Colon cancer arises from differentiated cell lineages in the context of inflammation

Mathijs P. Verhagen[1], Rosalie Joosten[1], Mark Schmitt[1,2], Niko Välimäki[3,4], Andrea Sacchetti[1], Kristiina Rajamäki[3,4], Jiahn Choi[5], Paola Procopio[2], Sara Silva[2], Berdine van der Steen[6], Thierry P.P. van den Bosch[1], Danielle Seinstra[1], Michail Doukas[1], Leonard H. Augenlicht[5], Lauri A. Aaltonen[3,7], and Riccardo Fodde[1]*

[1]Department of Pathology, Erasmus University Medical Center, Rotterdam, The Netherlands.

[2]Institute of Pharmacology, University of Marburg, Germany.

[3]Department of Medical and Clinical Genetics, University of Helsinki, Helsinki, Finland.

[4]Applied Tumor Genomics Research Program, Research Programs Unit, University of Helsinki, Helsinki, Finland.

[5]Department of Cell Biology, Albert Einstein College of Medicine, New York, U.S.A.

[6]Department of Otorhinolaryngology and Head & Neck Surgery, Erasmus University Medical Center, Rotterdam, The Netherlands.

[7]iCAN Digital Precision Cancer Medicine Flagship, University of Helsinki, Helsinki, Finland.

*to whom correspondence should be addressed at:

Erasmus MC, Department of Pathology, PO Box 2040, 3000 CA Rotterdam, The Netherlands

E-mail: r.fodde@erasmusmc.nl Tel: +31 10 7043896

Abstract

According to conventional views, colon cancer originates from stem cells. However, inflammation, a key risk factor for colon cancer, was shown to suppress intestinal stemness. Here, we employed Paneth cells (PCs) as a model to assess the capacity of differentiated lineages to trigger tumorigenesis in the context of inflammation. Upon inflammation, PC-specific *Apc* mutations led to intestinal tumors reminiscent not only of those arising in inflammatory bowel disease (IBD) patients but also of a larger fraction of sporadic colon cancers. The latter is likely due to the inflammatory consequences of Western-style dietary habits, the major colon cancer risk factor. Computational methods designed to predict the cell-of-origin of cancer confirmed that, in a substantial fraction of sporadic colon cancers the cells-of-origin are secretory lineages and not stem cells.

One-Sentence Summary: Secretory cell lineages trigger tumor formation in the context of the major etiologic colon cancer risk factors.

Paneth cells as the origin of intestinal tumors in the context of inflammation and loss of *Apc*.

The origin of the vast majority of cancers is thought to reside in stem- or progenitor-like cells which satisfy the needs for active proliferation, self-renewal, and differentiation capacity. This was demonstrated in the intestine where loss-of-function mutations in the *Apc* tumor suppressor gene successfully initiate adenoma formation only when they occur in *Lgr5*$^{+}$ CBCs. When the same *Apc* mutation is introduced in more committed and shorter-lived transit-amplifying cells, tumor formation is halted at early micro-adenoma stages(*1*). However, next to this "bottom-up" scenario, additional "top-down" models of intestinal tumorigenesis have been proposed where committed intestinal cells initiate cancer especially in the context of tissue injury and inflammation(*2, 3*). In the specific case of colon cancer, western style dietary habits and chronic inflammation, two of the major etiologic factors associated with increased risk for sporadic malignant disease in the digestive tract, are thought to induce specific cellular and molecular alterations of the intestinal epithelium which ultimately lead to the expansion of cell targets for tumor initiation and progression(*2, 4*). However, which specific cell lineages are capable of dedifferentiating upon tissue injury and which are the underlying mechanisms remain largely unclear.

Paneth cells are specialized secretory cells located at the very bottom of the crypt of Lieberkühn in the small intestine where they secrete antimicrobial peptides into the lumen(*5*). Moreover, they provide essential physical support and secrete signals to ensure *Lgr5*$^{+}$ stem cell function(*6*). In the colon, where Paneth cells are not present, *cKit*$^{+}$ Paneth-like cells, also known as deep crypt secretory (DCS) cells, play secretory and niche-like functional roles analogous to those of *bona fide* PCs in the small intestine(*7, 8*).

We first reported on the ability of Paneth cells to re-enter the cell cycle and de-differentiate upon irradiation and inflammation to acquire stem cell-like features and contribute to the tissue regenerative response(*9-11*). Consequently, we here questioned whether Paneth cells could be the origin of intestinal cancer upon inflammation. To this aim, we bred mice carrying Lox-alleles at genes most frequently mutated along the adenoma-to-carcinoma sequence, namely *Apc(12), Kras(13),* and *Trp53(14)*, each combined with a Cre specific for *Lgr5*$^{+}$ CBCs (*Lgr5*$^{CreERT2\text{-}EGFP}$)(*15*) or for Paneth cells (*Lyz1*^{CreERT2})(*16*). Following Cre activation by tamoxifen, DSS was administered through the drinking water to model inflammation (Fig. 1a). In the absence of DSS-induced inflammation, PC-specific single gene mutations did not give rise to intestinal tumors. In contrast, loss of *Apc* in *Lgr5*$^{+}$-CBCs transformed crypts into β-catenin^{hi} foci that grew into multiple adenomas 4-6 weeks after Cre induction (Fig. 1b). When single gene mutations were combined with DSS administration, *Apc* loss in Paneth cells resulted in increased nuclear and cytoplasmic β-catenin expression eventually leading to the formation of PC-derived adenomas (Fig. 1c). Of note, Paneth-specific *Kras* or *Trp53* mutations did not result in tumor formation even in the presence

45 of the inflammatory stimulus (Fig. 1d). However, the compound loss of *Apc* and oncogenic activation of
46 *Kras* in Paneth cells resulted in a striking increase in tumor multiplicity (6.1 fold) even in the absence of
47 DSS (6.9 fold)(Fig. 1e). The combination of *Apc* and *Tp53* mutations in PCs also led to an increase in tumor
48 multiplicity upon DSS administration (1.6 fold), though to a lesser extent when compared with the
49 compound *Apc*/*Kras*-mutant genotype, possibly indicating a distinct mechanism underlying tumor onset
50 in these mice. Indeed, phospho-histone H2A.X (Ser 139) IHC analysis confirmed an increase in DNA
51 damage and chromosomal instability in the *Tp53*-mutant tumors (Suppl. Fig. 1a). Targeting all three genes
52 in Paneth cells resulted in a very aggressive phenotype with high tumor multiplicity (10.1 fold) in the
53 absence of the inflammatory stimulus (Fig. 1e). When compared to *Apc*-driven tumors originated in PCs,
54 the histology of adenomas from mice in which two or three genes were targeted revealed a progressive
55 increase in dysplasia and invasive morphology (Suppl. Fig. 1b). Distribution of adenomas along the small
56 intestine also followed distinct patterns with a prevalence of duodenal tumors in compound *Apc*/*Kras*
57 tumors regardless of DSS (Suppl. Fig. 2a).
58 To validate the Paneth cell origin of the observed intestinal tumors, we bred *Lyz1*^{CreERT2} mice with R26^{LSL-}
59 ^{tdTomato} or R26$^{LSL-YFP}$ reporters and traced their lineage upon tamoxifen-driven targeting of the *Apc, Kras,*
60 and *Tp53* mutations. As shown in Figure 1f and Suppl. Fig. 2b, this confirmed the Paneth cell origin of the
61 corresponding tumors by capturing the process from microscopic lesions to adenoma formation.
62 Overall, these results demonstrate that Paneth cells can initiate intestinal adenomas upon genetic
63 ablation of *Apc* in the context of inflammation. In combination with *Apc* loss, activation of oncogenic *Kras*
64 or loss of *Tp53* function rescues the need for an inflammatory stimulus and results in increased PC-derived
65 tumor multiplicities and progression to malignant phenotype.

66 Next, we characterized lineage-specific markers in the PC- and CBC-originated tumors by
67 immunohistochemistry. Of note, while cells expressing the PC marker lysozyme (Lyz1) were notable in
68 *Lgr5*-derived tumors (*Lgr5*/*Apc*: 30.0 % ± 18.5 positive tumor cells), they were nearly absent in adenomas
69 that originated from Paneth cells (*Lyz1*/*Apc*: 0.48 % ± 1.16)(Fig. 1g-h). The opposite was observed for Dclk1
70 (doublecortin like kinase 1), a Tuft(*17*) and tumor stem cell marker(*18, 19*), that was more frequently
71 detected among PC-derived adenomas (*Lyz1*/*Apc*: 54.1 % ± 10.5) when compared with *Lgr5*-derived
72 tumors (*Lgr5*/*Apc*: 15.6 % ± 15.7)(Fig. 1i-j). Other lineage-specific markers for entero-endocrine (Chga),
73 goblet (Muc2), and stem cells (Olfm4) showed variable levels without clear-cut differences among tumors
74 with different cells-of-origin (Suppl. Fig. 2c). The increased *Dclk1* expression in PC-derived tumors is of
75 interest in view of its association with increased immune and stromal infiltration in colon cancer(*20*).

76 To confirm these results at the transcriptional level, expression of the *Lyz1* and *Dclk1* genes was analyzed
77 by RTqPCR (Fig. 1k). Indeed, *Lyz1* expression was lower in Paneth-derived tumors (*Lgr5/Apc* vs. *Lyz1/Apc*:
78 $\log_2$FC = 2.64, P_{val} = 7.5e-4) when compared to *Lgr5*-derived tumors. *Dclk1* expression was very low and
79 variable at the RNA level, and did not show significant differences across the groups.
80 To assess the relative activation levels of the Wnt signaling pathway among the different tumor groups,
81 we measured expression levels of *Axin2*, a well-established Wnt downstream target. *Axin2* expression was
82 higher in *Lgr5*- compared to PC-derived tumors (*Lgr5/Apc* vs. *Lyz1/Apc*: $\log_2$FC = 2.12, P_{val} = 0.017)(Fig. 1k).
83 Moreover, both *Kras* oncogenic activation and inflammation gradually increased *Axin2* levels in PC-
84 derived tumors, in agreement with the previously reported synergism between *Apc* and *Kras* mutations
85 in the activation of the Wnt pathway(*21*).
86 Thus, upon tumorigenesis, Paneth cells de-differentiate to a state that hampers secretory
87 differentiation leading to specific patterns of tumor histology and gene expression distinct from that of
88 *Lgr5*-derived tumors.

89 **Paneth cells dedifferentiate into revival stem cells upon enhanced Wnt signaling activation.**
90 To elucidate the mechanisms which underlie the conversion of PCs into cells of origin of small intestinal
91 tumors in the context of inflammation and/or of specific genetic hits, we combined scRNAseq analysis
92 with lineage tracing. To this aim, we induced the *Apc*, *Kras*, and *Tp53* genetic mutations in the $R26^{LSL-}$
93 ^{tdTomato}/*Lyz1*^{CreERT2} (or $R26^{LSL-YFP}$) reporter strains in the presence or absence of DSS (Fig. 2a). Subsequently,
94 cells were harvested from the intestinal epithelium, purified by FACS, and transcriptionally profiled by
95 scRNAseq (Methods; Suppl. Fig. 3). After preprocessing, we obtained the transcriptomes of 23.231
96 epithelial cells from 32 mice, distributed over the different lineages of the intestinal epithelium (Fig. 2b).
97 Close examination of the cells positive for the reporter genes revealed novel clusters of PCs that arise
98 upon DSS administration and the specific gene mutations, but were not observed among Paneth cells in
99 homeostatic conditions (PC cluster 1-4, Fig. 2c).
100 To characterize the novel PC-derived states, we performed differential expression analysis and identified
101 cluster-specific markers (Fig. 2d)(Suppl. Table 1). While PC Cluster 1 appeared at low frequency across
102 different genotypes, PC Cluster 2 arises directly upon exposure to the inflammatory stimulus. Both PC
103 Clusters 1 and 2 are characterized by increased expression of two markers of radio-resistant and secretory
104 progenitors with self-renewal capacity during regeneration, namely *Krt19(22)* and *Atoh1(23)*, while
105 increased expression of *Reg3b*, known for its protective role in the development of colitis and ileitis(*24*),
106 and *Cdkn1a* (p21), a marker of terminal differentiation in the intestine(*25*), was observed in Cluster 2
107 compared to 1.

108 PC Cluster 3 became apparent in mice carrying *Apc* mutations (7.23% ± 5.77 of traced cells) alone and in
109 combination with DSS treatment (16.40% ± 2.10 of traced cells), and in double and triple mutant (AP, AK,
110 AKP) animals, though not in mice carrying single *Kras* or *Tp53* mutations. PC Cluster 3 showed increased
111 expression of *Gif* (gastric intrinsic factor), *Cd81*, a tetraspanin family member known to mark the response
112 to γ–irradiation and correlated with the expression of ISC- and proliferation genes(*26*), and *Prom1*
113 (prominin 1, also known as Cd133), a well-established colon cancer stem cell marker(*27*).
114 PC Cluster 4 consisted of cells from mice where double (AK) and triple (AKP) mutations were targeted to
115 Paneth cells (23.25% ± 11.23 and 52.26% ± 28.23 of traced cells, respectively). Increased expression of
116 *Anxa2* (annexin 2), a functional marker of inflammatory response, and *Clu* (clusterin), previously shown
117 to earmark revival stem cells (RSCs) upon γ-irradiation(*28*), feature PC Cluster 4. Accordingly, evaluation
118 of the RSC signature showed elevated expression among the PC clusters (Fig. 2e), and *in situ* hybridization
119 analysis in PC-derived tumors from mice carrying compound mutations (AK and AKP)(Fig. 2f and Suppl.
120 Fig. 4a-b) confirmed the increased *Clu* expression. Finally, pathway analysis revealed the similarities
121 between the PC-derived Cluster 4 and RSCs, both earmarked by the activation of Yap1 signaling and
122 specific inflammatory pathways (Fig. 2g). Compared to RSCs, PC-derived and *Apc/Kras*-mutant cells from
123 Cluster 4 showed increased levels of Tgf-β and Wnt signaling (Suppl. Fig. 4c).
124 Thus, upon genetic targeting or inflammatory stimulus, PCs escape their homeostatic identity and acquire
125 distinct cellular features as shown by scRNAseq and FACS analysis (Suppl. Fig. 4d-e). Of note, DSS
126 treatment led to a lower expression of *Lgr5* and *Ascl2* in stem cells, as well as a lower association of the
127 CBC signature, confirming ours and others' previous observations according to which resident stem cells
128 lose their multipotency upon acute inflammation (Suppl. Fig. 4f)(*11*).
129 Collectively, these findings demonstrate that PCs efficiently dedifferentiate upon genetic targeting or
130 inflammatory stimulus leading to distinct cellular identities. During tumorigenesis driven by *Apc/Kras*, PCs
131 share features with the Yap1-dependent revival stem cell identity, and further activate Tgf-β and Wnt
132 signaling in their conversion to bona fide tumor cells.

133 **Transcriptomic comparison of Paneth- and *Lgr5*-derived tumors reveals a dichotomy in stem cell**
134 **phenotypes.**

135 To investigate the consequences of the cell-of-origin identity on the transcriptional profile of the
136 resulting intestinal tumors, we performed bulk RNA sequencing of macroscopically dissected lesions
137 originated from ISCs and PCs (Fig. 3a). Principal component analysis revealed that the major variance
138 component (61%) was attributed to differences in the cell-of-origin, while the impact of genotype or
139 inflammatory stimulus became notable in the second component of variation (10%)(Fig. 3b). Differential

140 expression analysis between tumors derived from PCs and ISCs in the same genetic and inflammatory
141 context (*Apc* and DSS) revealed tumor signatures specific for each cell-of-origin (Suppl. Table 2).
142 Gene set enrichment analysis (GSEA) indicated that tumors derived from ISCs were characterized by high
143 levels of Myc and Wnt signaling, while PC-derived adenomas showed higher levels of inflammatory
144 pathways indicative of infiltration from the tumor micro-environment (Fig. 3c, Suppl. Fig. 5). Of note, the
145 inflammatory characteristics of PC-derived tumors were observed also in mice where *Apc* and *Kras*
146 mutations were targeted to Paneth cells in the absence of DSS-driven inflammation (Fig. 3d), indicating
147 that specific mutant genotypes and the type of cell-of-origin can trigger tumor initiation by mimicking the
148 inflammatory context otherwise brought about by DSS.
149 Next, we employed the ISC index(*29*) (intestinal stem cell index; see Methods) to predict the relative
150 proportions of RSC and CBC stem cells in the intestinal tumors. In agreement with the scRNAseq analysis,
151 PC-derived tumors were RSC-enriched whereas the ISC-derived tumors consisted mainly of CBCs (Fig. 3e).
152 Notably, the highest RSC contribution was observed in tumors originated from *Apc/Kras*-mutant PCs in
153 the absence of inflammation, when compared with the equivalent genotype upon DSS administration.
154 The latter is of relevance to dissect the relative contribution of the inflammatory insult and of the somatic
155 mutations in the de-differentiation process leading to tumor initiation. As mentioned before, one of the
156 primary effects of various forms of tissue injury to the intestinal epithelium is the loss of resident stem
157 cells. In the Singh et al. (2022) study(*30*), ablation of *Lgr5*$^+$ ISCs was performed to study its consequences
158 by scRNAseq. Our analysis of these data sets revealed elevated expression of secretory genes in the
159 regenerated stem cells normally restricted to the Paneth cell lineage (Suppl. Fig. 6a). In parallel, upon
160 *Lgr5*$^+$ cell ablation, Paneth and goblet cells partially activate the RSC program (Suppl. Fig. 6b), indicating
161 that stem cell loss is sufficient to trigger plasticity and RSC reprogramming from PCs. Hence, in order to
162 assess whether the removal of the resident stem cells is sufficient to activate the lineage tracing capacity
163 of Paneth cells in the absence of the inflammatory injury, we implemented a similar DTR-based *Lgr5*$^+$
164 ablation experiments. As shown in Suppl. Fig. 6c-d, lineage tracing from Paneth cells was observed similar
165 to what previously shown upon inflammation(*11*).
166 Together, these results indicate that the cell-of-origin embodies the major source of inter-tumor
167 variability, and that the Paneth- or ISC-origin is reflected by the RSC- or CBC-like profile of the resulting
168 tumors, respectively. Notably, the inflammation-driven ISC loss activates de-differentiation at the
169 epithelial level.
170

171 **The transcriptional profile of Paneth-derived tumors mimics colitis-associated colorectal cancer.**

172 The small intestinal location of Paneth cells and of the tumors originating from them raise questions
173 on the relevance of our study for colon cancer. To explore the generalizability of our results, we first
174 analyzed bulk and single-cell RNAseq data from two distinct studies(*31, 32*) centered around the AOM/DSS
175 mouse model of colitis-driven colon cancer(*33*). This protocol relies on the oncogenic β-catenin mutations
176 caused by azoxymethane (AOM) which, in combination with the DSS-driven ulcerative colitis, result in
177 multiple adenocarcinomas in the distal colon. As shown in Suppl. Fig. 7a (by CiberSortx(*34*); see Methods),
178 based on their bulk RNAseq data(*31*) AOM/DSS-derived colon tumors encompass an abundant population
179 similar to the RSC-like PC Cluster 4. Moreover, analysis of the scRNAseq data(*32*) confirmed that
180 AOM/DSS-derived colon tumors, when compared to their sporadic counterparts, were distinct in terms of
181 the qualitative and quantitative composition of their tumor microenvironment, namely a pronounced
182 presence of infiltrating immune cells and tumor-associated fibroblasts (Suppl. Fig. 7b). Accordingly, tumor
183 cells derived in the context of AOM/DSS share transcriptional similarity with Paneth-derived tumor cells
184 and RSCs (Suppl. Fig. 7c-d).
185 Hence, notwithstanding the small intestinal location of the PC-derived tumors, their gene expression
186 signatures and overall inflammatory TME profiles are characteristic of colitis-associated carcinoma in the
187 mouse.

188 Next, in view of the marked differences in the transcriptional profiles between mouse intestinal tumors
189 with distinct cells of origin, we questioned whether similar differences distinguish human sporadic colon
190 cancers from those that arose in the context of IBD. To this aim, RNAseq profiles of human microsatellite
191 stable sporadic colon cancers (sCRC, N = 38) and from patients suffering from inflammatory bowel disease
192 (IBD-CRC, N = 14) were interrogated(*35*). GSEA of the most differentially expressed genes ($\log_2$FC > 5, P_{adj}
193 < 0.01) from the mouse tumors (*Lyz*T signature, N = 27 genes; *Lgr5*T signature, N = 40 genes) revealed a
194 significant association between the *Lyz*T signature and IBD-CRC (NES 1.67, P_{adj} 6.2e-3), while the *Lgr5*T
195 profile was significantly associated with sCRC (NES -1.62, P_{adj} 0.013)(Fig. 3f, Suppl. Fig. 8a-c, Suppl. Table
196 3). Evaluation of the hallmarks from the molecular signature database(*36*) revealed gene sets common to
197 PC-derived tumors and IBD-related CRCs (Interferon α/γ, inflammatory response, Il6/Il2 signaling, Kras,
198 Complement, Allograft rejection), and to *Lgr5*-derived tumors and sporadic CRCs (Myc targets, G2m
199 Checkpoint, E2F targets and Wnt β-catenin signaling)(Fig. 3g, Suppl. Fig. 8d).
200 We then compared the differentially expressed genes from PC- versus *Lgr5*-derived mouse tumors and
201 human IBD-CRC versus sCRC (Paneth/IBD-CRC, N = 49 genes, Lgr5/sCRC: N = 27 genes) and visualized their
202 expression across different cell types based on a large scRNAseq CRC study(*37*) (Fig. 3h). Of note, the

203 markers shared between PC-derived tumors and IBD-CRC were dominantly expressed in myofibroblasts
204 (e.g. *ITGAM*, *SLC1A3*), T cells (e.g. *SELPLG, LAX1*), and stromal cells (e.g. *CHRDL1, RELN*). In contrast, *Lgr5*-
205 derived tumors/sCRC markers were mostly observed in epithelial cells (e.g. *HOXB6, HOXB8, HOXB9,*
206 *AXIN2, ASCL2*), indicating the difference in stromal composition among these tumors (Fig. 3h).
207 Comparison of a set of gene signatures (Suppl. Table 4) in a large colon cancer cohort(*38*) confirmed the
208 presence of two distinct identities (Fig. 3i-j): a colitis-like identity enriched with RSCs and prevalent in
209 CMS4 (67%) and CMS1 (36%) tumors, and a sporadic-like identity enriched with CBCs and common in
210 CMS2 (55%) and CMS3 tumors (52%)(Fig. 3k-l, Suppl. Fig. 8e). Survival analysis revealed significant
211 differences in relapse free survival between the sporadic- and colitis-like CRC groups (Pvalue = 5.2e-
212 05)(Fig. 3m).
213 Thus, transcriptional signatures derived from small intestinal mouse tumors originated from Paneth cells
214 significantly overlap with those from human colon cancer that arose in the context of IBD, possibly
215 revealing a common cell-of-origin in secretory lineages.

216 **Western-style diet triggers an inflammatory response in intestinal secretory cells leading to their cell**
217 **cycle activation and dedifferentiation.**

218 The relative representation of patient-derived colon cancers whose expression profiles are reminiscent
219 of the PC-derived mouse tumors (~25%; Figure 3j) vastly exceeds the expected proportion of colon cancers
220 arising in patients with an history of clinically manifest IBD (1-2%)(*39*). One possible explanation for this
221 apparent discrepancy may come from the fact that WSD habits, often in combination with chronic over-
222 nutrition and sedentary lifestyles, have been associated with a state of chronic metabolic inflammation,
223 termed 'metaflammation'(*40*). In particular, a link between consumption of a diet high in fat and sugar
224 and Paneth cell dysfunction was recently demonstrated(*41*). Moreover, in a recent report by J.C. ad L.H.A.,
225 a purified mouse diet that mimics WSD habits and underlies the increased risk of colon cancer (NWD1)(*42*)
226 was shown to induce a low degree of chronic intestinal inflammation and other mechanisms that define
227 pathogenesis of human IBD(*43*). Therefore, we hypothesized that etiologic drivers of colon cancer other
228 than IBD, including widespread western-style dietary habits, may underlie dedifferentiation and tumor
229 onset mechanisms similar to those observed upon inflammation.
230 To provide support for this hypothesis, we first fed C57BL6/J mice for 3 months with the western-style
231 (NWD1) and control (AIN76A) diets and compared the transcriptional response of Paneth cells with that
232 obtained upon DSS administration (Fig. 4a). We examined genes upregulated upon inflammatory stimuli
233 (DSS signature; see Methods), which showed variable but overall increased levels in NWD1-fed mice when
234 compared to those on the control diet (Fig. 4b). Indeed, GSEA confirmed that the DSS signature was

235 associated significantly with Paneth cells exposed to NWD1 (NES 2.99, $P_{adj} < 0.001$), indicating that
236 western-style dietary habits trigger an inflammatory-like response in Paneth cells (Suppl. Fig. 9a). At the
237 gene ontology level, western-style diet activated signaling pathways related to cell cycle (G2M
238 Checkpoint) and proliferation (mitotic spindle, Myc targets; Suppl. Fig. 9b), suggesting that Paneth cells
239 re-enter the cell cycle upon long-term exposure to western-style diet, similar to what is observed upon
240 DSS-driven inflammation(*11*).
241 In view of the previously reported acquisition of stem-like features by PCs upon inflammation(*11*), we
242 next employed the organoid reconstitution assay (ORA)(*10, 44*) to assess whether similar effects are
243 exerted by the western-style diet. We first co-incubated Paneth and *Lgr5*$^{+}$ cells from the AIN76A- and
244 NWD1-fed mice in all 4 combinations. As shown in Figure 4c, Paneth cells from NWD1-fed animals
245 significantly improved organoid formation independently of their reconstitution with *Lgr5*$^{+}$ cells from
246 NWD1- or AIN76A-fed mice, possibly indicative of a paracrine effect enhancing the well-established niche
247 (ISC-supporting) role of Paneth cells. However, single (i.e. non reconstituted with *Lgr5*$^{+}$ CBCs) Paneth cells
248 from NWD1-fed mice formed organoids more efficiently when compared with either PCs from AIN76A-
249 fed mice or with *Lgr5*$^{+}$ ISCs from both groups of animals. These *ex vivo* results were further validated by
250 lineage tracing analysis of Paneth cells in R26$^{LSL\text{-}YFP}$*Lyz1*^{CreERT2} NWD1-fed mice that revealed extended YFP-
251 labeled ribbons thus confirming their de-differentiation and acquisition of stem-like features induced by
252 the western-style dietary cues (Fig. 4d).
253 To zoom in on the primary transcriptional response of Paneth cells to NWD1, we took advantage of the
254 scRNAseq data generated by J.C., L.H.C. and collaborators immediately upon exposure to NWD1(*43*)(Fig.
255 4a). Within 4 days of switching mice to the NWD1 diet, a subset of Paneth cells became apparent whose
256 transcriptional profile strongly associated with the DSS signature (labeled as "diet response cells" in Fig.
257 4e). Mirroring our previous observations obtained immediately upon DSS inflammatory stimulus, these
258 WSD-responsive cells increased their transcriptomic diversity as measured by CytoTRACE(*45*) (Fig. 4f; see
259 Methods). Already after the short exposure to NWD1, the diet-responsive cells acquired stem cell markers
260 while retaining some secretory features (Fig. 4g). Comparative pathway analysis between the
261 transcriptional response of Paneth cells to DSS and NWD1 revealed similar upregulation of Wnt, Myc,
262 Hedgehog, and G2M checkpoint signaling pathways (Fig. 4h).
263 As the observed WSD-driven gene expression changes are likely exerted through epigenetic modifications,
264 we next analyzed scATAC data obtained in the framework of the same Choi et al. (2023) study(*43*). Similar
265 to what observed by scRNAseq analysis, we identified a group of NWD1 diet-responsive Paneth cells with
266 significant epigenetic modifications including a main cluster (52%) on mouse chromosome 11 (synthenic

267 with human chromosomes 17 and 5) that encompasses a considerable fraction of genes encoding for
268 members of the Wnt, Pi3k/Akt and cell cycle pathways (Fig. 4i-k, Suppl. Fig. 9c-f). The latter were
269 previously shown by our laboratory to underlie Paneth cell dedifferentiation and acquisition of stem-like
270 features upon DSS-driven inflammation(*11*).
271 In order to assess the effects of the NWD1 diet on the mouse colon, we looked for proliferating Paneth-
272 like cells, also known as deep crypt secretory (DCS) cells(*7, 10*). Inflammation-driven cell cycle activation
273 in these allegedly post-mitotic lineages was previously observed in small intestinal Paneth cells of DSS-
274 treated mice and of Crohn's disease patients(*11*). By co-staining colonic tissues with the secretory lineage
275 marker WGA (wheat germ agglutinin)(*46*) and Ki67, a significant increase in the number of proliferating
276 secretory cells located at the crypt bottom was observed both in mice fed with the NWD1 diet and, as a
277 positive control, in those administered DSS in the drinking water (Fig. 5a-c).
278
279 **Secretory lineages as the cell of origin of human colon cancer**
280 Collectively, our results reveal an alternative bottom-up route to intestinal tumorigenesis originating
281 from Paneth cells in the murine small intestine and, allegedly, from secretory lineages in the human colon
282 triggered by inflammatory cues as in IBD or through western-style dietary factors. To validate the
283 relevance of our mouse study in patient-derived colon cancer, we took advantage of novel computational
284 methods(*47-49*) developed to predict the cell of origin of tumors by matching the mutational density along
285 the cancer genome with the profiles of epigenetic modifications characteristic of normal cell types (Suppl.
286 Fig. 10a)(*50*). While tumors from patients with an IBD history have a similar genome-wide tumor
287 mutational burden compared to their sporadic counterparts (Suppl. Fig. 10b), the presence of regions
288 ("genomic windows", see Methods) that are differentially mutated between sCRC and IBD-CRC is
289 suggestive of alternative mutational patterns (Suppl. Fig. 10c-d). Hence, we compared by COOBoostR(*49*)
290 (see Methods) the individual mutational landscapes with the epithelial cell types of the colon to compute
291 the putative cells of origin. As shown in Fig. 6a, while the majority of the sporadic colon cancers appear to
292 originate from stem cells (52%), among the IBD-related cases goblet (40%) and Best4 (28%) cells represent
293 the prevalent cells of origin. Of note, a substantial proportion of the sporadic cases is also predicted to
294 originate from goblet (22%) and Best4 cells (17%). Best4 cells, named after the specific expression of the
295 bestrophin 4 gene (*BEST4*), form a newly identified and yet only partially characterized intestinal epithelial
296 lineage with dual absorptive and secretory features(*51-53*). IHC and scRNAseq analysis in a cohort of
297 ulcerative colitis patients confirmed the presence of actively proliferating goblet and Best4 cells ($MUC2^+$

298 and BEST4$^+$; Fig. 5e-f), indicative of the primary effects of inflammation in these otherwise post-mitotic
299 cells, preceding their acquisition of stem-like features.
300 These results are indicative of the fraction of sporadic colon cancers whose expression profiles are
301 reminiscent of the PC-derived tumors and IBD-CRCs (24.7%; Figure 3j). Furthermore, the availability of
302 RNAseq data from a subset of the patient-derived tumors allowed us to confirm the prevalence of RSC-
303 like expression profiles and the high proportion of CMS1 and 4 cases among the sporadic and IBD cases
304 predicted to originate from non-stem, and in particular goblet, cells (Fig. 5g-h, Suppl. Fig. 10e).

305 *Lgr5*$^+$ stem cells have been established as the origin of intestinal tumors(*1*). However, whether the
306 same holds true in the context of inflammation has been challenged by their loss of lineage tracing
307 capacity and multipotency upon a broad spectrum of tissue injuries ranging from DSS-driven
308 inflammation(*11*) to γ-irradiation(*9, 22*), and western-style diet(*54*). Our results show that Paneth cell de-
309 differentiation and tumor onset in the mouse small intestine models sporadic colon cancer not only in IBD
310 patients but also in a substantial proportion of the cases predicted to originate in differentiated cell types
311 and not in resident ISCs. Both the differential gene expression profiles and inflammatory and stromal
312 features of the resulting tumors warrant a novel stratification of colon cancer for improved clinical
313 management. Preliminary investigations (*data not shown*) revealed that candidate markers differentially
314 expressed between PC- and *Lgr5*$^+$-initiated mouse tumors such as DCLK1 and HOXB9 were not equally
315 discriminative among patient-derived colon cancers. More systematic and high-throughput screening
316 approaches will likely lead to a novel stratification of the cases based on their cell of origin and specific
317 etiologic factors such as western-style diets and chronic inflammation.
318 Overall, the increased colon cancer risk exerted by inflammation may reflect on a 'trade-off' effect where
319 the chronic nature of the injury continuously stimulates de-differentiation of committed lineages into
320 stem-like cells in support of the tissue's regenerative response, thus resulting in the enlargement of the
321 pool of potential targets for tumor onset. Likewise, the same inflammation-inducing risk factors are bound
322 to affect tumor progression towards malignancy thus resulting into a distinct group of colon cancers with
323 specific prognosis and response to therapy (Fig. 5i).
324 Relative to the debate on the contribution of extrinsic risk factors versus the rate of stem cell division
325 to cancer development(*55, 56*), our results indicate that colon cancer etiologic factors such as
326 inflammation and poor dietary habits are likely to result in quantitative and qualitative alterations of the
327 stem cell niche which ultimately predispose to neoplastic transformation. The resulting subset of tumors
328 follow distinct evolutionary paths when compared to Wnt/Myc-driven *Lgr5*-derived tumorigenesis, and

329 are characterized by an inflammatory tumor phenotype more prone to infiltration from the tumor micro-
330 environment.

331 **Methods**

332 **Mice**

333 The following inducible Cre strains encompassing were employed: *Lgr5*$^{CreERT2-EGFP}$ (#008875, Jackson
334 Lab(*15*)) and p*Lys*^{CreERT2} (kind gift from the Clevers Lab(*16*)). For lineage tracing experiments, mice were
335 crossed with R26$^{LSL-YFP}$ mice (#006148, Jackson Laboratories) or R26$^{LSL-tdTomato}$ mice (#007908, Jackson
336 Laboratories). To target specific mutations in PCs, the above strains were further crossed with *Apc*^{15lox}
337 (#029275, Jackson Laboratories)(*12*), *Kras*$^{LSL-G12D}$ (#008179, Jackson Laboratories)(*13*), and *Trp53*^{flox}
338 (#008462, Jackson Laboratories)(*14*). For ablation experiments Lgr5$^{DTR-EGFP}$ mice were bred with c-Kit^{CreERT2}
339 and R26$^{LSL-tdTomato}$ animals. C-recombinase was activated by intraperitoneal injections of Tamoxifen (1 mg
340 dissolved in 100% EtOH and subsequently in sunflower oil; #T5648 and #S5007, Sigma) once or three times
341 in 4 days. Three days after Tamoxifen injection mice were treated once with diphteria toxin (DT) in order
342 to ablate Lgr5 positive cells. To induce acute intestinal inflammation, mice were administered 2-3%
343 dextran sodium sulfate (DSS) in their drinking water for 7 days (#0216011050, MP Biomedicals). DSS-
344 driven inflammation and the corresponding mechanisms underlying the consequent PC dedifferentiation
345 were described in previous studies(*11, 57, 58*).For diet experiments, mice were fed with AIN76A or NWD1
346 diet(*54*) and collected at various time points (4/8 days, 3/6/9 months). The formulation and composition
347 of the AIN76A and NWD1 diets are as outlined in Newmark et al. (Research Diets, New Brunswick, NJ)(*42*).
348 In brief, the western diets are based on AIN76A control diet (American Institute of Nutrition76A). NWD1
349 was adjusted to provide higher fat, and lower vitamin D_3, calcium donors to the single carbon pool and
350 fibre that produces mouse consumption levels similar to those common in segments of western societies
351 with high incidence of colorectal cancer. All mice here employed were inbred C57BL6/J.

352 For all experiments, mice were randomly assigned to experimental groups after matching for gender,
353 age of 8-12 weeks, and genotype. All protocols involving animals were approved by the Dutch Animal
354 Experimental Committee and in accordance with the Code of Practice for Animal Experiments in Cancer
355 Research established by the Netherlands Inspectorate for Health Protections, Commodities and
356 Veterinary Public Health. Animals were bred and maintained in the Erasmus MC animal facility (EDC)
357 under conventional specific pathogen-free conditions.

358

359 **Lineage tracing**

360 *Lys*$^{CreERT2/R26LSL-YFP}$ mice were injected three times with Tamoxifen (1 mg, i.p.) on consecutive days. One
361 week after the last Tamoxifen injection, intestinal tissues were harvested for lineage tracing analysis. To
362 this aim, tissue samples were first dissected and washed with PBS, and then fixed for 2 hours at 25°C with

363 4% buffered formaldehyde solution (Klinipath). Tissues were cryo-protected in 30% sucrose (Sigma)
364 overnight at 4°C, embedded in OCT (KP cryocompound, Klinipath), frozen on dry ice, and sectioned at -
365 20°C. Tissues were cut in 4-8 µm thick sections. These sections were incubated in PBS containing Alexa
366 568 Phalloidin (1:100; Invitrogen) or Alexa 633 Phalloidin (1:100; Invitrogen), and DAPI (Sigma) for 30
367 minutes at 25°C and washed in PBS-T. Tissues were mounted in Vectashield Mounting Medium (Vector
368 Labs) and imaged with a LSM700 confocal microscope (Zeiss). Images were processed with ImageJ.
369 Lineage tracing frequency was quantified by dividing the number of intestinal crypt/villus axes containing
370 Yfp^+ ribbons (encompassing at least 5 Yfp^+ labelled cells) by the total number of counted Yfp labelled
371 crypts. Lineage tracing analysis was performed on 3 mice for each diet type; in each mice at least 60 YFP
372 labelled crypts were analyzed.

373 **Immunohistochemistry on mouse tissues**

374 Tissues were fixed in 4% PFA overnight at 4°C and embedded in paraffin. The 4 µm sections were dewaxed
375 with Xylene and hydrated in consecutive rounds of 70% and 100% EtOH. Antigen retrieval was performed
376 in the 2100 Retriever pressure cooker (BioVendor) at pH 9 with Tris-EDTA buffer. After a 10 minutes
377 incubation at RT with 3% hydrogen peroxidase, tissues were blocked with 5% skim milk powder (Milipore)
378 in PBS-Tween. The following primary antibodies were employed by overnight incubation at 4°C: β-catenin,
379 #610154 (BD Biosciences); γ-H2AX, #9718 (Cell Signaling); Gfp, #A-11122 (ThermoFisher); Olfm4, #D6Y5A
380 (Cell Signaling); Lyz1, #A0099 (Dako); Dclk1, #ab37994 (Abcam); Muc2, #sc-15334 (Santa Cruz); Chga,
381 #NB120-15160 (Novus Bio); Ecad, #610182 (BD Biosciences); Cd3, #ab5690 (Abcam); F4/80, #D2S9R (Cell
382 Signaling); αSMA, #ab5690 (Abcam). Slides were washed twice with PBS-Tween and incubated for 30
383 minutes at room temperature with the Rabbit/Mouse EnVision kits (K4001/K4007, Dako). Slides were
384 counterstained with hematoxylin and, dehydrated with subsequent 70% and 100% EtOH and mounted
385 with Pertex (00811, Histolab). Whole slides were scanned with the Nanozoomer (Hamamatsu) and
386 analyzed with NDP viewer v2 (Hamamatsu).
387

388 **Immunohistochemistry analysis of human tissues**

389 Chromogenic double labelling was performed on 4-µm thick whole slide sections from FFPE tissue blocks,
390 on a validated and accredited automated slide stainer (Benchmark ULTRA System, VENTANA Medical
391 Systems, Tucson, AZ, USA) according to the manufacturer's instructions. Briefly, following
392 deparaffinization and heat-induced antigen retrieval for 40 minutes at 97°C, the tissue samples were
393 incubated with KI67 (MIB-1 antibody, #790-4286; Ventana) for 32 min at 37°C, followed by Ultraview
394 detection (#760-500, Ventana). After the stripping step, MUC2 (CCP58, #M7313, Dako) or BEST4

395 (HPA058564, Sigma-Aldrich) was incubated for 32 min at 37°C and followed by detection with Ultraview
396 Red (#760-501, Ventana) Counterstain was done by hematoxylin II for 12 min and a blue colouring reagent
397 for 8 min. Each tissue slide contained a fragment of FFPE pancreas or tonsil as an on-slide positive control.
398

399 **Immunofluorescence**

400 For cryo-sections, tissue samples were first dissected and washed with PBS, to then be fixed for 2 hours
401 at 25°C with 4% buffered formaldehyde solution (Klinipath). Tissues were cryo-protected in 30% sucrose
402 (Sigma) overnight at 4°C, embedded in OCT (KP cryocompound, Klinipath), frozen on dry ice, and sectioned
403 at −20°C. Tissues were cut in 4-8 µm thick sections. Immunofluorescence was performed using antibodies
404 against Olfm4 (1:200; CST), Mki67 (1:200; Novus biologicals), Lyz1 EC3.2.1.17 (1:1000; Dako), and c-Kit
405 (D13A2)(1:100; CST). Tissue sections were washed and blocked with 5% BSA in PBS-Tween for 1 hour to
406 then be incubated with the primary antibody overnight at 4°C. Slides were washed twice with PBS-Tween
407 and slides were incubated for 2 hours with Goat anti-Rabbit IgG (H+L) secondary Antibody, Alexa Fluor®
408 488 conjugate, (1:250; Life Technologies) and Wheat Germ Agglutinin (WGA)™ 647 conjugate (1:500,
409 ThermoFisher Scientific). Tissues were counterstained with DAPI (Sigma) and where indicated with F-actin
410 (#SC001, Spirochrome) for 30 min. at 25°C and washed in PBS-T. Tissues were mounted in VECTASHIELD
411 HardSet Antifade Mounting Medium (Vector Labs) and imaged with a Zeiss LSM700 confocal microscope.
412

413 ***In situ* hybridization**

414 In situ hybridization was performed by RNAscope™ Multiplex Fluorescent V2 Assay (ACDBio) technology
415 in combination with the RNA-Protein Co-Detection kit (ACDBio), according to the manufacturer's
416 instructions. In brief, 4 µm paraffin embedded sections were dehydrated, blocked for 10 minutes at room
417 temperature with H_2O_2, and processed with 1X Antigen Retrieval co-detection buffer for 15 minutes at
418 99°C. Sections were incubated with Ecad (1:250, #610182, BD Biosciences) overnight at 4°C. Hybridization
419 was performed with a probe targeting mouse *Clu* (Cat#427891, ADCBio), and the probe signal was
420 developed with Vivid dye 520 nm (1:1500). Sections were counterstained by DAPI and anti-mouse Alexa
421 633 (1:250, Invitrogen) and mounted with ProLong™ Gold Mountant (ThermoFisher). Samples were
422 imaged on the Stellaris 5 Confocal Microscope (Leica) and images were processed with ImageJ.
423

424 **Image analysis**

425 Images were scanned by Nanozoomer and imported in QuPath(*59*) as IHC images. Within intestinal
426 tumors, regions of interest (ROI) were randomly assigned and consisted of 300-600 cells. Next, the ROIs

427 were processed with positive cell selection plugin using default settings based on Cell DAB OD mean. The
428 percentage of positive cells per ROI was exported and visualized in R, where statistical tests were
429 performed. Fluorescent images were processed with ImageJ.
430
431 **Quantification of Ki67/WGA cells in murine colon**
432 The percentage of $Ki67^+/WGA^+$ double positive cells was quantified by dividing double positive cells by the
433 total amount of WGA^+ in the lower half of the colonic crypt. Only clearly detectable double positive cells,
434 as determined by the analysis of z-stacks were counted. For each quantification, at least 30 crypts of at
435 least 3 experimental mice were employed.
436
437 **Quantitative real-time PCR**
438 RNA was converted into cDNA using the High Capacity RNA-to-cDNA kit (Applied Biosystems). Samples
439 were run in duplicates on the 7500 Real Time system (Applied Biosystems). Quantitative PCR was
440 performed using the TaqMan assay (Applied Biosystems) according to instructions of the manufacturer.
441 Samples were normalized to the beta-actin (*Actb*) house-keeping gene. The gene-specific Taqman probes
442 were: Mm02619580_g1 (*Actb*), Mm00443610_m1 (*Axin2*), Mm00657323 (*Lyz1*), Mm01545303 (*Dclk1*).
443 Data were processed and visualized in R and statistical tests were employed to assess significance with
444 one-way ANOVA and Tukey post hoc test.
445
446 **Isolation of mouse intestinal crypts and cell dissociation**
447 Mouse small intestines were flushed with cold PBS, removed from fat and scraped with a glass slide to
448 remove intestinal villi. Resected sample were then sectioned into 5-10 mm segments and washed with
449 cold PBS before incubation for 30 minutes at 4°C in cold PBS supplemented with 6 mM EDTA. After an
450 additional PBS wash, crypts were detached from the muscle layer by 4 rounds of harsh pipetting with cold
451 PBS. Crypts were treated for 10 minutes at room temperature with DNAse (2000 U/ml-1, Thermo
452 Scientific) in Advanced DMEM/F-12 medium and passed through a 70 µm strainer. Purified crypts were
453 dissociated with 1 mL of pre-warmed TryplE and DNAse for 3 minutes at 37°C. Single cells were pelleted
454 in 10 mL ADF, resuspended in 3 mL 5% FCS HBSS and manually counted.
455
456 **Fluorescence activated cell sorting**
457 Cell suspensions were blocked with TruStain Fc blocking reagent (Biolegend) for 10 minutes at 4°C. After
458 washing with 5% FCS HBSS, cells were stained for 30 minutes at 4°C in 100 µl with Lin antibodies (CD31-

459 BV421 #563356, CD45-BV421 #563890, TER119-BV421 #563998, BD Biosciences), CD24-APC (#1109070,
460 Sony Biotechnology), and cKit-PE (#105808, Biolegend). For scRNAseq experiments, the antibody mix was
461 supplemented with 2 µl Hashing Antibody (Biolegend TotalSeq^{TM} Hashtag antibodies: #155831, #155833,
462 #155837, #155843, #155849) to label the cells with oligo barcodes. After 3 washes with 3.5 mL 5% FCS
463 HBSS, cells were filtered with a 40 µm strainer and resuspended in 1 mL 5% FCS HBSS with 1 µg/mL DAPI
464 (D9542, Sigma-Aldrich). FACS analysis and sorting were performed using a FACSAriaIII (BD Biosciences).
465 DAPI and BV421-conjugated antibodies were detected using a 405nm laser and 450/40 BP filter. GFP and
466 BB515-conjugated antibodies were detected using a 488nm laser and 502LP + 530/30 BP filters. tdTomato
467 and PE-conjugated antibodies were detected using a 561nm laser and 582/15 BP filter. APC-conjugated
468 antibodies were detected using a 633nm laser and 660/20 BP filter. Samples were sorted for epithelial
469 (Lin^{neg}), Paneth-enriched (SSC^{hi}CD24^{hi}cKit^{hi}), and traced cells (Yfp^{hi}/td-Tomato^{hi}) according to previously
470 established gates(*9-11*). To optimize the sorting for single cell experiments, the initial single cell gate was
471 followed by a more stringent, population specific single cell gate(*9*). FACS strategy was visualized with
472 FlowJo V10.
473
474 **Organoid Reconstitution Assay**
475 Whole crypts were extracted from the small intestine of *Lgr5*-EGFP mice fed for 3 month (from weaning)
476 with either AIN76A or NWD1 diets(*54*). After cellular dissociation and FACS, single *Lgr5*$^{+}$ and Paneth cells
477 were sorted in separate Eppendorf LoBind Tubes. Reconstitution of *Lgr5*-EGFP^{hi} stem cells with PCs
478 (SSC^{hi}CD24^{hi}cKit^{hi}) was performed by co-pelleting sorted cells at 300 g for 5 minutes in Eppendorf LoBind
479 Tubes, and by incubating them for 15 minutes at 25°C as previously described(*10, 44*). The cells were then
480 resuspended in 30 µL Matrigel (Corning), and cultured in 96 Well Flat Bottom dishes (Corning).
481 Additionally, Paneth and Lgr5 positive cells were plated singularly where indicated and cultured in
482 standard ENR medium(*60*) supplemented with 10 µM Y-27632 (Sigma) and 1 mM Jagged-1 (AnaSpec).
483 Cells were plated in triplicates and the generated organoids where counted 9 days after plating.
484 Experiments were repeated at 3 independent moments.
485
486 **Droplet-based scRNAseq**
487 Sorted single cells were mixed in equal sample proportions to a pool of 25-50k cells in 5% FCS HBSS to a
488 final concentration of ±1000 cells/µl. The cell pool was recounted manually with Trypan blue and between
489 10-15k cells were loaded into a 10X Genomics Chip (Next GEM Chip G). The Gel Bead-In EMulsions were
490 generated with the Chromium Controller (10X Genomics) using Single Cell 3' v3 chemistry with

491 FeatureBarcoding. After generation of cDNA and quality check with Bioanalyzer (Agilent), libraries were
492 sequenced with the Novaseq 6000 (Illumina) till a depth of 20-40k reads/cell. Raw data was processed
493 with CellRanger count (cellranger-7.0.0)(*61*) and aligned to the mm10 mouse reference genome with the
494 addition of the transgenes Yfp and Td-tomato.
495
496 **Analysis of scRNAseq: inflammation and genetic targeting of Paneth cells**
497 For each single cell run, filtered gene-cell matrices were imported in R using Seurat (v 4.1.1)(*62*). After a
498 centered log ratio transformation (CLR) of the hashtag oligo counts, cells were assigned back to their
499 mouse of origin using HTODemux (positive quantile = 0.99). Cells assigned as doublets or negative for any
500 of the barcodes were filtered from the data. Cells from mice with different gender but same genotype and
501 experimental group were further separated based on the expression *Xist*. After preprocessing, the five
502 runs were merged into a single Seurat object. To remove putative doublets from the sorting, doublets
503 were simulated and called by doubletfinder_v3 (nExp = 0.01)(*63*). After removal of doublets, low quality
504 cells were removed by filters based on the percentage of mitochondrial genes (> 10%), on nFeature_RNA
505 (< 200), and nCount_RNA (< 500). Next, data were integrated with the single cell transform (SCT) pipeline
506 using reciprocal PCA dimension reduction to find integration anchors. Integrated data was further
507 processed by dimension reduction with Uniform Manifold Approximation and Projection (UMAP) on the
508 first 30 principal components. Cells were clustered using shared nearest neighbor (SNN) modularity
509 optimization (resolution = 0.6) and annotated according to canonical marker genes. Traced cells were
510 defined as cell with a non-zero value for either td-Tomato or Yfp. Signatures were evaluated using the
511 Addmodulescore function and visualized in experimental groups or cell clusters using Violin Plots.
512 Pathway level evaluation was done using the GSVA package(*64*) on summarized expression profiles using
513 Average Expression by cell cluster.
514
515 **Bulk RNAseq of intestinal tumors**
516 Intestinal tumors were macro-dissected from intestinal tissues and cut into 1-2 mm small fragments which
517 were then re-suspended in 500 µl TRIzol (#15596026, Ambion) for RNA isolation. RNA quality was first
518 evaluated by Nanodrop ND-1000 (ThermoFisher) and further purified with the TURBO DNA-free Kit
519 (Invitrogen). Samples were sequenced with a read length of 150 bp (PE150) using the DNA nanoball (DNB-
520 seq) protocol till a depth of 25M reads/sample (BGI, Hong Kong). Adapter trimming and sequencing QC
521 was performed with SOAPnuke software(*65*) .
522

523 **Analysis of Lyz1/Lgr5-derived mouse tumors: bulk RNAseq**

524 The FASTQ files were aligned with STAR (v2.7.9a) to the mm10 mouse reference genome with Ensemble
525 gene annotations(*66*). After alignment, Sambamba (0.8.0) was applied to mark duplicates and perform
526 flagstat quality checks(*67*). Next, Subread (2.0.3) was used to count the primary strand-specific
527 alignment(*68*). Downstream analysis was performed in R using the DeSeq2 package (v 1.34.0)(*69*). In brief,
528 counts were log2 normalized and used for principal component analysis on the top 500 variable expressed
529 genes. Differential expression analysis was done following the DeSeq pipeline with Negative Binomial GLM
530 fitting and Wald statistics and genes were filtered using $log_2FC > 1.5$, $P_{adj} < 0.01$. Gene set enrichment
531 analysis was done with the fgsea package (v 1.20.0) using the HallMarks gene modules from the molecular
532 signature database(*36*). Gene set variation analysis was performed using the GSVA package (v 1.42.0)(*64*).
533 The relative proportion of RSC/CBC stem cell types was computed with the ISC index using default
534 settings(*29*).

535

536 **Analysis of Paneth cells upon exposure to inflammation and western-style diet: bulk RNAseq**

537 The FASTQ files were removed from TrueSeq adapters using Trimmomatic (v0.36). STAR(*66*) was used to
538 align the reads to the mm10 reference genome using Gencode annotation release M15 (GRCm38.p5).
539 Mapping quality was assessed with Sambamba Flagstat statistics. Count files were generated using
540 FeatureCounts (subread) and further processed in R using the DeSeq2 package. Differential expression
541 analysis were performed between the groups 'NWD1 diet' (n = 3) vs. 'AIN76 diet' (n = 3) with the Wald-
542 test. A heat map was produced using z-score scaled data and visualized with the ComplexHeatmap
543 package (v2.12.1).

544

545 **Analysis of IBD-CRC and sCRC RNAseq cohort**

546 RNA-seq material was previously described in Rajamäki et al.(*35*) In brief, 64 CRCs entered HiSeq LncRNA-
547 Seq library preparation and paired-end sequencing using Illumina HiSeqXTen. Salmon (version 0.12.0;
548 quant mode with ValidateMappings) was used to map raw sequences onto the human transcriptome
549 (Ensembl release 79). Gene-level quantification was done with DESeq2 (v1.18.1) followed by a variance
550 stabilizing transformation and Limma (v3.34.9) correction of sequencing batch effects(*69*). Only
551 microsatellite stable (MSS) tumors were included in the analyses. Differential gene expression was
552 calculated as ANOVA between 38 MSS sCRCs and 14 MSS IBD-CRCs. GSVA (v1.42.0)(*64*) analyses were ran
553 using default options (Gaussian kernel) and predefined gene sets from the molecular signature

554 database(*36*). Conensus Molecular Subtypes were mapped to the samples with the random forest
555 classifier of CMSClassifier R package.
556
557 **Cell of origin predictions from IBD-CRCs/sCRCs**
558 *Processing of Whole-Genome Sequencing (WGS) data of CRCs*
559 DNA was isolated from fresh-frozen tumor, normal colon, or blood. Paired tumor-normal WGS data was
560 generated from altogether 25 MSS IBD-CRCs and 257 MSS sCRCs. Libraries were prepared using TruSeq
561 Nano DNA HT or TruSeq PCR-Free Kit (Illumina), followed by paired-end sequencing using Illumina
562 platform (HiSeqXTen/HiSeq2000). More details are given in Rajamäki et al.(*35*). Sequence alignment to
563 the GRCh38 reference genome, other data preprocessing steps, and somatic variant calling were
564 performed with GenomeAnalysisToolkit GATK4 best practices workflow (version 4.0.4.0)(*70*). Mutations
565 were filtered based on allele fraction (AF) to extract clonal somatic variants. A genome-wide mean and
566 standard deviation was calculated for each tumor; somatic variants were then filtered to a minimum AF
567 of one standard deviation below the mean. On average, 85% of variants per tumor (IQR 83%-87%) were
568 kept-in as clonal somatic variants for the subsequent analysis. Similar to previous work(*47, 48*), the human
569 reference genome (hg38) was processed in non-overlapping windows of 1 million base pairs. We excluded
570 window-regions with <92% of bases within uniquely mapped 36-mers, regions overlapping telomeres and
571 centromeres and regions outside autosomes. Altogether 2387 windows remained for the analysis. To
572 identify differentially mutated windows, patients were normalized by the average of their windows and
573 IBD-CRCs vs sCRCs were compared with an Wilcoxon Rank Sum test ($P < 0.01$).
574
575 *Processing of scATAC data from human colon*
576 Normal colon (N = 9) samples from Becker et al.(*50*) were obtained from the public domain under GEO
577 identifier GSE201349. Signac(*71*) was used for processing and the FeatureMatrix function was used to
578 generate a count matrix based on the previously described 2387 windows. Cell labels were retrieved from
579 the original publication and reduced to 6 major classes by the aggregation of clusters from the same cell
580 type. To generate the embedding, batch correction was performed with IntegrateData, and UMAP was
581 computed on the 2:30 latent semantic indexing (LSI) components. To generate the chromatin mark
582 features per cell type, the counts from the scATAC data set were grouped with the AggregateExpression
583 function and equalized to 10M reads per cell type.
584
585 *Predictions with COOBoostR*

586 COOBoostR(*49*), an extreme gradient boosting-based (XGBoost) machine learning algorithm designed for
587 the prediction of cells of origin, was used to predict the putative cells of origin for each individual sample
588 from the WGS cohort (N = 282 tumors) based on the epigenetic profiles from the human colonic scATAC
589 data. We ran the algorithm with default values from XGBoost for the learning rate (mEta = 0.3) and the
590 maximum depth of the trees (mdepth = 6). Tumor predictions were aggregated on cohort level, and IBD-
591 CRC/sCRC were compared with a Fisher's exact test. For a subset of patients (N = 48 tumors), WGS and
592 RNAseq was available. Hence, for these tumors, CMS labels and signatures from the gene set variation
593 analysis (GSVA) were compared based on stratification by putative cell of origin.
594
595 **Analysis of human CRC bulk RNAseq cohort**
596 Log2 normalized expression of the meta cohort described in Guinney et al.(*38*) was obtained from the R2
597 Genomics Analysis and Visualization Platform and subsequently processed in R. Gene set variation analysis
598 was performed with the GSVA package using the Gaussian kernel(*64*). Subsequently, correlations of the
599 gene signatures (Suppl. Table 4) were visualized with ggcorr using Pearson Correlation. Next, samples
600 were classified as colitis-like, sporadic-like or as intermediate based on their GSVA scores for the IBD-CRC
601 and sCRC signatures. Using this classification, samples were visualized with their annotated Consensus
602 Molecular Subtype (CMS). Survival analysis was performed with the survival and survminer packages in R.
603 Multivariate analysis was performed with the log-rank test and pairwise comparisons were fitted with a
604 Cox proportional hazards regression model.
605
606 **Analysis of human scRNAseq cohorts**
607 *Colorectal cancer data set*
608 Filtered count matrices were retrieved from the gene expression omnibus (KUL cohort: GSE144735; SMC
609 cohort: GSE132465)(*37*) and processed in R with Seurat. After normalization and scaling of the Seurat
610 object, gene expression was summarized per cell type using AverageExpression based on the overlapping
611 differentially expressed genes of murine (Paneth vs. *Lgr5*) and human (IBD-CRC vs. sCRC) tumor groups.
612 The average expression was z-scored and visualized as a heat map using the ComplexHeatmap package.
613
614 *Ulcerative Colitis data set*
615 Data from Smillie et al.(*51*) was processed with the Seurat(*62*) workflow. A subset was taken
616 encompassing the epithelial cells according to annotation from the authors. Cells were defined as positive
617 when they had non-zero counts for that particular gene. Cells were aggregated on patient level and a t-

618 test was performed to compare number of double and single positive cells across the UC (N = 17) and
619 control group (N = 12).
620
621 **Analysis of mouse intestinal single cell data sets**
622 *Homeostasis and γ-irradiation*
623 Count matrices from Haber et al.(*72*) and Ayyaz et al.(*28*) were downloaded from GEO using GSE92332
624 and GSE117783, respectively. After preprocessing using the Seurat workflow, Seurat objects were merged
625 and gene expression was averaged by cell type and study using AverageExpression. Next, gene set
626 variation analysis (GSVA)(*64*) was computed using the Gaussian kernel and the output was visualized in a
627 heat map using the ComplexHeatmap package.
628
629 *AIN76A and NWD1 diets: scRNAseq*
630 The intestinal epithelia of mice fed with control (AIN76A) and western-style diets (4 days NWD1, 4 days
631 NWD1 + 4 days AIN recovery) were collected and sent for 10X Genomics scRNAseq analysis (N = 3 mice
632 per group)(*43*). After preprocessing with Seurat and batch correction using the integrated assay workflow,
633 the Seurat clusters were annotated according to canonical markers. CytoTRACE(*45*) was computed on the
634 whole data set using default settings. Next, Paneth cells (N = 991 cells) were taken from the data set, sub-
635 clustered, and visualized by UMAP with a split by the dietary condition. The DSS signature was computed
636 using AddModuleScore and projected on the UMAP embedding. The Paneth sub-cluster that appeared in
637 the NWD1 conditions and associated with the DSS signature was relabeled as 'Diet response'. Next,
638 marker genes were visualized using ViolinPlot based on the separation of diet response PCs and other PCs.
639 Last, Paneth cells from the diet data set were merged with the Paneth cells from the DSS data set. Average
640 expression was computed based on experimental condition, and GSVA analysis (Gaussian kernel) was
641 performed using the predefined gene sets from the Molecular Signature Database (Hallmark set)(*36, 64*).
642 A filter was applied to select pathways that were both upregulated in DSS and NWD1 condition (ΔGSVA =
643 0.2). The output was visualized as heat map using the ComplexHeatmap package.
644
645 *AIN76A and NWD1 diets: scATACseq*
646 The intestinal epithelia of mice fed with control (AIN76A) and western-style diets (3 months) were
647 collected and sent for 10X Genomics scATACseq analysis (N = 2 mice per group)(*43*). Processing was done
648 with Signac(*71*), and UMAP embedding was produced base on the 2:30 latent semantic indexing (LSI)
649 components. Unsupervised clustering was performed by a shared nearest neighbor (SNN) modularity

650 optimization based clustering algorithm (algorithm: 1, resolution: 1.2). Cell type predictions scores were
651 obtained by the integration with the AIN76A/NWD1 scRNAseq data set using the FindTransferAnchors
652 and TransferData functions. Two out of twenty-one clusters had high association to the Paneth cell
653 prediction score (cluster 5,20). Paneth cells were defined as part of either cluster 5/20 with the additional
654 condition of a non-zero prediction score for Paneth cells. Next, Paneth cells (N = 945 cells) were taken
655 from the data set, and visualized by UMAP with a split by the dietary condition. Cluster 5 represented cells
656 from both diets while cluster 20 was enriched in NWD1 and therefore relabeled as 'Diet response'. To
657 identify differentially accessible peaks, FindMarkers was used with logistic regression test. Peaks were
658 filtered based on $P < 0.05$, visualized with a RegionHeatmap and RIdeogram on the mouse genome, and
659 gene ontology was performed on the closest genes using the EnrichR prackage.
660
661 *Lgr5 ablation with Diphtheria toxin*
662 Count matrices were obtained from GEO (GSE183299) and processing was done with the Seurat
663 workflow(*62*). The revival stem cell (RSC) signature was evaluated using the Addmodulescore function and
664 visualized across experimental groups and cell clusters using the ComplexHeatmap package. A subset was
665 made of intestinal stem cells, and differential expression analysis was performed with FindMarkers. The
666 results were visualized with the EnhancedVolcano package, where marker genes of Paneth cells were
667 highlighted.
668
669 *AOM/DSS-derived tumors*
670 Data from Vega et al.(*32*) was downloaded through GEO (GSE134255) and processed with Seurat(*62*).
671 Differential abundance analysis was performed with MiloR(*73*) workflow using default parameters. The
672 revival stem cell (RSC) signature and the PC Cluster 4 signature were evaluated using the Addmodulescore
673 function and visualized with ViolinPlots.
674
675 **CiberSortx analysis of AOM/DSS adenomas**
676 CiberSortx(*34*) was leveraged to predict the fractions of various cell types in tumors arising in AOM/DSS
677 context. To this aim, count data from AOM/DSS-derived adenomas from Chen et al.(*31*) (GSE178145) were
678 converted to counts per million with Edger. The single cell count data from Paneth cell upon genetic
679 targeting were provided as reference and converted to a signature matrix in CiberSortx using default
680 parameters. For cell type imputation, CiberSortx was applied in relative mode using 100 permutations.
681

682 **Data availability**

683 All data relevant to this study are made available. Transcriptomic sequencing data has been deposited in
684 the Gene Expression Omnibus (GEO), and is available using the following identifiers: GSE221819 (bulk RNA
685 sequencing of murine tumors); GSE221820 (single cell RNA sequencing of genetically targeted Paneth
686 cells); GSE221818 (bulk RNA sequencing of sorted Paneth cells treated with control and western-style
687 diet). Additional data sets referenced in this study are publicly available in the Synapse and GEO
688 repositories. Previously published scRNAseq studies were employed relative to the mouse small intestine
689 in homeostasis (Haber et al.(*72*), GSE92332), upon irradiation (Ayyaz et al.(*28*), GSE117783), *Lgr5* ablation
690 (Singh et al.(*30*) GSE183299), and upon feeding with western-style diet (Choi et al. (*43*), scRNAseq:
691 GSE188577, scATACseq: GSE228006). Mouse colonic scRNAseq data from AOM/DSS tumors was obtained
692 from Vega et al. (*32*) (GSE134255) and bulk RNAseq was from Chen et al.(*31*) (GSE178145). Human
693 colorectal cancer Bulk RNAseq and scRNAseq studies were retrieved from Guinney et al.(*38*)
694 (syn2623706), Rajamäki et al.(*35*) and Lee et al.(*37*) (KUL cohort: GSE144735; SMC cohort: GSE132465),
695 respectively. Human scRNAseq data from Ulcerative Colitis patients (Smillie et al.(*51*)) was downloaded
696 from the Broad DUOS platform after a DTA agreement. Single cell ATAC data from human colon was
697 obtained from Becker et al.(*50*) (GSE201349).

698

699 **Figure Legends**

700 **Figure 1. Paneth cells as the cell-of-origin of intestinal cancer.**

701 **a.** Cre-Lox strategy taimed at the targeting of *Apc*, *Kras*, and *Tp53* mutations in intestinal stem cells (*Lgr5*$^+$
702 ISCs) and Paneth cells (*Lyz1*$^+$ PCs). **b.** and **c.** β-catenin IHC analysis of intestinal tumors initiated from *Lgr5*$^+$
703 ISCs (**b**) and PCs (**c**). The asterisks indicate *Lgr5*$^+$ ISCs and PCs with enhanced cytoplasmic and nuclear β-
704 catenin accumulation; tumor foci and adenomas are indicated by dashed lines. **d-e.** Tumor multiplicity
705 was calculated according to tumor-bearing animals (**d**) and by tumor number per genotype (**e**) in the
706 presence/absence of DSS and based on Swiss roll counts. Error bars denote standard deviations of the
707 mean. P values denote one-way ANOVA and Tukey post-hoc tests for group comparisons. **f.** Lineage
708 tracing analysis of Paneth cells (labelled by YFP) at different stages of tumor-initiation and progression. **g-**
709 **j.** Left panels: Lyz1 (**g**) and Dclk1 (**i**) IHC analysis *Lgr5*$^+$ ISCs and PCs-derived adenomas. Top-right inlets (**h-**
710 **j**): quantification of number of *Lyz1*- (**h**) and *Dclk1*-positive (**j**) tumor cells. P values depict one-way ANOVA
711 and Tukey post-hoc tests for group comparisons. **k.** *Lyz1*, *Dclk1*, and *Axin2* qPCR expression analysis across
712 different adenoma genotypes. P values represent one-way ANOVA and Tukey post-hoc tests for group
713 comparisons.

714

715 **Figure 2. Paneth cells dedifferentiate into a revival stem cell (RSC) identity.**

716 **a.** Schematics of the experimental approach. After genetic targeting of Paneth cells, intestinal crypts were
717 extracted, and the isolated cells labeled with hashing antibodies and sorted according to three different
718 sorting strategies: epithelium, PC-enriched and PC-traced cells. **b.** UMAP embedding of the different cell
719 clusters/lineages (left panel), annotated according to the expression of canonical marker genes (right
720 panel). **c.** Bar plot of the distribution of traced cells across the different mouse genotypes and
721 experimental conditions. **d.** Violin plots representing marker genes of the newly identified Paneth-derived
722 cell clusters (PC Cluster 1-4). **e.** Association analysis of the revival stem cell (RSC) signature with PC Cluster
723 1-4. **f.** RNA *in situ* hybridizations of the *Clu* gene in small tumors derived from Paneth cells upon compound
724 targeting of *Apc* and *Kras* mutations. **g.** Gene sets variation analysis among the Haber et al.(*72*), Ayyaz et
725 al.(*28*), and the present studies.

726

727 **Figure 3. Paneth-derived adenomas have an inflammatory phenotype mimicking colitis-associated**
728 **colon cancer.**

729 **a**. Schematics of the experimental approach to compare PC- and *Lgr5*-derived adenomas. **b**. Principal
730 Component Analysis (PCA) plot showing that the cell-of-origin is the dominant discriminator of variance.

731 **c**. Bar plot summarizing the gene set enrichment analysis between Paneth- (*Lyz1 Apc* DSS) and ISC- (*Lgr5*
732 *Apc* DSS) derived tumors. Pathways were filtered based on Pval < 0.05 and abs NES > 0.5. **d**. Subset of
733 inflammatory pathways, visualized as heat map based on values from the gene set variation analysis. **e**.
734 Box plots showing results of the stem cell index. P value depicts result of one-way ANOVA. **f**. Gene set
735 enrichment analysis showing significant but opposite association between the *Lyz1* tumor signature and
736 IBD-CRCs, and between the *Lgr5* tumor signature and sCRC. **g**. Heatmap showing GSVA scores, averaged
737 per tumor group, of pathways with similar patterns between the murine and human tumor groups. **h**.
738 Heatmap highlighting the differentially expressed genes (log2FC >1.5, Padj < 0.01) shared between the
739 Paneth/IBD-tumors and the *Lgr5*/sCRC-tumors. Values denote z-scores of average expression per cell
740 type. **i-m**. Two distinct sporadic colon cancer identities become apparent upon analysis of a large cohort
741 of CRC tumors (N = 3232 samples). **i.** Heatmap showing Pearson Correlations of the GSVA scores. **j-k**.
742 Scatter plot showing two distinct clusters of sporadic- and colitis-like in all (**j**) colon cancers and (**k**)
743 stratified according to their consensus molecular subtype (CMS). Grey lines indicate contours lines,
744 dashed lines show thresholds to classify tumors in colitis-like, sporadic-like, and intermediate groups. **l**.
745 Stacked bar plot analysis showing the distribution of consensus molecular subtypes (CMS1 to 4) across
746 the colitis-like and sporadic-like colon cancers. **m**. Kaplan-Meier survival analysis for relapse free survival.
747 P values denote result of log-rank test and cox regression models for univariate analyses. Hazard ratios
748 (HR) and confidence intervals (CI) are displayed for pairwise comparisons.
749
750 **Figure 4. Western-style diet triggers an inflammatory response leading to Paneth cell dedifferentiation.**
751 **a.** Schematics of the experimental approach designed to investigate the consequences of short- and long-
752 term exposure to western-style diet (NWD1) vs. control (AIN76A) diets. **b**. Heatmap showing z-scored DSS
753 signature (DSS vs. Control; Padj < 0.05, Log2FC > 0.25) in Paneth cells exposed to DSS or NWD1. **c**. Organoid
754 multiplicities derived either from single ISCs or PCs, and from reconstituted doublets (L: Lgr5+ ISCs, P:
755 Paneth cells). Pooled data from N = 4 independent experiments. P values were calculated using one-way
756 ANOVA and Tukey tests for group comparisons. Error bars depict standard deviations from the mean. **d**.
757 Representative image of lineage tracings from a NWD1-fed Lyz1-Yfp mouse. Scale bar: 50 µm. P value
758 depicts result of the student t-test and error bar represents the standard deviation. Data from N = 3 mice.
759 **e**. UMAP showing Paneth cells from AIN76A- and NWD1-fed mice (N = 3 mice per condition). The DSS
760 signature portrayed on UMAP embedding, highlights a subcluster of Paneth cells responsive to the NWD1
761 diet. **f**. Violin plots showing different levels of the DSS signature (top) and CytoTRACE score (bottom)
762 between the PCs responsive to the NWD1 diet and other Paneth cells. P values depict significance values

763 of Wilcoxon test. **g**. Violin plots representing marker genes of PCs responsive to the NWD1 diet, showing
764 co-expression of stem and secretory markers. **h**. Heatmap visualization of gene set variation analysis,
765 indicating pathways that are activated in Paneth cells after exposure to DSS or NWD1. **i**. UMAP plot of
766 Paneth cell subset in scATAC data set of AIN (N = 2) and NWD1 (N = 2) treated mice. **j**. Heat map listing
767 differential peaks between Diet response Paneth cells and other PCs. **k**. Ideogram displaying the
768 distribution along the mouse chromosomes of the differential peaks observed upon diet response (red)
769 cell when compared to those characteristic of Paneth cells (blue).
770
771 **Figure 5. Inflammation activates distinct cells of origin in the colon.**
772 **a.** Colonic tissues from either untreated mice or from mice administered 3% DSS for 7 days, as well as
773 from mice fed with the AIN76A or NWD1 synthetic diets for 3 months were analyzed for the presence of
774 proliferative deep crypt secretory (DSC) cells. WGA (wheat germ agglutinin) was used to stain DSC
775 cells(*46*), and Ki67 to mark proliferative cells. Tissues were counterstained by DAPI (nucleus). Asterisks
776 mark WGA/Ki67 double positive cells. **b-c**. Quantification of WGA^+/$Ki67^+$cells (**b**) and total WGA^+ cells (**c**)
777 in the lower colonic crypt of the mice, as shown in (**a**). Scale bar = 20 µm. A minimum of 50 crypts from 3
778 different mice were analysed. **d**. Barplot showing the predicted cell-of-origin in a cohort of IBD-CRC (N =
779 25) and sCRC (N = 257)(*35*) based on the COOBoostR computational approach(*49*) (see Methods). P value
780 depicts result of the Fisher exact test. **e**. MUC2, BEST4, and Ki67 IHC analysis of colonic tissues obtained
781 from controls and IBD patients. Asteriks mark examples of double positive cells. **f**. Boxplot showing
782 percentage of cycling ($Ki67^+$) $MUC2^+$ and $REG4^+$ cells in Ulcerative Colitis patients and controls. scRNAseq
783 data are from Smillie et al.(*51*). Positivity was defined per cell by the presence of at least one read for that
784 particular marker. Subsequently, cells were aggregated per patient to calculate percentages. **g**. Box plot
785 denoting differences in the stem cell index based on stratification of predicted cell-of-origin in a subset of
786 IBD-CRC and sCRC cases for which RNAseq data were available(*35*). P value depict the result of the t-test.
787 **h**. Mapping of Consensus Molecular Subtypes (CMS) on tumor samples stratified according to their
788 predicted cell-of-origin. P value depicts result of the Fisher exact test. **i**. Graphic abstract of the model
789 arising from the present study. Colon cancer can be initiated either from stem (CBC) or differentiated cells,
790 the latter in response to inflammatory cues. RSC reprogramming is activated in support of the
791 regenerative response. During this process, actively dividing RSCs expand the cell targets for tumor
792 initiation and progression, leading to an alternative route to tumorigenesis earmarked by an inflammatory
793 phenotype.
794

795

796 **Supplementary Figure Legends**

797 **Supplementary Figure 1. Immuno-histologic analysis of Paneth-derived tumors.**

798 **a.** Representative IHC images of Paneth-derived tumors from different genetic backgrounds, stained with
799 phospho-histone H2A.X (Ser 139). **b.** Top three panels: IHC analysis of Paneth-derived intestinal tumors
800 with different genetic backgrounds (as indicated), stained for β-catenin. Bottom: hematoxylin and eosin
801 histological stains of Paneth-derived tumors from *Lyz1*/*Apc*/*Kras*/*Tp53* mice.

802

803 **Supplementary Figure 2. Characterization of PC-derived small intestinal adenomas.**

804 **a**. Stacked bar plot displaying the localization and distribution of PC-derived adenomas across the small
805 intestine. **b**. Immunofluorescence imaging of intestinal crypts from *Lyz1*/tdTomato and *Lyz1-AK*/tdTomato
806 mice. **c**. IHC analysis of the lineage-specific ChgA, Muc2, and Olfm4 markers in PC- and *Lgr5*-derived
807 adenomas.

808

809 **Supplementary Figure 3. FACS gating strategy for Paneth cells enrichment and isolation.**

810 **a.** FACS gating strategy established for the isolation of live, single, epithelial cells (top). Subsequently,
811 Paneth cells were enriched using the $SSC^{hi}CD24^{hi}$ gate and further purified based on high cKit levels
812 ($SSC^{hi}CD24^{hi}cKit^{hi}$). Alternatively, traced cells were sorted based on positive expression for Yfp or td-
813 Tomato when the appropriate reporter mice were employed. **b.** Negative control employed to establish
814 reference levels based on unstained samples. **c.** Bar plot of the percentage of Paneth cells obtained by the
815 distinct gating strategies. P values depict results of one-way ANOVA and Tukey tests.

816

817 **Supplementary Figure 4. Paneth cells dedifferentiate upon DSS-driven inflammation.**

818 **a-b.** In situ hybridization (ISH) analysis of Clu mRNA (green) expression at various stages during PC-derived
819 tumorigenesis in *Lyz1-Apc*/*Kras* and (**b**) *Lyz1-Apc*/*Kras*/*Tp53* mice. **c.** Bar plot showing filtered pathways
820 (Pval < 0.05, abs NES > 0.5) of the gene set enrichment analysis comparing RSCs with PC Cluster 4. **d.** FACS
821 analysis of lineage-traced (Yfp+) cells in wild type (*Lyz1*-Yfp) and AKP-mutant (*Lyz1-Apc*/*Kras*/*Tp53*) mice.
822 The latter clearly localize outside of the $CD24^{hi}cKit^{hi}$ gate where wild type PCs reside. **e.** Bar plot relative
823 to the percentages of lineage-traced cells evaluated by FACS using the $CD24^{hi}cKit^{hi}$ PC-specific gate (left),
824 or by scRNAseq analysis (% within Paneth cluster; right). P values denote significance of one-way ANOVA.
825 **f.** Left: Bar plots showing the average percentage of *Ascl2*$^{+}$ and *Lgr5*$^{+}$ stem cells in control and DSS-treated

826 animals. Right: Violin plots showing a decrease in the crypt base columnar (CBC) signature upon DSS
827 treatment. P values denote significance of one-way ANOVA and Tukey test for group comparisons.
828
829 **Supplementary Figure 5. Distinct TME features in CBC- and PC-derived small intestinal adenomas.**
830 IHC analysis of *Lgr5*$^{+}$- and *Lyz*-derived adenomas with the tumor microenvironment markers αSMA
831 (fibroblasts), F4/80 (macrophages), and Cd3 (T cells).
832
833 **Supplementary Figure 6. *Lgr5*$^{+}$ CBCs ablation triggers dedifferentiation of Paneth and goblet secretory**
834 **lineages.**
835 **a.** Differential expression analysis of intestinal stem cells in homeostasis and during regeneration
836 following *Lgr5*$^{+}$ CBCs ablation. scRNAseq are from Singh et al.(*30*). Several among the top 50 differentially
837 expressed genes in the regenerated ISCs are Paneth-specific markers (red). **b**. Heatmap showing
838 transcriptional activity of the RSC program upon *Lgr5*$^{+}$ CBCs ablation. **c-d**. Lineage tracing of DT-treated
839 *Lgr5*$^{DTR\text{-}EGFP}$;*c-Kit*^{CreERT2}; Rosa^{tdTomato} mice across different intestinal sections (c.) and double stained with
840 Lyz1/tdTomato (d.). Scale bar = 50 µM, 2 mice per group were analyzed.
841
842 **Supplementary Figure 7. AOM/DSS-derived adenomas share similarities with PC-derived tumors.**
843 **a**. Stacked bar plot of the cell type fractions in AOM/DSS adenomas as predicted by Cibersort analysis(*34*)
844 (see Methods). Bulk RNAseq data are from Chen et al.(*31*). **b**. Differential cell type abundance (MiloR(*73*);
845 see Methods) between AOM/DSS and *Lrig1-Apc* adenomas. scRNAseq data are from Vega et al.(*32*). **c**.
846 Comparison of RSC and PC Cluster 4 module score in tumor cells from AOM/DSS and *Lrig1-Apc* adenomas.
847 **d**. Venn diagram indicating the overlap between marker genes from PC Cluster 4 and the AOM/DSS-
848 specific squamous tumor cell subpopulation.
849
850 **Supplementary Figure 8. Differential gene expression analysis between IBD colon cancers and murine**
851 **PC-derived small intestinal adenomas.**
852 **a**. Volcano plot showing differentially expressed (Pval < 0.01, FC cutoff 1.5) genes between *Lgr5*- and
853 Paneth-derived intestinal tumors. **b**. Volcano plot highlighting differentially expressed (Pval < 0.01, FC
854 cutoff 1) genes between sporadic- and IBD-CRCs. **c**. Box plots relative to differentially expressed CBC (Lgr5-
855 derived) and RSC (PC-derived) marker genes between sporadic- and IBD-CRCs. P values denote results of
856 the t test. **d**. Bar plot relative to the gene set enrichment analysis between sCRC and IBD-CRC (abs NES >

857 0.5, Pval < 0.05). **e**. Box plots relative to stem cell index values across the colitis, sporadic, and intermediate
858 group of colon cancers. P value depicts result of one-way ANOVA.
859
860 **Supplementary Figure 9. The NWD1 western-style diet triggers the dedifferentiation of Paneth cells.**
861 **a**. GSEA plot of Paneth cells from NWD1- vs. AIN76A-fed mice. The DSS signature significantly associates
862 with the NWD1 diet (NES 2.99, Padj 2.5e-56). **b**. Bar plot showing pathways elevated (P < 0.05, NES > 0.5)
863 in Paneth cells from NWD1- when compared to AIN76A-fed animals. **c**. Prediction scores for Paneth (left)
864 and goblet (right) cells on UMAP embedding from the scATAC data (Choi et al.(*43*)). **d**. Unsupervised
865 clustering reveals two distinct clusters (c5, c20) that include Paneth cells. **e**. Bar plot listing gene ontology
866 results based on differential peaks from the diet responsive Paneth cells. **f**. Immunofluorescence imaging
867 of WGA and Olf4m showing intestinal crypts from mice fed with AIN/NWD1 diets. Asterisks earmark
868 double positive cells. Scale bar indicates 10 µm.
869
870 **Supplementary Figure 10. Epigenetic profiles of normal colonic lineages and mutational patterns of IBD-**
871 **related and sporadic colon cancers.**
872 **a**. UMAP plot displaying 1 Mb processed scATAC data of the human colon (Becker et al.(*50*)). **b**. Average
873 number of passenger mutations in IBD-related and sporadic colon cancers(*35*). **c**. Relative mutation
874 frequency in IBD-enriched (N = 78) and sCRC-enriched (N = 136) bins based on Wilcox test (p<0.01). **d**.
875 Barplot showing the genomic overview of differential bins across chromosomes. **e**. Boxplots relative to
876 the gene set variation analysis (GSVA) scores of tumors stratified according to their predicted cell of origin.
877 P values denote results of t tests.

878 **References**

879 1. N. Barker *et al.*, Crypt stem cells as the cells-of-origin of intestinal cancer. *Nature* **457**, 608-611 (2009).
880 2. S. Schwitalla *et al.*, Intestinal tumorigenesis initiated by dedifferentiation and acquisition of stem-cell-
881 like properties. *Cell* **152**, 25-38 (2013).
882 3. I. M. Shih *et al.*, Top-down morphogenesis of colorectal tumors. *Proc Natl Acad Sci U S A* **98**, 2640-2645
883 (2001).
884 4. W. Li *et al.*, The nutritional environment determines which and how intestinal stem cells contribute to
885 homeostasis and tumorigenesis. *Carcinogenesis* **40**, 937-946 (2019).
886 5. C. L. Bevins, N. H. Salzman, Paneth cells, antimicrobial peptides and maintenance of intestinal
887 homeostasis. *Nat Rev Microbiol* **9**, 356-368 (2011).
888 6. T. Sato *et al.*, Paneth cells constitute the niche for Lgr5 stem cells in intestinal crypts. *Nature* **469**, 415-
889 418 (2011).
890 7. M. E. Rothenberg *et al.*, Identification of a cKit(+) colonic crypt base secretory cell that supports Lgr5(+)
891 stem cells in mice. *Gastroenterology* **142**, 1195-1205 e1196 (2012).
892 8. N. Sasaki *et al.*, Reg4+ deep crypt secretory cells function as epithelial niche for Lgr5+ stem cells in
893 colon. *Proc Natl Acad Sci U S A* **113**, E5399-5407 (2016).
894 9. S. Roth *et al.*, Paneth cells in intestinal homeostasis and tissue injury. *PLoS One* **7**, e38965 (2012).
895 10. M. Schewe *et al.*, Secreted Phospholipases A2 Are Intestinal Stem Cell Niche Factors with Distinct Roles
896 in Homeostasis, Inflammation, and Cancer. *Cell Stem Cell* **19**, 38-51 (2016).
897 11. M. Schmitt *et al.*, Paneth Cells Respond to Inflammation and Contribute to Tissue Regeneration by
898 Acquiring Stem-like Features through SCF/c-Kit Signaling. *Cell Rep* **24**, 2312-2328 e2317 (2018).
899 12. E. C. Robanus-Maandag *et al.*, A new conditional Apc-mutant mouse model for colorectal cancer.
900 *Carcinogenesis* **31**, 946-952 (2010).
901 13. E. L. Jackson *et al.*, Analysis of lung tumor initiation and progression using conditional expression of
902 oncogenic K-ras. *Genes Dev* **15**, 3243-3248 (2001).
903 14. S. Marino, M. Vooijs, H. van Der Gulden, J. Jonkers, A. Berns, Induction of medulloblastomas in p53-
904 null mutant mice by somatic inactivation of Rb in the external granular layer cells of the cerebellum.
905 *Genes Dev* **14**, 994-1004 (2000).
906 15. N. Barker *et al.*, Identification of stem cells in small intestine and colon by marker gene Lgr5. *Nature*
907 **449**, 1003-1007 (2007).
908 16. J. H. van Es *et al.*, Enteroendocrine and tuft cells support Lgr5 stem cells on Paneth cell depletion. *Proc*
909 *Natl Acad Sci U S A*, (2019).

910 17.F. Gerbe *et al.*, Distinct ATOH1 and Neurog3 requirements define tuft cells as a new secretory cell type
911 in the intestinal epithelium. *J Cell Biol* **192**, 767-780 (2011).
912 18.Y. Nakanishi *et al.*, Dclk1 distinguishes between tumor and normal stem cells in the intestine. *Nat Genet*
913 **45**, 98-103 (2013).
914 19.C. B. Westphalen *et al.*, Long-lived intestinal tuft cells serve as colon cancer-initiating cells. *J Clin Invest*
915 **124**, 1283-1295 (2014).
916 20.X. Wu, D. Qu, N. Weygant, J. Peng, C. W. Houchen, Cancer Stem Cell Marker DCLK1 Correlates with
917 Tumorigenic Immune Infiltrates in the Colon and Gastric Adenocarcinoma Microenvironments. *Cancers*
918 *(Basel)* **12**, (2020).
919 21.K. P. Janssen *et al.*, APC and oncogenic KRAS are synergistic in enhancing Wnt signaling in intestinal
920 tumor formation and progression. *Gastroenterology* **131**, 1096-1109 (2006).
921 22.S. Asfaha *et al.*, Krt19(+)/Lgr5(-) Cells Are Radioresistant Cancer-Initiating Stem Cells in the Colon and
922 Intestine. *Cell Stem Cell* **16**, 627-638 (2015).
923 23.D. Castillo-Azofeifa *et al.*, Atoh1(+) secretory progenitors possess renewal capacity independent of
924 Lgr5(+) cells during colonic regeneration. *EMBO J* **38**, (2019).
925 24.R. Shindo *et al.*, Regenerating islet-derived protein (Reg)3beta plays a crucial role in attenuation of
926 ileitis and colitis in mice. *Biochem Biophys Rep* **21**, 100738 (2020).
927 25.A. L. Gartel *et al.*, p21 (WAF1/CIP1) expression is induced in newly nondividing cells in diverse epithelia
928 and during differentiation of the Caco-2 intestinal cell line. *Exp Cell Res* **227**, 171-181 (1996).
929 26.T. Sato *et al.*, Characterization of radioresistant epithelial stem cell heterogeneity in the damaged
930 mouse intestine. *Sci Rep* **10**, 8308 (2020).
931 27.L. Ricci-Vitiani *et al.*, Identification and expansion of human colon-cancer-initiating cells. *Nature* **445**,
932 111-115 (2007).
933 28.A. Ayyaz *et al.*, Single-cell transcriptomes of the regenerating intestine reveal a revival stem cell. *Nature*
934 **569**, 121-125 (2019).
935 29.E. G. Vasquez *et al.*, Dynamic and adaptive cancer stem cell population admixture in colorectal
936 neoplasia. *Cell Stem Cell* **29**, 1213-1228 e1218 (2022).
937 30.P. N. P. Singh, S. Madha, A. B. Leiter, R. A. Shivdasani, Cell and chromatin transitions in intestinal stem
938 cell regeneration. *Genes Dev* **36**, 684-698 (2022).
939 31.L. Chen *et al.*, Dynamic Chromatin States Coupling with Key Transcription Factors in Colitis-Associated
940 Colorectal Cancer. *Adv Sci (Weinh)* **9**, e2200536 (2022).

941 32.P. N. Vega *et al.*, Cancer-Associated Fibroblasts and Squamous Epithelial Cells Constitute a Unique
942 Microenvironment in a Mouse Model of Inflammation-Induced Colon Cancer. *Front Oncol* **12**, 878920
943 (2022).
944 33.T. Tanaka *et al.*, A novel inflammation-related mouse colon carcinogenesis model induced by
945 azoxymethane and dextran sodium sulfate. *Cancer Sci* **94**, 965-973 (2003).
946 34.C. B. Steen, C. L. Liu, A. A. Alizadeh, A. M. Newman, Profiling Cell Type Abundance and Expression in
947 Bulk Tissues with CIBERSORTx. *Methods Mol Biol* **2117**, 135-157 (2020).
948 35.K. Rajamaki *et al.*, Genetic and Epigenetic Characteristics of Inflammatory Bowel Disease-Associated
949 Colorectal Cancer. *Gastroenterology* **161**, 592-607 (2021).
950 36.A. Subramanian *et al.*, Gene set enrichment analysis: a knowledge-based approach for interpreting
951 genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**, 15545-15550 (2005).
952 37.H. O. Lee *et al.*, Lineage-dependent gene expression programs influence the immune landscape of
953 colorectal cancer. *Nat Genet* **52**, 594-603 (2020).
954 38.J. Guinney *et al.*, The consensus molecular subtypes of colorectal cancer. *Nat Med* **21**, 1350-1356
955 (2015).
956 39.P. Munkholm, Review article: the incidence and prevalence of colorectal cancer in inflammatory bowel
957 disease. *Aliment Pharmacol Ther* **18 Suppl 2**, 1-5 (2003).
958 40.A. Christ, M. Lauterbach, E. Latz, Western Diet and the Immune System: An Inflammatory Connection.
959 *Immunity* **51**, 794-811 (2019).
960 41.T. C. Liu *et al.*, Western diet induces Paneth cell defects through microbiome alterations and farnesoid
961 X receptor and type I interferon activation. *Cell Host Microbe* **29**, 988-1001 e1006 (2021).
962 42.H. L. Newmark *et al.*, Western-style diet-induced colonic tumors and their modulation by calcium and
963 vitamin D in C57Bl/6 mice: a preclinical model for human sporadic colon cancer. *Carcinogenesis* **30**, 88-
964 92 (2009).
965 43.J. Choi *et al.*, Dynamic Intestinal Stem Cell Plasticity and Lineage Remodeling by a Nutritional
966 Environment Relevant to Human Risk for Tumorigenesis. *Mol Cancer Res* **21**, 808-824 (2023).
967 44.M. Schewe, A. Sacchetti, M. Schmitt, R. Fodde, The Organoid Reconstitution Assay (ORA) for the
968 Functional Analysis of Intestinal Stem and Niche Cells. *J Vis Exp*, (2017).
969 45.G. S. Gulati *et al.*, Single-cell transcriptional diversity is a hallmark of developmental potential. *Science*
970 **367**, 405-411 (2020).
971 46.E. E. L. Nystrom *et al.*, An intercrypt subpopulation of goblet cells is essential for colonic mucus barrier
972 function. *Science* **372**, (2021).

973 47.K. Kübler *et al.*, Tumor mutational landscape is a record of the pre-malignant state. *bioRxiv*, 517565
974 (2019).
975 48.P. Polak *et al.*, Cell-of-origin chromatin organization shapes the mutational landscape of cancer. *Nature*
976 **518**, 360-364 (2015).
977 49.S. Yang *et al.*, COOBoostR: An Extreme Gradient Boosting-Based Tool for Robust Tissue or Cell-of-Origin
978 Prediction of Tumors. *Life (Basel)* **13**, (2022).
979 50.W. R. Becker *et al.*, Single-cell analyses define a continuum of cell state and composition changes in
980 the malignant transformation of polyps to colorectal cancer. *Nat Genet* **54**, 985-995 (2022).
981 51.C. S. Smillie *et al.*, Intra- and Inter-cellular Rewiring of the Human Colon during Ulcerative Colitis. *Cell*
982 **178**, 714-730 e722 (2019).
983 52.G. Ito *et al.*, Lineage-specific expression of bestrophin-2 and bestrophin-4 in human intestinal epithelial
984 cells. *PLoS One* **8**, e79693 (2013).
985 53.K. Parikh *et al.*, Colonic epithelial cell diversity in health and inflammatory bowel disease. *Nature* **567**,
986 49-55 (2019).
987 54.K. Peregrina *et al.*, Vitamin D is a determinant of mouse intestinal Lgr5 stem cell functions.
988 *Carcinogenesis* **36**, 25-31 (2015).
989 55.C. Tomasetti, B. Vogelstein, Cancer etiology. Variation in cancer risk among tissues can be explained by
990 the number of stem cell divisions. *Science* **347**, 78-81 (2015).
991 56.S. Wu, S. Powers, W. Zhu, Y. A. Hannun, Substantial contribution of extrinsic risk factors to cancer
992 development. *Nature* **529**, 43-47 (2016).
993 57.S. Yu *et al.*, Paneth Cell Multipotency Induced by Notch Activation following Injury. *Cell Stem Cell* **23**,
994 46-59 e45 (2018).
995 58.J. C. Jones *et al.*, Cellular Plasticity of Defa4(Cre)-Expressing Paneth Cells in Response to Notch
996 Activation and Intestinal Injury. *Cell Mol Gastroenterol Hepatol* **7**, 533-554 (2019).
997 59.P. Bankhead *et al.*, QuPath: Open source software for digital pathology image analysis. *Sci Rep* **7**, 16878
998 (2017).
999 60.T. Sato *et al.*, Single Lgr5 stem cells build crypt-villus structures in vitro without a mesenchymal niche.
1000 *Nature* **459**, 262-265 (2009).
1001 61.G. X. Zheng *et al.*, Massively parallel digital transcriptional profiling of single cells. *Nat Commun* **8**,
1002 14049 (2017).
1003 62.Y. Hao *et al.*, Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587 e3529 (2021).

1004 63.C. S. McGinnis, L. M. Murrow, Z. J. Gartner, DoubletFinder: Doublet Detection in Single-Cell RNA
1005 Sequencing Data Using Artificial Nearest Neighbors. *Cell Syst* **8**, 329-337 e324 (2019).
1006 64.S. Hanzelmann, R. Castelo, J. Guinney, GSVA: gene set variation analysis for microarray and RNA-seq
1007 data. *BMC Bioinformatics* **14**, 7 (2013).
1008 65.Y. Chen *et al.*, SOAPnuke: a MapReduce acceleration-supported software for integrated quality control
1009 and preprocessing of high-throughput sequencing data. *Gigascience* **7**, 1-6 (2018).
1010 66.A. Dobin *et al.*, STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21 (2013).
1011 67.A. Tarasov, A. J. Vilella, E. Cuppen, I. J. Nijman, P. Prins, Sambamba: fast processing of NGS alignment
1012 formats. *Bioinformatics* **31**, 2032-2034 (2015).
1013 68.Y. Liao, G. K. Smyth, W. Shi, featureCounts: an efficient general purpose program for assigning
1014 sequence reads to genomic features. *Bioinformatics* **30**, 923-930 (2014).
1015 69.M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data
1016 with DESeq2. *Genome Biol* **15**, 550 (2014).
1017 70.A. McKenna *et al.*, The Genome Analysis Toolkit: a MapReduce framework for analyzing next-
1018 generation DNA sequencing data. *Genome Res* **20**, 1297-1303 (2010).
1019 71.T. Stuart, A. Srivastava, S. Madad, C. A. Lareau, R. Satija, Single-cell chromatin state analysis with Signac.
1020 *Nat Methods* **18**, 1333-1341 (2021).
1021 72.A. L. Haber *et al.*, A single-cell survey of the small intestinal epithelium. *Nature* **551**, 333-339 (2017).
1022 73.E. Dann, N. C. Henderson, S. A. Teichmann, M. D. Morgan, J. C. Marioni, Differential abundance testing
1023 on single-cell data using k-nearest neighbor graphs. *Nat Biotechnol* **40**, 245-253 (2022).

1024

Acknowledgements

We thank E. Bindels and R. Hoogenboezem for their support to generate the single cell sequencing data, J.D. Windster for help with in situ hybridizations, W.S. van de Geer for analysis support of bulk sequencing data; W. Szymanska, E. Middendorp Guerra and E. de Vet for help with the IHC and scRNAseq analyses; V.H.A.E. Verhagen for discussions on XGBoost machine learning. This study was financially supported by the Dutch Cancer Society (KWF; project no. 11407), the Wereld Kanker Onderzoek Fonds and World Cancer Research Funds (WKOF/WCRF; projects no. 2014-1181 and IIG_FULL_2022_015), and in part by the National Cancer Institute (NCI; project no. R01CA214625, R01CA229216, and P30-013330), the National Institute on Aging (NIA; project no. P30 AG038072, 5T32AG023475-20), Academy of Finland (Finnish Center of Excellence Program 2018–2025 312041, Academy Professor grant no. 319083 and 320149), iCAN Digital Precision Cancer Medicine Flagship (320185), Cancer Society of Finland, Sigrid Jusélius Foundation (220002), and the Jane and Aatos Erkko Foundation.

Author contributions

M.P.V., R.J., M.S. and A.S. designed and performed experiments. M.P.V., J.C., N.V., L.A.A. and L.H.A. generated and analyzed sequencing data sets. K.R., P.P, S.S, B.v.d.S. and T.P.P.v.d.B. performed immunohistochemistry. D.S. and M.D. were responsible for pathological examination and IBD patient selection. M.P.V. and R.F. wrote the manuscript. R.F. supervised the study and was responsible for concept and design of the study.

Competing interests

The authors declare no competing interests.

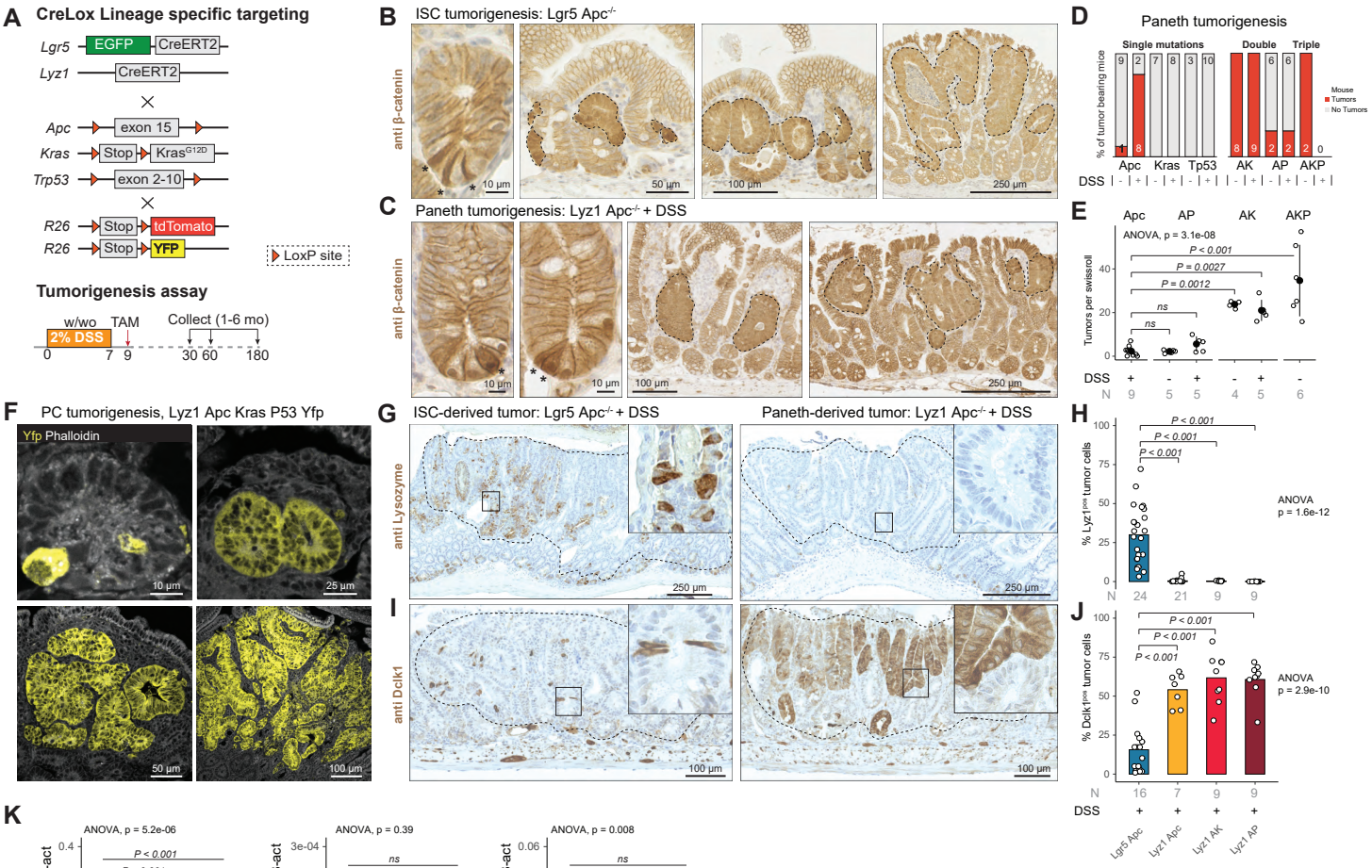

A
CreLox Lineage specific targeting
Lgr5
EGFP
CreERT2
Lyz1
CreERT2
Apc
exon 15
Kras
Stop
KrasG12D
Trp53
exon 2-10
R26
Stop
tdTomato
R26
Stop
YFP
LoxP site
Tumorigenesis assay
w/wo TAM
2% DSS
Collect (1-6 mo)
0 7 9 30 60 180
B
ISC tumorigenesis: Lgr5 Apc-/-
anti β-catenin
10 µm
50 µm
100 µm
250 µm
C
Paneth tumorigenesis: Lyz1 Apc-/- + DSS
anti β-catenin
10 µm
10 µm
100 µm
250 µm
D
Paneth tumorigenesis
Single mutations
Double
Triple
% of tumor bearing mice
Mouse
Tumors
No Tumors
Apc Kras Tp53 AK AP AKP
DSS
E
Apc AP AK AKP
ANOVA, p = 3.1e-08
P < 0.001
P = 0.0027
P = 0.0012
ns
ns
Tumors per swissroll
DSS
N 9 5 5 4 5 6
F
PC tumorigenesis, Lyz1 Apc Kras P53 Yfp
Yfp Phalloidin
10 µm
25 µm
50 µm
100 µm
G
ISC-derived tumor: Lgr5 Apc-/- + DSS
Paneth-derived tumor: Lyz1 Apc-/- + DSS
anti Lysozyme
250 µm
250 µm
H
P < 0.001
P < 0.001
P < 0.001
% Lyz1pos tumor cells
ANOVA
p = 1.6e-12
N 24 21 9 9
I
anti Dclk1
100 µm
100 µm
J
P < 0.001
P < 0.001
P < 0.001
% Dclk1pos tumor cells
ANOVA
p = 2.9e-10
N 16 7 9 9
DSS
Lgr5 Apc
Lyz1 Apc
Lyz1 AK
Lyz1 AP
K
ANOVA, p = 5.2e-06
P < 0.001
P < 0.001
P < 0.001
Lyz1 expression against B-act
N 5 3 8 8
DSS
Lgr5 Apc
Lyz1 Apc
Lyz1 Apc Kras
ANOVA, p = 0.39
ns
Dclk1 expression against B-act
N 5 3 7 8
DSS
Lgr5 Apc
Lyz1 Apc
Lyz1 Apc Kras
ANOVA, p = 0.008
ns
P = 0.022
P = 0.017
Axin2 expression against B-act
N 5 3 8 8
DSS
Lgr5 Apc
Lyz1 Apc
Lyz1 Apc Kras

Fig. 1

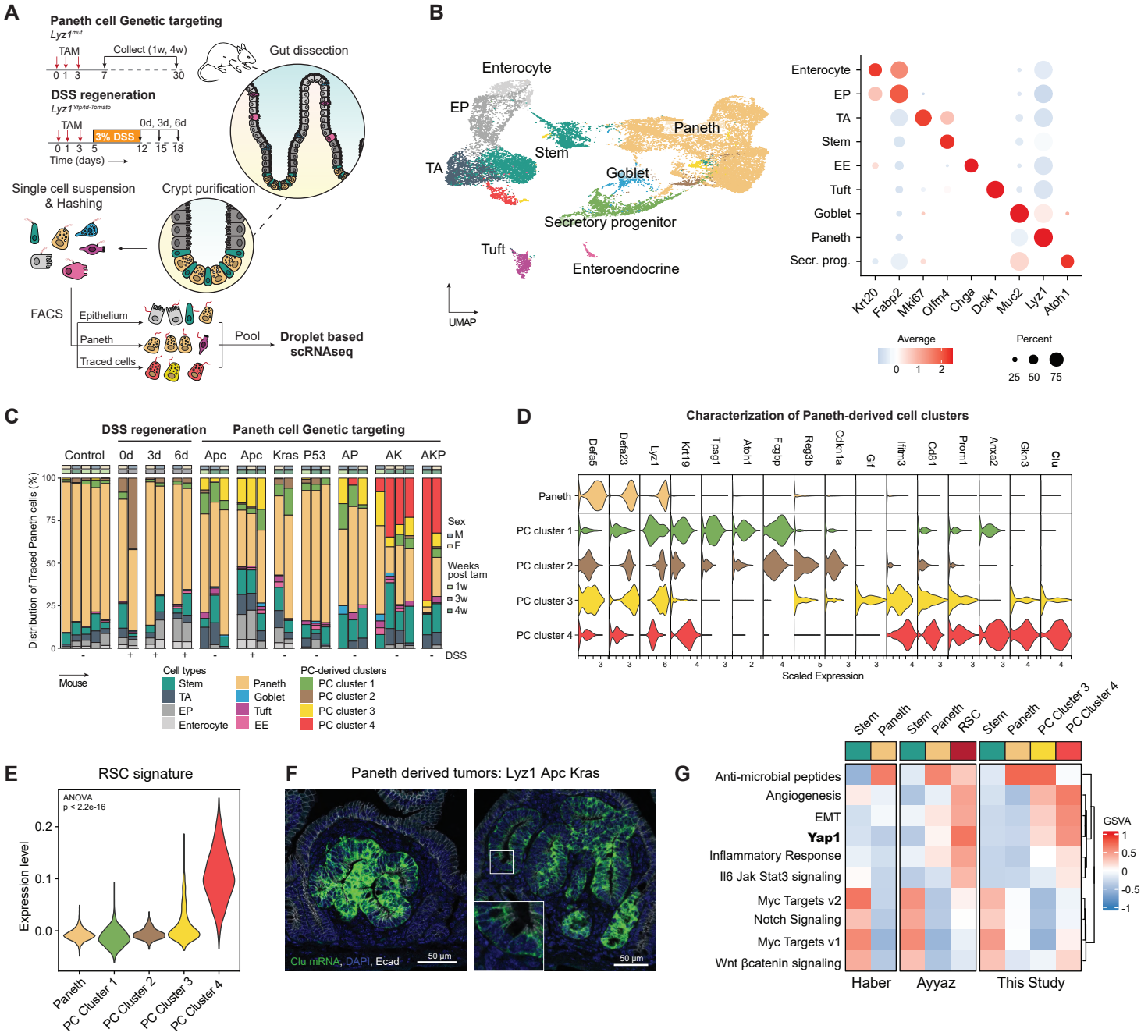

A
Paneth cell Genetic targeting
Lyz1mut
TAM
Collect (1w, 4w)
0 1 3 7 30
Gut dissection
DSS regeneration
Lyz1Yfp/td-Tomato
TAM
3% DSS
0d, 3d, 6d
0 1 3 5 12 15 18
Time (days)
Single cell suspension & Hashing
Crypt purification
FACS
Epithelium
Paneth
Traced cells
Pool
Droplet based scRNAseq
B
Enterocyte
EP
TA
Stem
Paneth
Goblet
Secretory progenitor
Tuft
Enteroendocrine
UMAP
Enterocyte
EP
TA
Stem
EE
Tuft
Goblet
Paneth
Secr. prog.
Krt20
Fabp2
Mki67
Olfm4
Chga
Dclk1
Muc2
Lyz1
Atoh1
Average
0 1 2
Percent
25 50 75
C
DSS regeneration
Paneth cell Genetic targeting
Control
0d
3d
6d
Apc
Apc
Kras P53
AP
AK
AKP
Distribution of Traced Paneth cells (%)
100
75
50
25
0
Sex
M
F
Weeks post tam
1w
3w
4w
DSS
Mouse
Cell types
Stem
TA
EP
Enterocyte
Paneth
Goblet
Tuft
EE
PC-derived clusters
PC cluster 1
PC cluster 2
PC cluster 3
PC cluster 4
D
Characterization of Paneth-derived cell clusters
Defa5
Defa23
Lyz1
Krt19
Tpsg1
Atoh1
Fcgbp
Reg3b
Cdkn1a
Glf
Ifitm3
Cd81
Prom1
Anxa2
Gkn3
Clu
Paneth
PC cluster 1
PC cluster 2
PC cluster 3
PC cluster 4
Scaled Expression
E
RSC signature
ANOVA
p < 2.2e-16
Expression level
0.2
0.1
0.0
Paneth
PC Cluster 1
PC Cluster 2
PC Cluster 3
PC Cluster 4
F
Paneth derived tumors: Lyz1 Apc Kras
Clu mRNA, DAPI, Ecad
50 µm
50 µm
G
Stem
Paneth
Stem
Paneth
RSC
Stem
Paneth
PC Cluster 3
PC Cluster 4
Anti-microbial peptides
Angiogenesis
EMT
Yap1
Inflammatory Response
Il6 Jak Stat3 signaling
Myc Targets v2
Notch Signaling
Myc Targets v1
Wnt βcatenin signaling
GSVA
1
0.5
0
-0.5
-1
Haber
Ayyaz
This Study

Fig. 2

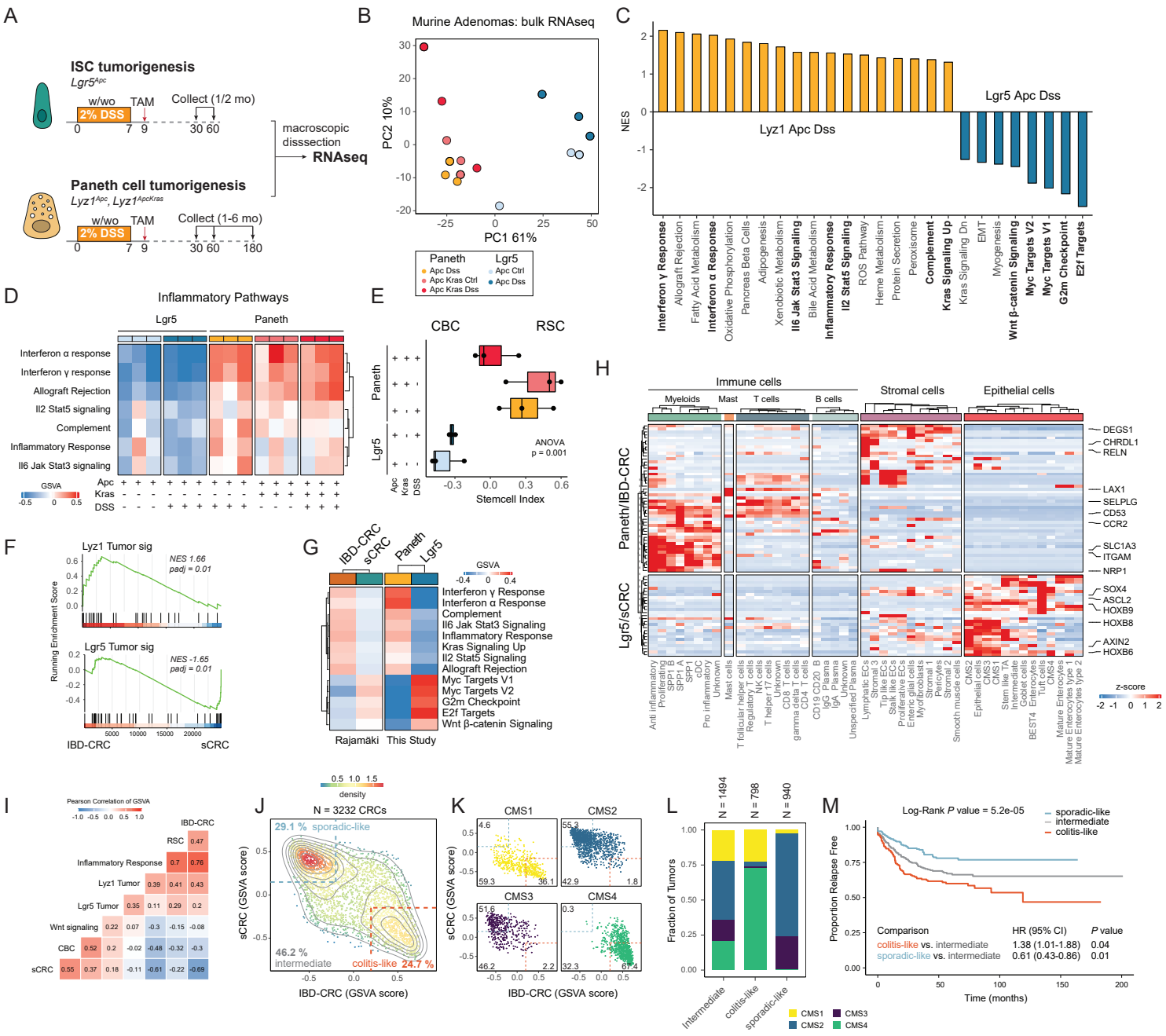

A
ISC tumorigenesis
Lgr5Apc
w/wo
TAM
2% DSS
Collect (1/2 mo)
0
7
9
30 60
macroscopic disssection
RNAseq
Paneth cell tumorigenesis
Lyz1Apc, Lyz1ApcKras
w/wo
TAM
2% DSS
Collect (1-6 mo)
0
7
9
30 60
180
B
Murine Adenomas: bulk RNAseq
PC2 10%
PC1 61%
Paneth
Apc Dss
Apc Kras Ctrl
Apc Kras Dss
Lgr5
Apc Ctrl
Apc Dss
C
NES
Lgr5 Apc Dss
Lyz1 Apc Dss
Interferon γ Response
Allograft Rejection
Fatty Acid Metabolism
Interferon α Response
Oxidative Phosphorylation
Pancreas Beta Cells
Adipogenesis
Xenobiotic Metabolism
Il6 Jak Stat3 Signaling
Bile Acid Metabolism
Inflammatory Response
Il2 Stat5 Signaling
ROS Pathway
Heme Metabolism
Protein Secretion
Peroxisome
Complement
Kras Signaling Up
Kras Signaling Dn
EMT
Myogenesis
Wnt β-catenin Signaling
Myc Targets V2
Myc Targets V1
G2m Checkpoint
E2f Targets
D
Inflammatory Pathways
Lgr5
Paneth
Interferon α response
Interferon γ response
Allograft Rejection
Il2 Stat5 signaling
Complement
Inflammatory Response
Il6 Jak Stat3 signaling
GSVA
Apc
Kras
DSS
E
CBC
RSC
Paneth
Lgr5
ANOVA
p = 0.001
Apc
Kras
DSS
Stemcell Index
H
Immune cells
Myeloids
Mast
T cells
B cells
Stromal cells
Epithelial cells
Paneth/IBD-CRC
Lgr5/sCRC
DEGS1
CHRDL1
RELN
LAX1
SELPLG
CD53
CCR2
SLC1A3
ITGAM
NRP1
SOX4
ASCL2
HOXB9
HOXB8
AXIN2
HOXB6
z-score
F
Lyz1 Tumor sig
NES 1.66
padj = 0.01
Lgr5 Tumor sig
NES -1.65
padj = 0.01
Running Enrichment Score
IBD-CRC
sCRC
G
IBD-CRC
sCRC
Paneth
Lgr5
GSVA
Interferon γ Response
Interferon α Response
Complement
Il6 Jak Stat3 Signaling
Inflammatory Response
Kras Signaling Up
Il2 Stat5 Signaling
Allograft Rejection
Myc Targets V1
Myc Targets V2
G2m Checkpoint
E2f Targets
Wnt β-catenin Signaling
Rajamäki
This Study
I
Pearson Correlation of GSVA
IBD-CRC
RSC
Inflammatory Response
Lyz1 Tumor
Lgr5 Tumor
Wnt signaling
CBC
sCRC
J
density
N = 3232 CRCs
29.1 % sporadic-like
46.2 % intermediate
colitis-like 24.7 %
sCRC (GSVA score)
IBD-CRC (GSVA score)
K
CMS1
CMS2
CMS3
CMS4
sCRC (GSVA score)
IBD-CRC (GSVA score)
L
N = 1494
N = 798
N = 940
Fraction of Tumors
Intermediate
colitis-like
sporadic-like
CMS1
CMS2
CMS3
CMS4
M
Log-Rank P value = 5.2e-05
sporadic-like
intermediate
colitis-like
Proportion Relapse Free
Comparison
HR (95% CI)
P value
colitis-like vs. intermediate
1.38 (1.01-1.88)
0.04
sporadic-like vs. intermediate
0.61 (0.43-0.86)
0.01
Time (months)

Fig. 3

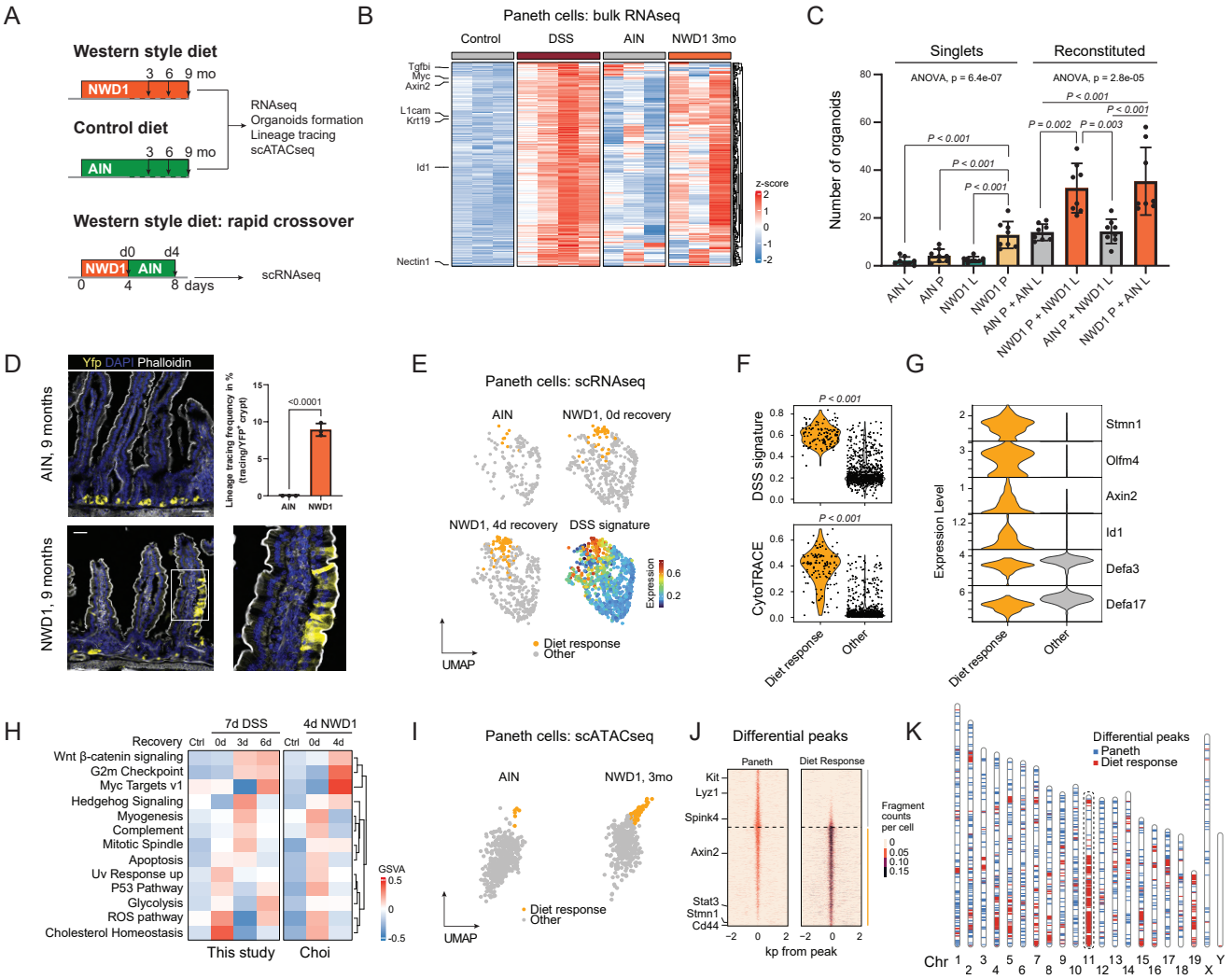

A
Western style diet
3 6 9 mo
NWD1
Control diet
3 6 9 mo
AIN
RNAseq
Organoids formation
Lineage tracing
scATACseq
Western style diet: rapid crossover
d0 d4
NWD1 AIN
0 4 8 days
scRNAseq
B
Paneth cells: bulk RNAseq
Control
DSS
AIN
NWD1 3mo
Tgfbi
Myc
Axin2
L1cam
Krt19
Id1
Nectin1
z-score
C
Singlets
Reconstituted
ANOVA, p = 6.4e-07
ANOVA, p = 2.8e-05
Number of organoids
AIN L
AIN P
NWD1 L
NWD1 P
AIN P + AIN L
NWD1 P + NWD1 L
AIN P + NWD1 L
NWD1 P + AIN L
D
Yfp DAPI Phalloidin
AIN, 9 months
NWD1, 9 months
Lineage tracing frequency in % (tracing/YFP+ crypt)
<0.0001
AIN
NWD1
E
Paneth cells: scRNAseq
AIN
NWD1, 0d recovery
NWD1, 4d recovery
DSS signature
Expression
Diet response
Other
UMAP
F
DSS signature
CytoTRACE
P < 0.001
Diet response
Other
G
Expression Level
Stmn1
Olfm4
Axin2
Id1
Defa3
Defa17
Diet response
Other
H
7d DSS
4d NWD1
Recovery
Ctrl 0d 3d 6d
Ctrl 0d 4d
Wnt β-catenin signaling
G2m Checkpoint
Myc Targets v1
Hedgehog Signaling
Myogenesis
Complement
Mitotic Spindle
Apoptosis
Uv Response up
P53 Pathway
Glycolysis
ROS pathway
Cholesterol Homeostasis
GSVA
This study
Choi
I
Paneth cells: scATACseq
AIN
NWD1, 3mo
Diet response
Other
UMAP
J
Differential peaks
Paneth
Diet Response
Kit
Lyz1
Spink4
Axin2
Stat3
Stmn1
Cd44
Fragment counts per cell
kp from peak
K
Differential peaks
Paneth
Diet response
Chr

Fig. 4

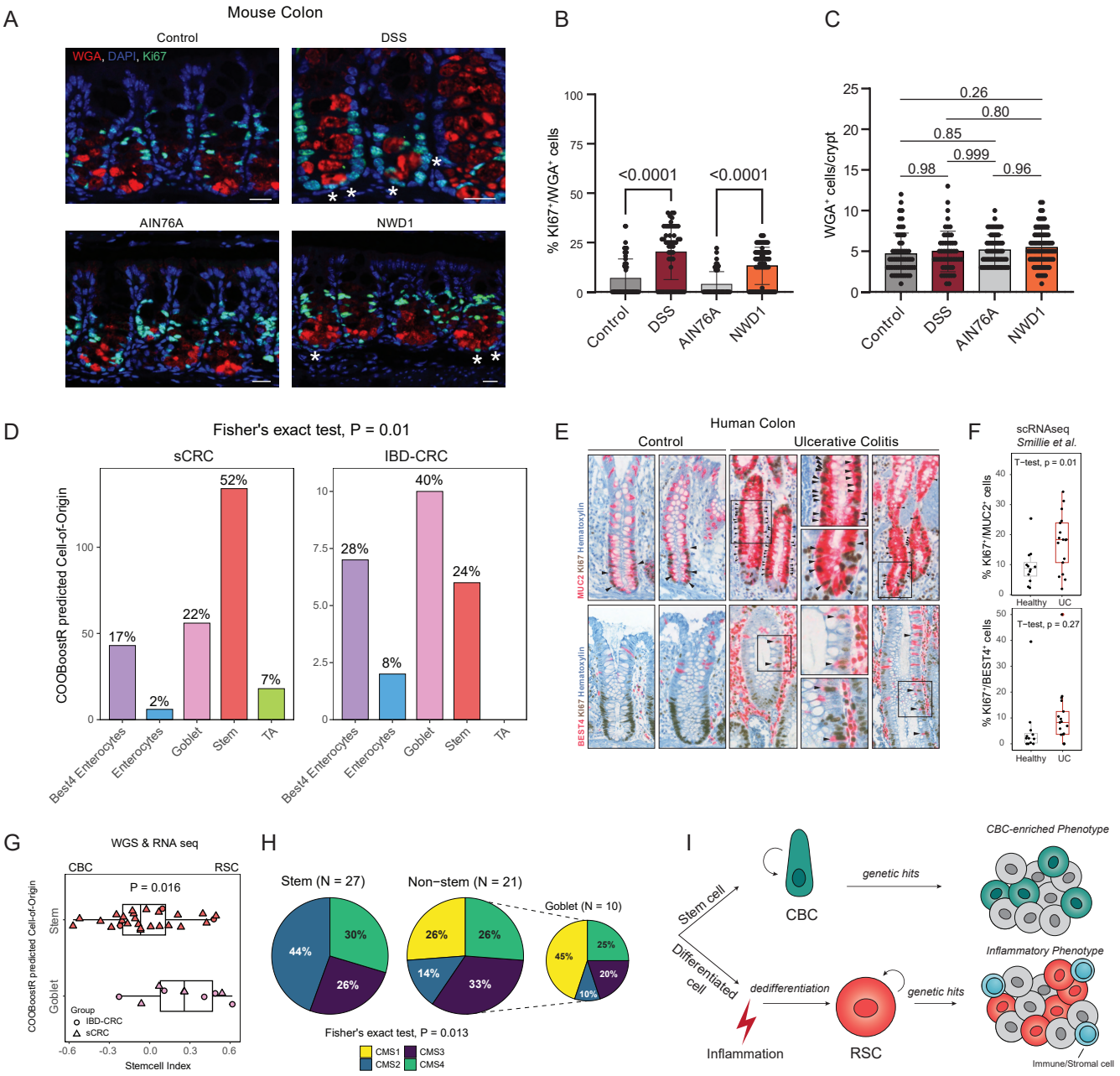
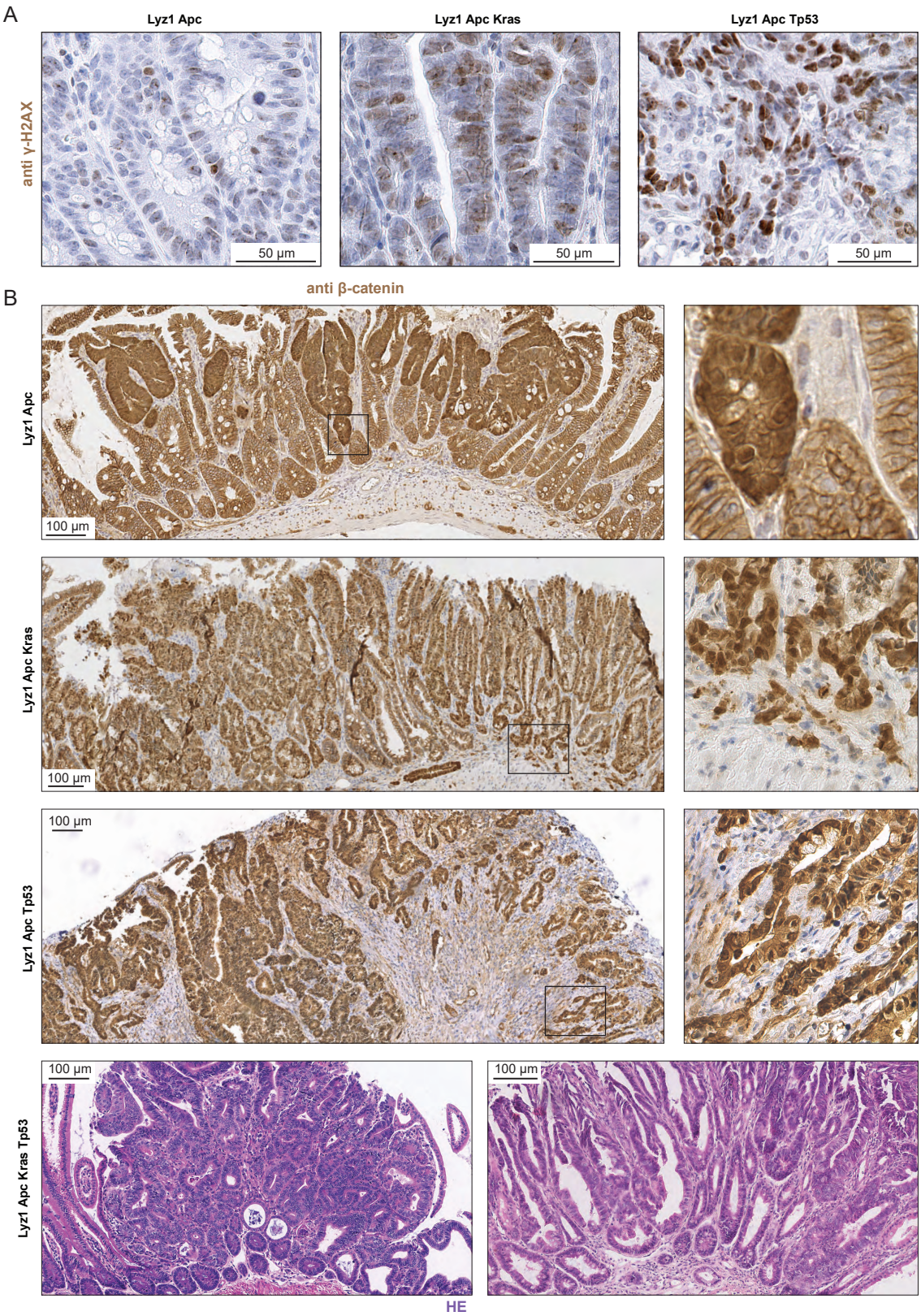
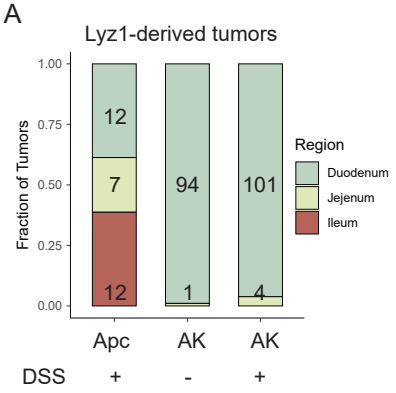

A
Mouse Colon
Control
DSS
WGA, DAPI, Ki67
AIN76A
NWD1
B
% KI67+/WGA+ cells
<0.0001
<0.0001
Control
DSS
AIN76A
NWD1
C
WGA+ cells/crypt
0.26
0.80
0.85
0.999
0.98
0.96
D
Fisher's exact test, P = 0.01
sCRC
IBD-CRC
COOBoostR predicted Cell-of-Origin
17%
2%
22%
52%
7%
28%
8%
40%
24%
Best4 Enterocytes
Enterocytes
Goblet
Stem
TA
E
Human Colon
Control
Ulcerative Colitis
MUC2 Ki67 Hematoxylin
BEST4 Ki67 Hematoxylin
F
scRNAseq
Smillie et al.
T-test, p = 0.01
% KI67+/MUC2+ cells
T-test, p = 0.27
% KI67+/BEST4+ cells
Healthy
UC
G
WGS & RNA seq
CBC
RSC
P = 0.016
Stem
Goblet
COOBoostR predicted Cell-of-Origin
Group
IBD-CRC
sCRC
Stemcell Index
H
Stem (N = 27)
Non−stem (N = 21)
Goblet (N = 10)
30%
26%
44%
26%
26%
14%
33%
25%
20%
45%
10%
Fisher's exact test, P = 0.013
CMS1
CMS2
CMS3
CMS4
I
Stem cell
Differentiated cell
CBC
genetic hits
CBC-enriched Phenotype
dedifferentiation
genetic hits
Inflammation
RSC
Inflammatory Phenotype
Immune/Stromal cell

Fig. 5

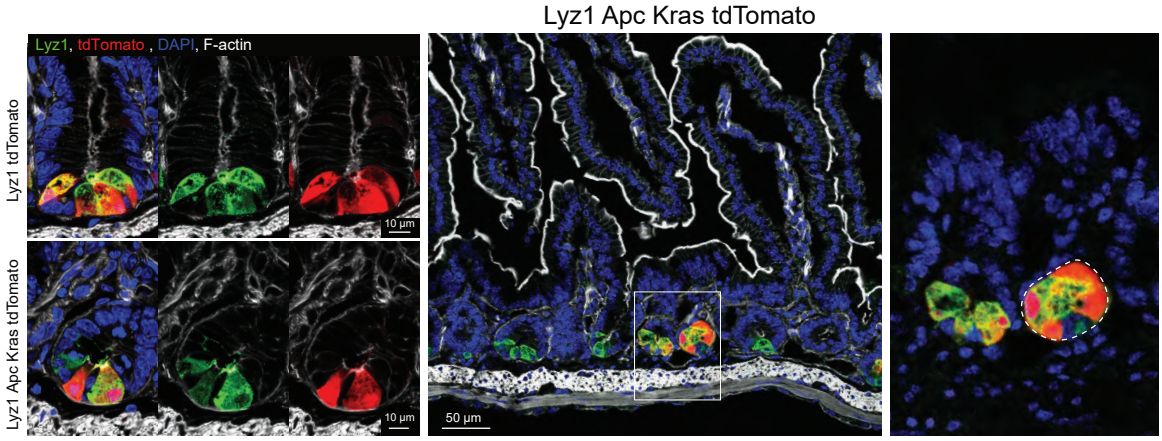

A
Lyz1 Apc
Lyz1 Apc Kras
Lyz1 Apc Tp53
anti γ-H2AX
50 µm
50 µm
50 µm
B
anti β-catenin
Lyz1 Apc
100 µm
Lyz1 Apc Kras
100 µm
Lyz1 Apc Tp53
100 µm
Lyz1 Apc Kras Tp53
100 µm
100 µm
HE

Suppl. Fig. 1

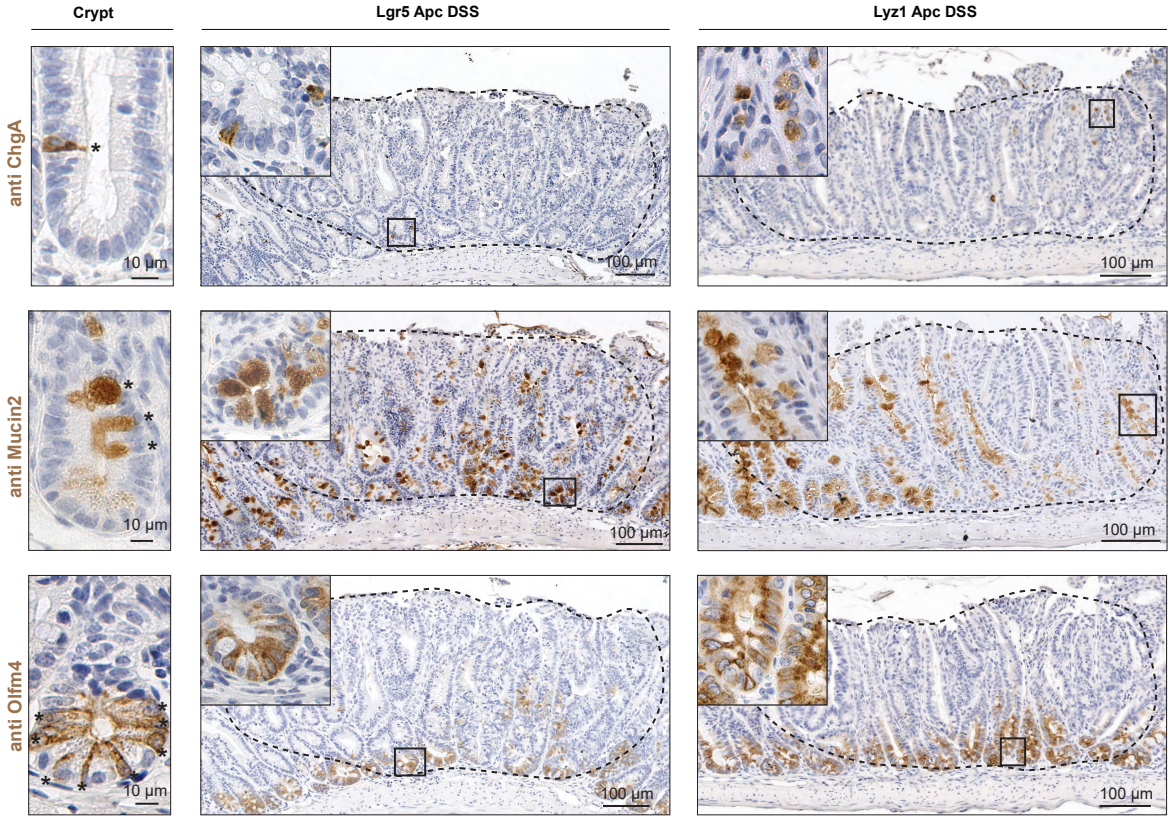
A



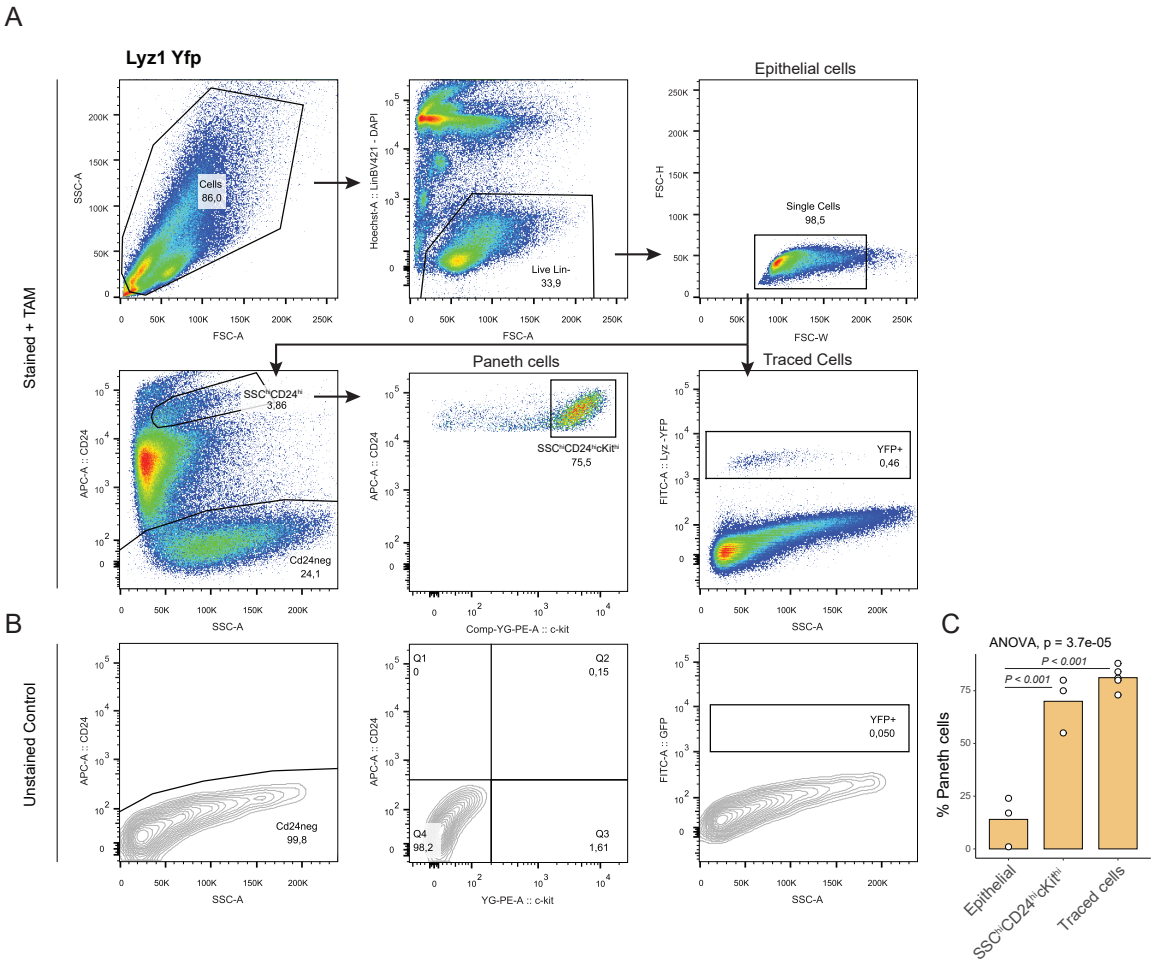
B



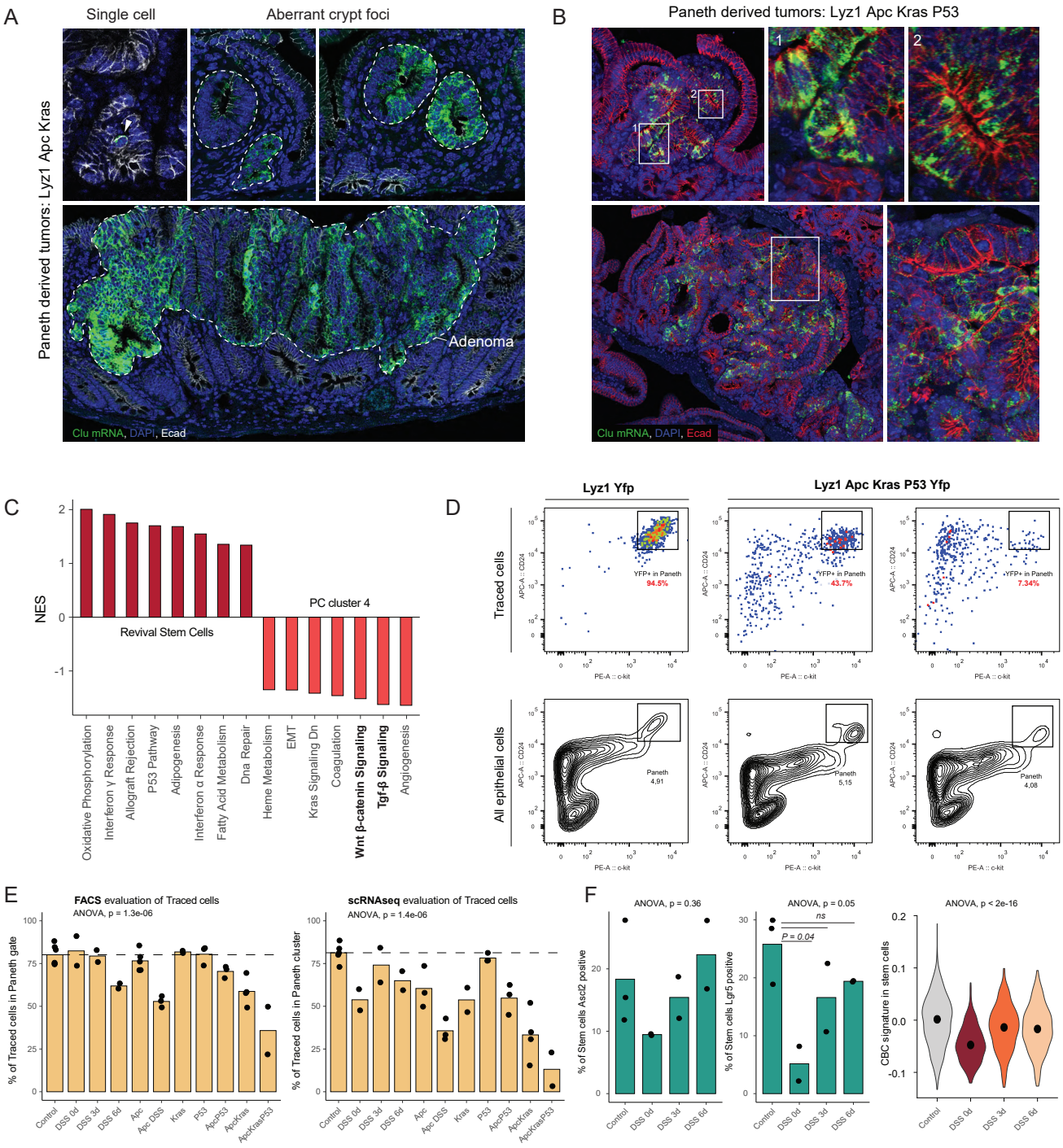
C



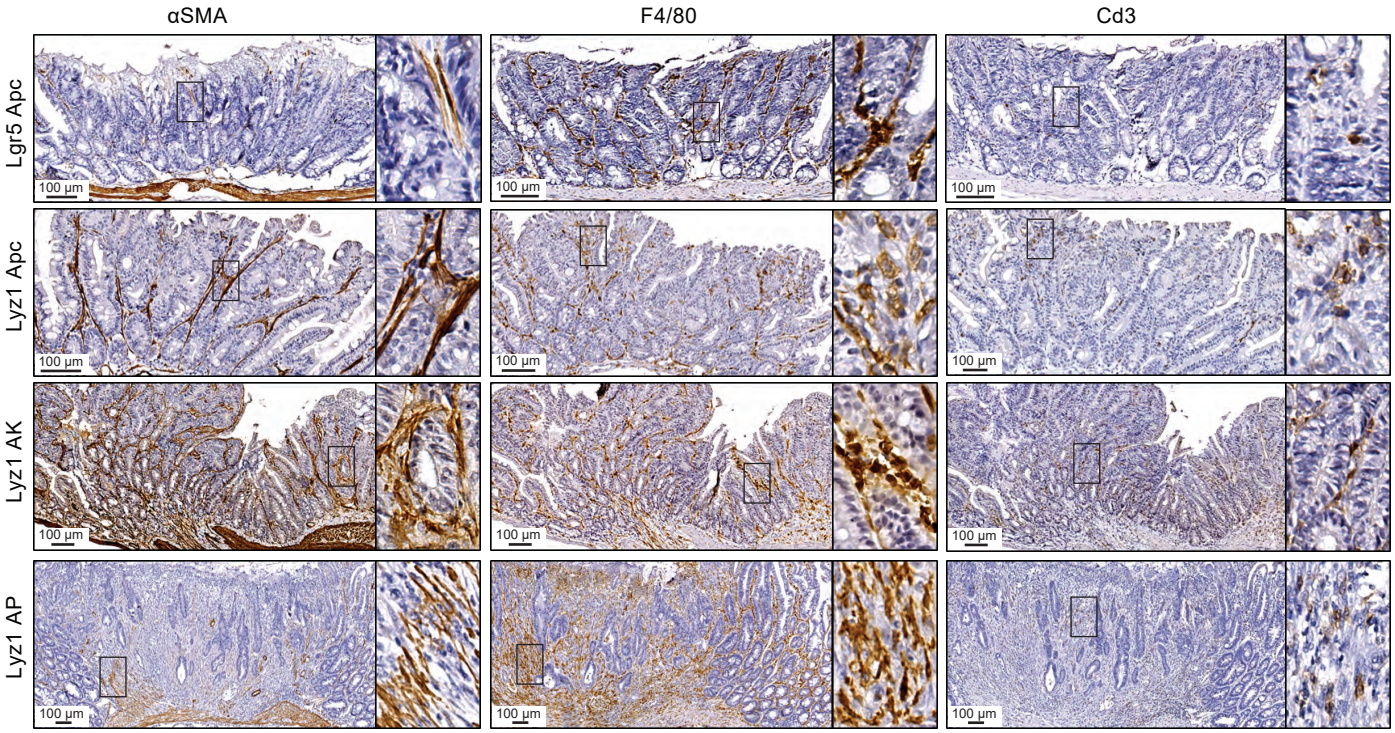
Suppl. Fig. 2


A
Lyz1 Yfp
Stained + TAM
Cells
86,0
SSC-A
FSC-A
Hoechst-A :: LinBV421 - DAPI
Live Lin-
33,9
Epithelial cells
Single Cells
98,5
FSC-H
FSC-W
Paneth cells
SSChiCD24hi
3,86
Cd24neg
24,1
APC-A :: CD24
SSC-A
SSChiCD24hicKithi
75,5
Comp-YG-PE-A :: c-kit
Traced Cells
YFP+
0,46
FITC-A :: Lyz-YFP
B
Unstained Control
Cd24neg
99,8
Q1
0
Q2
0,15
Q4
98,2
Q3
1,61
YG-PE-A :: c-kit
YFP+
0,050
FITC-A :: GFP
C
ANOVA, p = 3.7e-05
P < 0.001
P < 0.001
% Paneth cells
Epithelial
SSChiCD24hicKithi
Traced cells

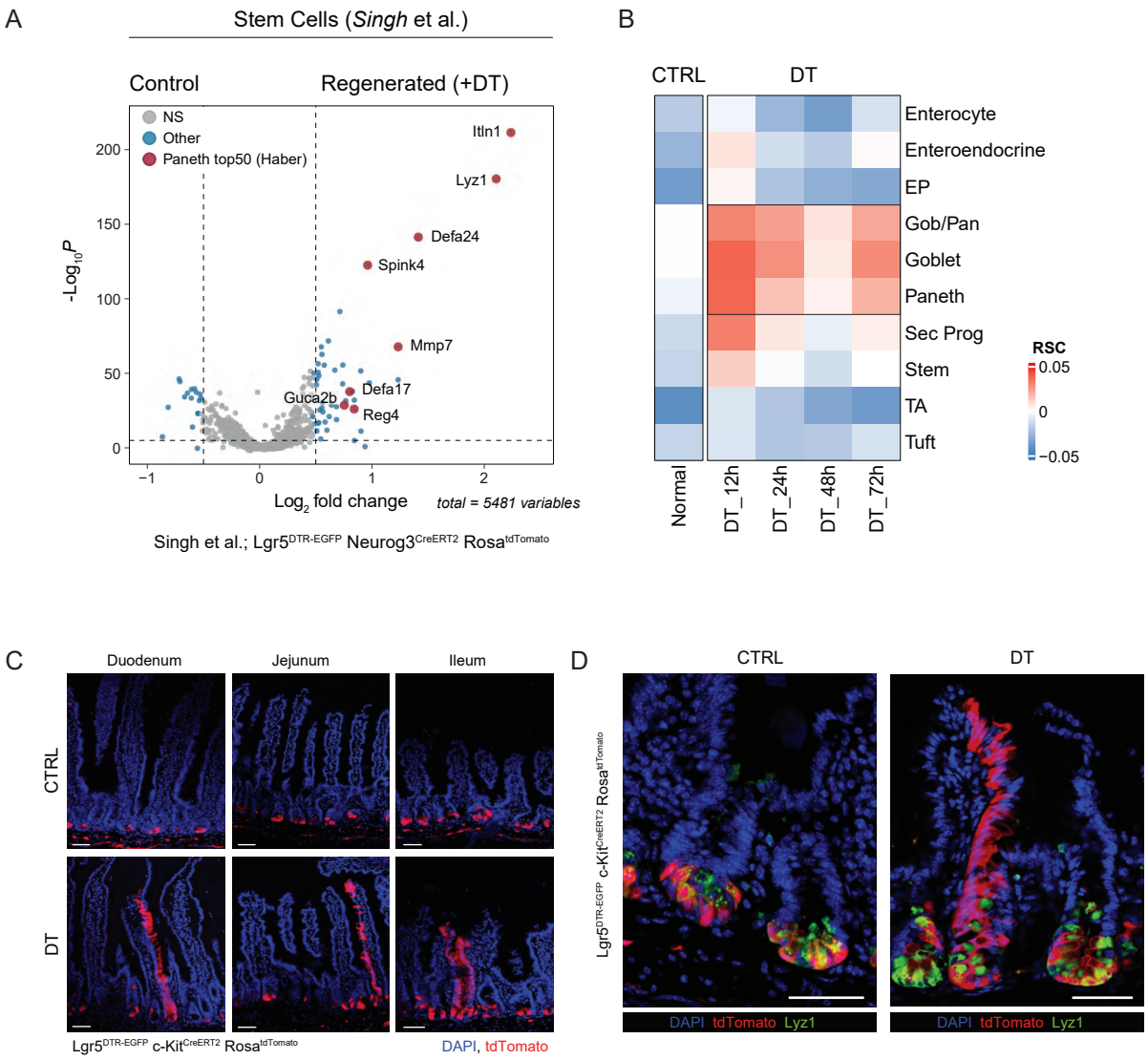
Suppl. Fig. 3


A
Single cell
Aberrant crypt foci
Paneth derived tumors: Lyz1 Apc Kras
Adenoma
Clu mRNA, DAPI, Ecad
B
Paneth derived tumors: Lyz1 Apc Kras P53
Clu mRNA, DAPI, Ecad
C
NES
Revival Stem Cells
PC cluster 4
Oxidative Phosphorylation
Interferon γ Response
Allograft Rejection
P53 Pathway
Adipogenesis
Interferon α Response
Fatty Acid Metabolism
Dna Repair
Heme Metabolism
EMT
Kras Signaling Dn
Coagulation
Wnt β-catenin Signaling
Tgf-β Signaling
Angiogenesis
D
Lyz1 Yfp
Lyz1 Apc Kras P53 Yfp
Traced cells
All epithelial cells
YFP+ in Paneth
94.5%
43.7%
7.34%
Paneth
4,91
5,15
4,08
E
FACS evaluation of Traced cells
ANOVA, p = 1.3e-06
% of Traced cells in Paneth gate
scRNAseq evaluation of Traced cells
ANOVA, p = 1.4e-06
% of Traced cells in Paneth cluster
Control
DSS 0d
DSS 3d
DSS 6d
Apc
Apc DSS
Kras
P53
ApcP53
ApcKras
ApcKrasP53
F
ANOVA, p = 0.36
% of Stem cells Ascl2 positive
ANOVA, p = 0.05
ns
P = 0.04
% of Stem cells Lgr5 positive
ANOVA, p < 2e-16
CBC signature in stem cells

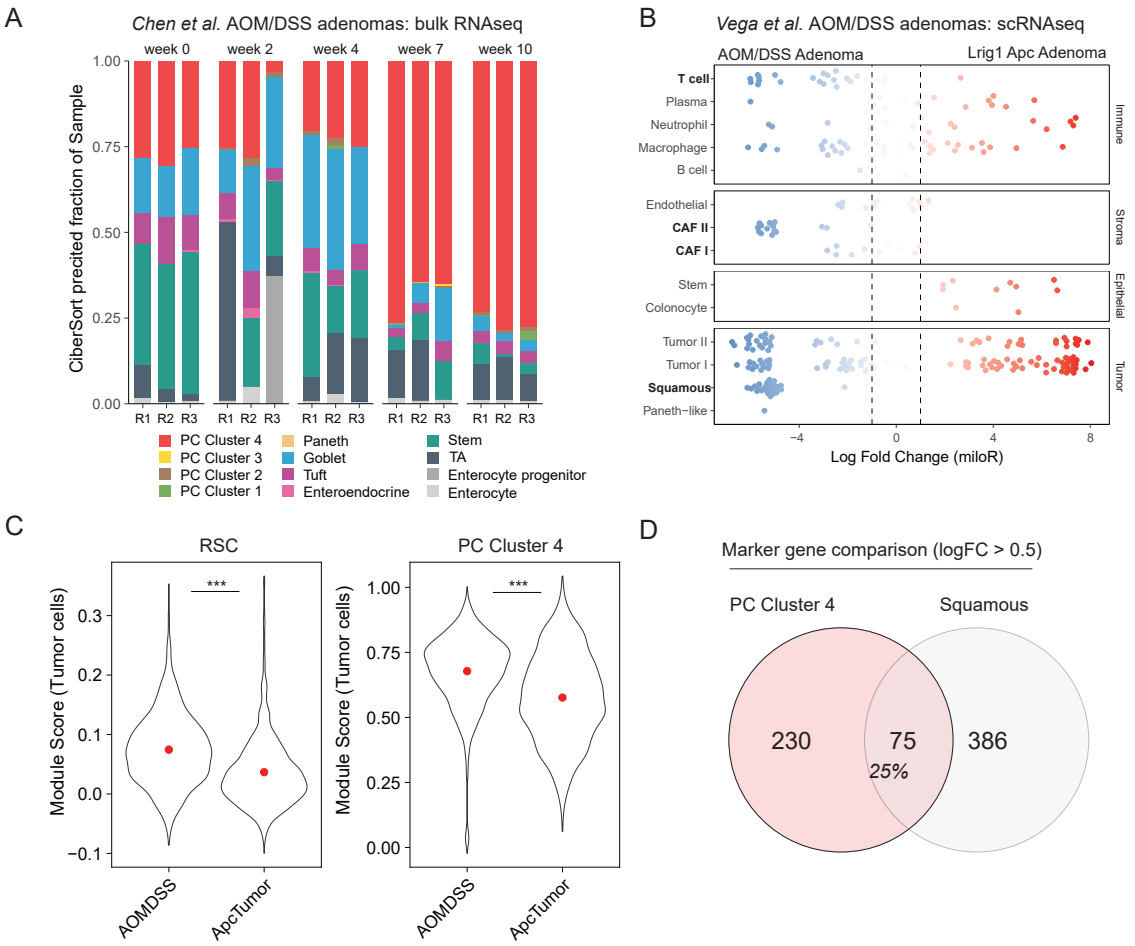
Suppl. Fig. 4



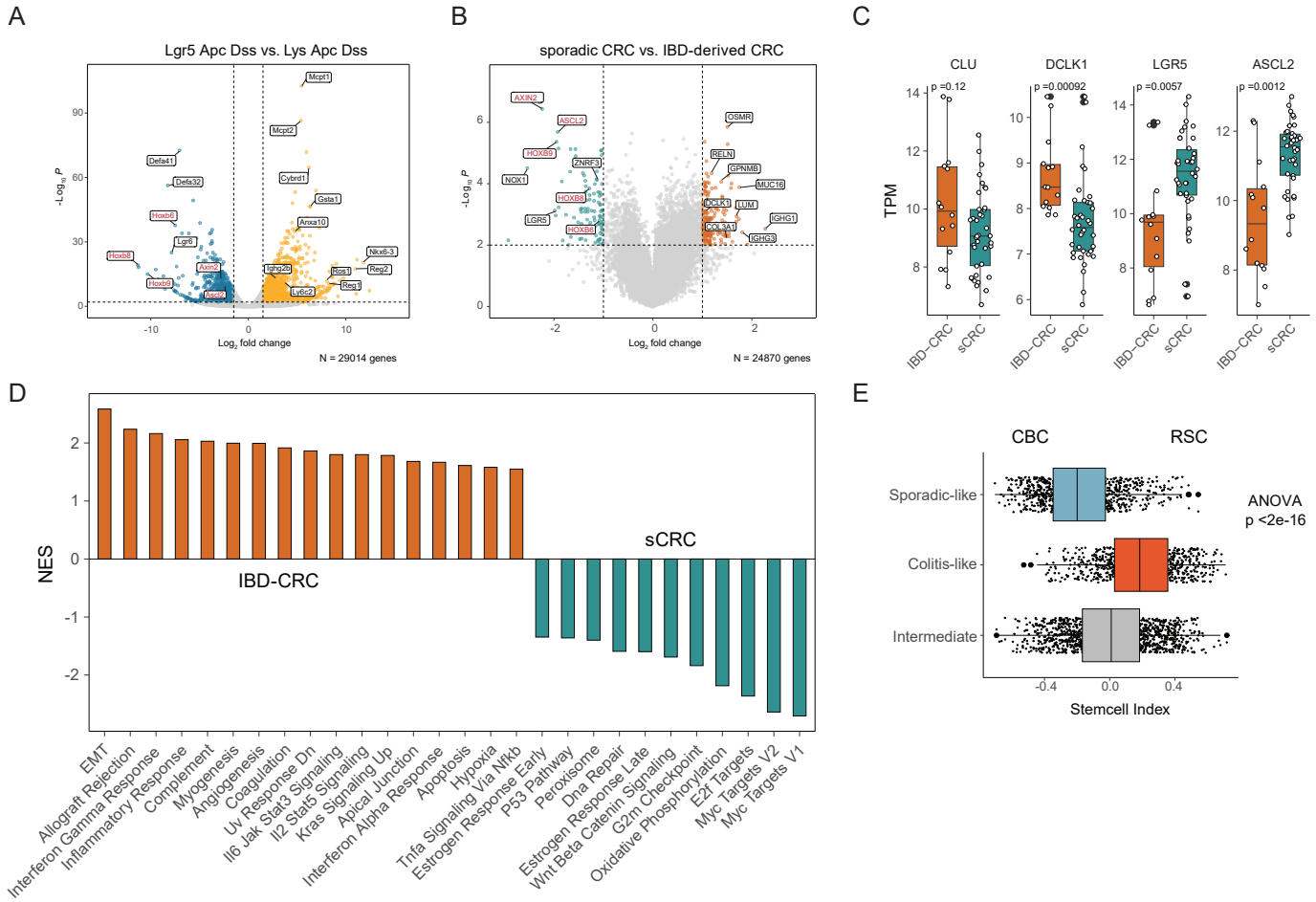
Suppl. Fig. 5


A
Stem Cells (Singh et al.)
Control
Regenerated (+DT)
NS
Other
Paneth top50 (Haber)
Itln1
Lyz1
Defa24
Spink4
Mmp7
Defa17
Guca2b
Reg4
-Log10P
Log2 fold change
total = 5481 variables
Singh et al.; Lgr5DTR-EGFP Neurog3CreERT2 RosatdTomato
B
CTRL
DT
Enterocyte
Enteroendocrine
EP
Gob/Pan
Goblet
Paneth
Sec Prog
Stem
TA
Tuft
RSC
0.05
0
−0.05
Normal
DT_12h
DT_24h
DT_48h
DT_72h
C
Duodenum
Jejunum
Ileum
CTRL
DT
Lgr5DTR-EGFP c-KitCreERT2 RosatdTomato
DAPI, tdTomato
D
CTRL
DT
Lgr5DTR-EGFP c-KitCreERT2 RosatdTomato
DAPI tdTomato Lyz1
DAPI tdTomato Lyz1

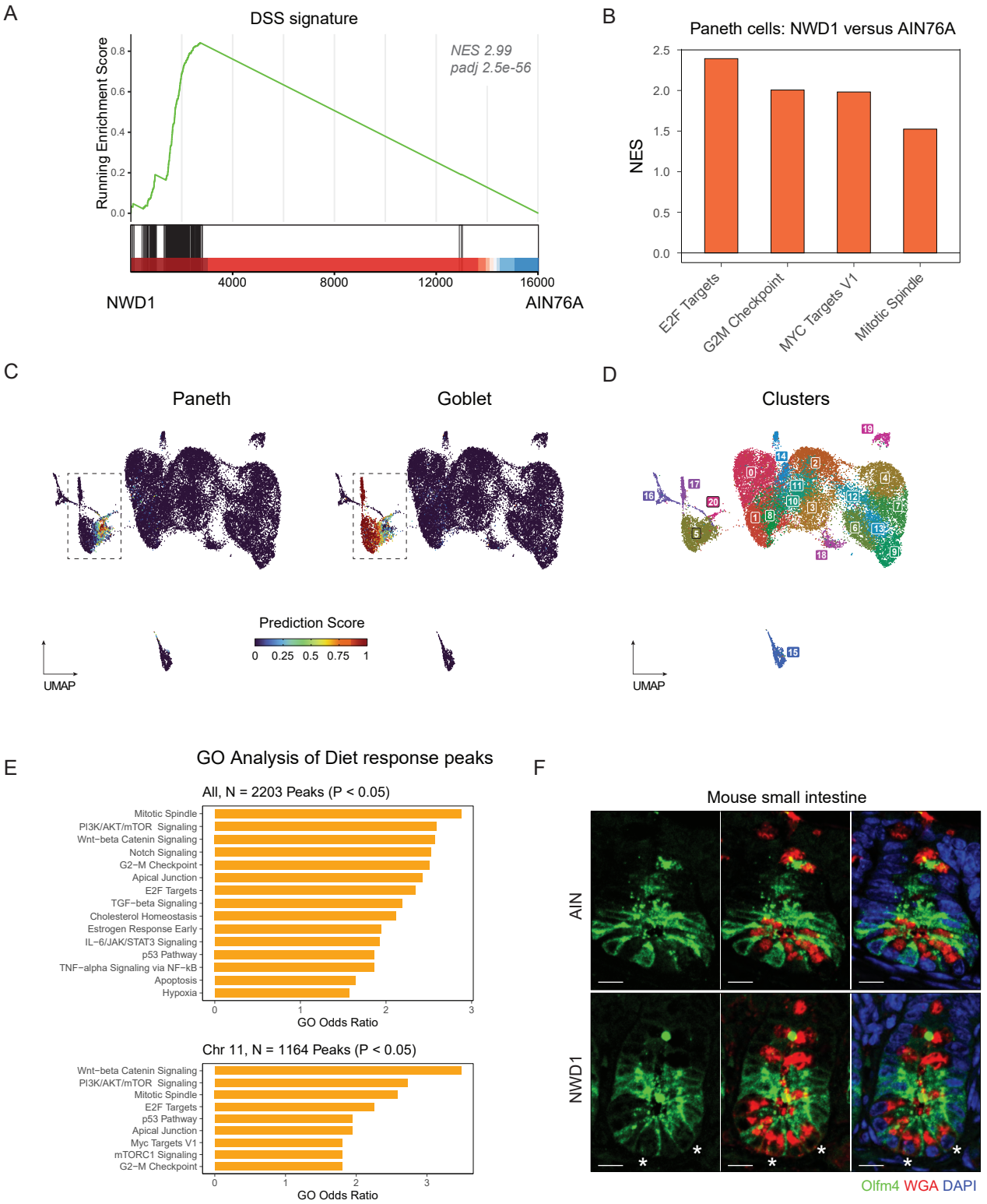
Suppl. Fig. 6


A
Chen et al. AOM/DSS adenomas: bulk RNAseq
week 0
week 2
week 4
week 7
week 10
CiberSort precited fraction of Sample
1.00
0.75
0.50
0.25
0.00
R1 R2 R3
PC Cluster 4
PC Cluster 3
PC Cluster 2
PC Cluster 1
Paneth
Goblet
Tuft
Enteroendocrine
Stem
TA
Enterocyte progenitor
Enterocyte
B
Vega et al. AOM/DSS adenomas: scRNAseq
AOM/DSS Adenoma
Lrig1 Apc Adenoma
T cell
Plasma
Neutrophil
Macrophage
B cell
Endothelial
CAF II
CAF I
Stem
Colonocyte
Tumor II
Tumor I
Squamous
Paneth−like
Immune
Stroma
Epithelial
Tumor
−4
0
4
8
Log Fold Change (miloR)
C
RSC
PC Cluster 4
Module Score (Tumor cells)
***
AOMDSS
ApcTumor
D
Marker gene comparison (logFC > 0.5)
PC Cluster 4
Squamous
230
75
386
25%

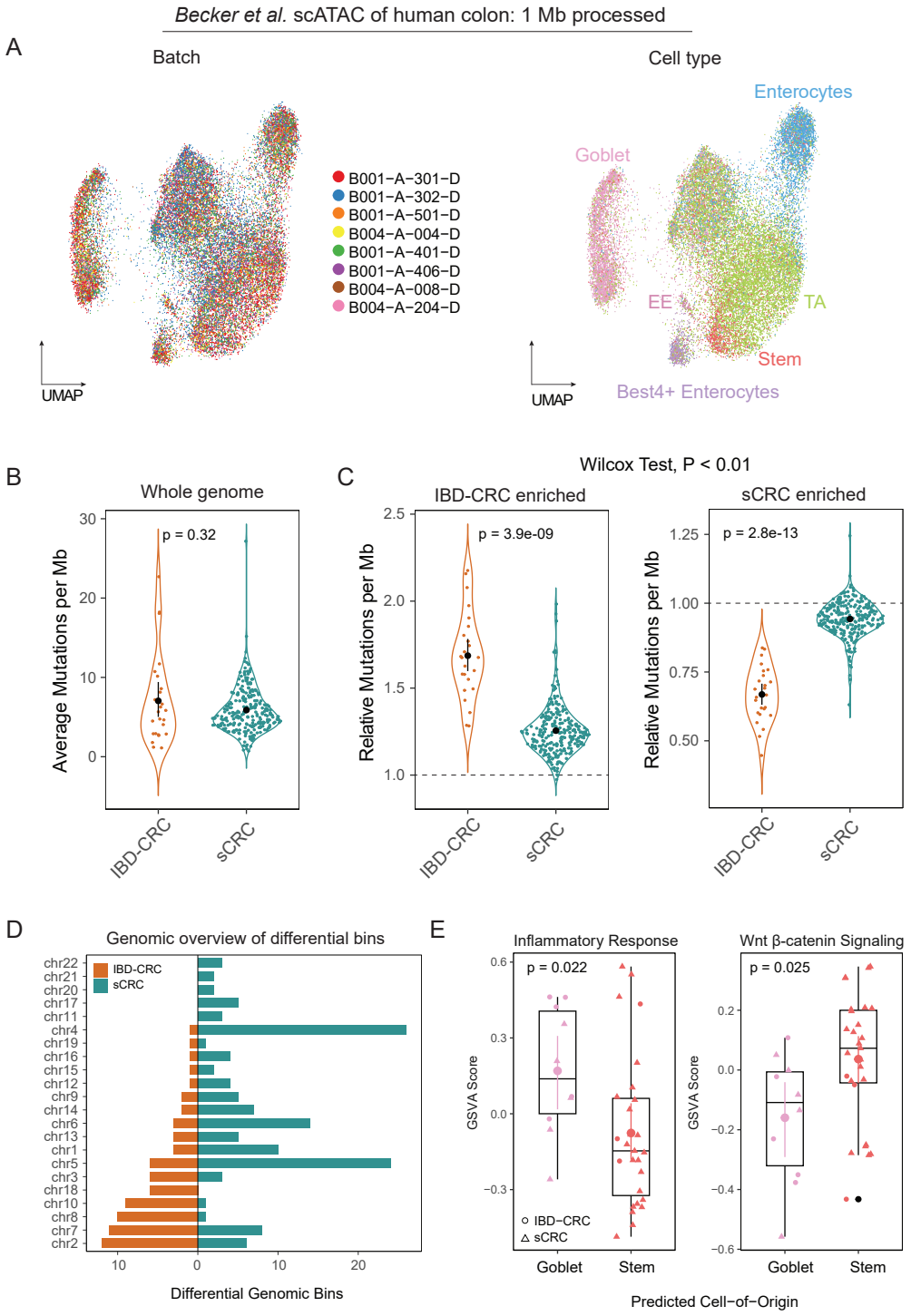
Suppl. Fig. 7


A
Lgr5 Apc Dss vs. Lys Apc Dss
Mcpt1
Mcpt2
Defa41
Defa32
Cybrd1
Gsta1
Hoxb6
Lgr8
Anxa10
Hoxb8
Nkx6-3
Axin2
Ighg2b
Reg3
Reg2
Hoxb9
Reg1
Ascl2
Ly6c2
-Log10 P
Log2 fold change
N = 29014 genes
B
sporadic CRC vs. IBD-derived CRC
AXIN2
ASCL2
OSMR
HOXB9
ZNRF3
RELN
NOX1
GPNMB
MUC16
HOXB8
DCLK1
LUM
LGR5
IGHG1
HOXB6
COL3A1
IGHG3
-Log10 P
Log2 fold change
N = 24870 genes
C
CLU
DCLK1
LGR5
ASCL2
p =0.12
p =0.00092
p =0.0057
p =0.0012
TPM
IBD-CRC
sCRC
D
NES
IBD-CRC
sCRC
EMT
Allograft Rejection
Interferon Gamma Response
Inflammatory Response
Complement
Myogenesis
Angiogenesis
Coagulation
Uv Response Dn
Il6 Jak Stat3 Signaling
Il2 Stat5 Signaling
Kras Signaling Up
Apical Junction
Interferon Alpha Response
Apoptosis
Hypoxia
Tnfa Signaling Via Nfkb
Estrogen Response Early
P53 Pathway
Peroxisome
Dna Repair
Estrogen Response Late
Wnt Beta Catenin Signaling
G2m Checkpoint
Oxidative Phosphorylation
E2f Targets
Myc Targets V2
Myc Targets V1
E
CBC
RSC
Sporadic-like
Colitis-like
Intermediate
ANOVA
p <2e-16
Stemcell Index

Suppl. Fig. 8


A
DSS signature
Running Enrichment Score
NES 2.99
padj 2.5e-56
NWD1
AIN76A
B
Paneth cells: NWD1 versus AIN76A
NES
E2F Targets
G2M Checkpoint
MYC Targets V1
Mitotic Spindle
C
Paneth
Goblet
Prediction Score
UMAP
D
Clusters
UMAP
E
GO Analysis of Diet response peaks
All, N = 2203 Peaks (P < 0.05)
Mitotic Spindle
PI3K/AKT/mTOR Signaling
Wnt−beta Catenin Signaling
Notch Signaling
G2−M Checkpoint
Apical Junction
E2F Targets
TGF−beta Signaling
Cholesterol Homeostasis
Estrogen Response Early
IL−6/JAK/STAT3 Signaling
p53 Pathway
TNF−alpha Signaling via NF−kB
Apoptosis
Hypoxia
GO Odds Ratio
Chr 11, N = 1164 Peaks (P < 0.05)
Wnt−beta Catenin Signaling
PI3K/AKT/mTOR Signaling
Mitotic Spindle
E2F Targets
p53 Pathway
Apical Junction
Myc Targets V1
mTORC1 Signaling
G2−M Checkpoint
GO Odds Ratio
F
Mouse small intestine
AIN
NWD1
Olfm4 WGA DAPI

Suppl. Fig. 9


Becker et al. scATAC of human colon: 1 Mb processed
A
Batch
B001−A−301−D
B001−A−302−D
B001−A−501−D
B004−A−004−D
B001−A−401−D
B001−A−406−D
B004−A−008−D
B004−A−204−D
UMAP
Cell type
Enterocytes
Goblet
EE
TA
Stem
Best4+ Enterocytes
UMAP
B
Whole genome
p = 0.32
Average Mutations per Mb
IBD-CRC
sCRC
C
Wilcox Test, P < 0.01
IBD-CRC enriched
p = 3.9e-09
Relative Mutations per Mb
sCRC enriched
p = 2.8e-13
Relative Mutations per Mb
D
Genomic overview of differential bins
IBD-CRC
sCRC
Differential Genomic Bins
E
Inflammatory Response
p = 0.022
GSVA Score
IBD−CRC
sCRC
Goblet
Stem
Wnt β-catenin Signaling
p = 0.025
GSVA Score
Goblet
Stem
Predicted Cell−of−Origin

Suppl. Fig. 10