of replicates are
695 indicated in the main text or figure legends.
696 All statistical analyses and graphs were performed and produced using the following software
697 packages: Microsoft Office 365 Excel, GraphPad Prism 9, R x86 v4.0.3/4.1.1/4.3.` and Jupyter-
698 lab v3.3.0. Image analysis was performed using Leica LASX, ImageJ/Fiji, QuPath and Adobe
699 Photoshop.
700
701
702 **Figures**

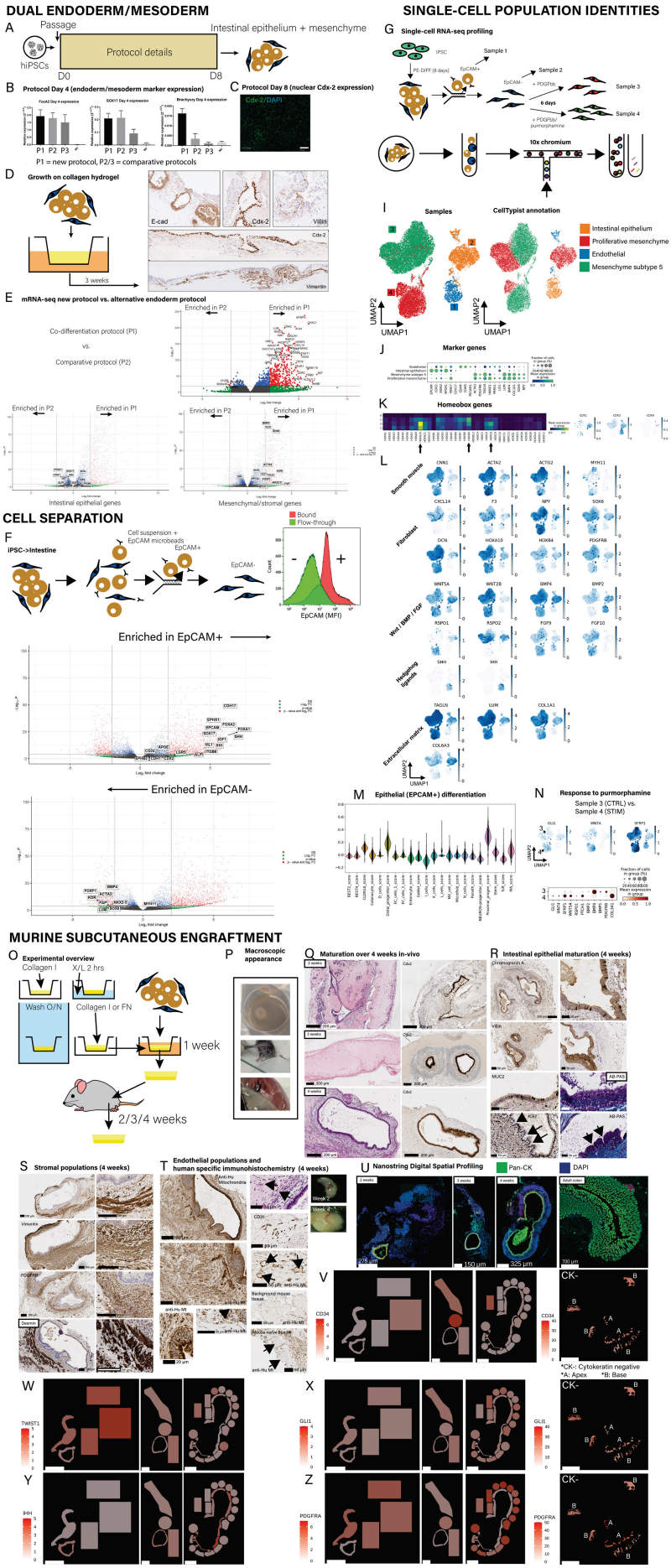


Fig. 1 Generating intestinal tissues from hiPSCs

Dual endoderm/mesoderm (A-E)

- **(A)** Overview of the differentiation protocol (protocol 1)
- **(B)** Expression of endoderm (FOXA2,SOX17)/mesoderm (Brachyury) markers by qRT-PCR at day 4 of each protocol (see methods) (n=3).
- **(C)** Cdx-2 (green) and DAPI (blue) of differentiated iPSC cultures at day 8 (protocol 1), representative of n=6 independent experiments.
- **(D)** Differentiated cells were seeded onto collagen hydrogels (2mg/mL) and cultured for 3 weeks and stained for E-cadherin, Cdx-2, villin and vimentin expression by immunohistochemistry, representative of n=3 independent experiments.
- **(E)** Cells were differentiated according to the protocols shown and then compared by mRNA-sequencing (n=6). Volcano plots illustrate top differentially expressed genes and specific epithelial and mesenchymal subsets.

Cell separation

- **(F)** Cells were differentiated according to protocol 1 and then separated on EpCAM expression (MACS) and profiled by flow cytometry/mRNA-sequencing. Representative flow cytometry plot following separation by MACS of separated populations (n=3). Volcano plots demonstrate top genes enriched in EpCAM+ and EpCAM- populations.

Single-cell population identities

- **(G)** Overview of study design.
- **(H)** Uniform manifold approximation and projection (UMAP) clustering of all samples (sample IDs correspond to 1G).
- **(I)** Automated annotation using the CellTypist algorithm of each cell population according to the “Developing Human Organs” model.
- **(J)** Dotplot of marker genes grouped by CellTypist annotations.
- **(K)** Left: Heatmap showing expression of Hox genes in each sample, arrows indicate posterior Hox genes. Right: feature plots of Caudal homeobox genes (Cdx1,2,4).
- **(L)** Feature plots of marker genes corresponding to smooth muscle and fibroblast lineages, Wnt/BMP/FGF and Hedgehog signalling pathways and extracellular matrix.
- **(M)** Stacked violin plot showing scores for mature intestinal epithelial cell populations (derived from Elmentaite et al. (18)) in the “intestinal epithelial” population.
- **(N)** Comparison of mesenchyme with (sample 4) and without (sample 3) purmorphamine treatment. Top: feature plots of stem-cell niche marker genes expressed in mesenchyme (21). Bottom: dotplot of key signalling and marker genes in each of the treated and control samples.

Murine subcutaneous engraftment

- **(O)** Experimental overview.
- **(P)** Macroscopic appearance of transplant before engraftment (top), during transplantation (middle) and at sacrifice (bottom).
- **(Q)** Images showing graft maturation over 2 – 4 weeks showing H+E stained tissue and Cdx-2 immunohistochemistry.
- **(R)** Intestinal epithelial maturation at 4 weeks demonstrated by immunohistochemistry for Chromogranin A, Villin, MUC2 and Ki67 and AB-PAS stain for mucus.
- **(S)** Stromal cell maturation and organisation at 4 weeks demonstrated by immunohistochemistry for smooth muscle actin, vimentin, PDGFRB and desmin.
- **(T)** Human origin of cell populations and functional human endothelium anastomosing with murine host shown by H+E stain, CD31 and human specific anti-human mitochondrial stain.

Nanostring Digital Spatial Profiling

- **(U)** Pan-cytokeratin and DAPI stained iPSC-derived transplants at 2, 3 and 4 weeks and adult human colon control tissue (n=1). All markers below correspond to the same tissues.

(**V-Z**) Endothelial cell marker, CD34, expression in each of the areas of illumination (Adult colon - CK- = cytokeratin negative cells, A: apex of crypt, B: base of crypt). (**W**) Mesoderm marker, TWIST1, expression across timepoints of *in vivo* growth. Gli1 (**X**), IHH (**Y**) and PDGFRA (**Z**) expression across timepoints of *in vivo* growth and spatial organisation in adult colon.

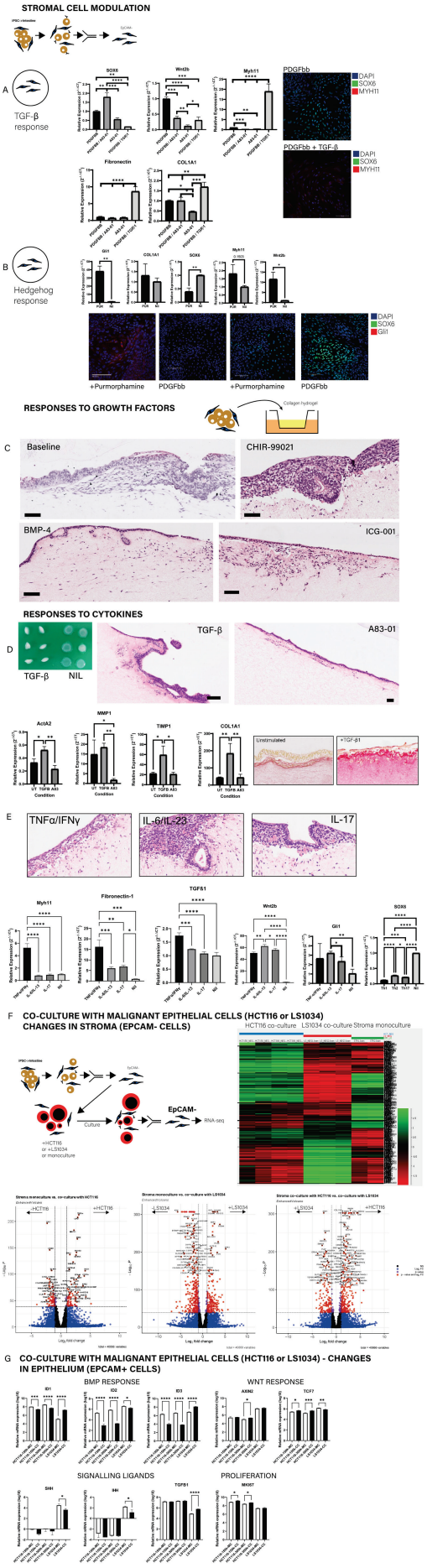


Fig. 2 Versatility of hiPSC-derived intestinal cell populations for modelling intestinal biology

(A) Separated EpCAM- cells were treated with PDGFbb (2ng/mL) and TGF-β1 (1ng/mL) or A83-01 (500nM) for 1 week and then assessed for gene expression (SOX6, Wnt2b, Myh11, fibronectin, Col1a1) by q-RTPCR (n=3).

(B) Separated EpCAM- cells were treated with PDGFbb (2ng/mL) and purmorphamine (10 μM) for 1 week and then assessed for gene expression (Gli1, Col1a1, SOX6, Myh11, Wnt2b) by q-RTPCR (n=3).

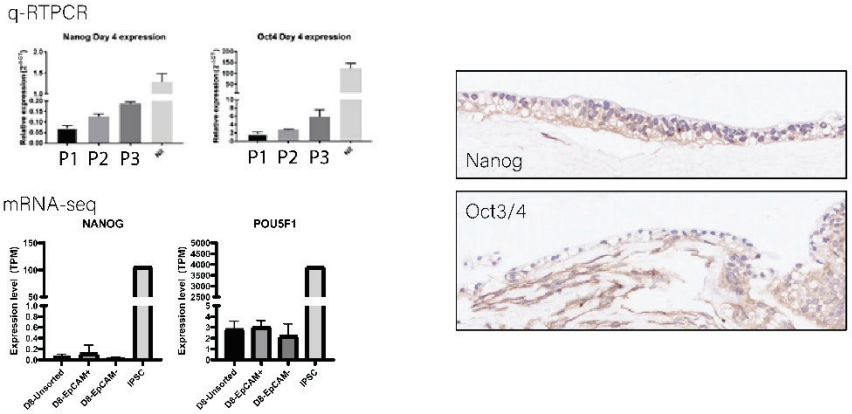
(C) hiPSC-derived intestinal cells were seeded onto collagen hydrogels and then cultured for three weeks in the presence of CHIR-99021 (5 μM), BMP-4 (5 ng/mL) or ICG-001 (1 μM). Representative H+E stained photomicrographs of at least two independent replicates. Scale bars = 100 microns.

(D/E) hiPSC-derived intestinal cells were seeded onto collagen hydrogels and then cultured for three weeks in the presence of TGF-β1 (1 ng/ML), A83-01 (500 nM), TNF-α+IFNγ (0.1ng/mL), IL-6/IL-23 (1ng/mL) or IL-17A (1ng/mL). Representative H+E or PSR stained photomicrographs of at least two independent replicates. Scale bars = 100 microns. Gene expression was evaluated by q-RTPCR.

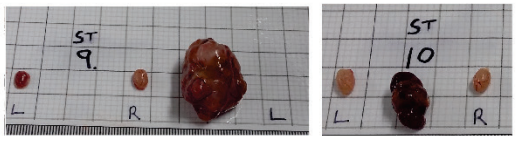
(F) mRNA-sequencing of EpCAM- cells following co-culture with HCT116 or LS1034 colorectal cancer cell lines. Top left: experimental overview, top right: hierarchical clustering and heatmap of gene expression in each condition (see Ext. Fig 4 for cluster details). Bottom: volcano plots of differential expression in the stroma in each condition.

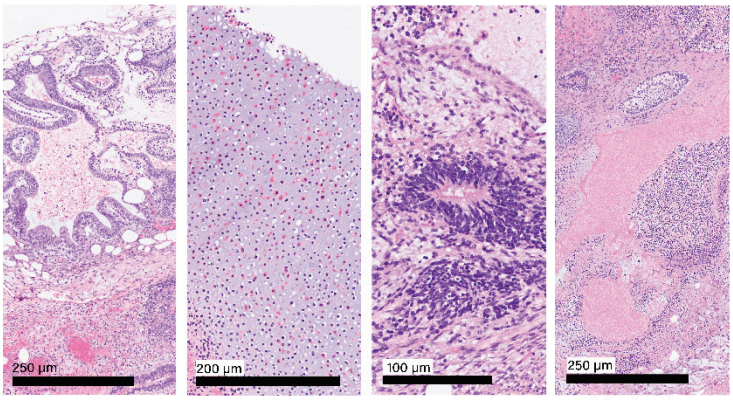
(G) Relative expression (log-transformed TPM) of selected genes in HCT116/LS1034 following mono-culture or co-culture with stroma. Significance determined by unpaired t-test between monoculture and co-culture for each condition (n=2 or 3).

Extended Figures and Supplementary material

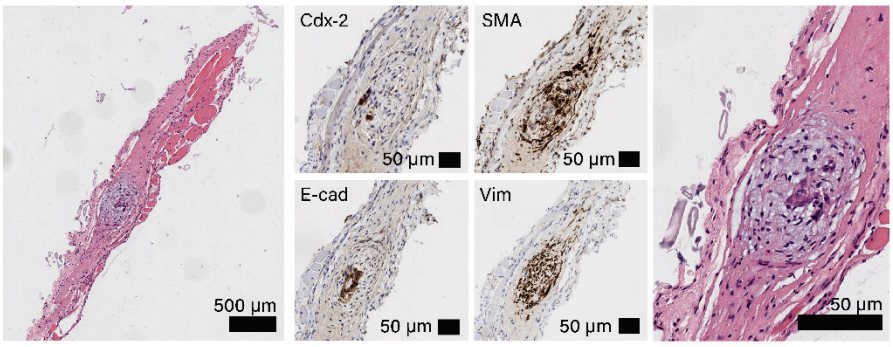


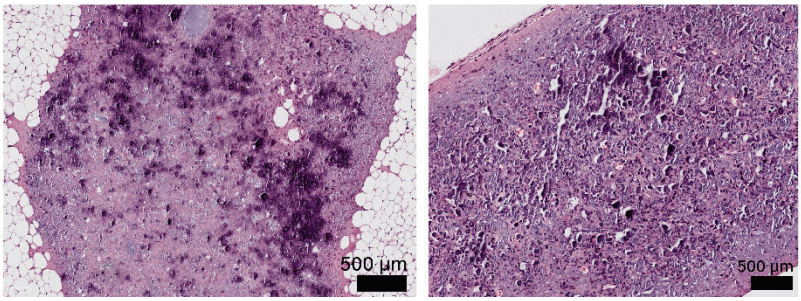
Undifferentiated iPSCs





Differentiated iPSCs





Ext. Fig. 1 Gene expression and in vivo teratoma assay demonstrates no residual pluripotency in hiPSC-derived intestinal cells.

mRNA profiling

q-RTPCR analysis of markers of pluripotency (Nanog/Oct4) after 4 days treatment with protocols indicated (see methods). Bulk mRNA sequencing was performed on cells differentiated using the novel protocol at day 8 and unseparated and separated EpCAM+ and EpCAM- populations, and undifferentiated hiPSCs. Differentiated cells cultured on collagen hydrogels probed using antibodies against Nanog and Oct3/4 (n=3), representative photomicrographs.

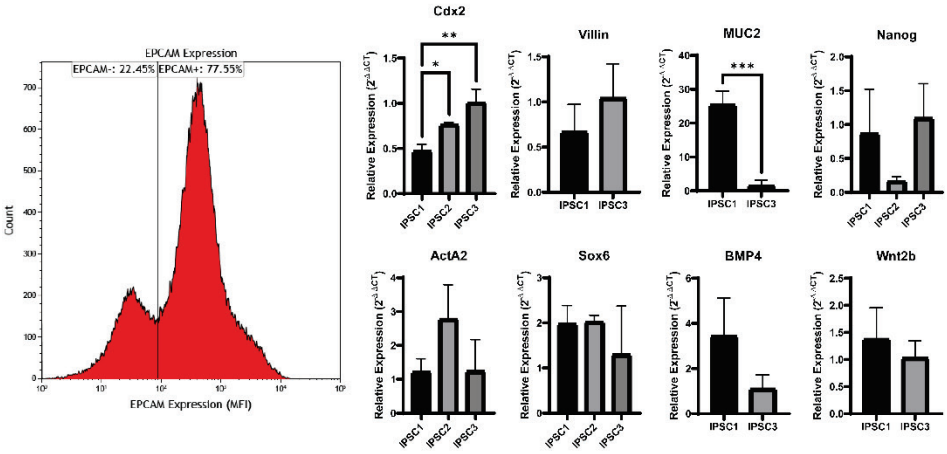
Teratoma assay

Upper row: macroscopic photographs of tumours around the testicular subcapsule grown in immunosuppressed (R2G2) mice following implantation with undifferentiated hiPSCs.

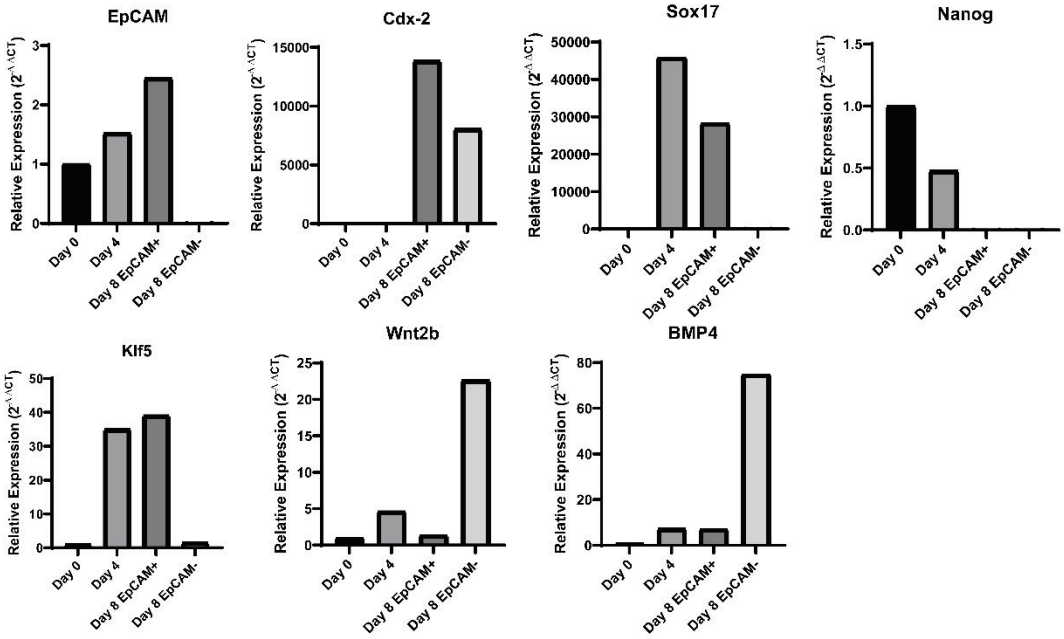
Middle row: Representative photomicrographs (of twelve images) of H&E stained sections demonstrating features of tissues derived from all three germ layers. From left to right: endodermal / glandular tissue; cartilage (mesoderm); neural blasteme / Homer-Wright rosette (ectoderm); lakes of necrosis. Scale bars represent distances indicated in individual photomicrographs.

Bottom two rows: All post-mortem tissues were examined by histology. Representative photomicrographs from each specimen where any possible engrafted tissue are shown. Tissue sections were stained with Cdx-2, E-cadherin, SMA and vimentin of viable hiPSC-derived intestinal cells. Scale bars represent distances indicated in individual photomicrographs.

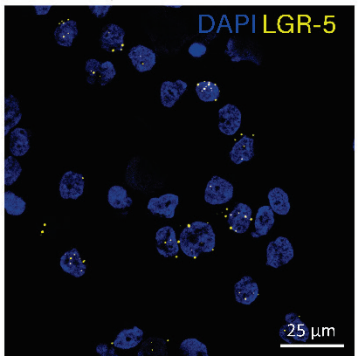
A Reproduction of protocol across independent iPSC lines



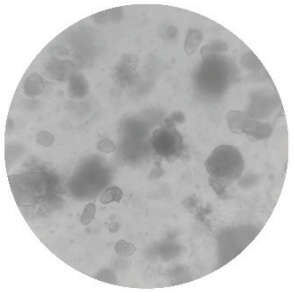
B Reproduction of protocol at an independent centre



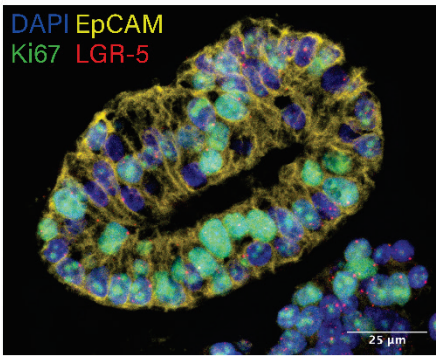
RNA-scope



Organoids



Immunofluorescence



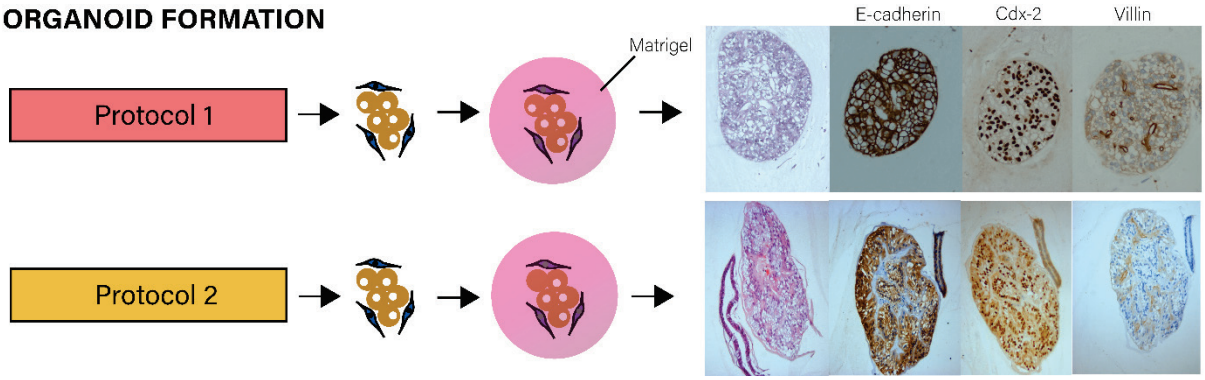
Ext. Fig. 2 The differentiation protocol is reproducible across different hiPSCs and laboratories

A Reproducibility over dual intestinal epithelial and mesenchymal differentiation across multiple hiPSC lines

The co-differentiation protocol was applied to three independent hiPSC lines derived from both males (1) and females (2). Cell populations showed both EpCAM+ and EpCAM- cells (representative flow cytometry histogram from one alternative iPSC line). Evidence of dual intestinal epithelial (Cdx2, Villin, MUC2) and mesenchymal differentiation (ActA2, Sox6, BMP4 and Wnt2b) was demonstrated as well as suppression of Nanog (comparison with hiPSC3, the hiPSC line used in other data figures).

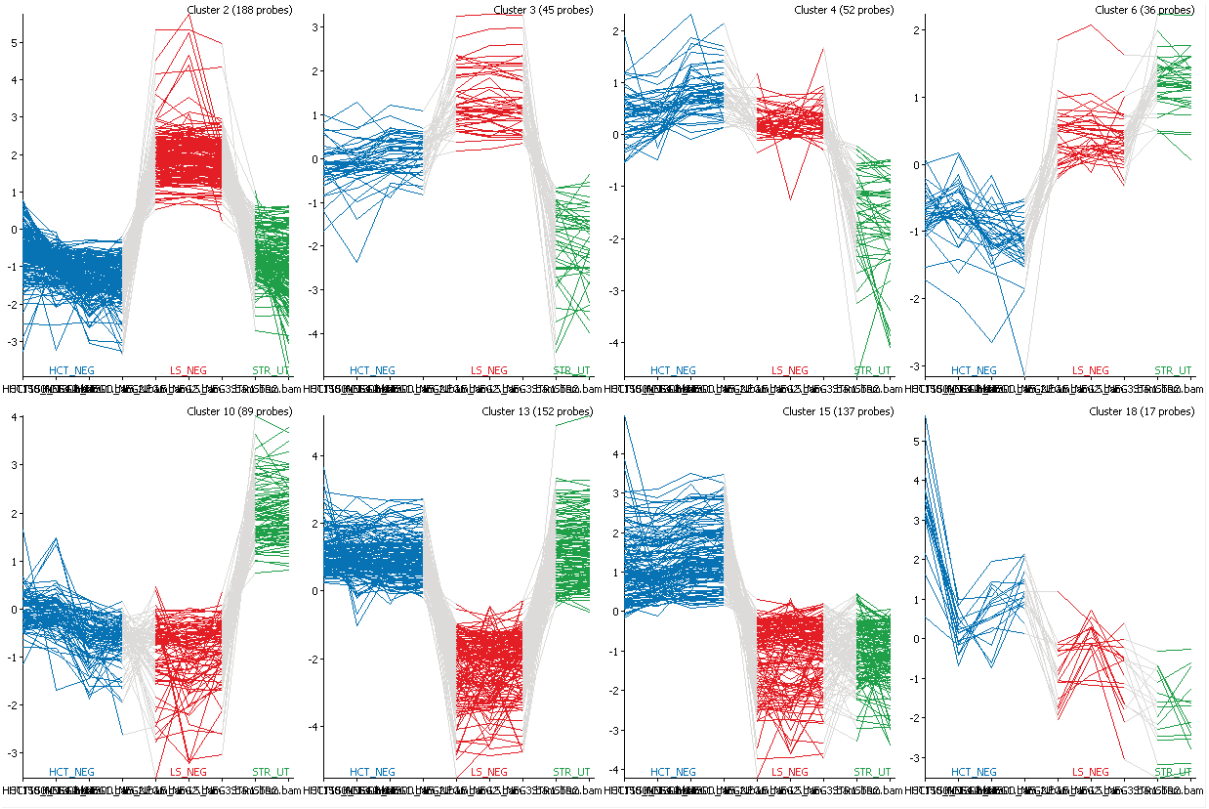
B Reproduction of the protocol at an independent centre

The co-differentiation protocol was applied to an independent hiPSC line by collaborators at an independent research facility (SM,SI,IT). Evidence of intestinal epithelial (EpCAM, Cdx-2, Sox17, Klf5) and mesenchymal (Wnt2b, BMP4) differentiation was shown as well as suppression of pluripotency (Nanog). EpCAM+ cells were evaluated for human LGR-5 mRNA by RNA-scope (25% across 3 independent replicates) and were capable of forming organoids with a mixture of cell types (EpCAM – yellow, Ki67 – green, LGR-5 – red, DAPI – blue), representative of n=6. After MACS separation, EPCAM^{+} cells were fixed on a coated slide using cytospin and stained with RNAScope probe for Human *LGR5* mRNA. 7500 cells from 3 independent replicates were counted; 25% of EpCAM+ were positive for LGR-5, while no cells were positive for LGR-5 at day 0.

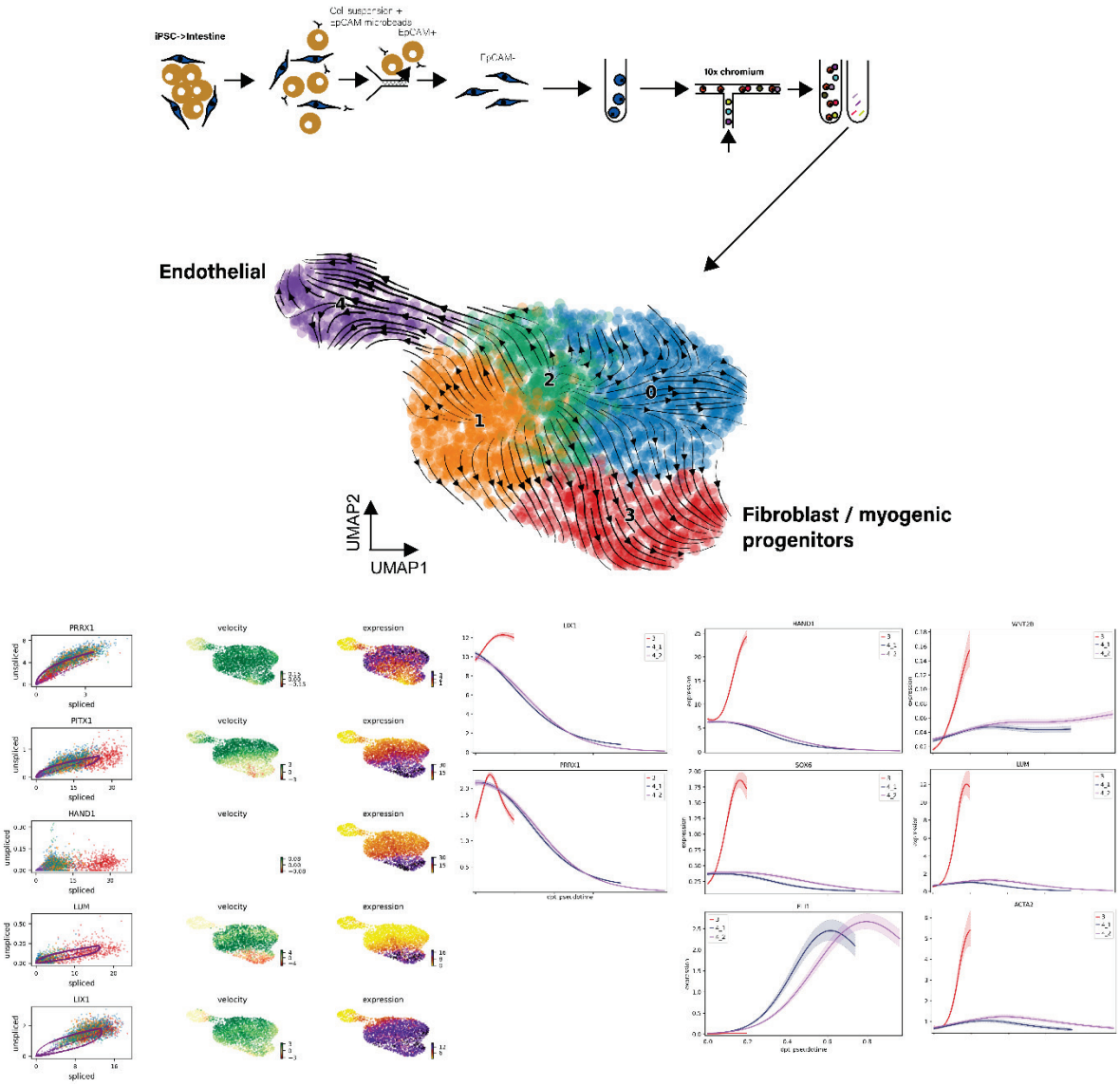


Ext. Fig 3 Embedding hiPSC-derived intestinal progenitors in Matrigel® demonstrates ability to form organoids.

Intestinal cell populations were embedded within Matrigel following differentiation for eight days with either Protocol 1 (top panel) or Protocol 2 (bottom panel). Sections of the resulting organoids were made and stained with H+E and immunohistochemistry for E-cadherin, Cdx-2 and Villin was performed. In both conditions, intestinal organoid structures were formed including formation of luminal structures, as demonstrated by brush-border villin positivity (n=3).



Ext. Fig. 4: Clusters of gene expression in stromal (EpCAM-) cells with monoculture (STR_UT) or co-culture with HCT116 (HCT_NEG) or LS1034 (LS_NEG), related to Fig. 2F. Genes in each cluster are detailed in Supplementary Table 1.



Ext. Fig 5: Trajectory pseudotime and RNA velocity modelling of early mesoderm lineage commitment to fibroblast/smooth muscle progenitors and endothelial progenitors (day 8 EpCAM- clusters isolated from the main single-cell RNA sequencing object).

869 **Suppl. Table 1: Gene expression clusters identified in EpCAM- cells following co-culture**

870 **/ monoculture with HCT116 or LS1034.**

871 See Excel file.

872

873 **Suppl. Table 2: culture medium and supplements**

874 Please contact the corresponding authors for further details.

875 **Suppl. Table 3: qRT-PCR primers**

Primer name	Sequence 5' – 3'
Cdx-2	Fwd: GGCAGCCAAGTGAAAACCA Rvs: AGCGACTGTAGTGAAACTCCT
Nanog	Fwd: CAATGGTGTGACGCAGGGAT Rvs: TGCACCAGGTCTGAGTGTTC
Oct3/4 (POU5F1)	Fwd: CAAAGCAGAAACCCTCGTGC Rvs: CTCGGACCACATCCTTCTCG
Sox17	Fwd: CGGGGACATGAAGGTGAAGG Rvs: ACGACTTGCCCAGCATCTTG
FoxA2	Fwd: GGGAGCGGTGAAGATGGA Rvs: TCATGTTGCTCACGGAGGAGTA
Brachyury	Fwd: GCTCACCAACAAGCTCAACG Rvs: AGTTGTCAGAATAGGATTGGGAG
Villin	Fwd: TGGTGTGGGAAGGGTTGTAG Rvs: GGGGGTGATGACCAGGTTTT
SOX6	Fwd: TAAGCAACTGATGAGGTCTC Rvs: AGGCGATGGTGTGGTAGTT
Wnt2b	Fwd: CTGACCTGATGCAGACGCAAG Rvs: AGGAGCCACCTGTAGCTCTCATGTA
BMP4	Fwd: GCCCGGAAGCTAGGTGAGT Rvs: CAGGAATCATGGTGTCTTGACAGA
Myh11	Fwd: AAACAGGGGATGGACGCAG Rvs: TGGTGGCATTCATATGGGCG
Gli1	Fwd: GTGCAAGTCAAGCCAGAACA Rvs: ATAGGGGCCTGACTGGAGAT
ActA2	Fwd: GCCAAGCACTGTCAGGAATC Rvs: TTGTCACACACCAAGGCAGT
COL1A1	Fwd: GATTCCCTGGACCTAAAGGTGC Rvs: AGCCTCTCCATCTTTGCCAGCA
FN1	Fwd: GGGCAACTCTGTCAACGAAG Rvs: GAGACATGCTTGTTCCTCTGG
MMP1	Fwd: AAGATGAAAGGTGGACCAACAATT Rvs: CCAAGAGAATGGCCGAGTTC
TIMP1	Fwd: CGG GGC TTC ACC AAG ACC Rvs: TCA GGC TAT CTG GGA CCG C
PBGD (Reference)	Fwd: GGAGCCATGTCTGGTAACGG Rvs: CCACGCGAATCACTCTCATCT

876

877 **Suppl. Table 4: Antibodies**

Antibody name	Manufacturer	Concentration	Species
Cdx-2 (EPR2764Y) (monoclonal)	ThermoFisher	1/100	Rabbit
SMA-α (ab5694) (polyclonal)	Abcam	1/400	Rabbit
E-cadherin (24E10) (monoclonal)	Cell Signalling	1/200	Rabbit
Villin (SP145) (monoclonal)	Abcam	1/200	Rabbit
Chromogranin A (ab45179) (polyclonal)	Abcam	1/500	Rabbit
Vimentin (D21H3) (monoclonal)	Cell Signalling	1/200	Rabbit
Nanog (D73G4) (monoclonal)	Cell Signalling	1/100	Rabbit
Oct4 (2750) (monoclonal)	Cell Signalling	1/100	Rabbit
MUC2 (EPR6145) (monoclonal)	Abcam	1/400	Rabbit
SOX6 (sc-393314) (monoclonal)	Santa-Cruz Biotechnology	1/100	Mouse
Gli1 (JF09-08) (monoclonal)	ThermoFisher	1/100	Rabbit
PDGFRB (G.290.3) (monoclonal)	Invitrogen	1/100	Rabbit
Human-specific mitochondria (ab92824) (monoclonal)	Abcam	1/1000	Mouse
Anti-Rabbit IgG (H+L), AF488	Invitrogen	1/100	Chicken
Anti-Rabbit IgG (H+L), AF568	Invitrogen	1/100	Donkey
Anti-Mouse IgG (H+L), AF594	Invitrogen	1/100	Goat

878

879 *CD31, desmin and Ki67 performed by Nottingham University Hospitals NHS Trust Cellular

880 Pathology Department, datasheets and SOPs available upon request.